MOBILE AND WIRELESS COMMUNICATIONS: PHYSICAL LAYER DEVELOPMENT AND IMPLEMENTATION

MOBILE AND WIRELESS COMMUNICATIONS: PHYSICAL LAYER DEVELOPMENT AND IMPLEMENTATION

Edited by

SALMA AIT FARES AND FUMIYUKI ADACHI

In-Tech
intechweb.org

Published by In-Teh

In-Teh
Olajnica 19/2, 32000 Vukovar, Croatia

www.intechweb.org
Additional copies can be obtained from:
publication@intechweb.org

First published January 2010
Printed in India

Technical Editor: Zeljko Debeljuh

Mobile and Wireless Communications: Physical layer development and implementation,
Edited by Salma Ait Fares and Fumiyuki Adachi

p. cm.
ISBN 978-953-307-043-8

Preface

Mobile and Wireless Communications have been one of the major revolutions of the late twentieth century. We are witnessing a very fast growth in these technologies where mobile and wireless communications have become so ubiquitous in our society and indispensable for our daily lives. The relentless demand for higher data rates with better quality of services to comply with state-of-the art applications has revolutionized the wireless communication field and led to the emergence of new technologies such as Bluetooth, WiFi, Wimax, Ultra wideband, OFDMA. Moreover, the market tendency confirms that this revolution is not ready to stop in the foreseen future.

Mobile and wireless communications applications cover diverse areas including entertainment, industrialist, biomedical, medicine, safety and security, and others, which definitely are improving our daily life. Wireless communication network is a multidisciplinary field addressing different aspects raging from theoretical analysis, system architecture design, and hardware and software implementations. While different new applications are requiring higher data rates and better quality of service and prolonging the mobile battery life, new development and advance research studies and systems and circuits design are necessary to keep pace with the market requirements. This book covers the most advanced research and development topics in mobile and wireless communication networks. It is divided into two parts with a total of thirty-four stand-alone chapters covering various areas of wireless communications of special topics including: physical layer and network layer, access methods and scheduling, techniques and technologies, antenna and amplifier design, integrate circuit design, applications and systems. These chapters present advanced novel and cutting-edge results and development related to wireless communication offering the readers the opportunity to enrich their knowledge in specific topics as well as to explore the whole field of rapidly emerging mobile and wireless networks. We hope that this book will be useful for the students, researchers and practitioners in their research studies.

This first part of the book addresses mainly the physical layer design of mobile and wireless communication and consists of sixteen chapters classified in four corresponding sections:

1. *Propagation Measurements and Channel Characterization and Modeling.*
2. *Multiple Antenna Systems and Space-Time Processing.*
3. *OFDM Systems.*
4. *Modeling and Performance Characterization.*

The first section contains three chapters related to Propagation Measurements and Channel Characterization and Modeling. The focus of the contributions in this section, are channel characterization in tunnels wireless communication, novel approach to modeling MIMO wireless communication channels and a review of the high altitude platforms technology.

The second section contains five chapters related to Multiple Antenna Systems and Space-Time Processing. The focus of the contributions in this section, are new beamforming and diversity combining techniques, and space-time code techniques for MIMO systems.

The third section contains four chapters related to OFDM Systems. This section addresses frequency-domain equalization technique in single carrier wireless communication systems, advanced technique for PAPR reduction of OFDM Signals and subcarrier allocation in OFDMA in cellular and relay networks.

The forth section contains four chapters related to Modeling and Performance Characterization. In this section, a unified data and energy model for wireless communication and a new system level mathematical performance analysis of mobile cellular CDMA networks are presented. In addition, novel approach to modeling MIMO wireless communication channels and a review of the in high altitude platforms technology are proposed.

Section 1: ***Propagation Measurements and Channel Characterization and modeling***

Chapter 1 presents a review of the theoretical early and recent studies done on communication within tunnels. The theory of mode propagation in straight tunnels of circular, rectangular and arched cross sections has been studied. Comparison of the theory with existing experimental measurements in real tunnels has been also covered.

Chapter 2 reviews the applications of the Thomson Multitaper, for problems encountered in communications. In particular it focuses on issues related to channel modeling, estimation and prediction for MIMO wireless communication channels.

Chapter 3 presents an overview of the HAP (High Altitude Platforms) concept development and HAP trails to show the worldwide interest in this emerging novel technology. A comparison of the HAP system has been given based on the basic characteristics of HAP, terrestrial and satellite systems.

Section 2: ***Multiple antenna systems and space-time processing***

Chapter 4 discusses the performance analysis of wireless communication systems where the receiver is equipped with maximal–ratio–combining (MRC), for performance improvement, in the Nakagami-m fading environment.

Chapter 5 presents a new approach using sequential blind beamforming to remedy both the inter-symbol interference and intra-symbol interference problems in underground wireless communication networks using jointly CMA, LMS and adaptive fractional time delay estimation filtering.

Chapter 6 examines the possibility of multiple HAP (High Altitude Platforms) coverage of a common cell area in WCDMA systems with and without space-time diversity techniques.

Chapter 7 presents an overview of space-time block codes, with a focus on hybrid codes, and analyzes two hybrid MIMO space-time codes with arbitrary number of STBC/ABBA and spatial layers, and a receiver algorithm with very low complexity.

Chapter 8 discusses the MIMO channel performance in the LOS environment, classified into two cases: the pure LOS propagation and the LOS propagation with a typical scatter. The MIMO channel capacity and the condition number of the matrix were also investigated.

Section 3: ***OFDM Systems***

Chapter 9 proposes an iterative optimization method of transmit/receive frequency domain equalization (TR-FDE) based on MMSE criterion, where both transmit and receive FDE weights are iteratively determined with a recursive algorithm so as to minimize the mean square error at a virtual receiver.

Chapter 10 proposes an enhanced version of the iterative flipping algorithm to efficiently reduce the PAPR of OFDM signal.

Chapter 11 discusses the problem of allocating resources to multiple users on the downlink in an LTE (Long Term Evolution) cellular communication system in order to maximize system throughput. A reduced complexity sub-optimal scheduler was proposed and found to perform quite well relative to the optimal scheduler.

Chapter 12 introduces the DTB (distributed transmit beamforming) approach to JCDS (joint cooperative diversity and scheduling) OFDMA-based relay network in multi source-destination pair's environment and highlights its potential to increase the diversity order and the system throughput performance. In addition, to trade-off a small quantity of the system throughput in return for significant improvement in the user throughput, a fixed cyclic delay diversity approach has been introduced at relay stations to the proposed JCDS-DTB.

Section 4: ***Modeling and Performance Characterization***

Chapter 13 presents the mathematical analysis of mobile cellular CDMA networks considering link unreliability in a system level analysis. Wireless channel unreliability was modeled by means of a Poisson call interruption process which allows an elegant teletraffic analysis considering both wireless link unreliability and resource insufficiency.

Chapter 14 presents a developed energy model to conduct simulations which describe the energy consumption by sending a well defined amount of data over a wireless link with fixed properties. The main aim in this study was to maximize the amount of successfully transmitted data in surroundings where energy is a scarce resource.

Chapter 15 reviews the performance strengths and weaknesses of various short range wireless communications e.g. RadioMetrix, IEEE 802.11a/b, IEEE 802.15.4, DECT, Linx, etc, which are commonly used nowadays in different RoboCup SSL wireless communication implementations. An adaptive error correction and frequency hopping scheme has been proposed to improve its immunity to interference and enhance the wireless communication performance.

Chapter 16 reviews the capacity dimensioning methods exploited for system capacity performance evaluation and wireless network planning used in development process of any generation of mobile communications system.

Editors

Salma Ait Fares
Graduate School of Engineering
Department of Electrical and Communication Engineering
Tohoku University, Sendai, Japan
Email: aitfares@mobile.ecei.tohoku.ac.jp

Fumiyuki Adachi
Graduate School of Engineering
Department of Electrical and Communication Engineering
Tohoku University, Sendai, Japan
Email: adachi@ecei.tohoku.ac.jp

Contents

Section 1: Propagation Measurements and Channel Characterization and modeling

Section 2: Multiple antenna systems and space-time processing

Section 3: OFDM Systems

Section 4: Modeling and Performance Characterization

1

Wireless Transmission in Tunnels

Samir F. Mahmoud
Kuwait University
Kuwait

1. Introduction

Study of electromagnetic wave propagation within tunnels was driven in the early seventies by the need for communication among workers in mine tunnels. Such tunnels are found in the form of a grid of crossing tunnels running for several kilometers, which called for reliable means of communication. Since then a great number of contributions have appeared in the open literature studying the mechanisms of communication in tunnels. Much experimental and theoretical work was done in USA and Europe concerning the development of wireless and wire communication in tunnels. A typical straight tunnel with cross sectional linear dimensions of few meters can act as a waveguide to electromagnetic waves at UHF and the upper VHF bands; i.e. at wavelengths in the range of a fraction of a meter to few meters. In this range, a tunnel is wide enough to support free propagation of electromagnetic waves, hence provides communication over ranges of up to several kilometers. At those high frequencies, the tunnel walls act as good dielectric with small loss tangent. For example at a frequency of 1000 MHz, the rocky wall with 10^{-2} Siemens/m conductivity and relative permittivity of about 10 has a loss tangent equal to 0.018. So the electromagnetic wave losses will be caused mainly by radiation or refraction through the walls with little or negligible ohmic losses as deduced by Glaser (1967, 1969).

Goddard (1973) performed some experiments on UHF and VHF wave propagation in mine tunnels in USA. His work was presented in the Thru-Earth Workshop held in Golden, Colorado in August 1973. Goddard’s experimental results show that small attenuation is attained in the UHF band. He also measured wave losses around corners and detected high coupling loss between crossing tunnels. The theory of mode propagation of electromagnetic waves in tunnels of rectangular cross section was reported by Emslie et al (1973) in the same workshop and later in (1975) in the IEEE, AP journal. Their presentation extended the theory given earlier by (Marcatelli and Schmeltzer, 1964) and have shown, among other things, that the modal attenuation decreases with the applied frequency squared. Chiba et al. (1978) presented experimental results on the attenuation in a tunnel with a cross section close to circular except for a flat base. They have shown that the measured attenuation closely match the theoretical attenuation of the dominant modes in a circular tunnel of equal cross section. Mahmoud and Wait (1974a) applied geometrical ray theory to obtain the fields of a dipole source in a tunnel with rectangular cross section of linear dimensions of several wavelengths. The same authors (Mahmoud and Wait, 1974b) studied the attenuation in a curved rectangular tunnel showing a considerable increase in attenuation due to the curvature.

Recent advances on wireless communication, in general, have revived interest in continued studies of free electromagnetic transmission in tunnels. Notable contributions have been made by Donald Dudley in USA and Pierre Degauque in France and their research teams.

In the next sections, we review the theoretical early and recent studies done on communication within tunnels. We start by reviewing the mode theory of propagation in a straight tunnel model with a circular cross-section. We obtain the propagation parameters of the important lower order modes in closed forms. Such modes dominate the total field at sufficiently distant points from the source. We follow by reviewing mode propagation in tunnels with rectangular cross section. Tunnels with arched cross sections are then treated as perturbed tunnel shapes using the perturbation theory to characterize their dominant modes. We also review studies made on wave propagation in curved tunnels and propagation around corners. We conclude by comparisons between theory and available experimental results.

2. Tunnel wall Characterization

Before treating modal propagation in tunnels, it will prove useful to characterize the tunnel wall as constant impedance surfaces for the dominant modes of propagation. This is covered in detail in [Mahmoud, 1991, chapter 3] for planar and cylindrical guides. Here we give a simplified argument for adopting the concept of constant impedance wall. By this term, we mean that the surface impedance of the wall is almost independent of the angle of wave incidence onto the wall. So, let us consider a planar surface separating the tunnel interior (air) from the wall medium with relative permittivity ε_r and permeability μ_0 where ε_r is usually >>1. At a given applied angular frequency ω, the bulk wavenumber in air is $k_0 = \omega\sqrt{\mu_0\varepsilon_0}$ and in the wall is $k = k_0\sqrt{\varepsilon_r} >> k_0$. We may define three right handed mutually orthogonal unit vectors $\hat{n}, \hat{t}, \hat{z}$ where $\hat{t}$ and $\hat{z}$ are tangential to the air-wall interface and $\hat{n}$ is normal to the interface and is directed into the wall. Now consider a wave that travels along the interface with dependence $\exp(j\omega t - j\beta z)$ in both the air and wall. For modes with $\beta < k_0$, one has $k_0^2 - \beta^2 < k_0^2$ and definitely $k_0^2 - \beta^2 << k^2$. Hence the transverse wavenumber in the wall (along $\hat{n}$) $k_n = \sqrt{k^2 - \beta^2}$ can be well approximated by the $k_n = \sqrt{k^2 - k_0^2}$ which is independent of β. If the wave is TE_n polarized , i.e. having zero E_n and nonzero H_n, the surface impedance of the wall is given by:

$$Z_s = E_t / H_z = \omega\mu_0 / \sqrt{k^2 - k_0^2} = \eta_0 / \sqrt{\varepsilon_r - 1} \tag{1}$$

Similarly if the wave is TM_n polarized, the surface admittance of the wall is

$$Y_s = H_t / E_z = -\omega\varepsilon / \sqrt{k^2 - k_0^2} = \varepsilon_r / \eta_0\sqrt{\varepsilon_r - 1} \tag{2}$$

where η_0 is the wave impedance in air(=120 π Ω).

The Z_s and Y_s in (1-2) are the constant impedance/admittance of the wall. When the air-wall interface is of cylindrical shape, the same Z_s and Y_s apply provided that $(k^2 - k_0^2)^{1/2} a >> 1$ ([Mahmoud, 1991), where a is the cylinder radius. This condition is satisfied in most tunnels at the operating frequencies.

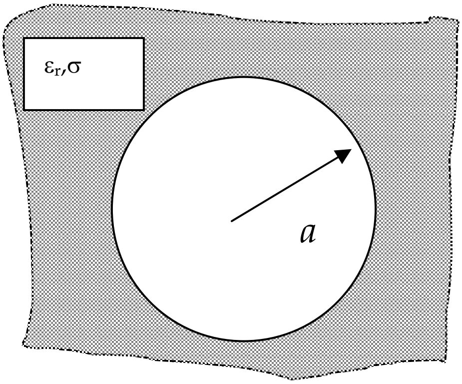


Fig. 1. A circular cylindrical tunnel of radius '*a*' in a host medium of relative permittivity ε_r and conductivity 'σ'.

3. Modal propagation in a tunnel with a circular cross Section

An empty tunnel with a circular cross section is depicted in Figure 1. The surrounding medium has a relative permittivity ε_r (assumed >>1) and conductivity σ S/m. The circular shape allows for rigorous treatment of the free propagating modes. The modal equation for the various propagating modes was derived by Stratton [1941] and solved approximately for the modal phase and attenuation constants of the low order modes in (Marcatili and Shmeltzer, 1964), (Glasier, 1969) and more recently by Mahmoud (1991). In the following, we derive modal solutions that are quite accurate as long as the tunnel diameter is large relative to the applied wavelength. Under these conditions the tunnel wall is characterized by the constant surface impedance and admittance given in (1-2). Specialized to the circular tunnel, they take the form:

$$E_\phi = \eta_0 \bar{Z}_s H_z \quad \text{and} \quad \eta_0 H_\phi = -\bar{Y}_s E_z \tag{3}$$

Here $\bar{Z}_s$ and $\bar{Y}_s$ are normalized impedance and admittance relative to η_0 and η_0^{-1} respectively. Explicitly:

$$\bar{Z}_s = 1/\sqrt{\varepsilon_r - 1 - i\sigma/\omega\varepsilon_0}, \tag{4a}$$

$$\bar{Y}_s = (\varepsilon_r - i\sigma/\omega\varepsilon_0)/\sqrt{\varepsilon_r - 1 - i\sigma/\omega\varepsilon_0} \tag{4b}$$

Because of the imperfectly reflecting walls, the allowed modes are generally hybrid with nonzero longitudinal field components E_z and H_z (Mahmoud, 1991, sec. 3.4). Thus:

$$E_z = jJ_n(k_\rho \rho)\sin n\phi \, \exp(-j\beta z) \tag{5a}$$

$$\eta_0 H_z = -j\Lambda J_n(k_\rho \rho)\cos m\phi \, \exp(-j\beta z) \tag{5b}$$

where Λ is the mode hybrid factor, β is the mode propagation factor, $J(.)$ is the Bessel function of first kind, and n is an integer *=0,1,2….* The transverse fields $E_\rho, E_\phi, H_\rho, and\, H_\phi$ are obtainable in terms of the longitudinal components through well-known relations as given in the Appendix. The boundary conditions at the tunnel wall $\rho=a$ require that $E_\phi = \eta_0 \overline{Z}_s H_z$ and $\eta_0 H_\phi = -\overline{Y}_s E_z$. Using (A2) & A(3) in the Appendix along with (5), these boundary conditions read:

$$F(u) + j\overline{Z}_s u^2 / \mathrm{v} = -\beta n / \Lambda k_0 \tag{6a}$$

$$F(u) + j\overline{Y}_s u^2 / \mathrm{v} = -\beta n \Lambda / k_0 \tag{6b}$$

where:

$$\begin{gathered} u = k_\rho a, \ \mathrm{v} = k_0 a, \\ \mathrm{v}^2 - u^2 = (\beta a)^2, \text{ and} \\ F_n(u) = \frac{u J_n'(u)}{J_n(u)} \equiv n \frac{J_{n-1}(u) - J_{n+1}(u)}{J_{n-1}(u) + J_{n+1}(u)} \end{gathered} \tag{7}$$

The prime denotes differentiation with respect to the argument and the last equality stems from identities in (A4).

Equations (6) lead to the modal equation for the propagation factor β and the mode hybrid factor Λ. Namely:

$$(F_n(u) + ju^2\overline{Y}_s / \mathrm{v})(F_n(u) + ju^2\overline{Z}_s / \mathrm{v}) - n^2(\beta / k_0)^2 = 0 \tag{8a}$$

$$\Lambda - \Lambda^{-1} = -j(\overline{Y}_s - \overline{Z}_s)u^2 k_0 / n\beta \mathrm{v} \tag{8b}$$

3.1 The ϕ- symmetric modes (n=0)

Considering first the ϕ-symmetric modes we set *n=0* in (8) which then reduces to two equations for the TM_{0m} and TE_{0m} modes; namely:

$$F_0(u) = -uJ_1(u) / J_0(u) = -ju^2\overline{Y}_s / \mathrm{v} \tag{9a}$$

$$F_0(u) = -uJ_1(u) / J_0(u) = -ju^2\overline{Z}_s / \mathrm{v} \tag{9b}$$

For the low order modes which are dominant at sufficiently high frequencies, $u/\mathrm{v} \ll 1$, hence the RHS of (9a-9b) are also <<1 and an approximate solution for the eigenvalue u is:

$$u_{0m} \cong x_{1m}(1 + j\overline{Y}_s / \mathrm{v}) \tag{10}$$

$$u_{0m} \cong x_{1m}(1 + j\overline{Z}_s / \mathrm{v}) \tag{11}$$

for TM_{0m} and TE_{0m} respectively. In the above x_{1m} is the mth root of the Bessel function J_1. Note that the second of (8) yields $\Lambda=0$ ($H_z=0$) or infinity ($E_z=0$) as expected for the TM_0 and TE_0 modes respectively. Since $\beta a = \sqrt{v^2 - u^2}$, we can use (10-11) to get the complex β factor. Of particular interest are the modal attenuation rates, which take the forms

$$\alpha_{0m} = \frac{x_{1m}^2}{k_0^2 a^3} \mathrm{Re}\begin{bmatrix} \overline{Z}_s \\ \overline{Y}_s \end{bmatrix} \quad \text{Neper/m} \tag{12}$$

The TE_0 modes are associated with Z_s and the TM_0 modes with Y_s. Since $\overline{Z}_s \ll \overline{Y}_s$ (see (4)), it is clear that the TE_0 modes have considerably less attenuation than the TM_0 modes; it is less by almost the factor ε_r.

3.2 Hybrid Modes

Next we consider the case $n>0$ for the hybrid modes. In the high frequency regime and for the low order modes, we use the following approximations in (8): $(\beta/k_0)^2 \sim 1$, and $u^2/\mathrm{v} \ll 1$. This leads to the approximate solutions:

$$u_{nm} = x_{n-1,m}(1 + j(\overline{Z}_s + \overline{Y}_s)/2\mathrm{v}) \tag{13}$$

and

$$u_{nm} = x_{n+1,m}(1 + j(\overline{Z}_s + \overline{Y}_s)/2\mathrm{v}) \tag{14}$$

The corresponding modes are designated as HE_{nm} and EH_{nm} modes respectively. Investigating the hybrid factor Λ (the second of eqn. 8), we find that it approaches +1 for HE and -1 for EH modes in the infinite frequency limit. This means that the modes are hybrid balanced modes in this limit. Further discussions on these modes are found in (Mahmoud,1991, Chap.5).

Using (13)-(14) along with the relation $\beta a = \sqrt{v^2 - u^2}$, we obtain the high frequency modal propagation constant. Again, the imaginary part of β gives the mode attenuation rate α ($\alpha=-\mathrm{Im}(\beta)$). The approximate modal attenuation factor α for the HE_{nm} modes is:

$$\alpha_{nm} = \frac{x_{n-1,m}^2}{2k_0^2 a^3} \mathrm{Re}\left[\overline{Y}_s + \overline{Z}_s\right] \qquad \text{(Neper/m)} \tag{15}$$

The attenuation rate for the other set of modes; EH_{nm} modes, is the same except that the Bessel root $x_{n-1,m}$ is replaced by $x_{n+1,m}$. The above formulae (12 and 15) show that the modal attenuation is inversely proportional to the frequency squared and the radius cubed. Note that these Formulae are restricted to the lower order modes of the tunnel at sufficiently high frequencies.

It is clear that the least attenuated mode of the *TE₀* and *TM₀* mode group is the *TE₀₁* mode, while the least attenuated hybrid mode is the *HE₁₁* mode. It is interesting to compare the attenuation rates of these two modes; namely the *TE₀₁* and the *HE₁₁* mode. Using (12) and (15), we get the ratio

$$\frac{\alpha_{TE01}}{\alpha_{HE11}} = \frac{x_{1,1}^2}{x_{0,1}^2}\frac{2}{\varepsilon_r + 1} = \frac{3.832^2}{2.405^2}\frac{2}{\varepsilon_r + 1} = \frac{5.078}{\varepsilon_r + 1} \tag{16}$$

where we have neglected the earth conductivity σ relative to $\omega\varepsilon$. As an example, for typical earth with ε_r =12, the above ratio amounts to 0.395, which means that the TE_{01} mode is the least attenuated mode in a typical circular tunnel.

In order to check the approximate closed forms (12)-(15) for the attenuation factors, we compare them with exact results presented recently by Dudley and Mahmoud [2006] in Table 1. The table lists the attenuation rates of some of the dominant modes in a circular tunnel of 2 meter radius and outer medium having ε_r=12 at 1 GHz. It is seen that the percentage error is less than ~2.7% for all the listed modes except the EH_{11} mode. This mode requires a higher frequency for the approximate attenuation to have a better accuracy.

Mode	Exact	Approximate	% Error
TE_{01}	1.098	1.096	0.18%
TE_{02}	3.716	3.673	1.16%
TE_{03}	7.937	7.724	2.68 %
HE_{11}	2.774	2.805	1.12%
HE_{21}	7.158	7.122	0.50%
HE_{31}	13.12	12.79	2.52%
TM_{01}	13.30	13.15	1.13%
EH_{11}	20.18	12.794	36.6%

Table 1. Comparison between Exact [Dudley & Mahmoud,2006] and approximate attenuation rates in dB/100meters at 1 GHz.

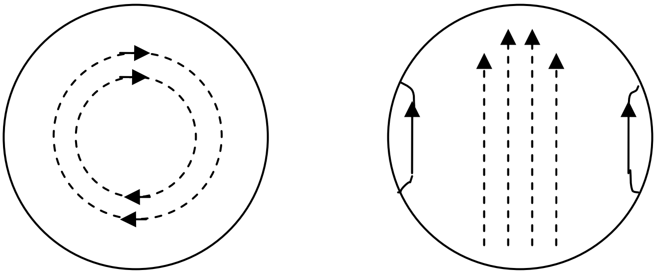
Exercise 1: Verify the approximate attenuation rates in Table 1, Column 3 by using (12) for the TE_{0m}/TM_{0m} modes and (15) for the HE_{nm} modes.

Exercise 2: Using any root finding software, verify the exact attenuation rates in column 2 of Table 1. To do so, you need to solve the modal equation (8) for the complex roots of *u*. You can use (10),(11) or (13) as initial guess for u of the TE_{0m}, TM_{0m} and HE_{nm} modes respectively. Once you get u, the complex β is obtained from $\beta a = \sqrt{v^2 - u^2}$. The mode attenuation rate α in *Neper/m* is the negative of the imaginary part of β. To convert to dB/100m, note that 1 *Neper/m*= 868.8 dB/100m.

3.3 Mode Excitation

In the above we have ordered the modes in ascending order of their attenuation rates. However, the actual level of the modes at a given distance from the source is determined also by their excitation factor, which in turn depends on their field distribution and the source type, location and orientation. The *E*-field distribution of the TE_{01} and the HE_{11} modes, which are the least attenuated modes, are sketched in Figure 2. The TE_{01} mode has a circumferential E_ϕ field, which vanishes at $\rho=0$ and is quite weak at the wall $\rho=a$ (being proportional to $J_1(u\rho/a)$). It follows that the TE_{01} mode can be excited by a circumferentially oriented linear dipole. For optimum excitation, the linear dipole should neither be at the center or very close to the wall. Alternatively, the TE_{01} mode can also be excited by a current loop placed in the cross section plane near the center of the tunnel. This will couple with H_z which is maximum at the tunnel center. Dudley (2005) has given rigorous treatment of TE_0 modes excitation by a loop, which is located coaxially with the tunnel.

On the other hand, the HE_{11} mode is almost linearly polarized as demonstrated in the Appendix (see equation A6). So, this mode is optimally excited by a linear dipole close to the center of the tunnel. A detailed rigorous treatment of the HE_{nm} mode excitation by a linear dipole is found in (Dudley and Mahmoud, 2006).



TE_{01} mode HE_{11} mode

Fig. 2. E-field lines of the lowest order modes in a circular tunnel.

Here we derive a simple formula for the excitation coefficient of the propagating modes in the tunnel. The source is assumed to be a small linear electric dipole of vector moment '$\vec{P}$' (Ampere-m) located at (ρ_1,ϕ_1) in the cross section at, say z=0 and oriented along an arbitrary direction in the transverse plane. The total excited fields $(\vec{E},\vec{H})$ are expressed as a sum over the natural modes in the tunnel, so

$$(\vec{E},\vec{H}) = \sum_r A_r^{\pm}\,(\vec{e}_r,\vec{h}_r)\exp(\mp j\beta_r z) \tag{17}$$

Where $(\vec{e}_r,\vec{h}_r)$, $r=1,2...$ are the normal modal fields ordered in an arbitrary manner. The A_r are the excitation coefficients. The $\pm$ signs correspond to the fields in the $z>0$ and $z<0$ respectively. To the above sum we should add a continuous of waves representing radiation through the surrounding host medium. However, this is normally heavily attenuated as demonstrated in (Dudley and Mahmoud, 2006) and hence will be omitted. The mode excitation coefficients are determined based on the orthogonality among the modes in a tunnel with constant impedance walls (Mahmoud, 1991, section 3.6) and the field discontinuity at the source. So we get ((Mahmoud, 1991, section 2.11).

$$A_r^+ = A_r^- = \frac{-\vec{P}.\vec{e}_r}{2\int_S (\vec{e}_r \mathrm{x} \vec{h}_r).\hat{z}\, ds} \tag{18}$$

where the integration is taken over the tunnel cross section bounded by the walls.

Exercise 3: For each of the linear electric dipoles shown in Figure 3, what are the important excited modes? [*HE11 for dipoles 1, and 2, Both HE11 and TE01 modes for dipoles 3 and 4*].

Exercise 4: Apply (18) to get the excitation coefficient for the HE11 mode when excited by a unit y-directed dipole moment; $\vec{P} = \hat{y}$ located at the center of the tunnel ρ=0. Note that $(\vec{e}_t, \vec{h}_t)$ are given by equations (A6-A7) in the Appendix. To simplify the problem, assume that the frequency is high enough so that $\beta \cong k_0$ *and* $\Lambda \cong +1$. Find the power launched in the HE_{11} mode.

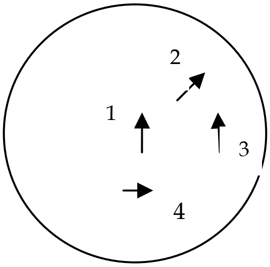


Fig. 3.

4. A rectangular tunnel

Modal propagation in rectangular waveguides with imperfectly conducting walls is considered to be an intractable problem in waveguide theory. In fact, as indicated by (Wait, 1967) long time ago, there is a fundamental difficulty with mode analysis in cylindrical waveguides with finite impedance walls and cross-sections other than circular. In such waveguides, a natural mode does not take the form of a wave with a single transverse wavenumber as is the case with waveguides with perfectly reflecting walls. In contrast, a waveguide mode is generally composed of a weighted sum of elementary waves having a single transverse wavenumber each. A comprehensive discussion of this phenomenon is reviewed by Mahmoud (1991 sec. 3.5) in relation to elliptical and rectangular waveguides. However, simple, but approximate modal solutions valid in the high frequency regime have been obtained by several authors. Marcatili &Scmeltzer (1964) obtain the x and y transverse wavenumber approximately by treating the rectangular waveguide as two parallel plates in the y and x directions respectively. Andersen et al. (1975) derive an aprproximate modal equation for the transverse wavenumber by considering the mode to be composed of four crossing plane waves interconnected at the tunnel walls by reflection matrices. The method accounts for the coupling between horizontal and vertical polarizations upon reflection.

Wait [1980] and Mahmoud [1991, sec. 6.2.2] obtain an approximate TM type modal solution based on the neglect of one weak boundary condition. In the following, we outline this formulation.

So, consider a rectangular guide of width *w* and height *h* and outer medium of complex relative permittivity ε_c ($=\varepsilon_r - i\sigma/\omega\varepsilon_0$) as depicted in Figure 4. The walls are characterized by constant impedance and admittance $\overline{Z}_s, \overline{Y}_s$ as given by (4). For a TM$_y$ mode, or a vertically polarized mode, E_x=0 and E_y may be given, for an even mode, by:

$$E_y = \cos(k_y y)\cos(k_x x)\exp(-j\beta z) \tag{19}$$

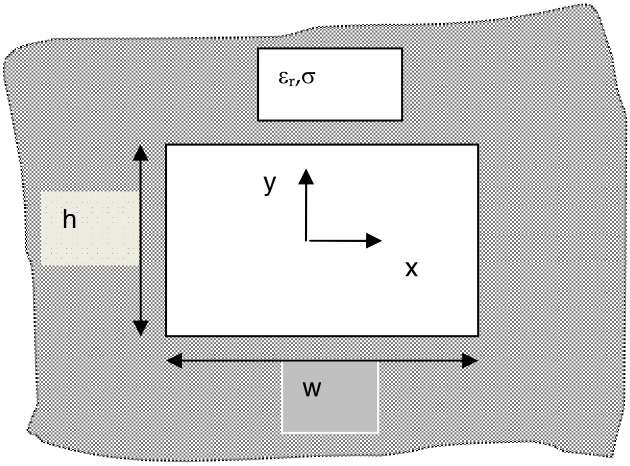


Fig. 4. A rectangular tunnel in a host medium

In the above: $k_y^2 + k_x^2 = k_0^2 - \beta^2$. Because the guide is oversized relative to the wavelength, both k_x and k_y are << k_0 and β for the low order modes. The E_z component is obtained from the divergence equation $\nabla.\vec{E} = 0$, hence:

$$j\beta E_z = \partial E_y / \partial y \tag{20}$$

which shows that E_z is of first order smallness relative to E_y. The magnetic field components are obtained as:

$$\eta_0 H_x = -\beta E_y / k_0, \eta_0 H_y \approx 0, and\, \eta_0 H_z = j\partial E_y / k_0 \partial x \tag{21}$$

where terms of second order smallness have been neglected (such as $k_x k_y / k_0^2$). Now the boundary condition at y=$\pm h/2$ requires that: $(\eta_0 H_x / E_z)|_{y=\pm b} = \pm\overline{Y}_s$, which reduces to:

$$k_y h \tan(k_y h/2) = jk_0 h/\overline{Y}_s \tag{22}$$

This is an equation for the y-wavenumber and its solution leads to a set of eigenvalues k_{yn}, n=1,2… Now we consider the side walls at $x=\pm w/2$. The boundary conditions at these two walls are

$$E_y = \pm\eta_0 \overline{Z}_s H_z \tag{23}$$

$$\eta_0 H_y = \mp\overline{Y}_s E_z \tag{24}$$

Using (19) and (21) in (23) leads to a modal equation for k_x.

$$k_x w \tan(k_x w/2) = jk_0 w/\overline{Z}_s \tag{25}$$

So (25) is an equation for k_x whose solution leads to a set of eigenvalues k_{xm}. This completes the modal solution except that we have not satisfied boundary condition (24). Fortunately however, H_y is of second order smallness for the lower order modes, hence this boundary condition can be safely neglected.

Approximate solutions of (22) and (25) for k_{yn} and k_{xm} in the high frequency regime, ($k_0 h >> \overline{Y}_s, k_0 w >> \overline{Z}_s$) are:

$$\begin{aligned} k_{yn}h &= n\pi[1 + j2\overline{Y}_s/k_0 h] \\ k_{xm}w &= m\pi[1 + j2\overline{Z}_s/k_0 w] \end{aligned}, \tag{26}$$

where m and n =1,3…are odd integers for the even modes considered. The corresponding mode attenuation rate is easily obtained as:

$$\alpha_{\text{VPmn}} = 2\pi^2 n^2 \operatorname{Re}(\overline{Y}_s)/k_0^2 h^3 + 2\pi^2 m^2 \operatorname{Re}(\overline{Z}_s)/k_0^2 w^3 \quad \text{Neper/m} \tag{27}$$

The attenuation rate of the corresponding horizontally polarized mode may be obtained from (27) by exchanging *w* and *h*. So:

$$\alpha_{HPmn} = 2\pi^2 n^2 \operatorname{Re}(\overline{Y}_s)/k_0^2 w^3 + 2\pi^2 m^2 \operatorname{Re}(\overline{Z}_s)/k_0^2 h^3 \quad \text{Neper/m} \tag{28}$$

These formulas agree with those derived by Emslie et al (1975). It is worth noting that like the circular tunnel, the attenuation of the dominant modes is inversely proportional to the frequency squared and the linear dimensions cubed. Comparing (27) and (28), we infer that the vertically polarized mode suffers higher attenuation than the horizontally polarized mode for *w*>*h*. Thus, for a rectangular tunnel with *w*>*h*, the first horizontally polarized mode; TM_{x11} is the lowest attenuated mode.

Exercise 5: Use (26) to derive (27). In doing so, note that $\alpha = -\operatorname{Im}[(k_0^2 - k_{xm}^2 - k_{yn}^2)^{1/2}] \cong (1/2k_0)\operatorname{Im}[k_{xm}^2 + k_{yn}^2]$. This, of course, is valid only for low order modes such that $m\pi/\text{w}$ and $n\pi/h$ are $<<k_0$. Compute the attenuation rate of the TM_{y11} and TM_{x11} modes in a tunnel having w=2h=4.3 meters at 1 GHz. Take $\varepsilon_r=10$ and $\sigma=0$. [*13.27 and 2.95 dB/100m*]

We can infer from the above discussion that the attenuation caused by the walls which are perpendicular to the major electric field is much higher than that contributed by the walls parallel to the electric field.

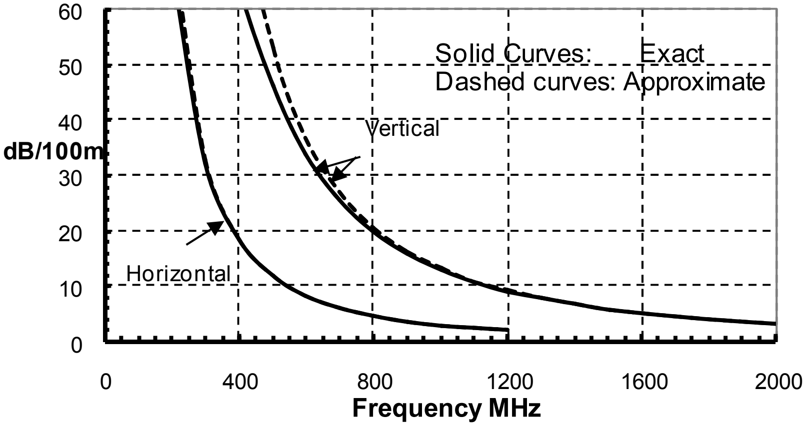


Fig. 5. Attenuation rates in dB/100m of VP and HP modes with m=n=1 in a rectangular tunnel of dimensions 4.3x2.15 m. ε_r=10.

The approximate attenuation rates given by (27-28) for the horizontally and vertically polarized (HP and VP) modes with m=n=1 are plotted versus the frequency in Figure 5. Here the tunnel dimensions are chosen as (w,h) = (4.3m, 2.15m) and ε_r=10. It is clear that the VP mode has considerably higher attenuation than its HP counterpart. The attenuation rates obtained by exact solution of equations (22) and (25) are also plotted for comparison. It is clear that both solutions coincide at the higher frequencies.

Ray theory:

When it is required to estimate the field at distances close to the source, the mode series becomes slowly convergent since it is necessary to include many higher order modes. As clear from the above argument, higher order modes are hard to analyze in a rectangular tunnel. In this case the ray series can be adopted for its fast convergence at short distances, say, of tens to few hundred meters from the source. At such distances, the rays are somewhat steeply incident on the walls, hence their reflection coefficients decrease quickly with ray order. Therefore, a small number of rays are needed for convergence.

A geometrical ray approach has been presented by (Mahmoud and Wait 1974a) where the field of a small linear dipole in a rectangular tunnel is obtained as a ray sum over a two-dimensional array of images. It is verified that small number of rays converges to the total field at sufficiently short range from the source. Conversely the number of rays required for convergence increase considerably in the far ranges, where only one or two modes give an accurate account of the field. The reader is referred to the above paper for a detailed discussion of ray theory in oversized waveguides.

5. Arched Tunnel

So far we have been studying tunnels with regular cross sections having either circular or rectangular shape. These shapes are amenable to analytical analysis that lead to full characterization of their main modes of propagation. However, most existing tunnels do not have regular cross sections and their study may require exhaustive numerical methods (Pingenot et al., 2006). In this section we consider cylindrical tunnels whose cross-section comprise a circular arch with a flat base as depicted in Figure 6. This can be considered as a circular tunnel whose shape is perturbed into a flat-based tunnel. So, we use the perturbation theory to predict attenuation and phase velocity of the dominant modes from those in a perfectly circular tunnel.

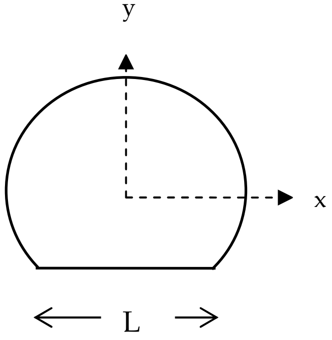


Fig. 6. An arched tunnel with radius *a* and flat base *L*.

5.1. Perturbation Analysis

We consider a cylindrical circular tunnel of radius '*a*' and cross section S_0 surrounded by a homogeneous earth of relative permittivity ε_r. Let us denote the vector fields of a given mode by $(\vec{E}_0, \vec{H}_0)\exp(-\gamma_0 z)$ where γ_0 is the longitudinal (along +z) propagation constant. Similarly, let $(\vec{E}, \vec{H})\exp(+\gamma z)$ be the vector fields of the corresponding mode in the perturbed tunnel of of area S (Figure 7). Note, however, that the mode is a backward mode; travelling in the (-z) direction. Both circular and perturbed tunnels have the same wall constant impedance Z and admittance Y. Now, we use Maxwell's equations that must be satisfied by both modal fields to get the reciprocity relation $\nabla.(\vec{E}_0 \times \vec{H} - \vec{E} \times \vec{H}_0) = 0$. Integrating over the infinitesimal volume between z and z+dz in the perturbed tunnel, we get, after some manipulations,

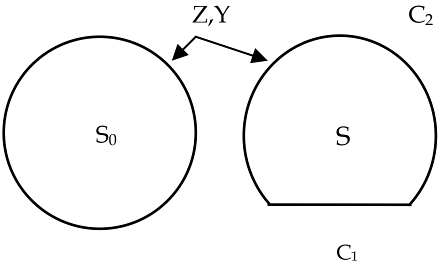


Fig. 7. A circular tunnel and a perturbed circular tunnel with a flat base. The walls are characterized by constant Z and Y.

$$\gamma - \gamma_0 = \frac{-\int_{C1} (\vec{E}_0 x\vec{H} - \vec{E} x\vec{H}_0).\hat{a}_n dC_1}{\iint_S (\vec{E}_0 x\vec{H} - \vec{E} x\vec{H}_0).\hat{a}_z dS} \tag{29}$$

where C_1 is the flat part of the cross- section contour, $\hat{a}_n$ is a unit vector along the outward normal to the wall (=- $\hat{a}_y$) and $\hat{a}_z$ is a unit axial vector. The integration in the denominator is taken over the cross section of the perturbed tunnel. In order to evaluate the numerator of (2), we use the constant wall impedance and admittance satisfied by the perturbed fields on the flat surface: $E_x = Z_s H_z$ and $H_x = -Y_s E_z$. Z_s and Y_s are given in (1-2). Using these relations, (29) reduces to:

$$\gamma - \gamma_0 = \frac{-2\int_{x=0}^{L/2} \left[E_{0x} H_z + E_z H_{0x} + Y_s E_z E_{0z} - Z_s H_z H_{0z} \right] dx}{\iint_S (\vec{E}_0 x\vec{H} - \vec{E} x\vec{H}_0).\hat{a}_z dS} \tag{30}$$

The integration in the numerator is taken over the flat surface of the perturbed tunnel. So far, the above result is rigorous, but cannot be used as such since the perturbed fields are not known. As a first approximation we can equate these fields to the backward mode fields in the un-perturbed (circular) tunnel. So we set: $H_z \approx H_{0z}, E_z \approx -E_{0z}$ in the numerator. In the denominator, the fields involved are the transverse fields (to z). So we use the approximations: $\vec{E} \mathrm{x} \hat{a}_z \approx \vec{E}_0 \mathrm{x} \hat{a}_z$ *and* $\vec{H} \mathrm{x} \hat{a}_z \approx -\vec{H}_0 \mathrm{x} \hat{a}_z$. Therefore (30) is approximated by:

$$\gamma - \gamma_0 \cong \frac{\int_{x=0}^{L/2} \left[E_{0x} H_{0z} - E_{0z} H_{0x} - Y_s E_{0z}^2 - Z_s H_{0z}^2 \right] dx}{\iint_S (\vec{E}_0 \mathrm{x} \vec{H}_0).\hat{a}_z dS} \tag{31}$$

Thus, for a given mode in the circular tunnel, (31) can be used to get the propagation constant of the perturbed mode in the corresponding perturbed tunnel. Of particular interest is the attenuation factor of the various modes. As a numerical example, we consider an arched tunnel of radius a=2meters with a flat base of width L. The surrounding earth has a relative permittivity ε_r=6. For an applied frequency f =500MHz, the modal attenuation factor, computed by (31), is plotted in Figure 8 for the perturbed TE_{01} and HE_{11} modes as a function of the L/a. Note that L/a=0 corresponds to a full circle and L/a=2 corresponds to a half circle. Generally the attenuation increases with L/a. The HE_{11} mode has two versions depending whether the polarization is horizontal (along x) or vertical (along y). Obviously the two modes are degenerate in a perfectly circular tunnel (L=0). However as L/a increases, this degeneracy breaks down in the perturbed tunnel. It is remarkable to see that the attenuation of the horizontally polarized HE_{11} mode becomes less than that of the vertically polarized mode in the perturbed tunnel. This agrees with measurements made by Molina et al. (2008).

It is interesting to study the effect of changing the frequency or the wall permittivity on the mode attenuation in the perturbed flat based tunnel. Further numerical results (not shown) indicate that the percentage increase of the attenuation relative to that in the circular tunnel is fairly weak on f and ε_r. Since the attenuation in an electrically large circular tunnel is inversely proportional to f^2, so will be the attenuation in the perturbed flat based tunnel.

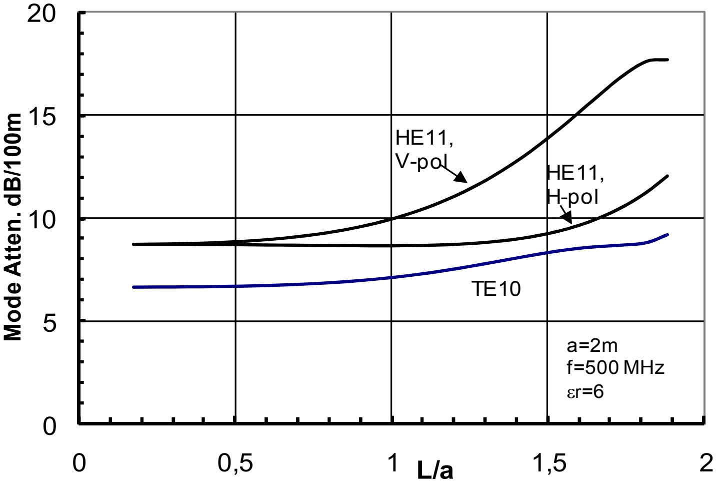


Fig. 8. Attenuation of the perturbed TE_{01} and HE_{11} modes in a flat based tunnel versus L/a. Note the difference between the attenuation of the VP and the HP versions of the HE_{11} mode.

5.2. An equivalent rectangular tunnel

As seen in Figure 8, the attenuation of the perturbed HE_{11} mode in the arched tunnel (with flat base) depends on the mode polarization; namely the vertically polarized HE_{11} mode is more attenuated than the horizontally polarized mode. The same observation is true for the HE_{11} mode in a rectangular tunnel whose height 'h' is less than its width 'w'. This raises the question whether we can model the flat based tunnel of Figure 6 by a rectangular tunnel. We will investigate this possibility in this section. To this end, let us start by comparing the attenuation of the HE_{nm} mode in tunnels with circular and a square cross sections. For the circular tunnel we have from (15)

$$\alpha\,|_{HEnm} = \frac{Z_s/\eta_0 + Y_s\eta_0}{2k_0^2a^3}\left[x_{n-1,m}\right]^2 \tag{32}$$

Where $x_{n-1,m}$ is the *m*th zero of the Bessel function $J_{n-1}(x)$. This formula is based on the condition: $k_0a >> x_{n-1,m}$. For the rectangular tunnel with width *'w'* and height *'h'* the attenuation of the HE_{nm} mode (with vertical polarization) is given by (27) which is repeated here

$$\alpha\,|_{HEnm} = \frac{2\pi^2}{k_0^2}\left[\frac{m^2Z_s/\eta_0}{w^3} + \frac{n^2Y_s\eta_0}{h^3}\right] \tag{33}$$

This is valid for electrically large tunnel, or when $k_0 >> m\pi/w$ and $n\pi/h$. Specializing this result for the HE_{11} mode in a square tunnel (*w*=*h* and *m*=*n*=*1*) , we get:

$$\alpha\,|_{HE11} = \frac{2\pi^2}{k_0^2w^3}\left[Z_s/\eta_0 + Y_s\eta_0\right] \tag{34}$$

Now compare the circular tunnel with the square tunnel for the HE_{11} mode. From (32) and (33), an equal attenuation occurs when

$$\mathrm{w} = \left(4\pi^2/2.4048^2\right)^{1/3} a = 1.897a \tag{33}$$

which means that the area of the equivalent square tunnel is equal to 1.145 times the area of the circular tunnel. This contrasts the work of (Dudley et.al, 2007) who adopted an equal area of tunnels. It is important to note that this equivalence is valid only for the HE_{11} mode in both tunnels; for other modes the attenuation in the circular and the square tunnels are generally not equal.

Now let us turn attention to the arched tunnel with flat base (Figure 6) for which we attempt to find an equivalent rectangular tunnel. We base this equivalence on equal attenuation of the HE_{11} mode in both tunnels. Let us maintain the ratio of areas as obtained from the square and circular tunnels; namely we fix the ratio of the equivalent rectangular area to the arched tunnel area to 1.145. Meanwhile we choose the ratio *h/w* equal to the arched tunnel height to its diameter. So, we write:

$$\begin{gathered} wh = 1.145[(\pi-\theta)a^2 + (La/2)\cos\theta], \text{ and} \\ h/w = (1+\cos\theta)/2 \end{gathered} \tag{34}$$

where $\theta = \mathrm{Arc}\sin(L/2a)$ is equal to half the angle subtended by the flat base L at the center of the circle.

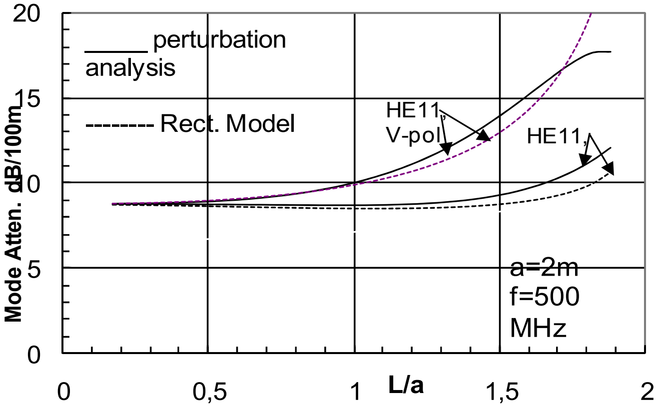


Fig. 9. Attenuation of the HE_{11} mode in the arched tunnel of Figure 6 using perturbation analysis and rectangular equivalent tunnel

Equation (34) defines the rectangular tunnel that is equivalent to the perturbed circular tunnel regarding the HE_{11} mode. In order to check the validity of this equivalence, we compare the estimated attenuation of the HE_{11} mode in the perturbed circular tunnel as obtained by perturbation analysis and by the equivalent rectangular tunnel in Figure 9. There is a reasonably close agreement between both methods of estimation for values of L/a between zero and ~1.82.

6. Curved Tunnel

Modal propagation in a curved rectangular tunnel has been considered by Mahmoud and Wait (1974b) and more recently by Mahmoud (2005). The model used is shown in Figure 10 where the curved surfaces coincide with $\rho=R-w/2$ and $\rho=R+w/2$ in a cylindrical frame (ρ,ϕ,z) with z parallel to the side walls. The tunnel is curved in the horizontal plane with assumed gentle curvature so that the mean radius of curvature R is $>>w$. The analysis is made in the high frequency regime so that $k_0w >>1$. The modes are nearly TE or TM to z with horizontal or vertical polarization respectively. The modal equations for the lower order TE_z and TM_z modes are derived in terms of the Airy functions and solved numerically for the propagation constant along the ϕ- direction. Numerical results are given in (Mahmoud, 2005) and are reproduced here in figures 11 and 12 for the dominant mode with vertical and horizontal polarization respectively. It is seen that wall curvature causes drastic increase of the attenuation especially for the horizontally polarized mode.
This can be explained by noting that the horizontal electric field is perpendicular to the curved walls, causing more attenuation to incur for this polarization.

Further study of the modal fields shows that these fields cling towards the outer curved wall casing increased losses in the wall. Besides, the mode velocity slows down.

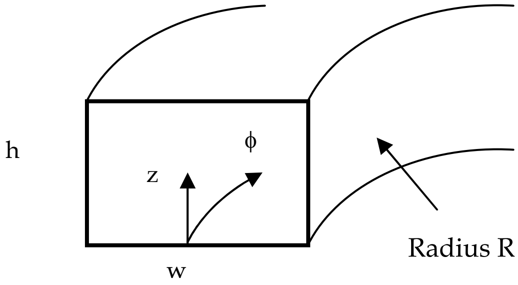


Fig. 10. A curved rectangular tunnel

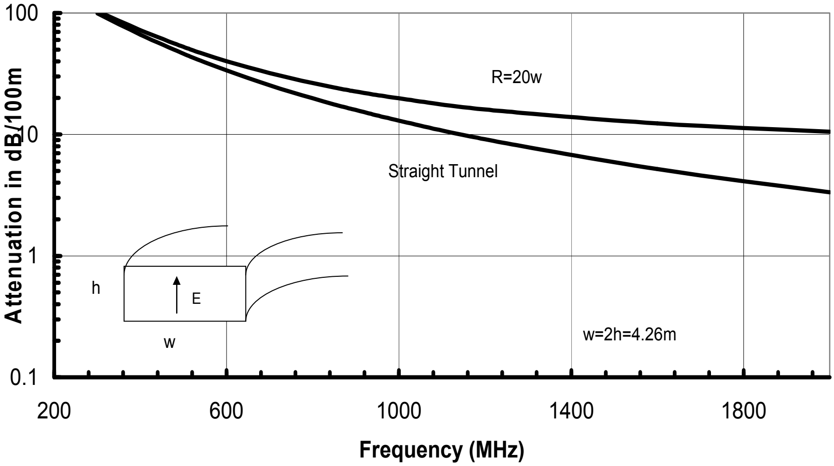


Fig. 11. Attenuation of TM_{y11}(VP) mode in a curved tunnel

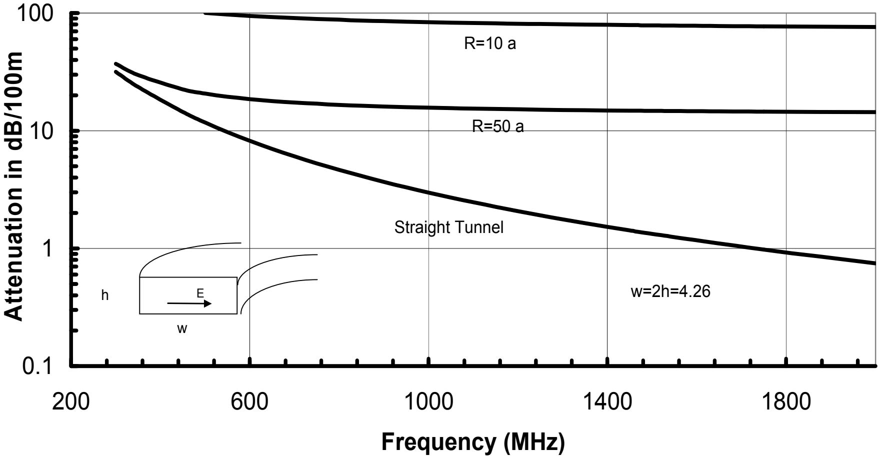


Fig. 12. Attenuation of TM_{x11} (HP) mode in a curved tunnel

7. Experimental work

Measurements of attenuation of the dominant mode in a straight rectangular mine tunnel were given by Goddard (1973). The tunnel cross section was *14x7* feet (or 4.26x2.13m) and the external medium had ε_r=10 and the attenuation was measured at 200, 450 and 1000 MHz. Emslie et al [25] compared these measurements with their theoretical values for the dominant horizontally polarized mode. Good agreement was observed at the first two frequencies, but the experimental values were considerably higher than the theoretical attenuation at the 1000 MHz. Similar trend has also been reported more recently by Lienard and Degauque (2000). The difference between the measured and theoretical attenuation at the 1000 MHz was attributed by Emslie et al.(1975) to slight tilt of the tunnel walls. Namely, by using a rather simple theory, it was shown that the increase of attenuation of the dominant mode due to wall tilt is proportional to the frequency and the square of the tilt angle. As a result, it was deduced that the high frequency attenuation of the dominant mode in a rectangular tunnel is governed mainly by the wall tilt.

Goddard (1973) has also measured the signal level around a corner and inside a crossed tunnel. The attenuation rate was very high for a short distance after which the attenuation approaches that of the dominant horizontally polarized mode. Emslie et al. (1975) have explained such behavior as follows. They argue that the crossed tunnel is excited by the higher order modes (or diffused waves in their terms) in the main tunnel. The modes excited in the crossed tunnel are mostly higher order modes with a small component of the dominant mode. These high order modes exhibit very high attenuation for a short distance after which the dominant mode becomes the sole propagating mode. So the signal level starts with a large attenuation rate which gradually decreases towards the attenuation rate of the dominant mode. The theory presented accordingly shows good agreement with measurements. More recently, Lee and Bertoni (2003) evaluated the modal coupling for tunnels or streets with L, T or cross junctions using hybrid ray-mode conversion. They argue that coupling occurs by rays diffracted at the corners into the side tunnels. It is found that the coupling loss is greatest at L-bends and least for cross junctions.

Chiba et.al (1975) have provided field measurements in one of the National Japanese Railway tunnels located in Tohoku. The tunnel cross section is an arch with a flat base as that depicted in Figure 6. The radius a=4.8 m, L=8.8m (L/a=1.83), the wall ε_r=5.5 and σ= 0.03 S/m. Field measurements were taken down the tunnel for different frequencies and polarizations. The attenuation of the dominant HE_{11} mode was then measured for both horizontal and vertical polarization at the frequencies 150, 470, 900, 1700, and 4000 MHz. We plot the predicted attenuation of the horizontally polarized HE_{11} mode in this same tunnel using both the perturbation analysis and the rectangular tunnel model in Figure 13. On top of these curves, the measured attenuation is shown as discrete dots at the above selected frequencies. The predicted attenuation shows the expected inverse frequency squared dependence. The measured attenuation follows the predicted attenuation except at the highest two frequencies (1700 and 4000 MHz) whence it is higher than predicted. This can be explained on account of wall roughness or micro-bending of the tunnel walls that affect the higher frequencies in particular.

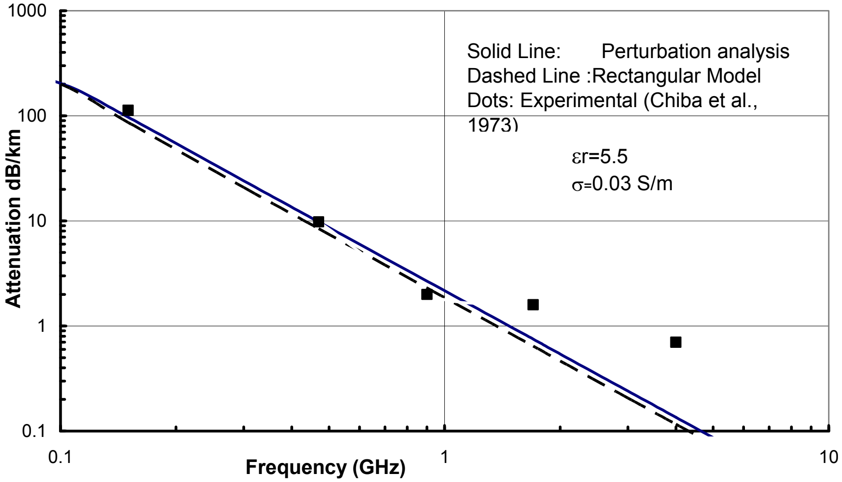


Fig. 13. Attenuation of the HE_{11} mode in the Japanese National Ralway tunnel by the perturbation analysis and the rectangular tunnel model versus measured values (as reported in (Chiba et al., 1973).

Measurements of the electric field down the Massif Central road tunnel south Central France have been taken by the research group in Lille University and the results are reported by Dudley et.al. (2007). The Massif Central tunnel has a flat based circular arch shape as in Figure 6 with radius a= 4.3 m and L=7.8 m; that is L/a= 1.81. The relative permittivity of the wall ε_r=5 and the conductivity σ= 0.01 S/m. The transmit and receive antennas were vertically polarized and the field measured down the tunnel at the frequencies 450 and 900 MHz are given in Figure 14. For the lower frequency, the field shows fast oscillatory behavior in the near zone, but at far distances from the source (greater than ~1800m), the field exhibits almost a constant rate of attenuation, which is that of the dominant HE_{11} (like) mode. We estimate the attenuation of this mode as 27.2 dB/km. At the 900 MHz frequency, there are two interfering modes that are observed in the range of 1500-2500m. One of these two modes must be the dominant HE_{11} mode. Some analysis is needed in this range that lead to an estimation of the attenuation of the HE_{11} mode, which we find as 6.8 dB/km.

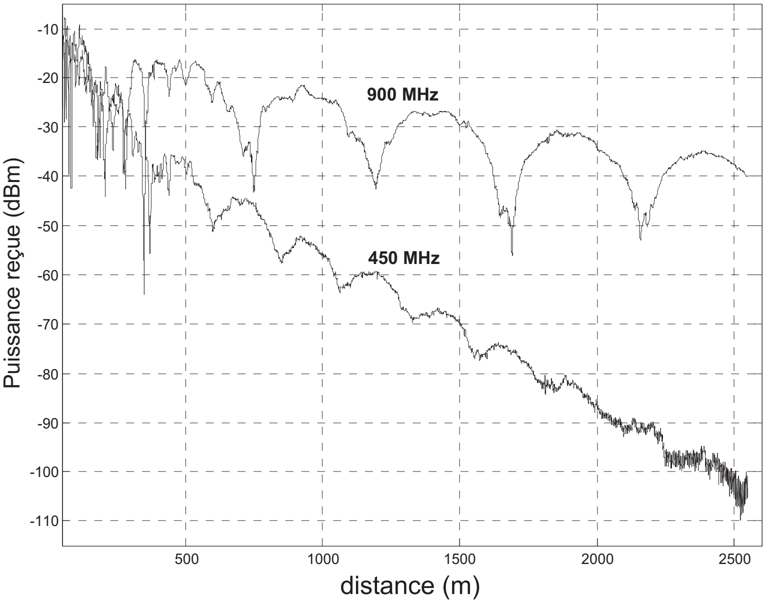


Fig. 14. Measured field down the Massif Central Tunnel in South France (Dudley et al., 2007) at 450 and 900 MHz.

A comparison between these measured attenuation rates and those predicted by the perturbation analysis or the equivalent rectangular tunnel (given in section 5) is made in Table 2. Good agreement is seen between predicted and measured attenuation although the measured values are slightly higher. This can be attributed to wall roughness and microbending.

	Perturbation Analysis (dB/km)	Equivalent Rectangular model	Measured Attenuation
f=450 MHz	22.0	24.1	27.2
f=900 MHz	5.52	6.03	6.8

Table 2. Measured versus predicted attenuation rates of the HE_{11} mode in the Massif Central Road Tunnel, South France.

8. Concluding discussion

We have presented an account of wireless transmission of electromagnetic waves in mine and road tunnels. Such tunnels act as oversized waveguides to UHF and the upper VHF waves. The theory of mode propagation in straight tunnels of circular, rectangular and arched cross sections has been covered and it is demonstrated that the dominant modes attenuate with rates that decrease with the applied frequency squared. We have also studied the increase of mode attenuation caused by tunnel curvature. Comparison of the theory with existing experimental measurements in real tunnels show good agreement except at the higher frequencies at which wall roughness, and microbending can increase signal loss over that predicted by the theory. While the higher order modes are highly attenuated and therefore contribute to signal loss, they can be beneficial in allowing the use of Multiple Input - Multiple Output (MIMO) technique to increase the channel capacity of tunnels. A detailed account of this important topic is found in (Lienard et al, 2003) and (Molina et al., 2008).

9. References

Andersen, J.B.; Berntsen, S. & Dalsgaard, P. (1975). Propagation in rectangular waveguides with arbitrary internal and external media, *IEEE Transaction on Microwave Theory and Technique,* MTT-23, No. 7, pp. 555-560.

Chiba, J.; Sato, J.R.; Inaba, T; Kuwamoto, Y.; Banno, O. & Sato, R. “ Radio communication in tunnels”, IEEE Transaction on Microwave Theory and Technique, MTT-26, No. 6, June 1978.

Dudley, D.G. (2004). Wireless Propagation in Circular Tunnels, *IEEE Transaction on Antennas and Propagation,* Vol. 53, n0.1, pp. 435-441.

Dudley, D.G. & Mahmoud, S.F. (2006). Linear source in a circular tunnel, *IEEE Transaction on Antennas and Propagation,* Vol. 54, n0.7, pp. 2034-2048.

Dudley, D.G., Martine Lienard, Samir F. Mahmoud and Pierre Degauque, (2007) “Wireless Propagation in Tunnels”, *IEEE Antenna and Propagation magazine*, Vol. 49, no. 2, pp. 11-26, April 2007.

Emslie, A.G.; Lagace, R.L. & Strong, P.F. (1973). Theory of the propagation of UHF radio waves in coal mine tunnels, *Proc. Through the Earth Electromagnetics Workshop*, Colorado School of mines, Golden, Colorado, Aug. 15-17.

Emslie, A.G.; Lagace R.L. & Strong, P.F. (1975). Theory of the propagation of UHF radio waves in coal mine tunnels, *IEEE Transaction on Antenn. Propagat.*, Vol. AP-23, No. 2, pp. 192-205.

Glaser, J.I. (1967). Low loss waves in hollow dielectric tubes, *Ph.D. Thesis,* M.I.T.

Glaser, J.I. (1969). Attenuation and guidance of modes in hollow dielectric waveguides, *IEEE Trans. Microwave Theory and Tech*, Vol. MTT-17, pp.173-174.

Goddard, A.E. (1973). Radio propagation measurements in coal mines at UHF and VHF, *Proc. Through the Earth Electromagnetics Workshop*, Colorado School of mines, Golden, Colorado, Aug. 15-17.

Lee, J. & Bertoni, H.L. (2003). Coupling at cross, T and L junction in tunnels and urban street canyons, *IEEE Transaction on Antenn. Propagat.*, Vol. AP-51, No. 5, pp. 192-205, pp.926-935.

Lienard, M. & Degauque, P. (2000). Natural wave propagation in mine environment, *IEEE Transaction on Antennas & propagate,* Vol-AP-48, No.9, pp.1326-1339.

Lienard, M.; Degauque, P.; Baudet, J. & Degardin, D. (2003). Investigation on MIMO Channels in Subway Tunnels, IEEE Journal on Selected Areas in Communication, Vol. 21, No. 3, pp.332-339.

Mahmoud, S.F. & Wait, J.R. (1974a). Geometrical optical approach for electromagnetic wave propagation in rectangular mine tunnels, *Radio Science* , Vol. 9, no. 12, pp. 1147-1158.

Mahmoud, S.F. & Wait, J.R. (1974a). Guided electromagnetic waves in a curved rectangular mine tunnel, *Radio Science*, May 1974, pp. 567-572.

Mahmoud, S.F. (1991) *Electromagnetic Waveguides; theory and applications*, IEE electromagnetic waves series 32, Peter peregrinus Ltd. on behalf of IEE.

Mahmoud, S.F. (2005). Modal propagation of high frequency e.m. waves in straight and curved tunnels within the earth, *Journal of electromagnetic wave applications* (JEMWA), Vol. 19, No. 12, pp. 1611-1627.

Marcatili, E.A.J. & Schmeltzer, R.A. (1964). Hollow metallic and dielectric waveguides for long distsnce optical transmission and lasers, *The Bell System Technical Journal*, Vol. 43, pp. 1783-1809, July 1964.

Molina-Garcia-Padro, J.M.; Lienard, M. , Nasr, A. & Degauque, P. (2008). On the possibility of interpreting field variations and polarization in arched tunnels using a model of propagation in rectangular or circular tunnels, *IEEE Transaction on Antenn. Propagat.*, Vol. AP-56, No. 4, pp. 1206-1211.

Molina-Garcia-Pardo, J. M.; Liénard, M.; Degauque, P.; Dudley, D.G. & and Juan-Llacer, M.(2008). MIMO Channel Characteristics in Rectangular Tunnels from Modal Theory, *IEEE Transactions on Vehicular Technology*, Vol 57, No. 3, pp.1974-1979.

Pingenot, J.; Rieben, R.N.; White, D.A.; & Dudley, D.G. (2006). Full wave analysis of RF signal attenuation in a lossy rough surface cave using high order Time Domain vector Finite Element method, *Journal of Electromagnetic Wave Applications,* Vol.20, no. 12, pp. 1695-1705.

Stratton, J.A. (1941). *Electromagnetic Theory*, McGraw Hill, New York.

Wait, J.R (1967). A fundamental difficulty in the analysis of cylindrical waveguides with impedance walls, *Electronic Letters*, Vol. 3, No. 2, pp. 87-88.

Wait, J.R. (1980). Propagation in rectangular tunnel with imperfectly conducting walls, *Electronic Letters*, Vol. 16, No. 13, pp. 521-522.

10. Appendix

The transverse fields of a hybrid mode in a circular tunnel are obtained from the axial fields by (Mahmoud, 1991, p. 7):

$$\begin{aligned} \vec{E}_t &= \frac{-j\beta}{k_\rho^2}\nabla_t E_z - \frac{j\omega\mu_0}{k_\rho^2}\nabla_t E_z \mathrm{x}\, \hat{z} \\ \vec{H}_t &= \frac{-j\beta}{k_\rho^2}\nabla_t H_z - \frac{j\omega\varepsilon_0}{k_\rho^2}\,\hat{\mathrm{z}}\,\mathrm{x}\nabla_t E_z \end{aligned} \tag{A1}$$

Where ∇_t is the transverse differential operator: $\nabla_t = (\partial / \partial\rho)\,\hat{\rho} + (\partial / \rho\partial\varphi)\hat{\varphi}$.

Using the axial fields in (5) we get the explicit expressions:

$$\begin{aligned} E_\rho &= \frac{k_0}{k_\rho}\left[(\beta / k_0)\, J_n'(k_\rho\rho) + \Lambda \frac{nJ_n(k_\rho\rho)}{k_\rho\rho} \right] \sin n\varphi e^{-j\beta z} \\ E_\varphi &= \frac{k_0}{k_\rho}\left[(\beta / k_0) \frac{nJ_n(k_\rho\rho)}{k_\rho\rho} + \Lambda J_n'(k_\rho\rho) \right] \cos n\varphi e^{-j\beta z} \end{aligned} \tag{A2}$$

And

$$\eta_0 H_\rho = \frac{k_0}{k_\rho}\left[-(\beta\Lambda/k_0)J_n'(k_\rho\rho) - \frac{nJ_n(k_\rho\rho)}{k_\rho\rho}\right]\cos n\varphi e^{-j\beta z}$$
$$\eta_0 H_\varphi = \frac{k_0}{k_\rho}\left[(\beta\Lambda/k_0)\frac{nJ_n(k_\rho\rho)}{k_\rho\rho} + J_n'(k_\rho\rho)\right]\sin n\varphi e^{-j\beta z} \tag{A3}$$

It is useful to express the transverse field in terms of Cartesian coordinates (x,y). Using the Bessel function identities:

$$nJ_n(u)/u = (J_{n-1}(u) + J_{n+1}(u))/2$$
$$J_n'(u) = (J_{n-1}(u) - J_{n+1}(u))/2 \tag{A4}$$

we get:

$$\frac{2k_\rho}{k_0}E_x = [\beta/k_0 + \Lambda]J_{n-1}(k_\rho\rho)\sin(n-1)\varphi + [-\beta/k_0 + \Lambda]J_{n+1}(k_\rho\rho)\sin(n+1)\varphi$$
$$\frac{2k_\rho}{k_0}E_y = [\beta/k_0 + \Lambda]J_{n-1}(k_\rho\rho)\cos(n-1)\varphi + [\beta/k_0 - \Lambda]J_{n+1}(k_\rho\rho)\cos(n+1)\varphi \tag{A5}$$

In the special case n=1, (A5) reduces to:

$$\frac{2k_\rho}{k_0}E_x = [-\beta/k_0 + \Lambda]J_2(k_\rho\rho)\sin 2\varphi$$
$$\frac{2k_\rho}{k_0}E_y = [\beta/k_0 + \Lambda]J_0(k_\rho\rho) + [\beta/k_0 - \Lambda]J_2(k_\rho\rho)\cos 2\varphi \tag{A6}$$

This shows that for the HE_{11} mode ($\Lambda \sim +1$), the field is almost y-polarized.
The corresponding magnetic field is

$$\frac{2k_\rho}{k_0}H_y = [\beta\Lambda/k_0 - 1]J_2(k_\rho\rho)\sin 2\varphi$$
$$\frac{2k_\rho}{k_0}H_x = -[\beta\Lambda/k_0 + 1]J_0(k_\rho\rho) + [\beta\Lambda/k_0 - 1]J_2(k_\rho\rho)\cos 2\varphi \tag{A7}$$

Equations (A6-A7) give the transverse fields of the HE_{1m} modes.

2

Wireless Communications and Multitaper Analysis: Applications to Channel Modelling and Estimation

Sahar Javaher Haghighi and Serguei Primak
Department of Electrical and Computer Engineering,
The University of Western Ontario
London, Ontario, Canada

Valeri Kontorovich
CINVESTAV
Mexico

Ervin Sejdić
Bloorview Research Institute
Institute of Biomaterials and Biomedical Engineering
University of Toronto
Toronto, Ontario, Canada

1. Introduction

The goal of this Chapter is to review the applications of the Thomson Multitaper analysis (Percival and Walden; 1993b), (Thomson; 1982) for problems encountered in communications (Thomson; 1998; Stoica and Sundin; 1999). In particular we will focus on issues related to channel modelling, estimation and prediction.

Sum of Sinusoids (SoS) or Sum of Cisoids (SoC) simulators (Patzold; 2002; SCM Editors; 2006) are popular ways of building channel simulators both in SISO and MIMO case. However, this approach is not a very good option when features of communications systems such as prediction and estimation are to be simulated. Indeed, representation of signals as a sum of coherent components with large prediction horizon (Papoulis; 1991) leads to overly optimistic results. In this Chapter we develop an approach which allows one to avoid this difficulty. The proposed simulator combines a representation of the scattering environment advocated in (SCM Editors; 2006; Almers et al.; 2006; Molisch et al.; 2006; Asplund et al.; 2006; Molish; 2004) and the approach for a single cluster environment used in (Fechtel; 1993; Alcocer et al.; 2005; Kontorovich et al.; 2008) with some important modifications (Yip and Ng; 1997; Xiao et al.; 2005).

The problem of estimation and interpolation of a moderately fast fading Rayleigh/Rice channel is important in modern communications. The Wiener filter provides the optimum solution when the channel characteristics are known (van Trees; 2001). However, in real-life applications basis expansions such as Fourier bases and discrete prolate spheroidal sequences (DPSS) have been adopted for such problems (Zemen and Mecklenbräuker; 2005; Alcocer-Ochoa et al.; 2006). If the bases and the channel under investigation occupy the same band, accurate and sparse representations of channels are usually obtained (Zemen and Mecklenbräuker; 2005). However, a larger number of bases is required to approximate the channel with the same accuracy when the bandwidth of the basis function is mismatched and larger than that of the signal. A bank of bases with different bandwidths can be used to resolve this particular problem (Zemen et al.; 2005). However, such a representation again ignores the fact that in some cases the band occupied by the channel is not necessarily centered around DC, but rather at some frequency different from zero. Hence, a larger number of bases is again needed for accurate and sparse representation. A need clearly exists for some type of overcomplete, redundant bases which accounts for a variety of scenarios. A recently proposed overcomplete set of bases called Modulated Discrete Prolate Spheroidal Sequences (MDPSS) (Sejdić et al.; 2008) resolves the aforementioned issues. The bases within the frame are obtained by modulation and variation of the bandwidth of DPSSs in such a way as to reflect various scattering scenarios.

2. Theoretical Background

2.1 Discrete Prolate Spheroidal Sequences

The technique used here was introduced first in (Thomson; 1982) and further discussed in (Percival and Walden; 1993b). We adopt notations used in the original paper (Thomson; 1982). Let $x(n)$ be a finite duration segment of a stationary process, $n = 0, \cdots, N-1$. It can be represented as

$$x(n) = \int_{-1/2}^{1/2} \exp\left(j2\pi f[n-(N-1)/2]\right) dZ(f) \tag{1}$$

according to the Cramer theorem (Papoulis; 1991). It is emphasized in (Thomson; 1982) that goal of the spectral analysis is to estimate moments of $Z(f)$, in particular its first and second moments from a finite sample $x(n)$. However, spectral analysis is often reduced only to considering the second centered moment

$$S(f)df = \mathcal{E}\left\{|dZ(f)|^2\right\} \tag{2}$$

well known as the spectrum (power spectral density). In the case of continuous spectrum the first moment of $dZ(f)$ is zero and it is usually not considered. However, in the case of a line or mixed spectrum one obtains

$$\mathcal{E}\left\{dZ(f)\right\} = \sum_k \mu_k \delta(f - f_k) df \tag{3}$$

where μ_k is the amplitude of the harmonic with frequency f_k.

The Discrete prolate spheroidal sequences (DPSS) are defined as solutions to the Toeplitz matrix eigenvalue problem (Thomson; 1982; Slepian; 1978)

$$\lambda_k(N,W)u_k(N,W;n) = \sum_{m=0}^{N-1} \frac{\sin\left(2W(n-m)\right)}{\pi(n-m)} u_k(N,W;m) \tag{4}$$

Their discrete Fourier transform (DFT) is known as Discrete Prolate Spheroidal Wave Functions (DPSWF) (Slepian; 1978)

$$U_k(N,W;f) = \sum_{n=0}^{N-1} u_k(N,W;n) \exp(-j2\pi n f) \tag{5}$$

In particular, if $f = 0$, equation (5) can be rewritten as

$$U_k(N,W;0) = U_k(0) = \sum_{n=0}^{N-1} u_k(N,W;n) \tag{6}$$

The DPSS are doubly orthogonal, that is, they are orthogonal on the infinite set $\{-\infty,...,\infty\}$ and orthonormal on the finite set $\{0,1,...,N-1\}$, that is,

$$\sum_{-\infty}^{\infty} v_i(n,N,W)v_j(n,N,W) = \lambda_i \delta_{ij} \tag{7}$$

$$\sum_{n=0}^{N-1} v_i(n,N,W)v_j(n,N,W) = \delta_{ij}, \tag{8}$$

where $i,j = 0,1,...,N-1$.
A number of spectral estimates, called eigen coefficients, are obtained using DPSS $u_k(N,W;n)$ as time-domain windows

$$Y_k(f) = \sum_{n=1}^{N-1} x(n)u_k(N,W;n)\exp(-j2\pi nf) \tag{9}$$

Since for a single line spectral component at $f = f_0$

$$\mathcal{E}\{Y_k(f)\} = \mu U_k(N,W;f-f_0) + \mu^* U_k(N,W;f+f_0) \tag{10}$$

and assuming that $2f_0 > W$, one obtains a simple approximation

$$\mathcal{E}\{Y_k(f_0)\} = \mu U_k(N,W;0) + \mu^* U_k(N,W;2f_0) \approx \mu U_k(N,W;0) \tag{11}$$

since $U_k(N,W;f)$ are maximally concentrated around $f = 0$ (Slepian; 1978). Complex magnitude μ could be estimated by minimizing error between the eigen spectrum $Y_k(f)$ and $\mu(f)U_k(N,W;0)$. The result of such minimization is a simple linear regression (Papoulis; 1991) of $Y_k(f)$ on $U_k(N,W;0)$ and is given by

$$\hat{\mu}(f) = \frac{1}{\sum_{k=0}^{K-1} |U_k(0)|^2} \sum_{k=0}^{K-1} U_k^*(0)Y_k(f) \tag{12}$$

The fact that the estimate $\hat{\mu}(f)$ is the linear regression allows to use standard regression F-test (Conover; 1998) for significance of the line component with amplitude $\hat{\mu}(f)$ at frequency f. This could be achieved by comparing the ration (Thomson; 1982)

$$F(f) = (K-1)\frac{|\hat{\mu}(f)|^2}{\epsilon^2(\hat{\mu},f)} \sum_{k=0}^{K-1} |U_k^*(0)|^2 \tag{13}$$

The location of the maximum (or local maxima) of $F(f)$ provides an estimate of the line component of the spectrum. The hypothesis that there is a line component with magnitude $\hat{\mu}(f_0)$ at $f = f_0$ is accepted if $F(f)$ has maximum at $f = f_0$ and

$$F(f_0) > F_\alpha(K) \tag{14}$$

where $F_\alpha(K)$ is the threshold for significance level α and $K-1$ degrees of freedom. Values of $F_\alpha(K)$ can be found in standard books on statistics (Conover; 1998).

Estimation of spectrum in the vicinity of a line spectrum component at $f = f_0$ is done according to the following equation

$$\hat{S}(f) = \frac{1}{K}\sum_{k=0}^{K-1} |Y_k(f) - \hat{\mu}V_k(f-f_0)|^2 \tag{15}$$

and

$$\hat{S}(f) = \frac{1}{K}\sum_{k=0}^{K-1} |Y_k(f)|^2 \tag{16}$$

otherwise.

It is recommended that the original sequence $x(n)$ is zero-padded to create a mesh of length $4-10$ times longer than the original N. This is essential to avoid missing a line spectrum component which is far from the grid frequency (Rao and Hamed; 2003).

2.2 Physical Model of Mobile-to-Mobile Channel

The propagation of electromagnetic waves in urban environments is a very complicated phenomenon (Beckmann and Spizzichino; 1963; Bertoni; 2000). The received waves are usually a combination of Line of Site (LoS) and a number of specular and diffusive components. Mobile-to-mobile communications (Sen and Matolak; 2008) introduce a new geometry of radio wave propagation, especially in urban settings. In such settings, antennas are located on a level well below rooftops or even tree-tops. Therefore propagation is dominated by rays reflected and diffracted from buildings, trees, other cars, etc. as in Fig. 1. Grazing angles are also often very small in such scenarios, therefore reflective surfaces cannot always be treated as smooth, resulting into LoS-like specular component, nor as very rough, resulting in a purely diffusive component (Beckmann and Spizzichino; 1963). It is also common to assume that, the resulting superposition of multiple reflected and diffused rays results into a spherically symmetric random process (Schreier and Scharf; 2003). However, it is shown in (Beckmann and Spizzichino; 1963), that there is an intermediate case which results in a partially coherent reflection, and, as such to improper complex Gaussian process, representing channel transfer function. When a number of reflected waves is not sufficient, the resulting distribution is highly non-Gaussian (Barakat; 1986). We defer investigation of such cases to future work and existing literature (Barakat; 1986), (Jakeman and Tough; 1987). Here we focus on the origin of the four-parametric distribution (Klovski; 1982) and estimation of its parameters.

2.2.1 Scattering from rough surfaces

Let us consider a rough surface of extent $2L \gg \lambda$ of the first Fresnel zone which is illuminated by a plane wave at the incidence angle $\theta_i = \pi/2 - \gamma$ as shown in Figs. 1-2.

We assume a simple Gaussian model of the surface roughness (Beckmann and Spizzichino; 1963), which is described by a random deviation $\zeta(x)$ from the mean level. The process $\zeta(x)$ has variance σ_r^2, the spatial covariance function $C(\Delta x) = \sigma_r^2 c(\Delta x)$, and the correlation length X (Beckmann and Spizzichino; 1963).

Let us consider first reflection from a surface portion of length X, equal to the correlation length of the roughness as shown in Fig. 2. Specular direction phases of elementary waves in specular direction have a random component

$$\eta_\phi = \frac{4\pi}{\lambda}\zeta(x)\cos\theta_i \tag{17}$$

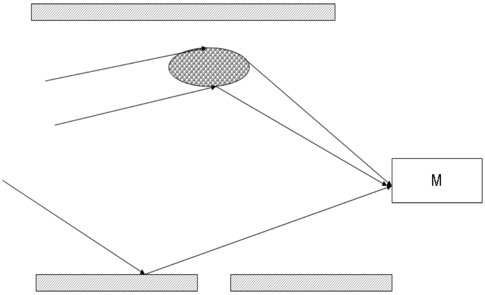


Fig. 1. Mobile-to-mobile propagation scenario. In addition to LoS and diffusive components (not shown) there are specular reflections from rough surfaces such as building facades and trees.

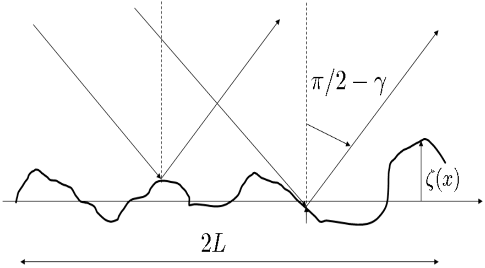


Fig. 2. Rough surface geometry. Size of the patch $2L$ corresponds to the size of the first Fresnel zone. Rough surface is described by a random process $\zeta(x)$.

Thus, the variance σ_ϕ^2 of the random phase deviation could be evaluated as

$$\sigma_\phi^2 = 16\pi^2 \frac{\sigma_r^2}{\lambda^2} \cos^2\theta_i \tag{18}$$

If $\sigma_\phi^2 \gg 4\pi^2$, *i.e.*

$$g = 2\frac{\sigma_r}{\lambda}\cos\theta_i \gg 1 \tag{19}$$

then the variation of phase is significantly larger then 2π. The distribution of the wrapped phase (Mardia and Jupp; 2000) is approximately uniform and the resulting wave could be considered as a purely diffusive component. However, in the opposite case of $0 < g \ll 1$ the variation of phase is significantly less then 2π and cannot be considered uniform. For a perfectly smooth surface $g = 0$ and the phase is deterministic, similar to LoS.

If the first Fresnel zone has extent $2L$, then there is approximately $N = 2L/X$ independent sections of the rough surface patches which contribute independently to the resultant field. Therefore, one can assume the following model for the reflected field/signal in the specular direction

$$\xi = \sum_{n=1}^{N} A_n \exp(j\phi_n) \tag{20}$$

where ϕ_n is a randomly distributed phase with the variance given by equation (18). If $\sigma_\phi^2 \gg 4\pi^2$ the model reduces to well accepted spherically symmetric diffusion component model; if $\sigma_\phi^2 = 0$, LoS-like conditions for specular component are observed with the rest of the values spanning an intermediate scenario.

Detailed investigation of statistical properties of the model, given by equation (20), can be found in (Beckmann and Spizzichino; 1963) and some consequent publications, especially in the field of optics (Barakat; 1986), (Jakeman and Tough; 1987). Assuming that the Central Limit Theorem holds, as in (Beckmann and Spizzichino; 1963), one comes to conclusion that $\xi = \xi_I + j\xi_Q$ is a Gaussian process with zero mean and *unequal* variances σ_I^2 and σ_Q^2 of the real and imaginary parts. Therefore ξ is an improper random process (Schreier and Scharf; 2003). Coupled with a constant term $m = m_I + jm_Q$ from the LoS type components, the model (20) gives rise to a large number of different distributions of the channel magnitude, including Rayleigh ($m = 0$, $\sigma_I = \sigma_Q$), Rice ($m \neq 0$, $\sigma_I = \sigma_Q$), Hoyt ($m \neq 0$, $\sigma_I > 0$ $\sigma_Q = 0$) and many others (Klovski; 1982), (Simon and Alouini; 2000). Following (Klovski; 1982) we will refer to the general case as a four-parametric distribution, defined by the following parameters

$$m = \sqrt{m_I^2 + m_Q^2}, \; \phi = \arctan \frac{m_Q}{m_I} \tag{21}$$

$$q^2 = \frac{m_I^2 + m_Q^2}{\sigma_I^2 + \sigma_Q^2}, \; \beta = \frac{\sigma_Q^2}{\sigma_I^2} \tag{22}$$

Two parameters, q^2 and β, are the most fundamental since they describe power ration between the deterministic and stochastic components (q^2) and asymmetry of the components (β). The further study is focused on these two parameters.

2.2.2 Channel matrix model

Let us consider a MIMO channel which is formed by N_T transmit and N_R received antennas. The $N_R \times N_T$ channel matrix

$$\mathbf{H} = \mathbf{H}_{LoS} + \mathbf{H}_{diff} + \mathbf{H}_{sp} \tag{23}$$

can be decomposed into three components. Line of sight component $\mathbf{H}_{LoS}$ could be represented as

$$\mathbf{H}_{LoS} = \sqrt{\frac{P_{LoS}}{N_T N_R}} \mathbf{a}_L \mathbf{b}_L^H \exp(j\phi_{LoS}) \tag{24}$$

Here P_{LoS} is power carried by LoS component, $\mathbf{a}_L$ and $\mathbf{b}_L$ are receive and transmit antenna manifolds (van Trees; 2002) and ϕ_{LoS} is a deterministic constant phase. Elements of both manifold vectors have unity amplitudes and describe phase shifts in each antenna with respect to some reference point[1]. Elements of the matrix $\mathbf{H}_{diff}$ are assumed to be drawn from proper (spherically-symmetric) complex Gaussian random variables with zero mean and correlation between its elements, imposed by the joint distribution of angles of arrival and departure (Almers et al.; 2006). This is due to the assumption that the diffusion component is composed of a large number of waves with independent and uniformly distributed phases due to large and rough scattering surfaces. Both LoS and diffusive components are well studied in the literature. Combination of the two lead to well known Rice model of MIMO channels (Almers et al.; 2006).

---

[1] This is not true when the elements of the antenna arrays are not identical or different polarizations are used.

Proper statistical interpretation of specular component $\mathbf{H}_{sp}$ is much less developed in MIMO literature, despite its applications in optics and random surface scattering (Beckmann and Spizzichino; 1963). The specular components represent an intermediate case between LoS and a purely diffusive component. Formation of such a component is often caused by mild roughness, therefore the phases of different partial waves have either strongly correlated phases or non-uniform phases.

In order to model contribution of specular components to the MIMO channel transfer function we consider first a contribution from a single specular component. Such a contribution could be easily written in the following form

$$\mathbf{H}_{sp} = \sqrt{\frac{P_{sp}}{N_T N_R}} \left[\mathbf{a} \odot \mathbf{w}_a\right] \left[\mathbf{b} \odot \mathbf{w}_b\right]^H \xi \tag{25}$$

Here P_{sp} is power of the specular component, $\xi = \xi_R + j\xi_I$ is a random variable drawn according to equation (20) from a complex Gaussian distribution with parameters $m_I + jm_Q$, σ_I^2, σ_Q^2 and independent in-phase and quadrature components. Since specular reflection from a moderately rough or very rough surface allows reflected waves to be radiated from the first Fresnel zone it appears as a signal with some angular spread. This is reflected by the window terms $\mathbf{w}_a$ and $\mathbf{w}_b$ (van Trees; 2002; Primak and Sejdić; 2008). It is shown in (Primak and Sejdić; 2008) that it could be well approximated by so called discrete prolate spheroidal sequences (DPSS) (Percival and Walden; 1993b) or by a Kaiser window (van Trees; 2002; Percival and Walden; 1993b). If there are multiple specular components, formed by different reflective rough surfaces, such as in an urban canyon in Fig. 1, the resulting specular component is a weighted sum of (25) like terms defined for different angles of arrival and departures:

$$\mathbf{H}_{sp} = \sum_{k=1} \sqrt{\frac{P_{sp,k}}{N_T N_R}} \left[\mathbf{a}_k \odot \mathbf{w}_{a,k}\right] \left[\mathbf{b}_k \odot \mathbf{w}_{b,k}\right]^H \xi_k \tag{26}$$

It is important to mention that in the mixture (26), unlike the LoS component, the absolute value of the mean term is not the same for different elements of the matrix $\mathbf{H}_{sp}$. Therefore, it is not possible to model them as identically distributed random variables. Their parameters (mean values) also have to be estimated individually. However, if the angular spread of each specular component is very narrow, the windows $\mathbf{w}_{a,k}$ and $\mathbf{w}_{b,k}$ could be assumed to have only unity elements. In this case, variances of the in-phase and quadrature components of all elements of matrix $\mathbf{H}_{sp}$ are the same.

3. MDPSS wideband simulator of Mobile-to-Mobile Channel

There are different ways of describing statistical properties of wide-band time-variant MIMO channels and their simulation. The most generic and abstract way is to utilize the time varying impulse response $\mathbf{H}(\tau, t)$ or the time-varying transfer function $\mathbf{H}(\omega, t)$ (Jeruchim et al.; 2000), (Almers et al.; 2006). Such description does not require detailed knowledge of the actual channel geometry and is often available from measurements. It also could be directly used in simulations (Jeruchim et al.; 2000). However, it does not provide good insight into the effects of the channel geometry on characteristics such as channel capacity, predictability, *etc.*. In addition such representations combine propagation environment with antenna characteristics into a single object.

An alternative approach, based on describing the propagation environment as a collection of scattering clusters is advocated in a number of recent publications and standards (Almers et al.; 2006; Asplund et al.; 2006). Such an approach gives rise to a family of so called Sum-Of-Sinusoids (SoS) simulators.

Sum of Sinusoids (SoS) or Sum of Cisoids (SoC) simulators (Patzold; 2002; SCM Editors; 2006) is a popular way of building channel simulators both in SISO and MIMO cases. However, this approach is not a very good option when prediction is considered since it represents a signal as a sum of coherent components with large prediction horizon (Papoulis; 1991). In addition it is recommended that up to 10 sinusoids are used per cluster. In this communication we develop a novel approach which allows one to avoid this difficulty. The idea of a simulator combines representation of the scattering environment advocated in (SCM Editors; 2006; Almers et al.; 2006; Molisch et al.; 2006; Asplund et al.; 2006; Molish; 2004) and the approach for a single cluster environment used in (Fechtel; 1993; Alcocer et al.; 2005) with some important modifications (Yip and Ng; 1997; Xiao et al.; 2005).

3.1 Single Cluster Simulator

3.1.1 Geometry of the problem

Let us first consider a single cluster scattering environment, shown in Fig. 2. It is assumed that both sides of the link are equipped with multielement linear array antennas and both are mobile. The transmit array has N_T isotropic elements separated by distance d_T while the receive side has N_R antennas separated by distance d_R. Both antennas are assumed to be in the horizontal plane; however extension on the general case is straightforward. The antennas are moving with velocities v_T and v_R respectively such that the angle between corresponding broadside vectors and the velocity vectors are α_T and α_R. Furthermore, it is assumed that the impulse response $H(\tau, t)$ is sampled at the rate F_{st}, *i.e.* $\tau = n/F_{st}$ and the channel is sounded with the rate F_s impulse responses per second, *i.e.* $t = m/F_s$. The carrier frequency is f_0. Practical values will be given in Section 4.

The space between the antennas consist of a single scattering cluster whose center is seen at the the azimuth ϕ_{0T} and co-elevation θ_T from the receiver side and the azimuth ϕ_{0R} and co-elevation θ_R. The angular spread in the azimuthal plane is $\Delta\phi_T$ on the receiver side and $\Delta\phi_R$ on the transmit side. No spread is assumed in the co-elevation dimension to simplify calculations due to a low array sensitivity to the co-elevation spread. We also assume that $\theta_R = \theta_T = \pi/2$ to shorten equations. Corresponding corrections are rather trivial and are omitted here to save space. The angular spread on both sides is assumed to be small comparing to the angular resolution of the arrays due to a large distance between the antennas and the scatterer (van Trees; 2002):

$$\Delta\phi_T \ll \frac{2\pi\lambda}{(N_T-1)d_T}, \Delta\phi_R \ll \frac{2\pi\lambda}{(N_R-1)d_R}. \tag{27}$$

The cluster also assumed to produce certain delay spread variation, $\Delta\tau$, of the impulse response due to its finite dimension. This spread is assumed to be relatively small, not exceeding a few sampling intervals $T_s = 1/F_{st}$.

3.1.2 Statistical description

It is well known that the angular spread (dispersion) in the impulse response leads to spatial selectivity (Fleury; 2000) which could be described by corresponding covariance function

$$\rho(d) = \int_{-\pi}^{\pi} \exp\left(j2\pi\frac{d}{\lambda}\phi\right) p(\phi)d\phi \tag{28}$$

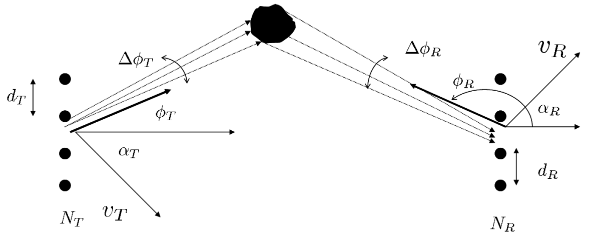


Fig. 3. Geometry of a single cluster problem.

where $p(\phi)$ is the distribution of the AoA or AoD. Since the angular size of clusters is assumed to be much smaller that the antenna angular resolution, one can further assume the following simplifications: a) the distribution of AoA/AoD is uniform and b) the joint distribution $p_2(\phi_T, \phi_R)$ of AoA/AoD is given by

$$p_2(\phi_T, \phi_R) = p_{\phi_T}(\phi_T) p_{\phi_R}(\phi_R) = \frac{1}{\Delta_{\phi_T}} \cdot \frac{1}{\Delta_{\phi_R}} \tag{29}$$

It was shown in (Salz and Winters; 1994) that corresponding spatial covariance functions are modulated sinc functions

$$\rho(d) \approx \exp\left(j\frac{2\pi d}{\lambda}\sin\phi_0\right)\operatorname{sinc}\left(\Delta\phi\frac{d}{\lambda}\cos\phi_0\right) \tag{30}$$

The correlation function of the form (30) gives rise to a correlation matrix between antenna elements which can be decomposed in terms of frequency modulated Discrete Prolate Spheroidal Sequences (MDPSS) (Alcocer et al.; 2005; Slepian; 1978; Sejdić et al.; 2008):

$$\mathbf{R} \approx \mathbf{W}\mathbf{U}\Lambda\mathbf{U}^H\mathbf{W}^H = \sum_{k=0}^{D} \lambda_k \mathbf{u}_k \mathbf{u}_k^H \tag{31}$$

where $\Lambda \approx \mathbf{I}_D$ is the diagonal matrix of size $D \times D$ (Slepian; 1978), $\mathbf{U}$ is $N \times D$ matrix of the discrete prolate spheroidal sequences and $\mathbf{W} = \operatorname{diag}\{\exp(j2\pi d/\lambda \sin n d_A)\}$. Here d_A is distance between the antenna elements, N number of antennas, $1 \le n \le N$ and $D \approx \lceil 2\Delta\phi\frac{d}{\lambda}\cos\phi_0\rceil + 1$ is the effective number of degrees of freedom generated by the process with the given covariance matrix $\mathbf{R}$. For narrow spread clusters the number of degrees of freedom is much less than the number of antennas $D \ll N$ (Slepian; 1978). Thus, it could be inferred from equation (31) that the desired channel impulse response $\mathbf{H}(\omega, \tau)$ could be represented as a double sum(tensor product).

$$\mathbf{H}(\omega, t) = \sum_{n_t}^{D_T}\sum_{n_r}^{D_R} \sqrt{\lambda_{n_t}\lambda_{n_r}}\mathbf{u}_{n_r}^{(r)}\mathbf{u}_{n_t}^{(t)^H} h_{n_t,n_r}(\omega, t) \tag{32}$$

In the extreme case of a very narrow angular spread on both sides, $D_R = D_T = 1$ and $\mathbf{u}_1^{(r)}$ and $\mathbf{u}_1^{(t)}$ are well approximated by the Kaiser windows (Thomson; 1982). The channel corresponding to a single scatterer is of course a rank one channel given by

$$\mathbf{H}(\omega, t) = \mathbf{u}_1^{(r)}\mathbf{u}_1^{(t)^H} h(\omega, t). \tag{33}$$

Considering the shape of the functions $\mathbf{u}_1^{(r)}$ and $\mathbf{u}_1^{(t)}$ one can conclude that in this scenario angular spread is achieved by modulating the amplitude of the spatial response of the channel on both sides. It is also worth noting that representation (32) is the Karhunen-Loeve series (van Trees; 2001) in spatial domain and therefore produces smallest number of terms needed to represent the process selectivity in spatial domain. It is also easy to see that such modulation becomes important only when the number of antennas is significant.

Similar results could be obtained in frequency and Doppler domains. Let us assume that τ is the mean delay associated with the cluster and $\Delta\tau$ is corresponding delay spread. In addition let it be desired to provide a proper representation of the process in the bandwidth $[-W : W]$ using N_F equally spaced samples. Assuming that the variation of power is relatively minor within $\Delta\tau$ delay window, we once again recognize that the variation of the channel in frequency domain can be described as a sum of modulated DPSS of length N_F and the time bandwidth product $W\Delta\tau$. The number of MDPSS needed for such representation is approximately $D_F = 2W\Delta\tau + 1$ (Slepian; 1978):

$$\mathbf{h}(\omega,t) = \sum_{n_f=1}^{D_f} \sqrt{\lambda_{nf}}\mathbf{u}_{n_f}^{(\omega)} h_{n_f}(t) \tag{34}$$

Finally, in the Doppler domain, the mean resulting Doppler spread could be calculated as

$$f_D = \frac{f_0}{c}\left[v_T \cos\left(\phi_{T0} - \alpha_T\right) + v_R \cos\left(\phi_{R0} - \alpha_R\right)\right]. \tag{35}$$

The angular extent of the cluster from sides causes the Doppler spectrum to widen by the folowing

$$\Delta f_D = \frac{f_0}{c}\left[v_T \Delta\phi_T v_T \left|\sin\left(\phi_{T0} - \alpha_T\right)\right| + v_R \Delta\phi_R \left|\sin\left(\phi_{R0} - \alpha_R\right)\right|\right]. \tag{36}$$

Once again, due to a small angular extent of the cluster it could be assumed that the widening of the Doppler spectrum is relatively narrow and no variation within the Doppler spectrum is of importance. Therefore, if it is desired to simulate the channel on the interval of time $[0 : T_{max}]$ then this could be accomplished by adding $D = 2\Delta f_D Tmax + 1$ MDPS:

$$\mathbf{h}_d = \sum_{n_d=0}^{D} \xi_{n_d} \sqrt{\lambda_{n_d}} \mathbf{u}_{n_d}^{(d)} \tag{37}$$

where ξ_{n_d} are independent zero mean complex Gaussian random variables of unit variance.

Finally, the derived representation could be summarized in tensor notation as follows. Let $\mathbf{u}_{n_t}^{(t)}$, $\mathbf{u}_{n_r}^{(r)}$, $\mathbf{u}_{n_f}^{(\omega)}$ and $\mathbf{u}_{n_d}^{(d)}$ be DPSS corresponding to the transmit, receive, frequency and Doppler dimensions of the signal with the "domain-dual domain" products (Slepian; 1978) given by $\Delta\phi_T \frac{d}{\lambda} \cos\phi_{T0}$, $\Delta\phi_R \frac{d}{\lambda} \cos\phi_{R0}$, $W\Delta\tau$ and $T_{max}\Delta f_D$ respectively. Then a sample of a MIMO frequency selective channel with corresponding characteristics could be generated as

$$\mathcal{H}_4 = \mathcal{W}_4 \odot \sum_{n_t}^{D_T}\sum_{n_r}^{D_R}\sum_{n_f}^{D_F}\sum_{n_d}^{d} \sqrt{\lambda_{n_t}^{(t)}\lambda_{n_r}^{(r)}\lambda_{n_f}^{(\omega)}\lambda_{n_d}^{(T)}}\,\xi_{n_t,n_r,n_f,n_d} \cdot {}_1\mathbf{u}_{n_t}^{(r)} \times {}_2\mathbf{u}_{n_r}^{(r)} \times {}_3\mathbf{u}_{n_f}^{(\omega)} \times {}_4\mathbf{u}_{n_d}^{(d)} \tag{38}$$

where $\mathcal{W}_4$ is a tensor composed of modulating sinusoids

$$\mathcal{W}_4 = {}_1\mathbf{w}^{(r)} \times {}_2\mathbf{w}^{(t)} \times {}_3\mathbf{w}^{(\omega)} \times {}_4\mathbf{w}^{(d)} \tag{39}$$

$$\begin{aligned}
\mathbf{w}^{(r)} &= \left[1, \exp\left(j2\pi\frac{d_R}{\lambda}\right), \cdots, \exp\left(j2\pi\frac{d_R}{\lambda}(N_R-1)\right)\right]^T \\
\mathbf{w}^{(t)} &= \left[1, \exp\left(j2\pi\frac{d_T}{\lambda}\right), \cdots, \exp\left(j2\pi\frac{d_T}{\lambda}(N_T-1)\right)\right]^T \qquad (40) \\
\mathbf{w}^{(\omega)} &= \left[1, \exp\left(j2\pi\Delta F\tau\right), \cdots, \exp\left(j2\pi\Delta F(N_F-1)\right)\right]^T \\
\mathbf{w}^{(d)} &= \left[1, \exp\left(j2\pi\Delta f_D T_s\right), \cdots, \exp\left(j2\pi\Delta f_D(T_{max}-T_s)\right)\right]^T
\end{aligned}$$

(41)

and $\odot$ is the Hadamard (element wise) product of two tensors (van Trees; 2002).

3.2 Multi-Cluster environment

The generalization of the model suggested in Section 3.1 to a real multi-cluster environment is straightforward. The channel between the transmitter and the receiver is represented as a set of clusters, each described as in Section (3.1). The total impulse response is superposition of independently generated impulse response tensors from each cluster

$$\mathcal{H}_4 = \sum_{k=0}^{N_c-1} \sqrt{P_k}\mathcal{H}_4(k), \; \sum_{k=1}^{N_c} P_k = P \tag{42}$$

where N_c is the total number of clusters, $\mathcal{H}_4(k)$ is a normalized response from the k-th cluster $||\mathcal{H}_4(k)||_F^2 = 1$ and $P_k \geq 0$ represents relative power of k-th cluster and P is the total power.

It is important to mention here that such a representation does not necessarily correspond to a physical cluster distribution. It rather reflects interplay between radiated and received signals, arriving from certain direction with a certain excess delay, ignoring particular mechanism of propagation. Therefore it is possible, for example, to have two clusters with the same AoA and AoD but a different excess delay. Alternatively, it is possible to have two clusters which correspond to the same AoD and excess delay but very different AoA.

Equations (38) and (42) reveal a connection between Sum of Cisoids (SoC) approach (SCM Editors; 2006) and the suggested algorithms: one can consider (38) as a modulated Cisoid. Therefore, the simulator suggested above could be considered as a Sum of Modulated Cisoids simulator.

In addition to space dispersive components, the channel impulse response may contain a number of highly coherent components, which can be modelled as pure complex exponents. Such components described either direct LoS path or specularly reflected rays with very small phase diffusion in time. Therefore equation (42) should be modified to account for such components:

$$\mathcal{H}_4 = \sqrt{\frac{1}{1+K}} \sum_{k=0}^{N_c-1} \sqrt{P_{ck}}\mathcal{H}_4(k) + \sqrt{\frac{K}{1+K}} \sum_{k=0}^{N_s-1} \sqrt{P_{sk}}\mathcal{W}_4(k) \tag{43}$$

Here N_s is a number of specular components including LoS and K is a generalized Rice factor describing ratio between powers of specular P_{sk} and non-coherent/diffusive components P_{ck}

$$K = \frac{\sum_{k=0}^{N_s-1} P_{sk}}{\sum_{k=0}^{N_c-1} P_{ck}} \tag{44}$$

While distribution of the diffusive component is Gaussian by construction, the distribution of the specular component may not be Gaussian. A more detailed analysis is beyond the scope of this chapter and will be considered elsewhere. We also leave a question of identifying and distinguishing coherent and non-coherent components to a separate manuscript.

4. Examples

Fading channel simulators (Jeruchim et al.; 2000) can be used for different purposes. The goal of the simulation often defines not only suitability of a certain method but also dictates choice of the parameters. One possible goal of simulation is to isolate a particular parameter and study its effect of the system performance. Alternatively, a various techniques are needed to avoid the problem of using the same model for both simulation and analysis of the same scenario. In this section we provide a few examples which show how suggested algorithm can be used for different situations.

4.1 Two cluster model

The first example we consider here is a two-cluster model shown in Fig. 4. This geometry is

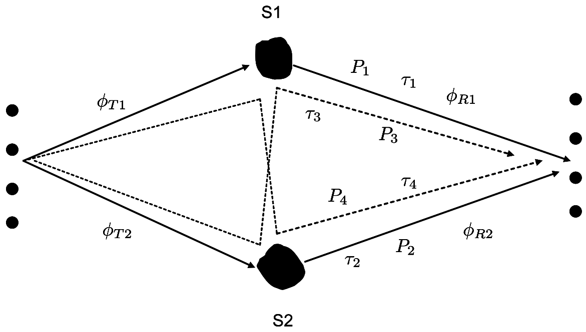


Fig. 4. Geometry of a single cluster problem.

the simplest non-trivial model for frequency selective fading. However, it allows one to study effects of parameters such as angular spread, delay spread, correlation between sites on the channel parameters and a system performance. The results of the simulation are shown in Figs. 5-6. In this examples we choose $\phi_{T1} = 20^o$, $\phi_{T2} = 20^o$, $\phi_{R1} = 0^o$, $\phi_{R2} = 110^o$, $\tau_1 = 0.2\mu s$, $\tau_2 = 0.4\mu s$, $\Delta\tau_1 = 0.2\mu s$, $\Delta\tau_2 = 0.4\mu s$.

4.2 Environment specified by joint AoA/AoD/ToA distribution

The most general geometrical model of MIMO channel utilizes joint distribution $p(\phi_T, \phi_R, \tau)$, $0 \le \phi_T < 2\pi$, $0 \le \phi_R < 2\pi$, $\tau_{min} \le \tau \le \tau_{max}$, of AoA, AoD and Time of Arrival (ToA). A few of such models could be found in the literature (Kaiserd et al.; 2006), (Andersen and Blaustaein; 2003; Molisch et al.; 2006; Asplund et al.; 2006; Blaunstein et al.; 2006; Algans et al.; 2002). Theoretically, this distribution completely describes statistical properties of the MIMO channel. Since the resolution of the antenna arrays on both sides is finite and a finite bandwidth of the channel is utilized, the continuous distribution $p(\phi_T, \phi_R, \tau)$ can be discredited to produce narrow "virtual" clusters centered at $[\phi_{Tk}, \phi_{Rk}, \tau_k]$ and with spread $\Delta\phi_{Tk}$, $\Delta\phi_{Rk}$ and $\Delta\tau_k$

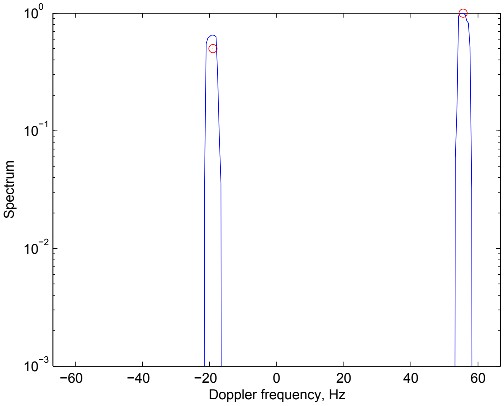


Fig. 5. PSD of the two cluster channel response.

respectively and the power weight

$$P_k = \frac{P}{4\pi^2(\tau_{max} - \tau_{min})} \times \int_{\tau_k - \Delta\tau_k/2}^{\tau_k + \Delta\tau_k/2} d\tau \int_{\phi_{Tk} - \Delta\phi_{Tk}/2}^{\phi_{Tk} + \Delta\phi_{Tk}/2} d\phi_T \int_{\phi_{Rk} - \Delta\phi_{Rk}/2}^{\phi_{Rk} + \Delta\phi_{Rk}/2} p(\phi_T, \phi_R, \tau) d\phi_R \quad (45)$$

We omit discussions about an optimal partitioning of each domain due to the lack of space. Assume that each virtual cluster obtained by such partitioning is appropriate in the frame discussed in Section 3.1.

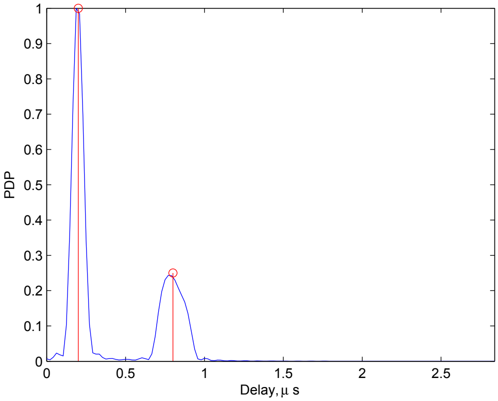


Fig. 6. PDP of the two cluster channel response.

As an example, let us consider the following scenario, described in (Blaunstein et al.; 2006). In this case the effect of the two street canyon propagation results into two distinct angles of arrival $\phi_{R1} = 20^o$ and $\phi_{R2} = 50^o$, AoA spreads roughly of $\Delta_1 = \Delta_2 = 5^o$ and exponential PDP corresponding to each AoA (see Figs. 5 and 6 in (Blaunstein et al.; 2006)). In addition, an almost uniform AoA on the interval $[60 : 80^o]$ corresponds to early delays. Therefore, a simplified model of such environment could be presented by

$$p(\phi_R, \tau) = \sqrt{P_1}\frac{1}{\Delta_1}\exp\left(-\frac{\tau - \tau_1}{\tau_{s1}}\right)u(\tau - \tau_1) + \sqrt{P_2}\frac{1}{\Delta_2}\exp\left(-\frac{\tau - \tau_2}{\tau_{s2}}\right)u(\tau - \tau_2) + \sqrt{P_3}\frac{1}{\Delta_3}\exp\left(-\frac{\tau - \tau_3}{\tau_{s3}}\right)u(\tau - \tau_3) \quad (46)$$

where $u(t)$ is the unit step function, τ_{sk}, $k = 1,2,3$ describe rate of decay of PDP. By inspection of Figs. 5-6 in (Blaunstein et al.; 2006) we choose $\tau_1 = \tau_2 = 1.2$ ns, $\tau_3 = 1.1$ ns and $\tau_{s1} = \tau_{s2} = \tau_{s3} = 0.3$ ns. Similarly, by inspection of the same figures we assume $P_1 = P_2 = 0.4$ and $P_3 = 0.2$. To model exponential PDP with unit power and average duration τ_s we represent it with a set of $N \geq 1$ rectangular PDP of equal energy $1/N$. The k-th virtual cluster then extends on the interval $[\tau_{k-1} : \tau_k]$ and has magnitude $P_k = 1/N\Delta\tau_k$ where $\tau_0 = 0$

$$\begin{aligned} \tau_k &= \tau_s \ln\frac{N-k}{N}, \, k = 1, ..., N-1 && (47) \\ \tau_N &= \tau_{N-1} + \frac{1}{N\tau_{N-1}}, \, k = N && (48) \\ \Delta\tau_k &= \tau_k - \tau_{k-1} && (49) \end{aligned}$$

Results of numerical simulation are shown in Figs. 7 and 8. It can be seen that a good agreement between the desired characteristics is obtained.

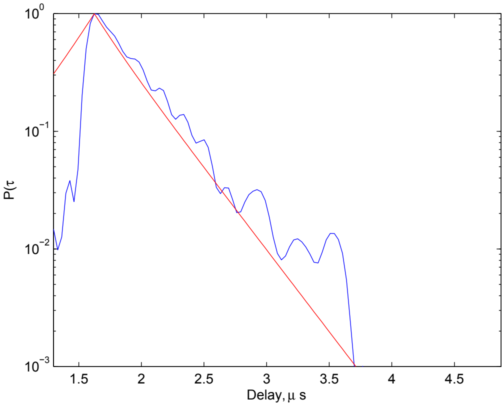


Fig. 7. Simulated power delay profile for the example of Section 4.2.

Similarly, the same technique could be applied to the 3GPP (SCM Editors; 2006) and COST 259 (Asplund et al.; 2006) specifications.

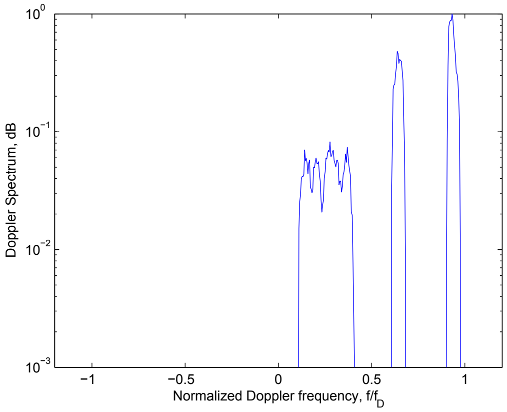


Fig. 8. Simulated Doppler power spectral density for the example of Section 4.2.

5. MDPSS Frames for channel estimation and prediction

5.1 Modulated Discrete Prolate Spheroidal Sequences

If the DPSS are used for channel estimation, then usually accurate and sparse representations are obtained when both the DPSS and the channel under investigation occupy the same frequency band (Zemen and Mecklenbräuker; 2005). However, problems arise when the channel is centered around some frequency $|\nu_o| > 0$ and the occupied bandwidth is smaller than $2W$, as shown in Fig. 9.

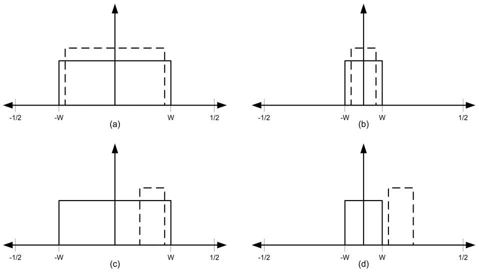


Fig. 9. Comparison of the bandwidth for a DPSS (solid line) and a channel (dashed line): (a) both have a wide bandwidth; (b) both have narrow bandwidth; (c) a DPSS has a wide bandwidth, while the channel's bandwidth is narrow and centered around $\nu_o > 0$; (d) both have narrow bandwidth, but centered at different frequencies.

In such situations, a larger number of DPSS is required to approximate the channel with the same accuracy despite the fact that such narrowband channel is more predictable than a wider band channel (Proakis; 2001). In order to find a better basis we consider so-called Modulated Discrete Prolate Spheroidal Sequences (MDPSS), defined as

$$M_k(N,W,\omega_m;n) = \exp(j\omega_m n)v_k(N,W;n), \tag{50}$$

where $\omega_m = 2\pi\nu_m$ is the modulating frequency. It is easy to see that MDPSS are also doubly orthogonal, obey the same equation (7) and are bandlimited to the frequency band $[-W+\nu : W+\nu]$.

The next question which needs to be answered is how to properly choose the modulation frequency ν. In the simplest case when the spectrum $S(\nu)$ of the channel is confined to a known band $[\nu_1;\nu_2]$, *i.e.*

$$S(\nu) = \begin{cases} \gg 0 & \forall \nu \in [\nu_1, \nu_2] \text{ and } |\nu_1| < |\nu_2| \\ \approx 0 & \text{elsewhere} \end{cases}, \tag{51}$$

the modulating frequency, ν_m, and the bandwidth of the DPSS's are naturally defined by

$$\nu_m = \frac{\nu_1 + \nu_2}{2} \tag{52}$$

$$W = \left| \frac{\nu_2 - \nu_1}{2} \right|, \tag{53}$$

as long as both satisfy:

$$|\nu_m| + W < \frac{1}{2}. \tag{54}$$

In practical applications the exact frequency band is known only with a certain degree of accuracy. In addition, especially in mobile applications, the channel is evolving in time. Therefore, only some relatively wide frequency band defined by the velocity of the mobile and the carrier frequency is expected to be known. In such situations, a one-band-fits-all approach may not produce a sparse and accurate approximation of the channel. To resolve this problem, it was previously suggested to use a band of bases with different widths to account for different speeds of the mobile (Zemen et al.; 2005). However, such a representation once again ignores the fact that the actual channel bandwidth $2W$ could be much less than $2\nu_D$ dictated by the maximum normalized Doppler frequency $\nu_D = f_D T$.

To improve the estimator robustness, we suggest the use of multiple bases, better known as frames (Kovačević and Chabira; 2007), precomputed in such a way as to reflect various scattering scenarios. In order to construct such multiple bases, we assume that a certain estimate (or rather its upper bound) of the maximum Doppler frequency ν_D is available. The first few bases in the frame are obtained using traditional DPSS with bandwidth $2\nu_D$. Additional bases can be constructed by partitioning the band $[-\nu_D;\nu_D]$ into K subbands with the boundaries of each subband given by $[\nu_k;\nu_{k+1}]$, where $0 \leq k \leq K-1$, $\nu_{k+1} > \nu_k$, and $\nu_0 = -\nu_D$, $\nu_{K-1} = \nu_D$. Hence, each set of MDPSS has a bandwidth equal to $\nu_{k+1} - \nu_k$ and a modulation frequency equal to $\nu_m = 0.5(\nu_k + \nu_{k+1})$. Obviously, a set of such functions again forms a basis of functions limited to the bandwidth $[-\nu_D;\nu_D]$. It is a convention in the signal processing community to call each basis function *an atom*. While particular partition is arbitrary for every level $K \geq 1$, we can choose to partition the bandwidth into equal blocks to reduce the amount of stored precomputed DPSS, or to partition according to the angular resolution of the receive antenna, *etc*, as shown in Fig. 10.

Representation in the overcomplete basis can be made sparse due to the richness of such a basis. Since the expansion into simple bases is not unique, a fast, convenient and unique projection algorithm cannot be used. Fortunately, efficient algorithms, known generically as pursuits (Mallat; 1999; Mallat and Zhang; 1993), can be used and they are briefly described in the next section.

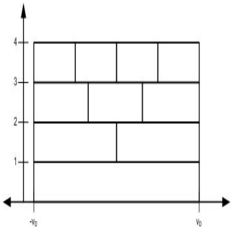


Fig. 10. Sample partition of the bandwidth for $K = 4$.

5.2 Matching Pursuit with MDPSS frames

From the few approaches which can be applied for expansion in overcomplete bases, we choose the so-called matching pursuit (Mallat and Zhang; 1993). The main feature of the algorithm is that when stopped after a few steps, it yields an approximation using only a few atoms (Mallat and Zhang; 1993). The matching pursuit was originally introduced in the signal processing community as an algorithm that decomposes any signal into a linear expansion of waveforms that are selected from a redundant dictionary of functions (Mallat and Zhang; 1993). It is a general, greedy, sparse function approximation scheme based on minimizing the squared error, which iteratively adds new functions (i.e. basis functions) to the linear expansion. In comparison to a basis pursuit, it significantly reduces the computational complexity, since the basis pursuit minimizes a global cost function over all bases present in the dictionary (Mallat and Zhang; 1993). If the dictionary is orthogonal, the method works perfectly. Also, to achieve compact representation of the signal, it is necessary that the atoms are representative of the signal behavior and that the appropriate atoms from the dictionary are chosen.

The algorithm for the matching pursuit starts with an initial approximation for the signal, $\widehat{x}$, and the residual, R:

$$\widehat{x}^{(0)} = 0 \tag{55}$$

$$R^{(0)} = x \tag{56}$$

and it builds up a sequence of sparse approximation stepwise by trying to reduce the norm of the residue, $R = \widehat{x} - x$. At stage k, it identifies the dictionary atom that best correlates with the residual and then adds to the current approximation a scalar multiple of that atom, such that

$$\widehat{x}^{(k)} = \widehat{x}^{(k-1)} + \alpha_k \phi_k \tag{57}$$

$$R^{(k)} = x - \widehat{x}^{(k)}, \tag{58}$$

where $\alpha_k = \langle R^{(k-1)}, \phi_k \rangle / \|\phi_k\|^2$. The process continues until the norm of the residual $R^{(k)}$ does not exceed required margin of error $\epsilon > 0$: $||R^{(k)}|| \leq \epsilon$ (Mallat and Zhang; 1993). In our approach, a stopping rule mandates that the number of bases, χ_B, needed for signal approximation should satisfy $\chi_B \leq \lceil 2N\nu_D \rceil + 1$. Hence, a matching pursuit approximates the signal using χ_B bases as

$$x = \sum_{n=1}^{\chi_B} \langle x, \phi_n \rangle \phi_n + R^{(\chi_B)}, \tag{59}$$

where ϕ_n are χ_B bases from the dictionary with the strongest contributions.

6. Numerical Simulation

In this section, the performance of the MDPSS estimator is compared with the Slepian basis expansion DPSS approach (Zemen and Mecklenbräuker; 2005) for a certain radio environment. The channel model used in the simulations is presented in Section 2.2 and it is simulated using the AR approach suggested in (Baddour and Beaulieu; 2005). The parameters of the simulated system are the same as in (Zemen and Mecklenbräuker; 2005): the carrier frequency is 2 GHz, the symbol rate used is 48600 $1/s$, the speed of the user is 102.5 km/h, 10 pilots per data block are used, and the data block length is $M = 256$. The number of DPSS's used in estimation is given by $\lceil 2M\nu_D \rceil + 1$. The same number of bases is used for MDPSS, while $K = 15$ subbands is used in generation of MDPSS.

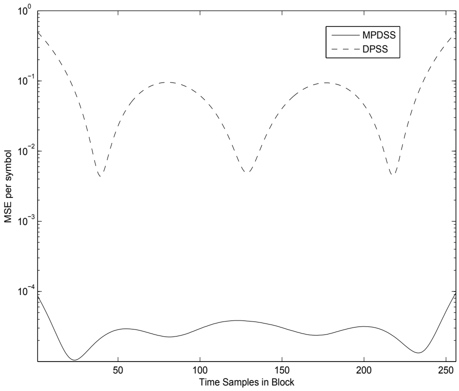


Fig. 11. Mean square error per symbol for MDPSS (solid) and DPSS (dashed) mobile channel estimators for the noise-free case.

As an introductory example, consider the estimation accuracy for the WSSUS channel with a uniform power angle profile (PAS) with central AoA $\phi_0 = 5$ degrees and spread $\Delta = 20$ degrees. We used 1000 channel realizations and Fig. 11 depicts the results for the considered channel model. The mean square errors (MSE) for both MDPSS and DPSS estimators have the highest values at the edges of the data block. However, the MSE for MDPSS estimator is several orders of magnitude lower than the value for the Slepian basis expansion estimator based on DPSS.

Next, let's examine the estimation accuracy for the WSSUS channels with uniform PAS, central AoAs $\phi_1 = 45$ and $\phi_1 = 75$, and spread $0 < \Delta \leq 2\pi/3$. Furthermore, it is assumed that the channel is noisy. Figs. 12 and 13 depict the results for $SNR = 10$ dB and $SNR = 20$ dB, respectively.

The results clearly indicate that the MDPSS frames are a more accurate estimation tool for the assumed channel model. For the considered angles of arrival and spreading angles, the MDPSS estimator consistently provided lower MSE in comparison to the Slepian basis expansion estimator based on DPSS. The advantage of the MDPSS stems from the fact that these bases are able to describe different scattering scenarios.

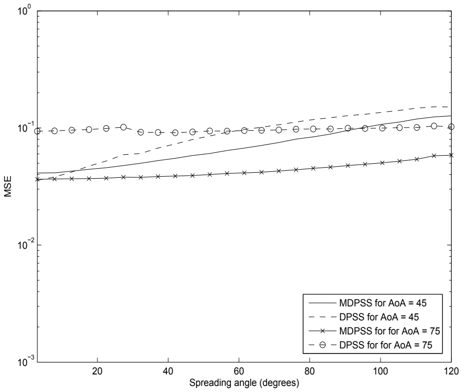


Fig. 12. Dependence of the MSE on the angular spread Δ and the mean angle of arrival for $SNR = 10$ dB.

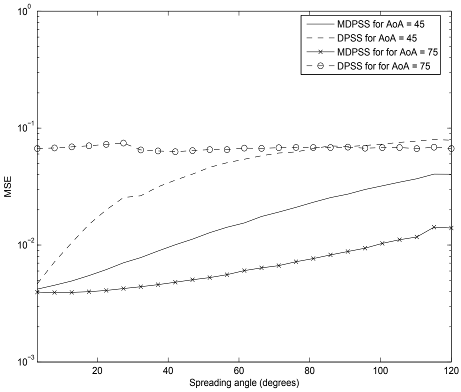


Fig. 13. Dependence of the MSE on the angular spread Δ and the mean angle of arrival for $SNR = 20$ dB.

7. Conclusions

In this Chapter we have presented a novel approach to modelling MIMO wireless communication channels. At first, we have argued that in most general settings the distribution of the in-phase and quadrature components are Gaussian but may have different variance. This was explained by an insufficient phase randomization by small scattering areas. This model leads to a non-Rayleigh/non-Rice distribution of magnitude and justifies usage of such generic distributions as Nakagami or Weibull. It was also shown that additional care should be taken when modelling specular components in MIMO settings.

Furthermore, based on the assumption that the channel is formed by a collection of relatively small but non-point scatterers, we have developed a model and a simulation tool to represent such channels in an orthogonal basis, composed of modulated prolate spheroidal sequences.

Finally MDPSS frames are proposed for estimation of fast fading channels in order to preserve sparsity of the representation and enhance the estimation accuracy. The members of the frame were obtained by modulation and bandwidth variation of DPSS's in order to reflect various scattering scenarios. The matching pursuit approach was used to achieve a sparse representation of the channel. The proposed scheme was tested for various mobile channels, and its performance was compared with the Slepian basis expansion estimator based on DPSS. The results showed that the MDPSS method provides more accurate estimation than the DPSS scheme.

8. References

Alcocer, A., Parra, R. and Kontorovich, V. (2005). An orthogonalization approach for communication channel modelling, *Proc. VTC-2005, Fall*.

Alcocer-Ochoa, A., Kontorovitch, V. Y. and Parra-Michel, R. (2006). A MIMO channel simulator applying the universal basis, *Proc. of 64th IEEE Vehicular Technology Conference (VTC-2006)*, Montreal, QC, Canada.

Algans, A., Pedersen, K. and Mogensen, P. (2002). Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading, *IEEE J. Sel. Areas Commun.* **20**(3): 523–531.

Almers, P., Bonek, E., Burr, A., Czink, N., Debbah, M., Degli-Esposti, V., Hofstetter, H., Kyösti, P., Laurenson, D., Matz, G., Molisch, A. F., Oestges, C. and Özcelik, H. (2006). Survey of channel and radio propagation models for wireless mimo systems.

Andersen, J. and Blaustaein, N. (2003). *Multipath Phenomena in Cellular Networks*, Artech House, 2002.

Asplund, H., Glazunov, A., Molisch, A., Pedersen, K. and Steinbauer, M. (2006). The COST 259 directional channel model Part II: Macrocell, *IEEE Trans. Wireless Commun.* **5**(12): 3434–3450.

Baddour, K. E. and Beaulieu, N. C. (2005). Autoregressive modeling for fading channel simulation, *IEEE Transactions on Wireless Communications* **4**(4): 1650 – 1662.

Barakat, R. (1986). Weak-scatterer generalization of the k-density function with application to laser scattering in atmospheric turbulence, *J. Opt. Soc. Am. A* **3**(4): 401–409.

Beckmann, P. and Spizzichino, A. (1963). *The scattering of electromagnetic waves from rough surfaces*, Pergamon Press, New York, USA.

Bertoni, H. (2000). *Radio Propagation For Modern Wireless Systems*, Prentice Hall PTR, Upper Saddle River, NJ.

Blaunstein, N., Toeltsch, M., Laurila, J., Bonek, E., Katz, D., Vainikainen, P., Tsouri, N., Kalliola, K. and Laitinen, H. (2006). Signal power distribution in the azimuth, elevation and time delay domains in urban environments for various elevations of base station antenna, *IEEE Trans. Antennas Propag.* **54**(10): 2902.

Bronez, T. (1992). On the performance advantage of multitaper spectral analysis, *IEEE Trans. Signal Process.* **40**(12): 2941–2946.

Conover, W. (1998). *Practical Nonparametric Statistics*, third edn, Wiley, New York.

Fancourt, C. and Principe, J. (2000). On the relationship between the Karhunen-Loeve transform and the prolate spheroidal wave functions, *Proc. ICASSP 00*, Istambul, Turkey, pp. 261–264.

Fechtel, S. (1993). A novel approach to modeling and efficient simulation offrequency-selective fading radio channels, *IEEE J. Sel. Areas Commun.* **11**(3): 422–431.

Fleury, B. (2000). First- and second-order characterization of direction dispersion and space selectivity in the radio channel, *IEEE Trans. Inf. Theory* **46**(6): 2027–2044.

Jakeman, E. and Tough, R. (1987). Generalized k distribution: a statistical model for weak scattering, *J. Opt. Soc. Am. A* **4**(9): 1764–1772.

Jeruchim, M. C., Balaban, P. and Shanmugan, K. S. (2000). *Simulation of Communication Systems: Modeling, Methodology and Techniques*, New York: Springer.

Kaiserd, T., Bourdoux, A., Boche, H., Fonollosa, J., Andersen, J. and Utschick, W. (2006). *Smart Antennas: State of the Art*, Hindawi, NY.

Kirsteins, I., Mehta, S. and Fay, J. (2006). Adaptive separation of unknown narrowband and broadband time series, *Proc. ICASSP 98*, Seatlle, WA, USA, pp. 2525–2528.

Klovski, D. (1982). *Digital Data Transmission over Radiochannels*, Sviaz, Moscow, USSR, (in Russian).

Kontorovich, V., Primak, S., Alcocer-Ochoa1, A. and Parra-Michel, R. (2008). Mimo channel orthogonalisations applyinguniversal eigenbasis, *Proc. IET Signal processing* **2**(2): 87–96.

Kovačević, J. and Chabira, A. (2007). Life beyond bases: The advent of the frames (part I), *IEEE Signal Processing Magazine* **24**(4): 86–104.

Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing*, second edn, Academic Press, San Diego.

Mallat, S. G. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing* **41**(12): 3397–3415.

Mardia, K. and Jupp, P. (2000). *Directional statistics*, Wiley, New York.

Molisch, A. F., Asplund, H., Heddergott, R., Steinbauer, M. and Zwick, T. (2006). The COST 259 directional channel model Part I: overview and methodology, *IEEE Trans. Wireless Commun.* **5**(12): 3421–3433.

Molish, A. (2004). A generic model for MIMO wireless propagation channels in macro- and microcells, *IEEE Trans. Signal Process.* **52**(1): 61–71.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, third edn, McGraw-Hill, Boston, MA.

Patzold, M. (2002). *Mobile Fading Channels*, John Wiley & Sons, New York.

Percival, D. B. and Walden, A. T. (1993a). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univeriate Techniques*, Cambridge University Press, New York.

Percival, D. and Walden, A. (1993b). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, New York.

Primak, S. and Sejdić, E. (2008). Application of multitaper analysis to wireless communications problems, *Proc. of First International Symposium on Applied Sciences on Biomedical and Communication Technologies (ISABEL 2008)*, Aalborg, Denmark.

Proakis, J. (2001). *Digital Communications*, fourth edn, McGraw-Hill, New York.

Rao, A. and Hamed, K. (2003). Multi-taper method of analysis of periodicities in hydrologic data, *Journal of Hydrology* **279**: 125–143.

Rappaport, T. (2002). *Wireless Communications: Principles and practice*, Prentice Hall, Upper Saddle River.

Salz, J. and Winters, J. (1994). Effect of fading correlation on adaptive arrays in digital mobile radio, *IEEE Trans. Veh. Technol.* **43**(4): 1049–1057.

Schreier, P. and Scharf, L. (2003). Second-order analysis of improper complex random vectors and processes, *IEEE Trans. Signal Process.* **51**(3): 714–725.

SCM Editors (2006). UMTS: Spatial channel model for MIMO simulations, *Tech. report 25.996*.

Sejdić, E., Luccini, M., Primak, S., Baddour, K. and Willink, T. (2008). Channel estimation using DPSS based frames, *Proc. of IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008)*, Las Vegas, USA, pp. 2849–2852.

Sen, I. and Matolak, D. (2008). Vehiclevehicle channel models for the 5-GHz band, *IEEE Trans. Intell. Transp. Syst.* **9**(26): 235–245.

Simon, M. and Alouini, M.-S. (2000). *Digital Communication over Fading Channels:A Unified Approach to Performance Analysis*, John Wiley & Sons, New York.

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V-The discrete case, *Bell System Technical Journal* **57**: 1371–1430.

Stoica, P. and Sundin, T. (1999). On nonparametric spectral estimation, *Journal of Circuits, Systems, and Signal Processing* **18**(2): 169–181.

Thomson, D. (1998). Multiple-window spectrum estimates for non-stationary data, *Proc. of Ninth IEEE SP Workshop on Statistical Signal and Array Processing, 1998*, pp. 344–347.

Thomson, D. (2000). Multitaper Analysis of Nonstationary and Nonlinear Time Series Data, *Nonlinear and Nonstationary Signal Processing* .

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis, *Proceedings of the IEEE* **70**(9): 1055–1096.

van Trees, H. (2001). *Detection, Estimation, and Moduclation Theory: Part I*, first edn, John Wiley & Sons, New York.

van Trees, H. (2002). *Detection, Estimation and Modulation Theory: Optimal Array Processing*, first edn, Wiley-Interscience, New York.

Xiao, H., Burr, A. and White, G. (2005). Simulation of time-selective environment by 3GPP spatial channel model and analysis on the performance evaluation by the CMD metric, *Proc. International Wireless Summit (IWS 05)*, Aalborg, Denmark, pp. 446–450.

Yip, K. and Ng, T. (1997). Karhunen-Loeve expansion of the WSSUS channel output and its application to efficient simulation, *IEEE J. Sel. Areas Commun.* **15**(4): 640–646.

Zemen, T., Hofstetter, H. and Steinbck, G. (2005). Slepian subspace projection in time and frequency for time-variant channel estimation, *14th IST Mobile and Wireless Summit*, Dresden, Germany.

Zemen, T. and Mecklenbräuker, C. F. (2005). Time-variant channel estimation using discrete prolate spheroidal sequences, *IEEE Transactions on Signal Processing* **53**(9): 3597–3607.

Zemen, T., Mecklenbräuker, C. F. and Fleury, B. (2006). Time-variant channel prediction using time-concentrated and band-limited sequences, *Proc. of 2006 IEEE International Conference on Communications (ICC '06)*, Vol. 12, Istanbul, Turkey, pp. 5660–5665.

3

High Altitude Platforms for Wireless Mobile Communication Applications

Zhe Yang and Abbas Mohammed
Blekinge Institute of Technology
Sweden

1. Introduction

The wireless communications field has experienced an extraordinary development during the past two decades or so. New wireless technologies give people more convenience and freedom to connect to different communication networks. It is thought that the demand for the capacity increases significantly when the next generation of multimedia applications are combined with future wireless communication systems.

Wireless communication services are typically provided by terrestrial and satellite systems. The successful and rapid deployment of both wireless networks has illustrated the growing demand for broadband mobile communications. These networks are featured with high data rates, reconfigurable support, dynamic time and space coverage demand with considerable cost. Terrestrial links are widely used to provide services in areas with complex propagation conditions and in mobile applications. Satellite links are usually used to provide high speed connections where terrestrial links are not available. In parallel with these well established networks, a new alternative using aerial high altitude platforms (HAPs) has emerged and attracted international attentions (Mohammed et al., 2008).

Communications platforms situated at high altitudes can be dated to the last century. In 1960 a giant balloon was launched in USA. It reflected broadcasts from the Bell laboratories facility at Crawford Hill and bounced the signals to long distance telephone call users. This balloon can be regarded as an ancestor of HAPs. Traditional applications of airships have been restricted in entertainment purposes, meteorological usage, and environment surveillance due to safety reasons. However, in the past few years, technological advancements in communications from airships has given a promising future in this area (Karapantazis & Pavlidou, 2005).

HAPs are airships or planes, operating in the stratosphere, at altitudes of typically 17-22 km (around 75,000 ft) (Collela et al., 2000; Hult et al., 2008b; Mohammed et al., 2008; Yang & Mohammed, 2008b). At this altitude (which is well above commercial aircraft height), they can support payloads to deliver a range of services: principally communications and remote sensing. A HAP can provide the best features of both terrestrial masts (which may be subject to planning restrictions and/or related environmental/health constraints) and satellite systems (which are usually highly expensive) (Cost 297, 2005; Mohammed et al., 2008). This makes HAP a viable competitor/complement to conventional terrestrial infrastructures and satellite systems. Thus HAPs are regarded as a future candidate for next generation systems, either as a stand-alone system or integrated with other satellite or terrestrial systems.

HAPs as a new solution for delivering wireless broadband, have been recently proposed for the provision of fixed, mobile services and application as shown in Fig. 1 (Grace et al., 2001b; Yang & Mohammed, 2008a). HAPs can act as base-stations or relay nodes, which may be effectively regarded as a very tall antenna mast or a very Low-Earth-Orbit (LEO) satellite (Thornton et al., 2001). This modern communication solution has advantages of both terrestrial and satellite communications (Djuknic et al., 1997; Steele, 1992). It is a good technique for serving the increasing demand of Broadband Wireless Access (BWA) by using higher frequency allocations especially in mm-wavelength and high-speed data capacity. HAPs are also proposed to provide other communication services, i.e. the 3rd generation (3G) services. The International Telecommunication Union (ITU) gave licensed a frequency band around 2 GHz for IMT-2000 service (Grace et al., 2001a). An optical inter-platform link can be established between platforms to expand a network in the altitude to cover a large area. A broadband access link between the platform and user on the ground can be established to support different applications. Fixed, mobile and portable user terminals can be supported by the system. With assistance of terrestrial networks, HAPs can also provide the telecommunication services or the backbone for terrestrial networks in remote areas.

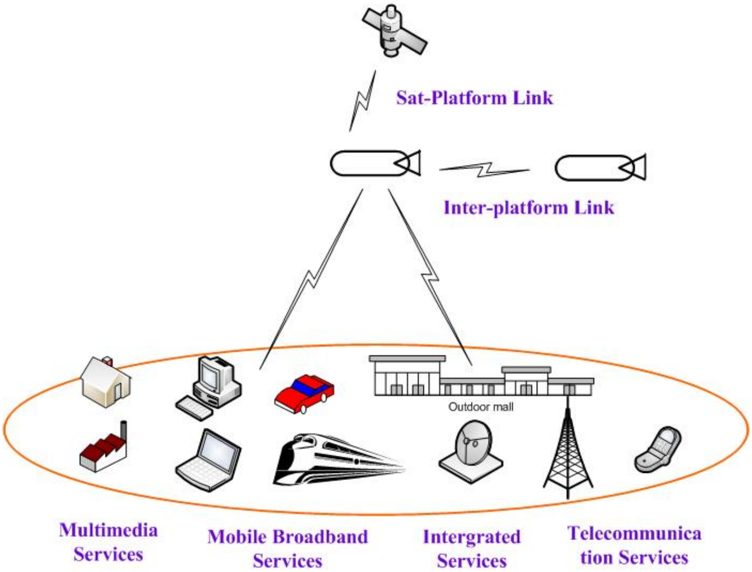


Fig. 1. HAP system deployed at 17~22 km above the ground

The full chapter is organized as follows: in section 1, we give an introduction to the HAP concept in wireless communications. In section 2, an overview of communication applications from HAPs, frequency allocations, well-known HAP research activities and trials are given. In section 3, main characteristics of HAP system are summarized and scenarios of HAP deployment are discussed. Finally, conclusions and future research are given in section 4.

2. Applications, Research and Trials of HAP Systems

2.1 Applications and frequency allocations of HAP System

HAPs have been proposed to deliver modern broadband services, i.e. high-speed internet, High-Definition Television (HDTV), Local Multi-Point Distribution (LMDS), Multi-Channel Multimedia Distribution Service (MMDS), and Wireless Interoperability for Microwave Access (WiMAX). All these services require wide bandwidth and high capacity. Generally these applications can be thought to equip base stations onboard, and based on well established terrestrial system design experience, but they are facing new challenges, e.g. cell structures, handover controls and dynamic channel assignment.

BWA services operate in the higher frequency bands, i.e. the mm-wave bands at several GHz, to provide the required radio frequency bandwidth allocation. The frequency bands allocated for LMDS in most countries in the world are around 30 GHz. ITU has assigned frequency bands of 47-48 GHz to HAPs worldwide. The 28-31 GHz bands have also assigned to HAP in some regions.

- On the ITU 1997 World Radio communication Conference (WRC-97), the ITU passed the RESOLUTION 122 to use the bands 47.2-47.5 GHz and 47.9-48.2 GHz for HAPs to provide the Fixed Service (FS) (ITU-R, 2003). WRC-2000 adopted the revision of RESOLUTION 122 to allow HAPs utilizing the bands 18-32 GHz, 27.5-28.35 GHz and 31-31.3 GHz in interested countries on non-interference and non-protection basis, which extended the previous RESOLUTION 122 (ITU-R, 2003). At the recent WRC-03 in ITU 2003, ITU gave the temporary RESOLUTION 145 [COM5/17] for potential using of the bands 27.5-28.35 GHz and HAPs in the FS (ITU-R, 2003).

HAPs may be one of the most important infrastructures for International Mobile Telecommunications (IMT-2000) 3G service, since HAPs can offer new means to provide IMT-2000 service with minimal network infrastructure. IMT-2000 standard has included provision for base-station deployment from HAPs and still needs further study before the deployment from HAPs in the areas of cell planning and antenna development. Employing access techniques such as Code Division Multiple Access (CDMA), Wideband-CDMA (W-CDMA) based IMT-2000 and CDMA based universal mobile telecommunications system (UMTS) from HAPs to provide 3G communications have been examined. (Foo et al., 2002; Hult et al., 2008a)

- RESOLUTION 221 was adopted by WRC-00 in ITU 2000 to approve HAPs providing IMT-2000 in the bands 1885-1980 MHz, 2010-2025 MHz and 2110-2170 MHz in with explicit region restrictions (ITU-R, 2003).

- RESOLUTION 734, which proposed HAPs to operate in the frequency range of 3-18 GHz, was adopted by WRC-2000 to allow these studies. It is noted that the range of 10.6 to 18 GHz range was not allocated to match the RESOLUTION 734.

2.2 HAP research and trails in the World

Many countries and organizations have made significant efforts in the research of HAPs system and its applications. Some well-known projects are listed below:

- The US Lockheed Martin compnay has won a contract from US Defense Advanced Research Projects Agency (DARPA) and the US Air Force (USAF) to build a high-altitude airship demonstrator featuring radar technology powerful enough to detect a car hidden under a canopy of trees from a distance of more than 300 km. Lockheed's Skunk Works division will build and fly a demonstrator aircraft with a scaled-down sensor system in fiscal year 2013 (Flightglobal, 2009).
- Since 2005 the EU Cost 297 action has been established in order to increase knowledge and understanding of the use of HAPs for delivery of communications and other services. It is now the largest gathering of research community with interest in HAPs and related technologies (Cost 297, 2005; Mohammed et al., 2008).
- CAPANINA of the European Union (EU) - The primary aim of CAPANINA is to provide technology that will deliver low-cost broadband communications services to small office and home users at data rates up to 120 Mbit/s. Users in rural areas will benefit from the unique wide-area, high-capacity coverage provided by HAPs. Trials of the technology are planned during the course of the project. Involving 13 global partners, this project is developing wireless and optical broadband technologies that will be used on HAPs (Grace et al., 2005).
- SkyNet project in Japan - A Japanese project lanuched at the beginning in 1998 to develop a HAP and studying equipments for delivery of broadband and 3G communications. This aim of the project was the development of the on-board communication equipment, wireless network protocols and platforms (Hong et al., 2005)
- European Space Agency (ESA) - has completed research of broadband delivery from HAPs. Within this study a complete system engineering process was performed for aerostatic stratospheric platforms. It has shown the overall system concept of a stratospheric platform and a possible way for its implementation (ESA, 2005).
- Lindstrand Balloons Ltd. (LBL) - The team in this company has been building lighter-than-air vehicles for almost 21 years. They have a series of balloon developments including Stratospheric Platforms, Sky Station, Ultra Long Distance Balloon (ULDB-NASA) (Lindstrand Balloons Ltd, 2005).
- HALE - The application of High-Altitude Long Endurance (HALE) platforms in emergency preparedness and disaster management and mitigation is led by the directorate of research and development in the office of critical infrastructure protection and emergency preparedness in Canada. The objective of this project has been to assess the potential application of HALE-based remote sensing technologies to disaster management and mitigation. HALE systems use advanced aircraft or balloon technologies to provide mobile, usually uninhabited, platforms operating at altitudes in excess of 50,000 feet (15,000 m) (OCIPEP, 2000).

- An US compnay Sanswire Technologies Inc. (Fort Lauderdale, USA) and Angel Technologies (St. Louis, USA) carried out a series of research and demonstrations for HAP practical applications. The flight took place at the Sanswire facility in Palmdale, California, on Nov. 15, 2005. These successful demonstrations represent mature steps in the evolution of Sanswire's overall high altitude airship program.
- Engineers from Japan have demonstrated that HAPs can be used to provide HDTV services and IMT-2000 WCDMA services successfully.

A few HAP trails have been carried out in the EU CAPANINA project to demonstrate its capabilities and applications (CAPANINA, 2004).

- In 2004, the first trial was in Pershore, UK. The trial consisted of a set of several tests based on a 300 m altitude tethered aerostat. Though the aerostat was not situated at the expected altitude it have many tasks of demonstrations and assessments, e.g. BFWA up to 120 Mbps to a fixed user using 28 GHz band, end-to-end network connectivity, high speed Internet, Video On Demand (VoD) service, using a similar platform-user architecture as that of a HAP.
- In October 2005, the second trial was conducted in Sweden. A 12,000 cubic meter balloon, flying at an altitude of around 24 km for nine hours, was launched. It conducted the RF and optical trials. Via Wi-Fi the radio equipment has supported date rates of 11 Mbps at distances ranging up to 60 km. This trial is a critical step to realize the ultimate term aim of CAPANINA to provide the 120 Mpbs data rate.

3. HAP Communication System and Deployment

3.1 Advantages of HAP system

HAPs are regarded to have several unique characteristics compared with terrestrial and satellite systems, and are ideal complement or alternative solutions when deploying next generation communication system requiring high capacity. Typical characteristics of these three systems are shown in Table 1.

Subject	**HAPs**	**Terrestrial**	**Satellite**
Cell radius	3~7 km	0.1~2 km	50 km for LEO
BS Coverage area radius	Typical 30 km ITU has suggested 150 km	5 km	A few hundred km for LEO
Elevation angles	High	Low	High
Propagation delay	Low	Low	Noticeable
Propagation Characteristic	Nearly Fress Space Path Loss (FSPL)	Well established, typically Non FPSL	FPSL with rain
BS power supply	Fuel (ideally solar)	Electricity	Solar
BS maintenance	Less complexity in terms of coverage area	Complex if multiple BSs needed to update	Impossible
BS cost	No specific number but supposed to be economical in terms of coverage area	Well established market, cost depending on the companies	5 billion for Iridium, Very expensive
Operational Cost	Medium (mainly airship maintenance)	Medium ~ High in terms of the number of BSs	High
Deployment complexity	Low (especially in remote and high density population area)	Medium (more complex to deploy in the city area)	High

Table 1. System characteristics of HAP, terrestrial and satellite systems.

The novel HAP has features of both terrestrial and satellite communications and has the advantages of both communication systems (Djuknic et al., 1997). The advantages include large coverage area, high system capacity, flexibility to respond to traffic demands etc. The main advantages can be summarized as following:

- Large-area coverage - HAPs are often considered to have a coverage radius of 30 km by virtue of their unique location (Djuknic et al., 1997; Grace et al., 2001b; Tozer & Grace, 2001). Thus, the coverage area is much larger than comparable terrestrial systems that are severely constrained by obstructions. HAPs can yield significant link budget advantage with large cells at the mm-wave bands where LOS links are required.
- Rapid deployment - A HAP can be quickly deployed in the sky within a matter of hours. It has clear advantages when it is used in disaster or emergency scenarios.
- Broadband capability - A HAP offers line of sight (LOS) propagation or better propagation non line of sight (NLOS) links owing to its unique position. A proportion of users can get a higher communication quality as low propagation delay and low ground-based infrastructure ensure low blocking from the HAP.
- Low cost - Although there is no direct evidence of HAP operation cost, it is believed that the cost of HAP is going to be considerably cheaper than that of a satellite (LEO or geostationary orbit (GEO)) because HAPs do not require expensive launch and maintenance. HAPs, can be brought down, repaired quickly and replaced readily for reconfiguration, and may stay in the sky for a long period. Due to the large coverage area from HAP, a HAP network should be also cheaper than a terrestrial network with a large number of terrestrial base stations.

3.2 HAP system deployment

Depending on different applications, HAP are generally proposed to have three communication scenarios with integration into terrestrial or satellite systems (Karapantazis & Pavlidou, 2005).

3.2.1 Terrestrial-HAP-Satellite system

The network architecture is shown in Fig. 2. It is composed of links between HAPs, satellite and terrestrial systems. It can provide fault tolerance, and thus support a high quality of service (QoS). Broadcasting and broadband services can be delivered from the platform. Inter-platform communications can be established for extending coverage area.

3.2.2 Terrestrial-HAP system

HAPs have been suggested by ITU to provide the 3G telecommunication services. HAP system is considered to be competitive in the cost compared to deploying a number of terrestrial base stations. In the architecture shown in Fig. 3, HAPs are considered to project one or more macro cells and serve a large number of high-mobility users with low data rates. Terrestrial systems can provide service with high data rates or in areas where NLOS propagation is mostly prevailing. The HAP network can be connected to terrestrial network through a gateway. Due to its wide coverage area and competitive cost of deployment, HAPs could be employed to provide services for areas with low population density, where it could expensively deploy fibre or terrestrial networks.

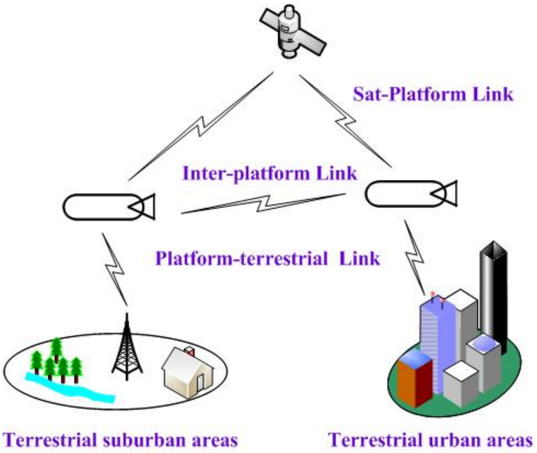


Fig. 2. Integrated Satellite-HAP-Terrestrial system

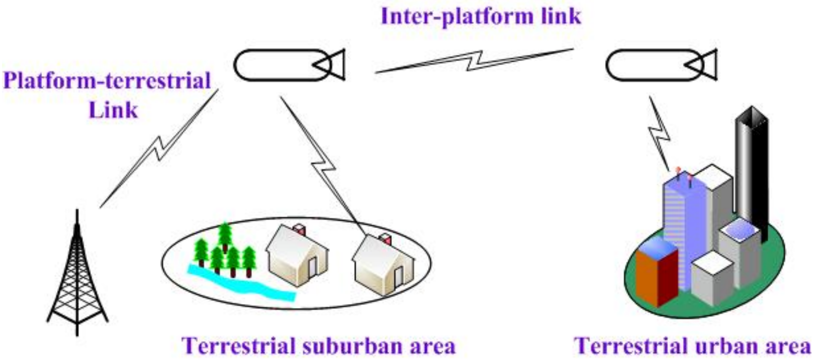


Fig. 3. HAP-Terrestrial system

3.2.3 A stand-alone HAP system

HAPs are potential to be a stand-alone system in many applications, e.g. broadband for all, environment and disaster surveillance. The architecture is shown in Fig. 4. In rural or remote areas, it is rather expensive and inefficient to deploy terrestrial systems. Furthermore, a satellite system is costly to be launched because of small traffic demand. HAPs system may be deployed economically and efficiently. A backbone link could be established by fibre network or satellites depending on applications.

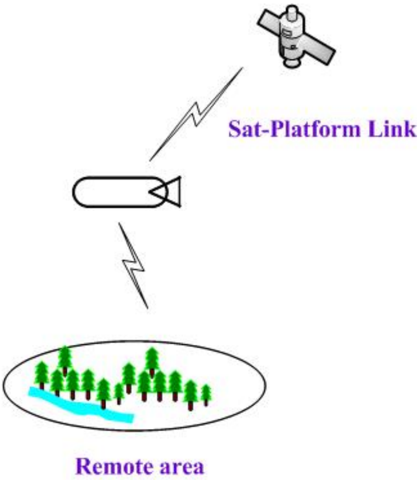


Fig. 4. A stand-alone HAP system

4. Conclusions and Future Research

In this chapter, an overview of the HAP concept development and HAP trails has been introduced to show the worldwide interest in this emerging novel technology. A comparison of the HAP system has been given based on the basic characteristics of HAP, terrestrial and satellite systems. Main advantages of HAPs for wireless communication applications in rural areas were wide coverage area, high capacity and cost-effective deployment. Three scenarios of HAP communication have been illustrated.

It is extremely beneficial to investigate other possibilities of providing mobile services from HAPs since this would provide an important supplemental HAP application under the goal "Broadband for All". Previous HAP application investigations in the CAPANINA project mainly addressed the fixed-wireless application in the mm-wave band at 30/31 GHz or even higher. Delivery of mobile services from HAPs enables HAPs to exploit the highly profitable mobile market. The IEEE802.16e standard and beyond provide both stationary and mobile services. To extend the HAP capabilities to support full operations under the WiMAX standards brings a more competitive service especially in the mobile service field. Some 3G HAP mobile communication studies have also been carried out in the 2 GHz band. High Speed Downlink Packet Access (HSDPA), which is usually regarded as an enhanced version of W-CDMA, and 3GPP Long Term Evolution (LTE) with MIMO and/or adaptive antenna systems capabilities for achieving higher data rates and improved system performance are also attractive directions for further investigations.

5. References

CAPANINA. (2004). CAPANINA project. from http://www.capanina.org/

Collela, N. J., Martin, J. N., & Akyildiz, I. F. (2000). The HALO Network. *IEEE Communications Magazine, 38*(6), 142-148.

Cost 297. (2005). Cost 297 Action Overview. 2005, from http://www.hapcos.org/overview.php

Djuknic, G. M., Freidenfelds, J., & Okunev, Y. (1997). Establishing Wireless Communications Services via High-Altitude Aeronautical Platforms: A Concept Whose Time Has Come? *IEEE Commun. Mag., 35*(9), 128-135.

ESA. (2005). Hale Aerostatic Platforms. from http://www.esa.int/SPECIALS/GSP/SEMD6EZO4HD_0.html

Flightglobal. (2009). Lockheed to build high-altitude airship http://www.flightglobal.com/articles/2009/04/30/325876/lockheed-to-build-high-altitude-airship.html.

Foo, Y. C., Lim, W. L., & Tafazolli, R. (2002, 24-28 September). *Centralized Downlink Call Admission Control for High Altitude Platform Station UMTS with Onboard Power Resource Sharing*. Vehicular Technology Conference,VTC 2002-Fall.

Grace, D., Daly, N. E., Tozer, T. C., Burr, A. G., & Pearce, D. A. J. (2001a). Providing Multimedia Communications from High Altitude Platforms. *Intern. J. of Sat. Comms.*(No 19), 559-580.

Grace, D., Mohorcic, M., Oodo, M., Capstick, M. H., Pallavicini, M. B., & Lalovic, M. (2005). *CAPANINA - Communications from Aerial Platform Networks Delivering Broadband Information for All*. Paper presented at the IST Mobile Communications Summit, Dresden, Germany

Grace, D., Thornton, J., Konefal, T., Spillard, C., & Tozer, T. C. (2001b). *Broadband Communications from High Altitude Platforms - The HeliNet Solution*. Paper presented at the Wireless Personal Mobile Conference, Aalborg, Denmark

Hong, T. C., Ku, B. J., Park, J. M., Ahn, D.-S., & Jang, Y.-S. (2005). Capacity of the WCDMA System Using High Altitude Platform Stations. *International Journal of Wireless Information Networks, 13*(1).

Hult, T., Mohammed, A., & Grace, D. (2008a). WCDMA Uplink Interference Assessment from Multiple High Altitude Platform Configurations. *EURASIP Journal on Wireless Communications and Networking, 2008*.

Hult, T., Mohammed, A., Yang, Z., & Grace, D. (2008b). Performance of a Multiple HAP System Employing Multiple Polarization. *Wireless Personal Communications*.

ITU-R. (2003, 4 July). *Final Acts (Provisional)*. ITU WRC-03.

Karapantazis, S., & Pavlidou, F. (2005). Broadband communications via high-altitude platforms: a survey. *Communications Surveys & Tutorials, IEEE, 7*(1), 2-31.

Lindstrand Balloons Ltd. (2005). Lindstrand Balloons Ltd. from http://www.lindstrand.co.uk

Mohammed, A., Arnon, S., Grace, D., Mondin, M., & Miura, R. (2008). Advanced Communications Techniques and Applications for High-Altitude Platforms. *Editorial for a Special Issue, EURASIP Journal on Wireless Communications and Networking, 2008*.

Preparedness, O. o. C. I. P. a. E. (2000). Application of High-Altitude Long Endurance (HALE) Platforms in Emergency Preparedness and Disaster Management and Mitigation.

Steele, R. (1992). Guest Editorial-an Update on Personal Communications. *IEEE Communication Magazine*, 30-31.

Thornton, J., Grace, D., Spillard, C., Konefal, T., & Tozer, T. C. (2001). Broadband Communications from a High Altitude Platform - The European HeliNet Programme. *IEE Electronics and Communications Engineering Journal, 13*(3), 138-144.

Tozer, T. C., & Grace, D. (2001). High-Altitude Platforms for Wireless Communications. *IEE Electronics and Communications Engineering Journal, 13*(3), 127-137.

Yang, Z., & Mohammed, A. (2008a). *Broadband Communication Services from Platform and Business Model Design*. Paper presented at the IEEE Pervasive Computing and Communications (PerCom) Google PhD Forum HongKong

Yang, Z., & Mohammed, A. (2008b). *On the Cost-Effective Wireless Broadband Service Delivery from High Altitude Platforms with an Economical Business Model Design*. Paper presented at the IEEE 68th Vehicular Technology Conference, 2008. VTC 2008-Fall, Calgary Marriott, Canada

4

Performance of Wireless Communication Systems with MRC over Nakagami–*m* Fading Channels

Tuan A. Tran[1] and Abu B. Sesay[2]
[1]*SNC-Lavalin T&D Inc., Canada*
[2]*The University of Calgary, Canada*

1. Introduction

The Nakagami–*m* distribution (*m*–distribution) (Nakagami, 1960) received considerable attention due to its greater flexibility as compared to Rayleigh, log-normal or Rician fading distribution (Al–hussaini & Al–bassiouni, 1985; Aalo, 1995; Annamalai et al., 1999; Zhang, 1999; Alouini et al., 2001). The distribution also includes Rayleigh and one-sided Gaussian distributions as special cases. It can also accommodate fading conditions that are widely more or less severe than that of the Rayleigh fading. Nakagami–*m* fading is, therefore, often encountered in practical applications such as mobile communications.

This chapter discusses the performance analysis of wireless communication systems where the receiver is equipped with maximal–ratio–combining (MRC), for performance improvement, in the Nakagami-*m* fading environment. In MRC systems, the combined signal–to–noise ratio (SNR) at the output of the combiner is a scaled sum of squares of the individual channel magnitudes of all diversity branches. Over Nakagami-*m* fading channels, the combined output SNR of the MRC combiner is a sum of, normally, correlated Gamma random variables (r.v.'s). Therefore, performance analysis of this diversity–combining receiver requires knowledge of the probability density function (PDF) or the moment generating function (MGF) of the combined SNR. The PDF of the sum of Gamma r.v.'s has also long been of interest in mathematics (Krishnaiah & Rao, 1961; Kotz & Adams, 1964; Moschopoulos, 1985) and many other engineering applications.

The current research progress in this area is as follows. The characteristic function (CF) of the sum of identically distributed, correlated Gamma r.v.'s is derived in (Krishnaiah & Rao, 1961) and (Kotz & Adams, 1964). Then, the PDF of the sum of statistically independent Gamma r.v.'s with non–identical parameters is derived in (Moschopoulos, 1985). The results derived in (Krishnaiah & Rao, 1961; Kotz & Adams, 1964; Moschopoulos, 1985) are used for performance analysis of various wireless communication systems in (Al–hussaini & Al–bassiouni, 1985; Aalo, 1995; Annamalai et al., 1999; Zhang, 1999; Alouini et al., 2001) and references therein. In (Win et al., 2000), the CF of a sum of arbitrarily correlated Gamma r.v.'s with non–identical but integer fading orders is derived by using a so-called virtual branch technique. This technique is also used in (Ghareeb & Abu-Surra, 2005) to derive the

CF of the sum of arbitrarily correlated Gamma r.v.'s. In (Alouini et al., 2001), using the results derived in (Moschopoulos, 1985), the PDF of the sum of arbitrarily correlated, non–identically distributed Gamma r.v.'s but with identical fading orders (both integer as well as non–integer) is derived. Performance of an MRC receiver for binary signals over Nakagami–*m* fading with arbitrarily correlated branches is analyzed in (Lombardo et al., 1999) for the case of identical fading orders *m*'s (both integer as well as non–integer). The distribution of multivariate Nakagami–*m* r.v.'s is recently derived in (Karagiannidis et al., 2003a) also for the case of identical fading orders. The joint PDF of Nakagami-*m* r.v.'s with identical fading orders using Green's matrix approximation is derived in (Karagiannidis et al., 2003b). A generic joint CF of the sum of arbitrarily correlated Gamma r.v.'s with non–identical and non-integer fading orders is recently derived in (Zhang, 2003).

For a large number of diversity branches the virtual branch technique proposed in (Win et al., 2000) has a high computational complexity since the eigenvalue decomposition (EVD) is performed over a large matrix. Although the joint CF derived in (Zhang, 2003) is very general, it does not offer an immediate simple form of the PDF and therefore analyzing some performance measures can be complicated. In this chapter, we provide some improvements over the existing results derived in (Win et al., 2000) and (Zhang, 2003). Firstly, we transform the correlated branches into multiple uncorrelated virtual branches so that the EVDs are performed over several small matrices instead of a single large matrix. Secondly, we derive the exact PDF of the sum of arbitrarily correlated Gamma r.v.'s, with non-identical and half-of-integer fading orders, in the form of a single Gamma series, which greatly simplifies the analysis of many different performance measures and systems that are more complicated to analyze by the CF– or MGF–based methods. Note that parts of this chapter are also published in (Tran & Sesay, 2007).

The chapter is organized as follows. Section 2 describes the communication signal model. We derive the MGF and PDF of the sum of Gamma r.v.'s in Section 3. In Section 4, we address the application of the derived results to performance analysis of wireless communication systems with MRC or space-time block coded (Su & Xia, 2003) receivers. Numerical results and discussions are presented in Section 5 followed by the conclusion in Section 6.

The following notations are used throughout this chapter: $E\{x\}$ denotes the statistical average of random variable x; lowercase, bold typeface letters, e.g. $\mathbf{x}$, represent vectors; uppercase, bold typeface letters, e.g. $\mathbf{X}$, represent matrices; $\triangleq$ denotes the definition; $\mathbf{I}_m$ denotes an $m \times m$ identity matrix; $\mathbf{x}^T$ and $\mathbf{X}^T$ denote the transpose of vector $\mathbf{x}$ and matrix $\mathbf{X}$, respectively; $\|\mathbf{x}\|^2 \triangleq \mathbf{x}^T\mathbf{x}$; $\lceil x \rceil$ denotes the smallest integer greater than or equal x; $P(\cdot|x)$ denotes the statistical conditional function given random variable x; $j \triangleq \sqrt{-1}$ denotes the complex imaginary unit; $|x|^2 \triangleq xx^*$, where x^* denotes the complex conjugate of x; $Q(x)$ denotes the Q-function, defined as $Q(x) \triangleq (1/\sqrt{2\pi})\int_x^\infty \exp(-z^2/2)dz$, and $\mathrm{erfc}(x)$ denotes the complementary error function, defined as $\mathrm{erfc}(x) \triangleq (2/\sqrt{\pi})\int_x^\infty \exp(-z^2)dz$; $x = f^{-1}(y)$ denotes the inverse function of function $y = f(x)$.

2. Communication Signal Model

Consider a wireless communication system equipped with one transmit antenna and L receive antennas and assume perfect channel estimation is attained at the receiver. The low-pass equivalent received signal at the kth receive antenna at time instant t is expressed by

$$r_k(t) = \alpha_k(t)e^{j\varphi_k(t)}s(t) + w_k(t), \tag{1}$$

where $\alpha_k(t)$ is an amplitude of the channel from the transmit antenna to the kth receive antenna. In (1), $\alpha_k(t)$ is an m- Nakagami distributed random variable (Nakagami, 1960), $\varphi_k(t)$ is a random signal phase uniformly distributed on $[0,2\pi)$, $s(t)$ is the transmitted signal that belongs to a signal constellation Ξ with an averaged symbol energy of $\overline{E}_s \triangleq E\{|s(t)|^2\}$, and $w_k(t)$ is an additive white Gaussian noise (AWGN) sample with zero mean and variance σ_w^2. The overall instantaneous combined SNR at the output of the MRC receiver is then given by

$$\eta(t) \triangleq \frac{\overline{E}_s}{\sigma_w^2}\sum_{k=1}^{L}\alpha_k^2(t) = \frac{\overline{E}_s}{\sigma_w^2}\gamma(t), \tag{2}$$

where $\gamma(t) \triangleq \sum_{k=1}^{L} x_k(t)$ with $x_k(t)$ being defined as $x_k(t) \triangleq \alpha_k^2(t)$. From now on the time index t is dropped for brevity. Since α_k is an m- Nakagami distributed random variable, the marginal PDF of x_k is a Gamma distribution given by (Proakis, 2001)

$$p_{X_k}(x_k) = \frac{x_k^{m_k-1}}{\Gamma(m_k)}\left(\frac{m_k}{\Omega_k}\right)^{m_k}\exp\left(-\frac{m_k x_k}{\Omega_k}\right), \tag{3}$$

where

$$\Omega_k = E\{x_k\} \text{ and } m_k = \Omega_k^2 / E\{(x_k - \Omega_k)^2\} \geq 1/2. \tag{4}$$

In (4), the Ω_k's and m_k's are referred to as fading parameters in which the m_k's are referred to as *fading orders*, and $\Gamma(\cdot)$ is the Gamma function (Gradshteyn & Ryzhik, 2000). Finding the PDF or MFG of $\gamma \triangleq \gamma(t)$, which is referred to as the received SNR coefficient, is essential to the performance analysis of diversity combining or space-time block coded receivers of wireless communication systems which is addressed in this chapter.

3. Derivation of the Exact MGF and PDF of γ

3.1 Moment Generating Function

In this section, we derive the MGF of γ for the case $m_k = n_k / 2$ with n_k being an integer and $n_k \geq 1$. First, without loss of generality, assume that the x_k's are indexed in increasing fading orders m_k's, i.e., $m_1 \leq m_2 \leq \cdots \leq m_L$. Let $\mathbf{z}_k$ denote a $2m_k \times 1$ vector defined as $\mathbf{z}_k \triangleq [z_{k,1}, z_{k,2}, \ldots, z_{k,2m_k}]^T$, $k = 1, \ldots, L$, where the $z_{k,i}$'s are independently and identically distributed zero-mean real Gaussian random variables with variances of $E\{z_{k,i}^2\} = \Omega_k / 2m_k$. The random variables x_k's, $1 \le k \le L$, are then constructed by $x_k \triangleq \sum_{i=1}^{2m_k} z_{k,i}^2 = \|\mathbf{z}_k\|^2$. Therefore, the received SNR coefficient γ is expressed by $\gamma \triangleq \sum_{k=1}^{L} \|\mathbf{z}_k\|^2$. Following (Win et al., 2000), the elements of the vectors $\mathbf{z}_k$'s, $k = 1, \ldots, L$, are constructed such that their correlation coefficients are given by

$$E\{z_{k,i} z_{l,w}\} = \begin{cases} \Omega_k / 2m_k, & \text{if } i = w \text{ and } k = l \\ \rho_{k,l} \sqrt{\Omega_k \Omega_l / 4m_k m_l}, & \text{if } k \ne l \text{ but } i = w \text{ and } 1 \le i, w \le 2\min\{m_k, m_l\} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

and $0 \le \rho_{k,l} \le 1$. Here, $\rho_{k,l}$ is the normalized correlation coefficient between $z_{k,i}$ and $z_{l,w}$. The correlation coefficient between two branches, x_k and x_l, is related to $\rho_{k,l}$ though

$$\begin{aligned} \rho_{x_k x_l} &\triangleq \frac{E\{(x_k - \Omega_k)(x_l - \Omega_l)\}}{\sqrt{\operatorname{var}(x_k)\operatorname{var}(x_l)}} \\ &= \rho_{k,l}^2 \sqrt{\frac{\min(m_k, m_l)}{\max(m_k, m_l)}}. \end{aligned} \tag{6}$$

Further analysis is complicated by the fact that $\rho_{k,l} \ne 0$ even for some $l \ne k$. However, we observe from (5) that the correlation coefficient $\rho_{k,l} = 0$ for both $l = k$ and $l \ne k$ as long as $w \ne i$. We exploit this fact to rearrange the r.v.'s in the received SNR coefficient γ as follows. Let $\mathbf{v}_w$ denote an $L_w \times 1$ ($1 \le L_w \le L$ with $L_1 = L$) vector, which is defined as $\mathbf{v}_w \triangleq [z_{L-L_w+1,w}, z_{L-L_w+2,w}, \ldots, z_{L,w}]^T$ for $w = 1, 2, \ldots, 2m_L$, where the vector length L_w depends on the fading order m_w. The indexing is selected such that if $L - L_w + 1 > 2m_w$ then $z_{w,i} = 0$ and is removed from the vector $\mathbf{v}_w$. Also, let γ_w's denote *new* r.v.'s defined by $\gamma_w \triangleq \sum_{i=L-L_w+1}^{L} z_{i,w}^2 = \|\mathbf{v}_w\|^2$ for $w = 1, 2, \ldots, 2m_L$. Therefore, the random variables γ_w's are formed by summing all the wth elements of the random variables x_k's, $k = 1, 2, \ldots, L$. From (5), we note that the r.v.'s $z_{k,i}$ and $z_{l,w}$ are uncorrelated if $i \ne w$. Furthermore, since the r.v.'s $z_{k,i}$ and $z_{l,w}$ are Gaussian by definition, they are also statistically independent if $i \ne w$. Consequently, the newly formed r.v.'s γ_w*'s* are also statistically independent. From the definitions of the vectors $\mathbf{z}_k$ and $\mathbf{v}_w$, we have

$$\gamma = \sum_{w=1}^{2m_L} \|\mathbf{v}_w\|^2 = \sum_{w=1}^{2m_L} \gamma_w. \tag{7}$$

In the sum of γ, we have grouped the wth, $w = 1, 2, \ldots, 2m_L$, elements of x_k, $k = 1, 2, \ldots, L$, together so that different groups in the sum are statistically independent. Therefore, such a rearrangement of the elements of the Gamma random variables in the sum of γ actually transforms L correlated branches into $2m_L$ independent branches. The wth new independent branch is a sum of L_w correlated Gamma variables with a *common* fading order of 0.5. Let $\Phi_{\gamma_w}(s)$ denote the MGF of γ_w. Since the r.v.'s γ_w's are statistically independent, we have

$$\Phi_{\gamma}(s) = \prod_{w=1}^{2m_L} \Phi_{\gamma_w}(s). \tag{8}$$

Let $\mathbf{R}_{\mathrm{V},w}$ denote the correlation matrix of vector $\mathbf{v}_w$, where the $(k,l)^{\text{th}}$ element of $\mathbf{R}_{\mathrm{V},w}$ can be shown to be (Win et al., 2000)

$$\begin{aligned} \mathbf{R}_{\mathrm{V},w}(k,l) &\triangleq \frac{E\left\{\left(z_{kk,w}^2 - \frac{\Omega_{kk}}{2m_{kk}}\right)\left(z_{ll,w}^2 - \frac{\Omega_{ll}}{2m_{ll}}\right)\right\}}{\sqrt{\operatorname{var}\left(z_{kk,w}^2\right)\operatorname{var}\left(z_{ll,w}^2\right)}} \\ &= \rho_{kk,ll}^2, \end{aligned} \tag{9}$$

where $kk \triangleq L - L_w + k$, $ll \triangleq L - L_w + l$ for $k,\ l = 1,\ 2,\ldots,\ L_w$, and $\rho_{k,k}^2 = 1$. Since $\mathbf{R}_{\mathrm{V},1}$ is an $L \times L$ matrix, from the construction given in (5) and the definition of vector $\mathbf{v}_w$, we have

$$\mathbf{R}_{\mathrm{V},w} = \mathbf{R}_{\mathrm{V},1}(L - L_w + 1 : L, L - L_w + 1 : L),\ w = 2,3,\ldots,2m_L. \tag{10}$$

The Matlab notation $\mathbf{R}_{\mathrm{V},1}(k:l, m:n)$, denotes a sub matrix of the matrix $\mathbf{R}_{V,1}$ whose rows and columns are, respectively, the kth through lth rows and the mth through nth columns of the matrix $\mathbf{R}_{V,1}$. Let $\mathbf{\Theta}_w$ be an $L_w \times L_w$ positive definite matrix (i.e., its eigenvalues are positive) defined by

$$\mathbf{\Theta}_w = \operatorname{diag}\left(\frac{\Omega_{L-L_w+1}}{m_{L-L_w+1}}, \frac{\Omega_{L-L_w+2}}{m_{L-L_w+2}}, \ldots, \frac{\Omega_L}{m_L}\right)\sqrt{\mathbf{R}_{\mathrm{V},w}}, \tag{11}$$

where the square root operation in (11) implies taking the square root of each and every element of the matrix $\mathbf{R}_{\mathrm{V},w}$. The joint characteristic function (CF) of vector the $\mathbf{v}_w$ is given by (Krishnaiah & Rao, 1961; Kotz & Adams, 1964; Lombardo et al., 1999)

$$\begin{aligned} \psi_{\mathbf{v}_w}(t_1,\ldots,t_{L_w}) &= E\left\{\exp\left(j\textstyle\sum_{i=1}^{L_w} z_{L-L_w+i}^2 t_i\right)\right\} \\ &= \left[\det\left(\mathbf{I}_{L_w} - j\mathbf{T}_w\mathbf{\Theta}_w\right)\right]^{-1/2}, \end{aligned} \tag{12}$$

where $\mathbf{T}_w \triangleq \operatorname{diag}(t_1,\ldots,t_{L_w})$. Let $\{\lambda_{w,i} > 0\}_{i=1}^{L_w}$ denote the set of eigenvalues of the matrix $\mathbf{\Theta}_w$. Using (12), the CF of the r.v. γ_w is given by (Krishnaiah & Rao, 1961; Kotz & Adams, 1964)

$$\psi_{\gamma_w}(t) = \prod_{i=1}^{L_w}\left(1 - jt\lambda_{w,i}\right)^{-1/2}. \tag{13}$$

Therefore, the MGF of γ_w is given by

$$\Phi_{\gamma_w}(s) = \prod_{i=1}^{L_w} \left(1 - s\lambda_{w,i}\right)^{-1/2}. \tag{14}$$

Substituting (14) into (8) gives

$$\Phi_{\gamma}(s) = \prod_{w=1}^{2m_L} \prod_{i=1}^{L_w} \left(1 - s\lambda_{w,i}\right)^{-1/2}. \tag{15}$$

Remark: By working with the vectors $\mathbf{v}_w$'s instead of the vectors $\mathbf{z}_k$'s, we only deal with the sum of correlated Gamma random variables with an *identical fading order* of 0.5, which greatly simplifies the analysis. In the proposed method, the most computationally complex EVD's are performed over $L \times L$ matrices; all other EVD's are performed over $L_w \times L_w$ matrices with $L_w < L$. Therefore, the improvement over (Win et al., 2000) is that the EVD proposed in (Win et al., 2000) is performed over a $D_T \times D_T$ matrix, where $D_T \triangleq \sum_{k=1}^{L} 2m_k$. When the number of branches, L, is large and the m_k's are greater than 1, we have $D_T \gg L$ and thus the EVD in (Win et al., 2000) is much more computationally complex than the EVD's proposed in this chapter.

3.2 Probability Density Function

The PDF of γ, which is denoted by $p_\gamma(\gamma)$, is also desired for the cases where performance is harder to analyze by the MGF- or CF-based methods. The PDF of the sum of statistically independent, non-identically distributed Gamma random variables is derived in (Moschopoulos, 1985). This PDF is in the form of a single Gamma series and is thus desired by the performance analysis. Here, we apply the MGF derived in (15) for the case of *non-identical, non-integer* m_k's and *non-identical* Ω_k's to Eq. (2.9) in (Moschopoulos, 1985) to obtain the PDF of γ. First, let us re-write (15) as

$$\begin{aligned}\Phi_{\gamma}(s) &= \prod_{w=1}^{2m_L} \prod_{i=1}^{L_w} \left(1 - s\lambda_{w,i}\right)^{-1/2} \\ &= \prod_{i=1}^{M} \left(1 - s\lambda_i\right)^{-1/2},\end{aligned} \tag{16}$$

where $M \triangleq \sum_{w=1}^{2m_L} L_w$. The eigenvalues λ_i's in (16) are defined as

$$\lambda_i \triangleq \lambda_{w,i-L^{(w)}}, \quad \text{for} \quad L^{(w)} + 1 \le i \le L^{(w)} + L_w, \quad w = 1,\ 2,\ldots,\ 2m_L, \tag{17}$$

where $L^{(w)} \triangleq \sum_{k=1}^{w-1} L_k$ with $L^{(1)} = 0$. The MGF given in (16) is now in a form that can readily be applied to Eq. (2.9) in (Moschopoulos, 1985) to obtain the exact PDF of γ, which is given by

$$p_\gamma(\gamma)=c\sum_{k=0}^{\infty}\frac{\delta_k\gamma^{A_k-1}e^{-\gamma/\lambda_1}}{\Gamma(A_k)\lambda_1^{A_k}},\quad \gamma>0 \\ =0,\ \text{otherwise,} \tag{18}$$

where $A_k \triangleq k+M/2$, $c\triangleq\prod_{i=1}^{M}(\lambda_1/\lambda_i)^{1/2}$ and δ_k's are recursively computed by

$$\delta_{k+1}=\frac{1}{2(k+1)}\sum_{w=1}^{k+1}\delta_{k+1-w}\sum_{i=1}^{M}\left[1-(\lambda_1/\lambda_i)\right]^{w},\quad k=0,\ 1,\ 2,\ldots \tag{19}$$

with $\delta_0=1$. In (18), we assume, without any loss of generality, that $\lambda_1=\min_i\{\lambda_i\}$. If $\lambda_1\neq\min_i\{\lambda_i\}$, we can simply find the minimum eigenvalue and put it at the first position. A low-complexity computation of the parameters δ_{k+1} is discussed in Appendix I.

For practical numerical evaluations, it is desired to obtain a truncated version of (18) and the associated truncation error. The PDF in (18) can be truncated to give

$$p_\gamma(\gamma,K)=c\sum_{k=0}^{K}\frac{\delta_k\gamma^{A_k-1}e^{-\gamma/\lambda_1}}{\Gamma(A_k)\lambda_1^{A_k}},\ \gamma>0 \tag{20}$$

and 0 elsewhere. By applying the upper bound given by Eq. (2.12) in (Moschopoulos, 1985) for $p_\gamma(\gamma)$, the associated truncation error produced by (20) is upper-bounded by

$$\begin{aligned} e_K(\gamma) &\triangleq p_\gamma(\gamma)-p_\gamma(\gamma,K) \\ &\leq\frac{c\gamma^{\frac{M}{2}-1}e^{-\gamma(1-g)/\lambda_1}}{\lambda_1^{M/2}\Gamma\left(\frac{M}{2}\right)}-p_\gamma(\gamma,K), \end{aligned} \tag{21}$$

where $g\triangleq\max_{2\leq i\leq M}\left[1-(\lambda_1/\lambda_i)\right]$. From (21), we can choose K such that the error is in the desired regime.

Remark: *The upper bound of the PDF given by Eq. (2.12) in (Moschopoulos, 1985) is attained when* $\lambda_1=\cdots=\lambda_M$ *and is tight when* $\chi\triangleq\max_k\{\lambda_k\}/\min_k\{\lambda_k\}\approx1$. *However, when* $\chi\gg1$, *this upper bound becomes extremely loose and thus* (21) *cannot be used to determine* K *for a good truncation.*

In the case when χ is large, we propose that K be determined as follows. For a specific K, using Eq. (3.462–9) in (Gradshteyn & Ryzhik, 2000) the error of the area under the PDF due to truncation is given by

$$\begin{aligned} E_{er}(K) &\triangleq\int_0^{\infty}p_\gamma(\gamma)d\gamma-\int_0^{\infty}p_\gamma(\gamma,K)d\gamma \\ &=1-\int_0^{\infty}p_\gamma(\gamma,K)d\gamma \\ &=1-c\sum_{k=0}^{K}\delta_k. \end{aligned} \tag{22}$$

It is pointed out in (Moschopoulos, 1985) that the interchange of the integration and summation in (22) is justified due to the uniform convergence of $p_\gamma(\gamma)$. For a pre–

determined threshold of error $E_{er}(K) \le \varepsilon$, we can easily choose K from (22) such that this condition is satisfied. The advantage of this method compared to the bounding method derived in (Moschopoulos, 1985), given in (21), is that K can be determined for an arbitrary χ.

Note that the PDF derived in (18) is an extension of the PDF given by Eq. (5) derived in (Alouini et al., 2001) to the case of *non–identical* fading orders m_k's. The MGF and PDF of γ derived in (15) and (20), respectively, can be used for general performance analysis in wireless communication systems such as (a) outage probability, (b) bit error probability and (c) Shannon capacity analysis as shown in (Alouini et al., 2001).

4. Application to Performance Analysis of MRC Systems

4.1 Preliminaries

Bit error probability (BEP) analysis can be performed using either the PDF or MGF of the received SNR coefficient at the output of the MRC combiner.

Method 1**: **BEP analysis using PDF of the received SNR coefficient. It is shown in (Lu et al., 1999) that the conditional BEP, given the received SNR coefficient γ, of an $\bar{M}$-PSK or $\bar{M}$-QAM modulated system, is a function of

$$P_b(\xi \mid \gamma) \triangleq \operatorname{erfc}\left(\sqrt{\xi\gamma}\right), \tag{23}$$

where $\operatorname{erfc}(x)$ is the complementary error function (Gradshteyn & Ryzhik, 2000), ξ is a deterministic variable determined from the unfaded received SNR $\bar{E}_s / \sigma_w^2$ at each branch and the digital modulation scheme used, which will be discussed in more details. In (23), $P_b(\xi \mid \gamma)$ is considered as elementary conditional BEP based upon which the overall conditional BEP is calculated. First, using the integral representation of $\operatorname{erfc}(\sqrt{\xi\gamma})$ given by Eq. (7.4.11) in (Abramowitz & Stegun, 1972) and a change of variable, we arrive at

$$P_b(\xi \mid \gamma) = \tfrac{2}{\pi}\int_0^{\infty}\left(z^2+1\right)^{-1}\exp\left[-\xi\gamma(z^2+1)\right]dz. \tag{24}$$

Then, using the PDF $p_\gamma(\gamma)$ of the received SNR coefficient γ, the statistical average of the elementary BEP of the receiver is then computed by

$$\begin{aligned} P_b(\xi) &\triangleq E\left\{P_b(\xi \mid \gamma)\right\} \\ &= \tfrac{2}{\pi}\int_0^{\infty}\int_0^{\infty}\left(z^2+1\right)^{-1}\exp\left[-\xi\gamma(z^2+1)\right]p_\gamma(\gamma)dzd\gamma. \end{aligned} \tag{25}$$

Method 2: BEP analysis using MGF of the received SNR coefficient. The elementary BEP in (25) can also be manipulated as

$$\begin{aligned} P_b(\xi) &= \tfrac{2}{\pi}\int_0^{\infty}\left(\frac{1}{z^2+1}\right)E\left\{\exp[-\xi\gamma(z^2+1)]\right\}\,dz \\ &= \tfrac{2}{\pi}\int_0^{\infty}\left(\frac{1}{z^2+1}\right)\Phi_\gamma\left(-\xi\left(z^2+1\right)\right)\,dz. \end{aligned} \tag{26}$$

If a closed-form solution to (26) is not available, we can resort to numerical analysis with a high degree of accuracy using the well-known Gaussian-Chebyshev Quadrature (GCQ) given in (Abramowitz & Stegun, 1972). First, apply the substitution $y=(1-z^2)/(1+z^2)$ to (26) and then use an N-point integral GCQ, we arrive at

$$\begin{aligned} P_b(\xi) &= \frac{1}{N}\sum_{n=1}^{N}\Phi_\gamma\left(-\frac{2\xi}{\cos(\theta_n)+1}\right)+R_N(\xi) \\ &= \frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{M}\left(1+\frac{2\xi\lambda_i}{\cos(\theta_n)+1}\right)^{-1/2}+R_N(\xi), \end{aligned} \tag{27}$$

where $\theta_n \triangleq [(2n-1)\pi/2N]$ and $R_N(\xi)$ is the remainder given in (Abramowitz & Stegun, 1972). It is shown in (Annamalai et al., 1999; Zhang, 1999) and references therein that using (27) is very accurate even with only a small N.

4.2 Bit Error Probability Analysis

Method 1: BEP analysis using the PDF $p_\gamma(\gamma)$. Application of the PDF of the received SNR coefficient at the output of the MRC combiner $p_\gamma(\gamma)$, given in (18), to (25) gives the elementary BEP of the receiver as

$$P_b(\xi)=\frac{2c}{\pi}\sum_{k=0}^{\infty}\Delta_k\int_0^{\infty}\int_0^{\infty}(z^2+1)^{-1}\gamma^{A_k-1}\exp\left\{-\gamma\left[\xi(z^2+1)+1/\lambda_1\right]\right\}d\gamma dz, \tag{28}$$

where $\Delta_k \triangleq \{\delta_k/[\Gamma(A_k)\lambda_1^{A_k}]\}$. Like (22), the interchange of the integration and summation in (28) is justified due to the uniform convergence of $p_\gamma(\gamma)$ as shown in (Moschopoulos, 1985). Then, by using Eq. (3.462-.9) and Eq. (3.259-.3) in (Abramowitz & Stegun, 1972), we have

$$P_b(\xi)=c\sum_{k=0}^{\infty}f(k,\xi)B\left(\frac{1}{2},A_k+\frac{1}{2}\right){}_2F_1\left(A_k,\frac{1}{2};A_k+1,\frac{1}{1+\lambda_1\xi}\right), \tag{29}$$

where $B(\cdot,\cdot)$ is the beta function, ${}_2F_1(\cdot,\cdot\,;\,\cdot,\cdot)$ is the Gauss hypergeometric function given in (Gradshteyn & Ryzhik, 2000), and $f(k,\xi)\triangleq\delta_k/[\pi(1+\xi\lambda_1)^{A_k}]$. Since $1>1/(1+\lambda_1\xi)$, the Gauss hypergeometric function ${}_2F_1(\cdot,\cdot\,;\,\cdot,\cdot)$ in (29) converges. We then invoke the results derived in (Lu et al., 1999) and use the relationship $Q(x)=0.5\mathrm{erfc}(x/\sqrt{2})$ to obtain, after straightforward manipulations, the conditional BEP given the received SNR coefficient γ of a coherent $\bar{M}$-PSK modulated system with MRC receiver as

$$\begin{aligned} P_b^{PSK}(\cdot \mid \gamma) &\cong a_P \sum_{k=1}^{K_P} \operatorname{erfc}\left(\sqrt{\xi_k \gamma}\right) \\ &= a_P \sum_{k=1}^{K_P} P_b(\xi_k \mid \gamma), \end{aligned} \tag{30}$$

where $\bar{M}$ is the size of the signal constellation, $a_P \triangleq 1/\max(b,2)$ with $b \triangleq \log_2 \bar{M}$ is the number of information bits per modulated symbol, deterministic variable ξ_k, as mentioned in (23), is defined by the branch unfaded received SNR $\bar{E}_s / \sigma_w^2$ and the modulation scheme as $\xi_k \triangleq \left\{\bar{E}_s \sin^2[(2k-1)\pi / \bar{M}] / \sigma_w^2\right\}$, and $K_P \triangleq \left\lceil \max(\bar{M}/4,1) \right\rceil$. The statistical average of (30) is finally computed by

$$\begin{aligned} \bar{P}_b^{PSK} &\triangleq E\left\{P_b^{PSK}(\cdot \mid \gamma)\right\} \\ &\cong E\left\{a_P c \sum_{k=1}^{K_P} P_b(\xi_k \mid \gamma)\right\} \\ &= a_P c \sum_{k=1}^{K_P} E\left\{P_b(\xi_k \mid \gamma)\right\} \\ &= a_P c \sum_{k=1}^{K_P} P_b(\xi_k), \end{aligned} \tag{31}$$

where $P_b(\xi_k \mid \gamma)$ and $P_b(\xi_k)$ are defined in (24) and (25), respectively. Note that from (31) we can also easily obtain the BEP of a coherent FSK modulated system by setting $\bar{M} = 2$ and replacing ξ_k with $\xi_k/2$. Similarly, invoking the results derived in (Lu et al., 1999) gives the approximation of conditional BEP, given the received SNR coefficient γ, of a coherent, square $\bar{M}$-QAM modulated system with MRC receiver as

$$\begin{aligned} P_b^{QAM}(\cdot \mid \gamma) &\cong a_Q \sum_{k=1}^{K_Q} \operatorname{erfc}\left(\sqrt{\zeta_k \gamma}\right) \\ &= a_Q \sum_{k=1}^{K_Q} P_b(\zeta_k \mid \gamma), \end{aligned} \tag{32}$$

where $a_Q \triangleq [2(\sqrt{\bar{M}}-1)/(b\sqrt{\bar{M}})]$, $K_Q \triangleq \left\lceil \sqrt{\bar{M}}/2 \right\rceil$ and $\zeta_k \triangleq \{3\bar{E}_s(2k-1)^2 / [2\sigma_w^2(\bar{M}-1)]\}$. Therefore, the statistical average BEP of a square $\bar{M}$-QAM modulated system is approximated by

$$\begin{aligned}\overline{P}_b^{QAM} &\triangleq E\left\{P_b^{QAM}(\cdot\mid\gamma)\right\}\\ &\cong E\left\{a_Q c\sum_{k=1}^{K_Q} P_b(\zeta_k\mid\gamma)\right\}\\ &= a_Q c\sum_{k=1}^{K_Q} P_b(\zeta_k),\end{aligned} \tag{33}$$

Where, again, $P_b(\zeta_k\mid\gamma)$ and $P_b(\zeta_k)$ are defined in (24) and (25), respectively.

Method 2: BEP analysis using the MGF $\Phi_\gamma(s)$. Consider now the option of using the MGF $\Phi_\gamma(s)$ to analyze the BEP of the MRC receiver. Substituting (16) into (27) gives the elementary BEP as

$$P_b(\xi)=\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{M}\left(1+\frac{2\xi\lambda_i}{\cos(\theta_n)+1}\right)^{-1/2}+R_N(\xi). \tag{34}$$

Therefore, for an $\bar{M}$-PSK modulated system with MRC receiver, the overall BEP is approximated by

$$\overline{P}_b^{PSK}\cong\frac{a_P}{N}\sum_{k=1}^{K_P}\sum_{n=1}^{N}\prod_{m=1}^{M}\left(1+\frac{2\xi_k\lambda_m}{\cos(\theta_n)+1}\right)^{-1/2}+R_P, \tag{35}$$

where λ_m's are defined in (17) and $R_P\triangleq a_P\sum_{k=1}^{K_P}R_N(\xi_k)$ is the overall remainder of the *N*-point GCQ operation. Similarly, for a square $\bar{M}$-QAM modulated system with MRC receiver, the overall BEP is approximated by

$$\overline{P}_b^{QAM}\cong\frac{a_Q}{N}\sum_{k=1}^{K_Q}\sum_{n=1}^{N}\prod_{m=1}^{M}\left(1+\frac{2\zeta_k\lambda_m}{\cos(\theta_n)+1}\right)^{-1/2}+R_Q, \tag{36}$$

where $R_Q\triangleq a_Q\sum_{k=1}^{K_Q}R_N(\zeta_k)$ is the overall remainder of the *N*-point GCQ operation.

4.3 Outage Probability Analysis

Outage probability is also a useful performance measure of the receiver. The outage probability P_{out} is defined as the probability that the BEP P_b exceeds a BEP threshold $\overline{P}_0$ or equivalently the probability that the overall instantaneous SNR η falls below a pre-determined SNR threshold, say $\overline{\eta}_0$, that is

$$\begin{aligned}P_{\text{out}}&\triangleq\Pr\left\{P_b\geq\overline{P}_0\right\}=\Pr\left\{\eta\leq\overline{\eta}_0\right\}\\ &=\Pr\left\{\gamma\leq\overline{\eta}_0\sigma_w^2/\overline{E}_s\right\},\end{aligned} \tag{37}$$

since $\eta\triangleq\gamma\overline{E}_s/\sigma_w^2$ as defined in (2), where $\overline{\eta}_0\triangleq f^{-1}(\overline{P}_0)$. Therefore, we have

$$P_{out} = \int_0^{\overline{\eta}} p_\gamma(\gamma)d\gamma, \tag{38}$$

where $\overline{\eta} \triangleq \overline{\eta}_0 \sigma_w^2 / \overline{E}_s$. Substituting (18) into (38) and using Eq. (3.381-1) in (Gradshteyn & Ryzhik, 2000) gives

$$\begin{aligned} P_{out} &= c\sum_{k=0}^{\infty} \frac{\delta_k}{\Gamma(A_k)}\left[\Gamma(A_k) - \Gamma(A_k, \overline{\eta}/\lambda_1)\right] \\ &= 1 - c\sum_{k=0}^{\infty} \frac{\delta_k \Gamma(A_k, \overline{\eta}/\lambda_1)}{\Gamma(A_k)}, \end{aligned} \tag{39}$$

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function (Gradshteyn & Ryzhik, 2000). A truncation version of (39) can be obtained by using (20) for practical computation purposes. Note that (39) is similar to the result derived in (Alouini et al., 2001), but for the case of *non-identical* and *non-integral* fading orders m_k's. A number of other performance measures as discussed in (Alouini et al., 2001) can also be easily analyzed using the PDF and MGF derived in this Chapter.

4.4 Application to Performance Analysis of Space-Time Block Coded Systems

In this section, we apply the derived results to analyze the performance of orthogonal space-time block coded (OSTBC) (Su & Xia, 2003) system over quasi-static frequency-flat Nakagami-*m* fading channels. Consider an OSTBC system equipped with N_t transmit-antennas and N_r receive-antennas with perfect channel estimation at the receiver. At the receiver, after a simple linear combining operation, the combined received signal, denoted by r_k, for the *k*th code symbol, denoted by s_k, is equivalent to

$$r_k = \left(\sum_{i=1}^{N_t}\sum_{l=1}^{N_t} |\alpha_{il}|^2\right) s_k + \sum_{i=1}^{N_t}\sum_{l=1}^{N_r} \alpha_{il} w_{il}, \tag{40}$$

where α_{il}'s are channel fading gains from the *i*th transmit-antenna to the *l*th receive-antenna, which are Nakagami-*m* distributed random variables with fading parameters m_{il}'s and Ω_{il}'s, and w_{il}'s are AWGN samples with a zero-mean and a variance σ_w^2. The fading orders m_{il}'s are integers or half of integers. The SNR of r_k, given the channel fading gains α_{il}'s, is then computed by

$$\eta_{r,k} = \frac{\overline{E}_s}{N_t \sigma_w^2} \sum_{i=1}^{N_t}\sum_{l=1}^{N_r} |\alpha_{il}|^2. \tag{41}$$

The overall received SNR $\eta_{r,k}$ in (41) is a sum of correlated Gamma random variables, which is in the form of the received SNR given in (2). Therefore, the application of the results derived in this chapter to performance analysis of OSTBC systems over quasi-static frequency-flat Nakagami-*m* fading channels is immediate.

5. Numerical Results and Discussions

5.1 Exponentially Correlated Branches

It is shown in (Lee, 1993) and (Zhang, 1999) that it is reasonable to place two adjacent antennas such that their correlation is from 0.6 to 0.7. Therefore, we choose the power-correlation between two adjacent antennas to be 0.6 and the correlations amongst the antennas follow the exponential rule, i.e., $\rho_{x_k x_l} = 0.6^{|k-l|}$, for this example. An example for an arbitrary case is shown in the next subsection. For this example, we consider an MRC receiver with $L = 4$ diversity branches and branch power correlation matrix $\mathbf{R_{X1}}$ in (42). Let $\mathbf{m}_k = [m_{k,1}, m_{k,2}, m_{k,3}, m_{k,4}]^T$ and $\mathbf{\Omega}_k = [\Omega_{k,1}, \Omega_{k,2}, \Omega_{k,3}, \Omega_{k,4}]^T$ denote the fading parameter vectors. We consider three cases with the following fading parameters:

Case 1: $\mathbf{m}_1 = [0.5, 0.5, 0.5, 0.5]^T$ and $\mathbf{\Omega}_1 = [0.85, 1.21, 0.92, 1.12]^T$,

Case 2: $\mathbf{m}_2 = [0.5, 1.0, 1.5, 2]^T$ and $\mathbf{\Omega}_2 = [1.15, 1, 0.92, 1.2]^T$, and

Case 3: $\mathbf{m}_3 = [1.5, 2.5, 3, 3.5]^T$ and $\mathbf{\Omega}_3 = [1.35, 1, 0.95, 1.15]^T$.

$$\mathbf{R_{X1}} = \begin{pmatrix} 1 & 0.6 & 0.36 & 0.216 \\ 0.6 & 1 & 0.6 & 0.36 \\ 0.36 & 0.6 & 1 & 0.6 \\ 0.216 & 0.36 & 0.6 & 1 \end{pmatrix}. \tag{42}$$

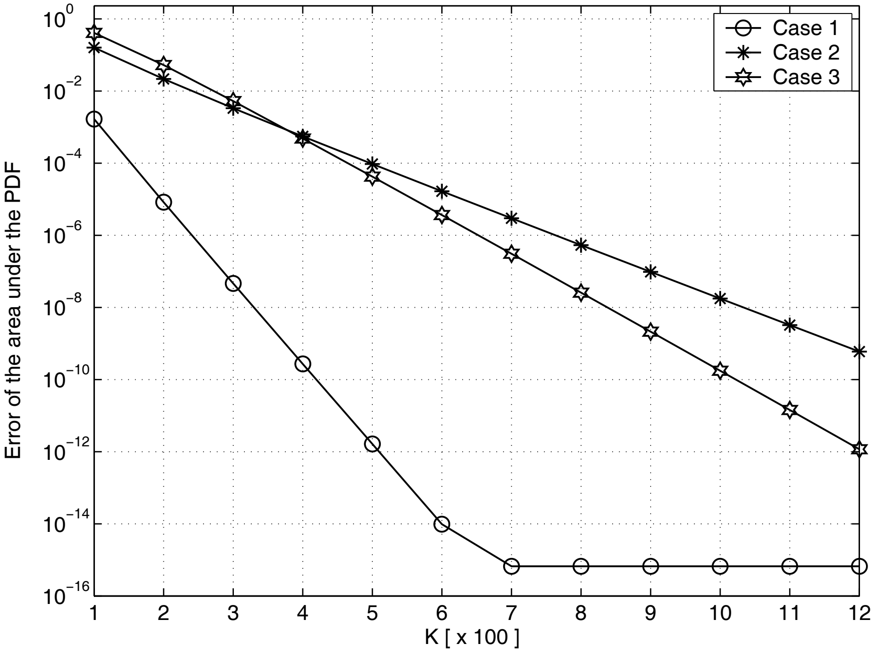


Fig. 1. The error of the areas under the PDFs, $E_{er}(K)$, due to truncation to $K+1$ terms with the branch power-correlation matrix $\mathbf{R_{X1}}$.

Note that these three cases do not necessarily reflect any particular practical system parameters. They, rather, demonstrate the generality of the derived results. These three cases cover a very wide range of fading conditions, from worse to better than the Rayleigh fading cases. They also represent a wide range of the variations of fading severities and gain imbalances among diversity branches. Using (6) and (9), the correlation matrix $\mathbf{R}_{\mathrm{V},1}$ relates to the branch power–correlation matrix $\mathbf{R}_{\mathbf{X1}}$ through (43).

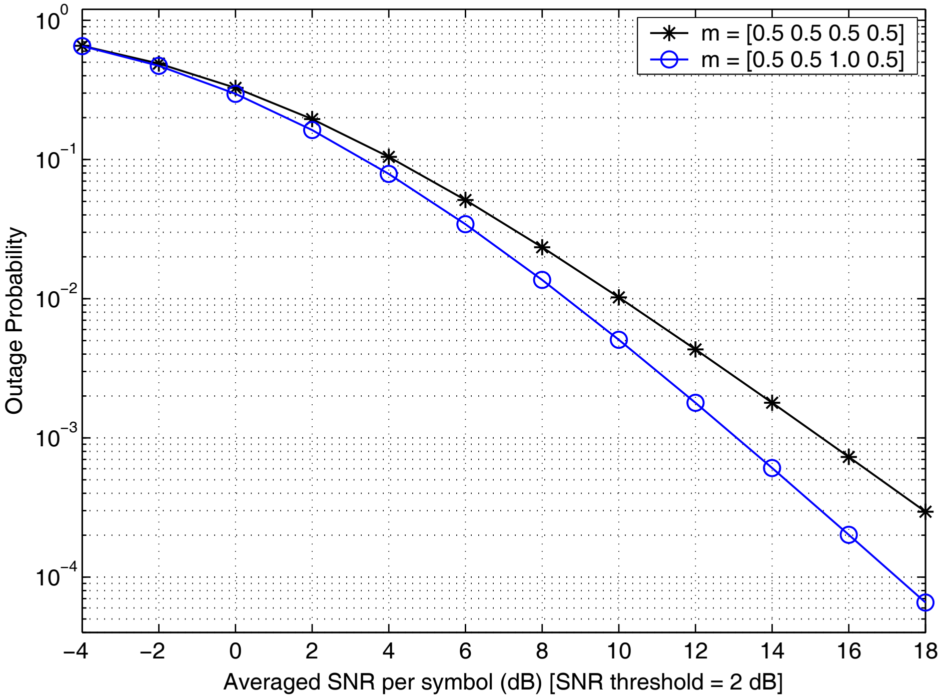


Fig. 2. Outage probability comparisons for different fading orders with the branch power-correlation matrix $\mathbf{R}_{\mathbf{X1}}$.

$$\mathbf{R}_{\mathrm{V},1}(k,l)=\mathbf{R}_{\mathbf{X1}}(k,l)\sqrt{\frac{\max(m_k,m_l)}{\min(m_k,m_l)}}. \tag{43}$$

The next step is to determine the truncation parameter K for each case. The truncation errors of the area under the PDFs computed by (22) for different values of K for the above cases are given in Figure 1. It is noted from Figure 1 that the convergence rate of the PDF depends heavily on the fading parameters (i.e., m_k's and Ω_k's). It is also evident from Figure 1 that using (22) we can easily choose a good truncation of the PDF for practical evaluation.

For a better view of the convergence of the infinite series in (18), we first compute the PDF in (18) with $K=100{,}000$ and then compute the PDF's for smaller values of K and compare those PDF's with that of the case $K=100{,}000$. We use sample mean-squared error (MSE) as the measure of the difference between different PDF's. The results are shown in Table 1. As observed from Table 1, the difference in MSE sense between the PDF computed for $K = 100{,}000$ and that computed for $K = 100$ is extremely small ($10^{-26.2482}$) and the convergence is fast. Note, however, that the convergence rate depends on the system parameters as evident from Figure 1.

K	MSE
1	$10^{-2.2726}$
10	$10^{-3.1351}$
20	$10^{-3.8496}$
30	$10^{-4.5195}$
50	$10^{-5.9959}$
70	$10^{-10.9634}$
100	$10^{-26.2482}$

Table 1. MSE's between the PDF's computed for different K's and the PDF computed for $K = 100{,}000$ with the branch power-correlation matrix $\mathbf{R}_{\mathbf{X}1}$.

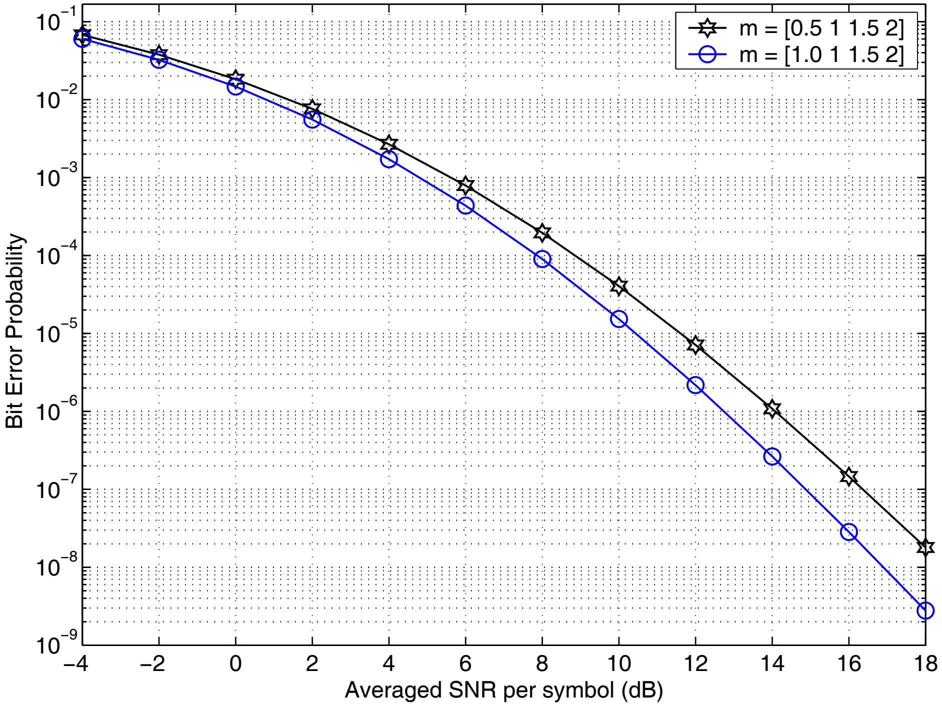


Fig. 3. Bit error probability comparisons for different fading orders with BPSK modulation, and branch power and branch correlation matrix of the Case 2 with the branch power-correlation matrix $\mathbf{R}_{\mathbf{X}1}$.

Practicality of the derived results: The question that remains to be answered is whether the performances of MRC wireless systems over Nakagami-*m* fading channels would vary significantly when the fading orders are changing to the nearest integers from half of integers.

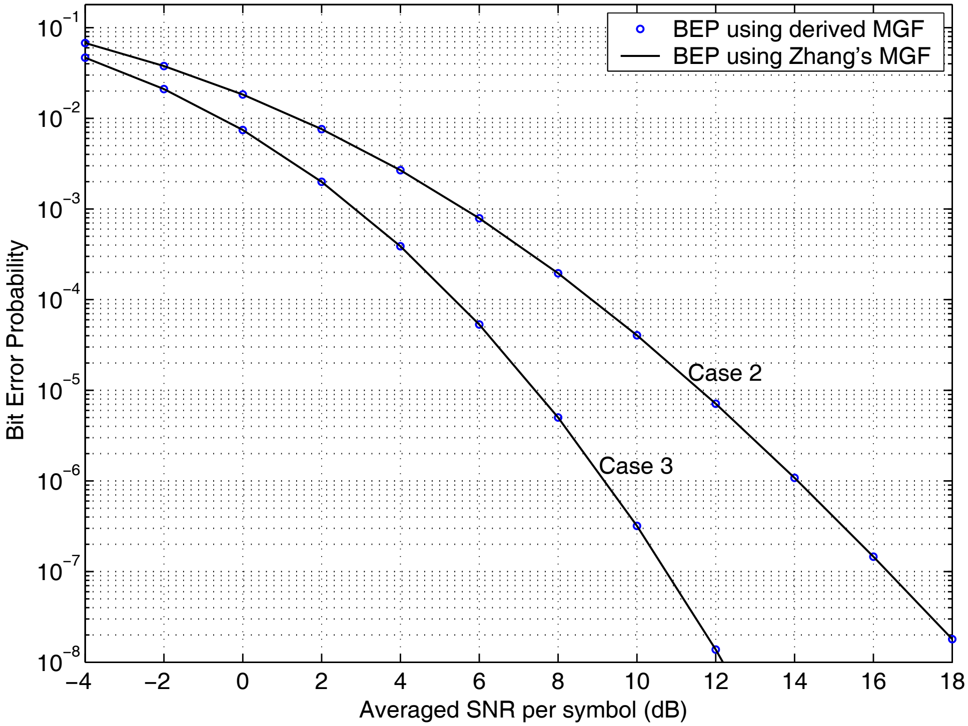


Fig. 4. Bit error probabilities of an MRC receiver derived in (35) using the MGF derived in (15) and the MGF given in (Zhang, 2003) for the Cases 2 and 3 with $L=4$ branches and the branch power-correlation matrix $\mathbf{R}_{\mathbf{X}1}$ for BPSK modulation.

We investigate this question through numerical results. Figure 2 presents the outage probabilities for different values of fading orders. The results presented in Figure 2 are for a SNR threshold $\overline{\eta}=2\text{dB}$, the branch power vector $\mathbf{\Omega}_1$ and branch correlation matrix given in (42). As seen in Figure 2, among the four fading orders only one is rounded from 0.5 to 1 but the difference in outage probability is already significant, especially at high SNR's. The same is observed for BEP performance given in Figure 3. These observations are also reported in (Ghareeb & Abu-Surra, 2005). Clearly, the derived results in this chapter for the case of half-of- integer fading orders *m*'s are practically useful.

Comparison to previous results: Figure 4 shows the BEP's of an MRC receiver with $L=4$ branches with the fading orders, branch powers and correlation matrix specified in Case 2 and Case 3 above for BPSK modulation. As seen in Figure 4, the BEP's derived in (35), using the MGF given in (15), very well match with those derived using Zhang's MGF in (Zhang, 2003). Note, however, that the MGF in (15) is derived by a totally different approach from that derived in (Zhang, 2003). Therefore, the approach presented in this chapter provides an alternative to, as well as a verification of, Zhang's approach (Zhang, 2003) for integer and half-of- integer fading orders.

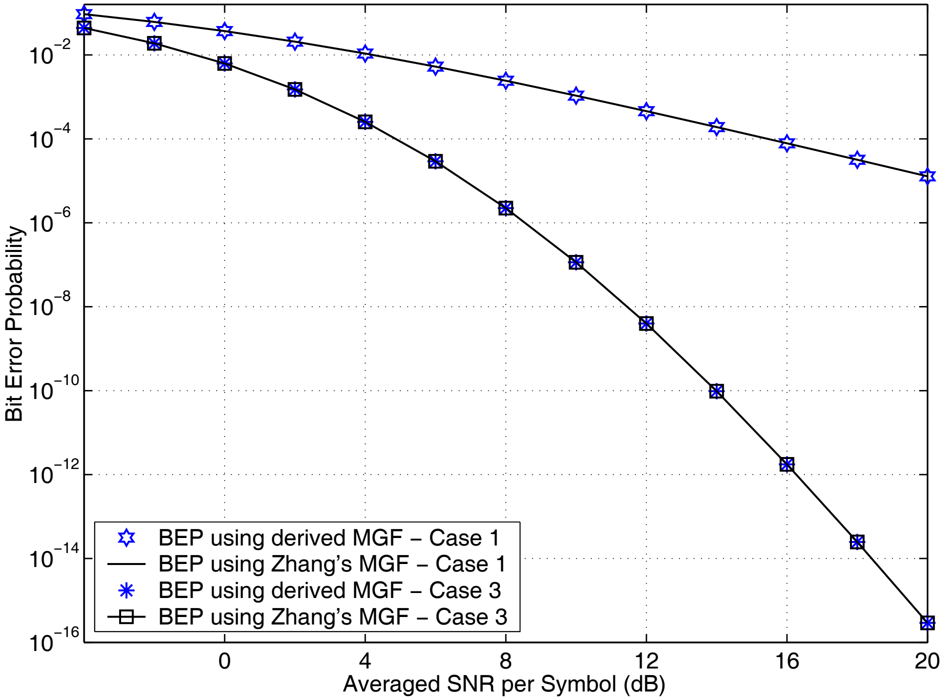


Fig. 5. Bit error probabilities of an MRC receiver derived in (35) using the MGF derived in (15) and the MGF given in (Zhang, 2003) for the Cases 1 and 3 with $L = 4$ branches and the branch power-correlation matrix $\mathbf{R}_{X2}$ for BPSK modulation.

5.2 Arbitrarily Correlated Branches

Consider now the case $L = 4$ diversity branches with an arbitrary branch power-correlation matrix $\mathbf{R}_{X2}$ given by

$$\mathbf{R}_{X2} = \begin{pmatrix} 1 & 0.5 & 0.2 & 0.05 \\ 0.5 & 1 & 0.5 & 0.2 \\ 0.2 & 0.5 & 1 & 0.5 \\ 0.05 & 0.2 & 0.5 & 1 \end{pmatrix}. \tag{44}$$

For the branch power-correlation matrix in (44), the farther the antennas are away from one another the smaller the branch correlation is. This corresponds to an arbitrary case as opposed to the previous example. In this example, we still consider the same fading parameters as that considered in Subsection 5.1 for brevity. Figure 5 shows the BEP's of an MRC receiver having four diversity branches with the fading orders, branch powers and correlation matrix specified in Case 1 and Case 3 in Subsection 5.1 also for BPSK modulation. It is seen in Figure 5 that the BEP's derived in (35), using the MGF given in (15), also very well match with those derived from Zhang's MGF in (Zhang, 2003). The results in Figure 5 demonstrate the generality of the derived results.

6. Conclusion

Most of the existing works dealing with multivariate Gamma r.v.'s in the literature are for the cases of *identical* and *integer* fading order. In this chapter, we have derived the MGF and PDF of the sum of *arbitrarily correlated* Gamma r.v.'s with *non–integer* and *non–identical* fading parameters. The new results remove most of the restrictions imposed on existing results. We also derived a simple method to determine a good truncation of the PDF and the associated truncation error for any fading parameters including the cases where the upper bound given in (Moschopoulos, 1985) is too loose and cannot be used. The derived PDF and MGF can be used to evaluate a number of performance measures of diversity combining receivers in wireless communications over Nakagami–*m* fading channels. The derived MGF offers a low complexity for the analysis and the PDF is in the form of a single Gamma series, which greatly simplifies the analyses of a number of performance measures in an MRC wireless system.

Appendix I: Low–Complexity Computation of $\delta_1,\ldots,\delta_K$

First, we recognize that in (19), the common terms between δ_k and δ_{k+1} are

$$X(w)=\sum_{i=1}^{M}\left[1-(\lambda_1/\lambda_i)\right]^{w},\ \ w=1,\ 2,\ldots,\ k. \tag{45}$$

Therefore, when computing δ_{k+1}, we can use $X(w)$, $w=1,\ 2,\ldots,\ k$, that we already computed for δ_k and need to compute only one more term, that is $X(k+1)$. Suppose that we need to truncate the PDF $p_\gamma(\gamma)$ at $K+1$ terms in (20). We then need to compute $\delta_1,\ \delta_2,\ldots,\ \delta_K$ (since $\delta_0=1$) according to (19). However, from (45) we note that $X(w)$, $w=1,\ 2,\ldots,\ K$, used to compute δ_K can be used to compute $\delta_1,\ \delta_2,\ldots,\ \delta_{K-1}$ as well. Therefore, if we use (19), we have to repeat many computations, which are already done. With this in mind, we propose a very simple algorithm to compute $\delta_1,\ \delta_2,\ldots,\delta_K$ as follows. First, for $w=1,\ 2,\ldots,\ K$, compute and store $X(w)$. Then, for $k=0,\ 1,\ldots,\ K-1$, compute

$$\delta_{k+1}=\frac{1}{2(k+1)}\sum_{w=1}^{k+1}X(w)\delta_{k+1-w}. \tag{46}$$

Using (46) significantly reduces the computational complexity of (19), especially for a large K.

7. References

Aalo, V. A. (1995). Performance of maximal–ratio diversity systems in a correlated Nakagami fading environment. *IEEE Trans. Commun.*, vol. 43, no. 8, pp. 2360–2369

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functions*, Dover

Al–hussaini, E. K. & Al–bassiouni, A. A. M. (1985). Performance of MRC diversity systems for the detection of signals with Nakagami fading. *IEEE Trans. Commun.*, vol. COM–33, no. 12, pp. 1315–1319

Alouini, M.–S., Abdi, A. & Kaveh, M. (2001). Sum of Gamma Variates and performance of wireless communications systems over Nakagami fading channels. *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1471–1480

Annamalai, A. , Tellambura, C. & Bhargava, V. K. (1999). Exact evaluation of maximal–ratio and equal–gain diversity receivers for M–ary QAM on Nakagami fading channels. *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1335–1344

Ghareeb, I. & Abu-Surra, S. (2005). Differential Detection of GMSK Signals with Postdetection MRC over Correlated and Unbalanced Nakagami-*m* Fading Channels. *IEE Proc., Commun.*, vol. 152, no. 2, pp. 221-228

Gradshteyn, I. S. & Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*, Academic Press, 6th edn.

J. Lu, J., Letaief, K. B., Chuang, J. C-I. & Liou, M. L. (1999). M-PSK and M-QAM BER computation using signal-space concepts. *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 181-184

Karagiannidis, G. K., Sagias, N. C. & Tsiftsis, T. A. (2006). Closed-form statistics for the sum of squared Nakagami-*m* variates and its applications. *IEEE Trans. Commun.*, vol. 54, no. 8, pp. 1353-1359.

Karagiannidis, G. K., Zogas, D. A. & Kotsopoulos, S. A. (2003a). On the multivariate Nakagami–*m* distribu-tion with exponential correlation. *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1240–1244

Karagiannidis, G. K., Zogas, D. A. & Kotsopoulos, S. A. (2003b). An Efficient Approach to Multivariate Nakagami-*m* Distribution using Green's Matrix Approximation. *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 883-889

Kotz, S. & Adams, J. W. (1964). Distribution of sum of identically distributed exponentially correlated Gamma–variables. *Ann. Instit. Math. Statist.*, vol. 35, pp. 277–283

Krishnaiah, P. R. & Rao, M. M. (1961). Remarks on a multivariate Gamma distribution. *Amer. Math. Monthly*, vol. 68, no. 4, pp. 342–346

Lee, W. C. (1993). *Mobile Communications Design Fundamentals*, John Wiley & Sons, 2nd edn.

Lombardo, P., Fedele, G. & Rao, M. M. (1999). MRC performance for binary signals in Nakagami fading with general branch correlation. *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 44–52

Moschopoulos, P. G. (1985). The distribution of the sum of independent Gamma random variables. *Ann. Inst. Statist. Math. (Part A)*, vol. 37, pp. 541–544

Nakagami, M. (1960). The *m*–distribution—A general formula for intensity distribution of rapid fading, In: *Statistical Methods in Radio Wave Propagation*, Pergamon Press, pp. 3–36

Proakis, J. G. (2001). *Digital Communications*, McGraw–Hill, Inc., 4th edn.

Su, W. & Xia, X.-G. (2003). On space-time block codes from complex orthogonal designs. *Wireless Personal Commun.*, vol. 25, no. 1, pp. 1-26

Tran, T. A. & Sesay, A. B. (2007). Sum of arbitrarily correlated Gamma variates and performance of wireless communication systems over Nakagami-*m* fading channels. *IET Communications*, vol. 1, no. 6, pp. 1133-1137

Win, M. Z., Chrisikos, G. & Winters, J. H. (2000). MRC performance for M–ary modulation in arbitrarily correlated Nakagami fading channels. *IEEE Commun. Lett.*, vol. 4, no. 10, pp. 301–303

Zhang, Q. T. (1999). Maximal–ratio combining over Nakagami fading channels with an arbitrary branch covariance matrix. *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1141–1150

Zhang, Q. T. (2003). A generic correlated Nakagami fading model for wireless communications. *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1745–1748

5

Sequential Blind Beamforming for Wireless Multipath Communications in Confined Areas

Salma Ait Fares[1], Tayeb Denidni[2], Sofiene Affes[2] and Charles Despins[2,3]
[1]*Tohoku University, Japan*
[2]*INRS-EMT, Canada*
[3]*Prompt-Quebec, Canada*

1. Introduction

The main cause of degradation of communication quality in harsh confined environments, such as underground mines, is multipath fading as it is typically more severe than co-channel interference (CCI) (Nerguizian et al., 2005; Chao & Chung, 1994). Usually underground communication systems are based on IEEE 802.11 standard using carrier sense multiple access with collision avoidance (CSMA-CA) that minimizes the CCI problem and leaves multipath fading as the main source of errors. This multipath phenomenon arises when a transmitted signal undergoes reflection from various obstacles in the propagation environment. This gives rise to multiple waves arriving at the receiver from different directions with different amplitudes, phases and time delays, which inflicts inter-symbol interference (ISI) and intra-symbol interference (isi) on the received signal and consequently time dispersion and fading (Saunders, 1999; Amca et al., 1999). To provide a remedy, adaptive equalizers (Amca et al., 1999), error-correcting codes (Sanada & Wang, 1997), diversity (Cozzo & Hughes, 2003), RAKE receivers for code division multiple access (CDMA) (Tanaka, 1994; Affes & Mermelstein, 2003), Coded Orthogonal Frequency Division Multiplexing (Stott, 2000), and adaptive antenna arrays (AAA) (McNeil et al., 2001; Bellofiore,a, 2002; Bellofiore,b, 2002) have been proposed. Among these, AAA techniques exploit spatial diversity by using multiple antennas. They have the potential of achieving high data-rates and increasing the capacity of mobile services by effectively reducing the multipath effect and CCI.

Adaptive antenna beamforming has been shown to be an effective mean for combating CCI and multipath propagation. Recently many different types of antenna beamforming have been developed (Lee & Choi, 2005; Youna & Un, 1994), and their design employs some information about the desired signal such as using a training sequence. Their use not only reduces the capacity of the system but also requires synchronization between the incident desired signal and the reference (sequence) one. Therefore, there is a strong demand for blind algorithms. For instance, the constant modulus algorithm (CMA) (Godara, 1997), applied in a blind AAA, is considered as a promising method in mobile communications for mitigating multipath fading and suppressing CCI signals. Unlike the least-mean-square (LMS) algorithm (Widrow, 1967), CMA-AAA does not need synchronization between the incident desired signal and the reference signal. It improves the received signal by suppressing not only the CCI, but also the delayed paths of the desired signal (Godara, 1997). However, suppressing the multipath rays of the desired signal wastes a part of the available power and requires additional degrees of freedom, i.e., more array elements. In addition, since these arriving paths are delayed replicas of an identical source, it is desirable to separate and combine the delayed paths instead of suppressing them for received power maximization.

Several methods have been proposed to separate and combine the delayed paths. Among them, we mention the spatial domain path-diversity methods (Tanabe et al., 2000; (Ogawa et al., 1999) for time division multiple access systems, based on spatial processing for efficient exploitation of the propagated energy. In these methods, multiple beams are formed to separate the direct path and the delayed paths by using training sequences. Paths with various time delay differences are separately extracted at the receiver to produce each of the multiple paths in the spatial domain. Then, the time delay difference is corrected, and the multipath signals are combined.

These aforementioned approaches have certain advantages and limitations, some of them require training sequences and all of them aim to solve the problem of ISI in multipath propagation environments, when the time path arrival (TPA) is an integer multiple of the sampling interval. Regarding the isi problem, when the TPA is non-integer multiple of the sampling interval, other approaches are used such as Rake receiver, applied in CDMA systems, to treat the path arrivals at the chip interval (Tanaka, 1994; Affes & Mermelstein, 2003) or straightforwardly, oversampling (Amca et al., 1999; Slock, 1994) is employed. However, over-sampling will complicate the structure of the analog-to-digital converter, and also may not synchronize correctly within the actual time path arrival τ.

Generally speaking, when the TPA is a non-integer multiple of the sampling interval, the power available in these paths is wasted and the fractional time delay estimation (FTDE) is called for to overrule the over-sampling solution. FTDE filters are employed in several signal processing applications such as the estimation of the time delay difference between signals received at two spatially separated antenna arrays (So et al., 1994). They consist in using a first-order or linear interpolator to implement a Fractional Delay Filter (FDF) (So et al., 1994). An FDF is a type of digital filter designed for band-limited interpolation, which can be implemented using a Finite Impulse Response (FIR) filter based on the truncated *sinc*-interpolation method (Valimaki & Laakso, 2000).

In this Chapter, we present a new approach using sequential blind spatial-domain path-diversity beamforming (SBB) to remedy both the ISI and isi problems using jointly CMA, LMS and adaptive FTDE filtering. This approach is designed to sequentially recover multipath rays by using multiple beamformings for received power maximization. First, the strongest path is extracted using the Modified-CMA (MCMA) (Oh & Chin, 1995; AitFares et al., 2004; AitFares et al., 2006 a; AitFares et al., 2008). Second, the paths coming with delays that are multiple integers of the sampling interval are estimated using Integer Delay- CMA SBB (ID-CMA SBB) filters adapted using LMS with the CMA delayed output as a reference signal (AitFares et al., 2004). Finally, the paths coming with fractional delays are estimated using Fractional Delay-CMA SBB (FD-CMA SBB) filters (AitFares et al., 2006 a) adapted using LMS and FTDE, the latter is implemented by an FIR *sinc*-interpolation filter (So et al., 1994). Furthermore, by relying on the common phase ambiguity characteristic presented in these extracted paths, additional enhancement is obtained by using a modified coherent

Maximum Ratio Combining (MRC) detector with hard Decision Feedback Identification (DFI) (Affes & Mermelstein, 2003; AitFares et al., 2006 b; AitFares et al., 2008) to constructively combine the different received paths for signal-to-noise ratio (SNR) maximization.

2. Modified CMA

Beamforming techniques based on blind adaptation present several comparative advantages with those based on a temporal or spatial reference. A system based on a temporal reference might carry out synchronization and demodulation, before starting weight adaptation. While a system based on a spatial reference requires strictly hardware calibration and depends on the direction of angle of arrival information of the desired signal. On the other hand, the system based on blind adaptation does not require any information related to the antenna array geometry or the training sequences.

Many communication signals have a constant modulus (CM) property, such as, FM, PM, FSK, PSK. If these transmitted signals are affected by multipath fading, noise and/or interference, the CM property may be lost. To ease this problem, the constant modulus algorithm (CMA) has been introduced as a suitable algorithm for compensating multipath fading of constant envelope signals and has originally been proposed particularly for blind channel equalization (Godara, 1997). Afterwards, CMA has been introduced as an adaptive FIR filter for phase-modulated signals, and studied and applied to array antennas as a narrow-band adaptive beamformer to recover a constant-module signal. In fact, the CMA algorithm works with the principle that the amplitude of a signal transmitted at constant modulus and received by the array antenna, should remain constant. Hence, the main aim of an adaptive beamforming based on CMA is to restore the average of the antenna array output, $y_{CMA}(k)$, to a constant modulus as illustrated in Figure 1. That can be accomplished by adjusting the vector of beamforming weight, $\mathbf{w}_{CMA}=[w_1, \ldots, w_N]^T$, by minimizing a certain cost function. In the reference of (Godara, 1997) which uses the CMA in order to carry out blind equalization, the criterion was to minimize the function $J^{(p)}$, referred to as the dispersion of order p ($p > 0$ integer), and defined by:

$$J^{(p)}(k) = E\left[\left(|y_{CMA}(k)|^P - R_{CMA,p}\right)^2\right], \tag{1}$$

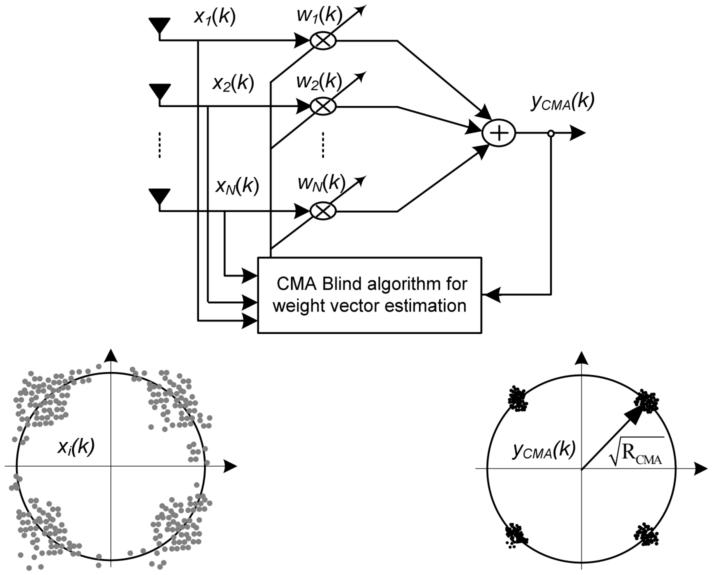
where E[•] is the expectation value of [•], y_{CMA} is the antenna array output using CMA beamforming and R_{CMA} is a constant that depends on the transmitted symbols a and defined by

$$R_{CMA,p} = \frac{E[|a|^{2p}]}{E[|a|^p]}, \tag{2}$$

where a is the transmitted symbol. The case generally used is for p=2, where the cost function, $J^{(2)}$, is indeed the quadratic error between the square amplitude of the array output signal and the constant $R_{CMA,2}$, from where the equation (1) becomes

$$J^{(2)}(k) = E\left[\left(|y_{CMA}(k)|^2 - R_{CMA,2}\right)^2\right], \tag{3}$$

where $R_{CMA,2} = \frac{E[|a|^4]}{E[|a|^2]}$. For simplicity, we denote $R_{CMA,2}$ simply as R_{CMA}.



Constellation of the received signal at the input of the array antenna receiver.

Constellation of the received signal at the CMA array antenna output.

Fig. 1. Adaptive beamforming using CMA blind algorithm.

A stochastic gradient method can be used to minimize this cost function by adjusting adaptively the beamforming weight vector, w_{CMA}, according to the following expression:

$$\boldsymbol{w}_{CMA}(k+1) = \boldsymbol{w}_{CMA}(k) - \mu \cdot e^*(k) \cdot \boldsymbol{x}(k), \tag{4}$$

where $\boldsymbol{x}(k)=[x_1, \ldots, x_N]^T$ represents the received signal vector and the error signal is given by

$$e(k) = y_{CMA}(k) \cdot (y_{CMA}^2(k) - R_{CMA}(k)). \tag{5}$$

The antenna array output is expressed by:

$$y_{CMA}(k) = \boldsymbol{w}_{CMA}^H(k) \cdot \boldsymbol{x}(k). \tag{6}$$

After convergence, the array output is reconstituted with a signal having a constant envelope on average, while the weight vector w_{CMA} converges to the Wiener solution.
However, from (3), since this cost function is invariable with an arbitrary phase rotation, ϕ, in the constellation of the estimated signal, i.e.

$$J_{CMA}(k) = E\left[\left(\left|y_{CMA}(k) \cdot e^{j\phi}\right|^2 - R_{CMA}\right)^2\right] = E[(|y_{CMA}(k)|^2 - R_{CMA})^2], \tag{7}$$

the constellation of the antenna array output may undergo an arbitrary phase rotation. In other words, the CMA algorithm restores on average the array output to a signal with a constant envelope while trying to place this output in a circle of radius $\sqrt{R_{CMA}}$ with a random phase, ϕ. Consequently, to bring back the constellation of the array output to its good position, a phase rotation mechanism is required in the CMA adaptive process. To solve the problem of the phase rotation of the estimated signal, using CMA, without recourse to any additional phase estimation, Modified–CMA (MCMA) (Oh & Chin, 1995) is used. In the MCMA version, the cost function, given by equation (1), is divided into real and imaginary parts, so that the envelope as well as the phase will be considered in the array output estimation.

2.1 Modified – CMA algorithm

The expression of the cost function of the MCMA is expressed as follows:

$$J_{MCMA}(k) = E\left[\left(\left|y_{MCMA\_R}(k)\right|^2 - R_R\right)^2\right] + E\left[\left(\left|y_{MCMA\_I}(k)\right|^2 - R_I\right)^2\right], \tag{8}$$

where the parameters R_R and R_I are real constants defined as:

$$R_R = \frac{E[a_R^4(n)]}{E[a_R^2(n)]}, \tag{9}$$

and

$$R_I = \frac{E[a_I^4(n)]}{E[a_I^2(n)]}, \tag{10}$$

where

$$a_R(n) = real(a(n)), \tag{11}$$

$$a_I(n) = imag(a(n)), \tag{12}$$

where $real(\bullet)$ and $imag(\bullet)$ denote, the real and imaginary parts of $(\bullet)$, respectively, and $a(n)$ represents the transmitted signal sequence, $a \in A_M=\{a_1, \ldots, a_M\}$, and

$$y_{MCMA\_R}(k) = real(y_{MCMA}(k)), \tag{13}$$

$$y_{MCMA_I}(k) = imag\big(y_{MCMA}(k)\big). \tag{14}$$

A stochastic gradient approach can be employed in order to minimize this modified cost function, $J_{MCMA}(k)$, by adjusting adaptively the weight vector, $\boldsymbol{w}_{MCMA}$, as

$$\boldsymbol{w}_{MCMA}(k+1) = \boldsymbol{w}_{MCMA}(k) - \mu \cdot e^*(k) \cdot \boldsymbol{x}(k), \tag{15}$$

where μ is a small positive step size.
The error signal is expressed as:

$$e(k) = e_R(k) + j.e_I(k), \tag{16}$$

where

$$e_R(k) = y_{MCMA\_R}(k) \cdot \big(y_{MCMA\_R}^2(k) - R_R\big), \tag{17}$$

$$e_I(k) = y_{MCMA\_I}(k) \cdot \big(y_{MCMA\_I}^2(k) - R_I\big). \tag{18}$$

In this case, the MCMA filter output will be expressed as:

$$y_{MCMA}(k) = \boldsymbol{w}_{MCMA}^H(k) \cdot \boldsymbol{x}(k). \tag{19}$$

In our proposed approach, we adopt the MCMA algorithm since it solves the problem of the phase rotation of the estimated signal at the array output without the need of a separate phase adaptation mechanism. The following sections propose two versions of sequential blind beamforming (ID-CMA SBB and FD-CMA SBB), based on spatial-domain path-diversity, in order to detect, respectively, the paths arriving with multiple integer delays of the sampling interval and the paths arriving with fractional time delays, while using the MCMA beamforming output as a reference signal.

3. ID-CMA SBB Approach

For a frequency selective fading channel, we propose, in this section, a new method of path synchronization, named ID-CMA SBB (Integer Delay-CMA SBB), based on spatial-domain path-diversity (AitFares et al., 2004; AitFares et al., 2008). This approach resolves the problems involved in the presence of the ISI generated by the presence of the paths arriving with multiple integer delays of the symbol duration. However, since these delayed paths are only delayed replicas of identical sources, we may detect them while making use of the first estimated path as a reference signal. In fact, the ID-CMA approach principle consists in estimating the strongest path by using MCMA algorithm, while other auxiliary beamformers are sequentially used to extract the remaining delayed paths, arriving with integer multiple delays of the symbol duration ($k.T_s$, k is an integer).

3.1. Signal model

Consider a uniform linear array of N omni-directional antenna elements receiving L multipath signals. The received signal $x_m(k)$ at the m-th antenna can be expressed as:

$$x_m(k) = \sum_{i=1}^{L} \alpha_i(k) \cdot s(k-\tau_i) \cdot e^{-j\pi(m-1)\sin(DoA_i)} \cdot e^{-2\pi f_D k} + \eta_m(k), \quad m=1, \ldots, N, \quad (20)$$

where, $\alpha_i(k)$ are the complex gains of the Rayleigh fading rays (with uniformly distributed phases φ_i between 0 and 2π) of the i-th path, $\alpha_i(k)e^{-2\pi f_D k}$ are of Jakes' model with f_D as maximum Doppler spread[1], $s(k)$ is the desired source sequence, drawn from alphabet members $A_M = \{a_1, \ldots, a_M\}$, L is the number of multipath signals, τ_i is the path delay for the i-th path, DoA_i is the direction of arrival of the i-th path and $\eta_m(k)$ is an additive white Gaussian noise process with variance σ_n^2 at the m-th receive antenna. For convenience, the array is assumed to be uniform and linear with inter-element spacing $d = \lambda/2$, where λ is the wavelength at the operating frequency.

To simplify the illustration of the proposed approach, we suppose receiving two paths (L=2) of which the first, of highest power, arrives at the array antenna with a time delay equal to zero, and the second with a time delay equal to T_s. Then equation (20) becomes:

$$x_m(k) = A_{1m} \cdot s(k) + A_{2m} \cdot s(k-T_s) + \eta_m(k), \quad m=1, \ldots, N, \quad (21)$$

where $$A_{im}(k) = \alpha_i(k) \cdot e^{-j\pi(m-1)\sin(DoA_i)} \cdot e^{-2\pi f_D k}, \quad m=1, \ldots, N, \; i = \{1, 2\}. \quad (22)$$

Figure 2 shows the new structure for spatial-domain path diversity to resolve multipath signals when TPAs are integer ($0, T_s$) multiples of the sampling interval. This method is implemented using two sequential beamformers. The first beamformer is used to estimate

---

[1] We chose here the model of a point-source propagation with no angular distribution to consider the worst-case scenario in terms of performance. However, the proposed algorithm can be applied to any channel type or angular distribution without limitation in principle.

the strongest path, and its weights (w_{MCMA}) are adapted using the MCMA. The output of this filter (y_{MCMA}) is fed into ID-CMA beamformer to construct the path arriving with an integer delay $\tau_2 = T_s$ using the LMS algorithm.

However, to ensure that the ID-CMA filter detects the path arriving with a specific integer delay, i.e. T_s, a signal canceller filter is used to extract the contribution of the already extracted path from the received signal vector. Finally, a path combination technique is used to combine these extracted paths for power maximization. This combination technique will be presented later in Section 5. In what follows, we will present the mathematical development of the shaded blocks, illustrated in Figure 2, constituting the ID-CMA SBB approach, particularly the cancellation filter of y_{MCMA} and the ID-CMA$_1$ filter.

By using equation (21), the MCMA filter output given by (19), becomes

$$y_{MCMA}(k) = \boldsymbol{w}_{MCMA}^H(k) \cdot \boldsymbol{A}_1(k) \cdot s(k) + \boldsymbol{w}_{MCMA}^H(k) \cdot \boldsymbol{A}_2(k) \cdot s(k - T_s) + \boldsymbol{w}_{MCMA}^H(k) \cdot \boldsymbol{\eta}(k), \quad (23)$$

where the vector $\boldsymbol{A}_i(k) = [A_{i1}(k), A_{i2}(k)]^T$ for i=1, 2, the element $A_{im}(k)$ is given by (22) for m=1, 2, and $\boldsymbol{\eta}(k) = [\eta_1(k), \ldots, \eta_N(k)]^T$.

After convergence, the MCMA beamforming output can be expressed as

$$y_{MCMA}(k) \cong \hat{s}(k) + \gamma_1(k), \quad (24)$$

where the noise scalar is given by

$$\gamma_1(k) = \boldsymbol{w}_{MCMA}^H(k) \cdot \boldsymbol{\eta}(k), \quad (25)$$

and $\hat{s}(k)$ represents an estimation of the transmitted signal $s(k)$.

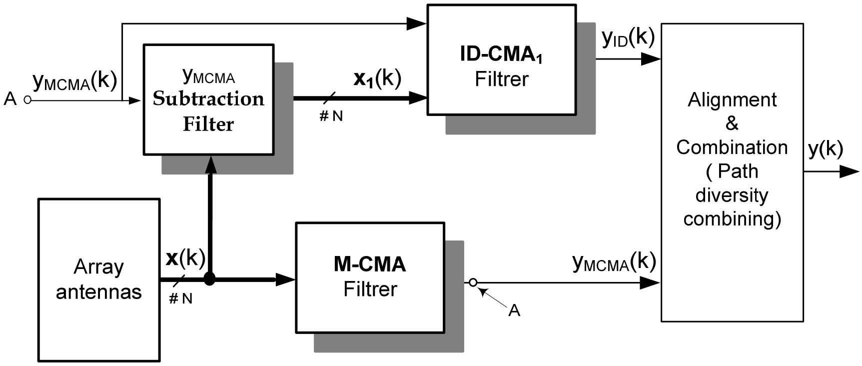


Fig. 2. The proposed ID-CMA SBB approach for extracting the direct path and delayed path arriving with an integer multiple of the sampling period equal to T_s.

3.2. First path cancelling

In order to ensure an efficient estimation of the second path using the ID-CMA$_1$ filter, the contribution of the first estimated path, $y_{MCMA}(k)$, should be removed from the received signal vector $x(k)$. To perform this extraction, a replica of the first detected path should be estimated, i.e., an estimate of the first path power and, then, it will be removed from the received signal vector $x(k)$. However, in what follows, we will study two techniques of cancellation of the signal contribution by employing the method of *correlation* (Yuan & Tsai, 2005) as depicted in Figure 3(a), or by using the *Adaptive Signal Canceller* (ASC) which is particularly applied in echo cancellation processing, as illustrated in Figure 3(b) (Furukawa, 1996).

A - *Correlation coefficient approach*

The purpose of this method consists in calculating the correlation coefficient at each antenna element between the received signal at the m-th antenna element $x_m(k)$ and the estimated signal $y_{MCMA}(k)$, which is regarded as a reference signal. This coefficient, provided by (26), represents the quotient of the correlation between the received signal at the m-th antenna element, and the estimated signal by the MCMA filter, and the autocorrelation of this estimated signal:

$$\rho_m = \frac{E[x_m(k)\cdot y^*_{MCMA}(k)]}{E[|y_{MCMA}(k)|^2]}, \quad \text{for } m = 1,\dots N. \tag{26}$$

By using (21), (24) and (25), and assuming all signals are decorrelated, this correlation coefficient may be simplified as

$$\begin{aligned}\rho_m &= \frac{E\left[\left(A_{1m}(k)s(k) + A_{2m}(k)s(k-T_s) + \eta_m(k)\right)\cdot\left(\hat{s}(k) + \sum_{m=1}^{N} w^*_{MCMA,m}(k)\cdot\eta_m(k)\right)^*\right]}{E\left[\left|\hat{s}(k) + \sum_{m=1}^{N} w^*_{MCMA,m}(k)\cdot\eta_m(k)\right|^2\right]} \\ &= \frac{E[A_{1m}(k)s(k)s^*(k)]}{E[|\hat{s}(k)|^2] + E\left[\left|\sum_{m=1}^{N} w^*_{MCMA,m}(k)\eta_m(k)\right|^2\right]} + \frac{E\left[\eta_m(k)\cdot\sum_{m=1}^{N} w_{MCMA,m}(k)\cdot\eta^*_m(k)\right]}{E[|\hat{s}(k)|^2] + E\left[\left|\sum_{m=1}^{N} w^*_{MCMA,m}(k)\eta_m(k)\right|^2\right]}.\end{aligned} \tag{27}$$

By computing the instantaneous value for the numerator and the denominator of (27) and, by assuming that the variance of $s(k)$ is equal to one and by neglecting the terms associated to the noise power, equation (27) can be approximated by:

$$\rho_m \cong A_{1m}(k). \tag{28}$$

The signal replica $z_m(k)$ of the first path estimated at the m-th antenna element is then generated by multiplying $\rho_m(k)$ by $y_{MCMA}(k)$ as follows:

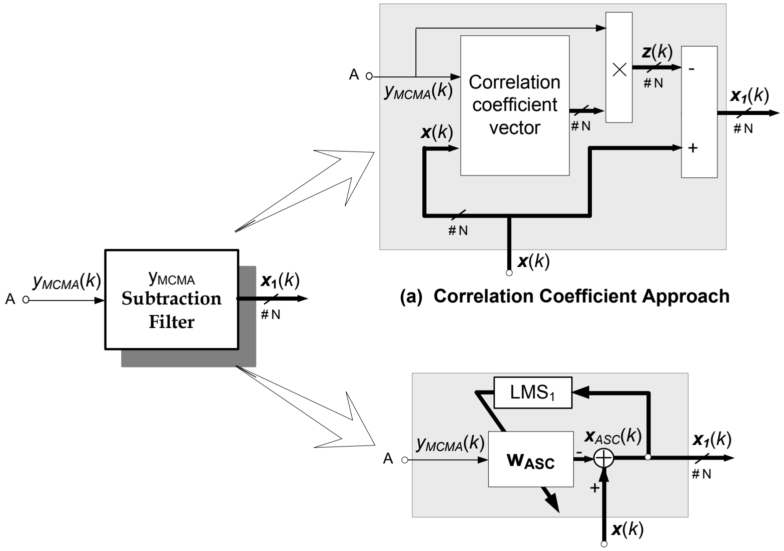
$$\begin{aligned} z_m(k) &= \rho_m(k)\cdot y_{MCMA}(k) \\ &= A_{1m}(k)\cdot\hat{s}(k) + A_{1m}(k)\cdot\gamma_1(k), \end{aligned} \tag{29}$$

where γ_1 is expressed by (25).

The second term on the right side of (29) is sufficiently small compared to the first term. Then, the signal received at the m-th antenna element, $x_{1m}(k)$, after extraction of the replica $z_m(k)$ from the received signal, $x_m(k)$, given by equation (21), will be expressed by:

$$\begin{aligned} x_{1m}(k) &= x_m(k) - z_m(k) \\ &= A_{2m}(k)\cdot s(k-T_s) + \eta_m(k), \qquad m = 1,\dots,N. \end{aligned} \tag{30}$$

Even if the resulting value of the signal vector $x_1(k) = [x_{11}, \dots, x_{1N}]^T$ has a noise term, its main component contains information about the desired signal, representing the second delayed path.



(a) Correlation Coefficient Approach

(b) ASC Filter

Fig. 3. Canceller Filter for the y_{MCMA} path using (a) the correlation method (b) ASC filter.

B - *Adaptive Signal Canceller (ASC) approach*

An alternative to remove the contribution of the first estimated path is to adjust, adaptively, the generation of the replica of the signal $y_{MCMA}(k)$, i.e., $z_m(k)$, expressed by equation (29) using the ASC filter. This real time adaptation consists in estimating the contribution of the signal related to the first path; therefore, it will be adaptively extracted to effectively remove the first path from the received signal. The principle of operation is illustrated in Figure 3(b) where the LMS algorithm is used to adjust the weights of the ASC filter.

The signal vector at the output of the ASC filter is expressed by

$$\boldsymbol{x}_{ASC}(k) = \boldsymbol{w}_{ASC}^{H}(k) \cdot y_{MCMA}(k), \tag{31}$$

where $\boldsymbol{w}_{ASC}(k)$ represents the weight vector of the ASC filter.

By using the vector signal $\boldsymbol{x}(k)$ as a reference signal for the LMS algorithm, the vector error signal, $\boldsymbol{x}_1(k)$, of the ASC filter is expressed by:

$$\boldsymbol{x}_1(k) = \boldsymbol{x}(k) - \boldsymbol{x}_{ASC}(k), \tag{32}$$

where $\boldsymbol{x}_{ASC}(k)$ represents the signal vector of the ASC filter output.

Furthermore, the weight vector of the filter ASC is adjusted by using the LMS algorithm as follows:

$$\boldsymbol{w}_{ASC}(k+1) = \boldsymbol{w}_{ASC}(k) + \mu_1 \cdot \boldsymbol{x}_1^{H}(k) \cdot y_{MCMA}(k), \tag{33}$$

where μ_1 is a small positive step size.

It should be mentioned that the signal subtraction can be made using different adaptive algorithms, namely RLS, SMI, etc. The LMS algorithm presents an acceptable convergence and is used here to ease the mathematical development of the ASC filter. In the following sections, we favored the use of the ASC filter in our proposed SBB owing to its adaptive nature using the LMS algorithm.

3.3. ID-CMA Filter

As explained earlier, a delayed replica of the signal $y_{MCMA}(k)$ will be used as a reference signal to estimate the second path using the ID-CMA$_1$ filter as shown in Figure 4. To implement this proposed scheme, we delayed the output $y_{MCMA}(k)$ by a delay equal to that of the 2nd path, which is assumed here equal to T_s, (according to (21)), and then fed it back to the ID-CMA$_1$ filter as a reference signal for the LMS algorithm. The signal output of this ID-CMA$_1$ filter is expressed by:

$$y_{ID}(k) = \boldsymbol{w}_{ID}^{H}(k) \cdot \boldsymbol{x}_1(k), \tag{34}$$

where $x_1(k)$ is the ASC filter output given by (32) and $w_{ID}(k)$ is the weight vector of the ID-CMA$_1$ filter adjusted here by using the LMS algorithm as follows

$$\boldsymbol{w}_{ID}(k+1) = \boldsymbol{w}_{ID}(k) + \mu_2 \cdot e_{ID}^{*}(k) \cdot \boldsymbol{x}_1(k), \tag{35}$$

where μ_2 represents a small positive step size. The error signal, e_{ID}, is computed by using a delayed replica of the signal $y_{MCMA}(k)$, i.e., $y_{MCMA}(k\text{-}1)$, as a reference signal for the LMS algorithm, and it is expressed by

$$e_{ID}(k) = y_{MCMA}(k-1) - y_{ID}(k), \tag{36}$$

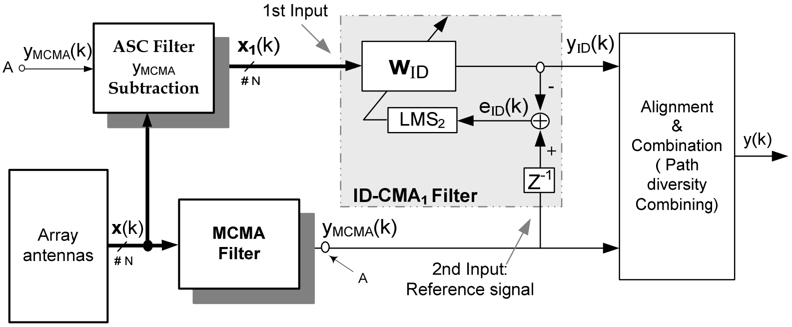


Fig. 4. ID-CMA Filter to estimation the second path arriving with delay equal to T_s.

4. FD-CMA SBB Approach

In addition to the presence of paths arriving with delays that are integer multiples of the sampling interval, some reflections in multipath propagation can arrive at the receiver between one period of two successive symbols, as illustrated in Figure 5. At the receiver side, these delayed signals can be combined in constructive or destructive manner. A constructive combination implies that the presence of this path diversity is exploited and, an improved performance is achieved. However, a destructive combination implies that the second delayed path is shifted by a certain phase and that its summation with the first path degrades the quality of the received signal. This type of multipath effect, known as the intra-symbol interference (isi), can cause significant amplitude attenuation of the received signal, thus leading to reception errors of the desired signal.

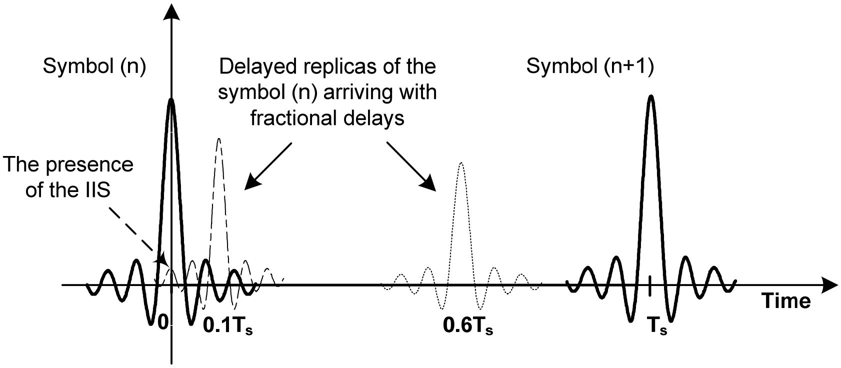


Fig. 5. Paths arrived at fractional delays.

On the other hand, at the sampling time, the information containing in this paths arriving with a delay less than the sampling interval (fractional), is generally lost and ignored by the receiver, thus leading to a reduction in the received signal power.

However, in order to maximize the received signal power, the presence of different paths, arriving with a fractional delay, may be exploited by detecting and combining them constructively. Indeed, some research works have proposed the spatio-temporal diversity techniques by using Rake receiver, applied to the CDMA system, to detect paths arriving at chip duration (Tanaka, 1994; Affes & Mermelstein, 2003), and also temporal diversity approach using an over-sampling (Slock, 1994). Other studies have also proposed a variable sampling frequency, $1/(T_s - \xi)$ where $0 \leq \xi < T_s$, to maximize the chance of detecting the path arriving with a fractional delay (Amca et al., 1999).

In what follows, we propose a space domain path-diversity technique, called FD-CMA SBB (Fractional Delay-CMA SBB), to detect the path arriving with a fractional delay, without recourse to an over-sampling neither a spatio-temporal processing or a specific application system (i.e., CDMA, TDMA). Indeed, this proposed approach is based on sequential blind CMA and nested LMS beamformings with adaptive fractional-time-delay estimation (FTDE) filter.

4.1. Signal model

To simplify the illustration of the proposed FD-CMA method, we suppose receiving two paths (L=2) where the first with highest power, arrives at the antenna arrays with a delay equal to zero, and second path arrives with a delay equal to $\tau < T_s$. So according to (20), the received signal at the m-th antenna element can be expressed as:

$$x_m(k) = A_{1m}s(k) + A_{2m}s(k-\tau) + \eta_m(k). \tag{37}$$

The FD-CMA SBB approach (AitFares et al., 2006 a) is illustrated in Figure 6 for this studied example. This approach consists first to estimate the first arriving path, $y_{MCMA}(k)$, using the

MCMA algorithm. The contribution of this estimate signal is removed from the received signal vector $x(k)$ by using the ASC filter. Then, using this estimated $y_{MCMA}(k)$, the fractional delay and its corresponding path are detected by the FD-CMA filter. Finally, a path combining stage is used to maximize the received power. In following, the mathematical development of the FD-CMA SBB filter is presented.

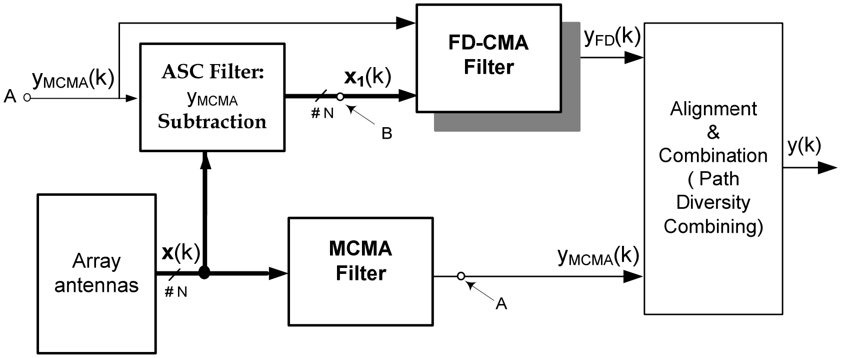


Fig. 6. The proposed FD-CMA approach to detect a path arriving with a fractional time delay and its corresponding.

4.2. FD-CMA filter

As mentioned above, the estimated signal $y_{MCMA}(k)$ will be exploited to detect the second path by using the FD-CMA filter as shown in Figure 6. After removing the contribution of the y_{MCMA} signal by the ASC filter, the signals, collected at points A and B in Figure 6, can be expressed by:

Point A: $$y_{MCMA}(k) \cong \hat{s}(k) + \gamma_1(k), \tag{38}$$

Point B: $$\boldsymbol{x}_1(k) \cong \boldsymbol{\beta}(k) \cdot s(k-\tau) + \boldsymbol{\gamma}_2(k), \tag{39}$$

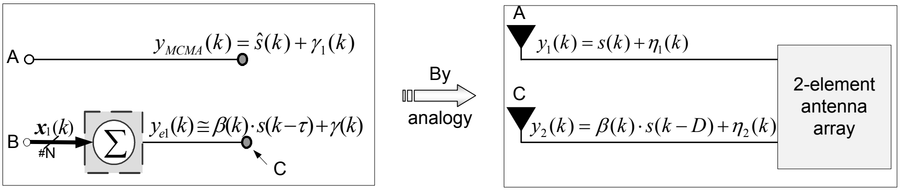
where $\hat{s}(k)$ is an estimation of the transmitted symbol $s(k)$, $\boldsymbol{\beta}(k)$ is a multiplicative factor vector representing the impulse response vector of the path arriving with the fractional delay τ, and $\gamma_1(k)$ and $\boldsymbol{\gamma}_2(k)$ are additive white Gaussian noises.

By summing the vector $x_1(k)$ at point B, the signal collected at point C, shown in Figure 7(a) can be expressed by:

Point C: $$y_{e1}(k) = \sum_{m=1}^{N} \boldsymbol{x}_{1_m}(k) \cong \beta(k) \cdot s(k-\tau) + \gamma(k), \tag{40}$$

where $\beta(k) = \sum_{m=1}^{N} \boldsymbol{\beta}_m(k)$, $\gamma(k) = \sum_{m=1}^{N} \boldsymbol{\gamma}_{2_m}(k)$ and the notation $\boldsymbol{v}_m(k)$ denotes the m-th element of the vector $\boldsymbol{v}(k)$.

Therefore, the signals received at points A and C, illustrated in Figure 7(a) can be regarded as two signals received by two spatially separated virtual antenna array elements, where the time delay τ represents the fractional time delay propagation between these two virtual antenna array elements, as shown in Figure 7(b), and the parameter D represents the time delay propagation between two antennas.



(a) (b)

Fig. 7. By adding the elements of the vector signal at point B, the two signals capted at points A and C are seen as two signals received by a virtual 2-element antenna array.

Consequently, by analogy with the fractional delay filter (FDF) proposed in (So et al., 1994), the fractional time delay between the signals $s(k)$ and $s(k\text{-}\tau)$ can be estimated using the fractional time delay estimation filter (FTDE) proposed in (So et al., 1994). This FTDE filter consists of a linear interpolator which can be implemented using an FIR filter based on the truncated *sinc*-interpolation method. In fact, this FIR filter in (So et al., 1994) is designed to estimate the fractional time delay D by minimizing the square error,$|e_h(k)|^2$, where the signal received by the 2nd antenna is taken as a reference signal for the LMS algorithm as shown in Figure 8.

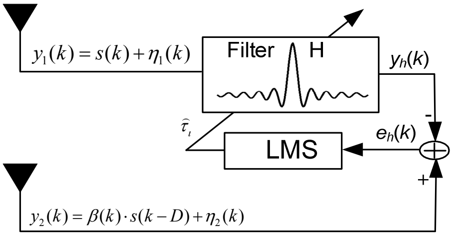


Fig. 8. FTDE filter used to estimate the fractional time delay propagation (D) between two spatially separated antennas.

We have developed this FTDE filter for complex signals, and subsequently we have adopted it in our work, to estimate the fractional time delay of the second path, as shown in Figure 9. However, instead of summing directly the signal x_1(k) at point B (as shown in Figure 9) to construct the reference signal of the FTDE filter, the beamforming filter w_{FD}, is inserted at point B to construct the FD-CMA filter and estimate the fractional time delay and its corresponding path as depicted in Figure 10. The weight vector of this filter (w_{FD}) is adapted using LMS algorithm (LMS_5 in Figure 10), where the first beamformer output, y_{MCMA}, is fed into the FTDE filter to generate a reference signal for the LMS algorithm, i.e., $y_{MCMA}(k - \hat{\tau})$. The weight adaptation for both FTDE and w_{FD} filters is nested in the sense that the output of the w_{FD} beamforming filter is used as a reference signal for FTDE filter and vice-versa. The delay block is inserted in order to align temporally the two signals and y_h and x_{1_D}, that we will explain later.

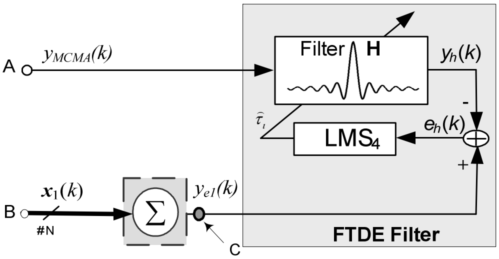


Fig. 9. FTDE filter for estimating the fractional delay of the signal received at point C.

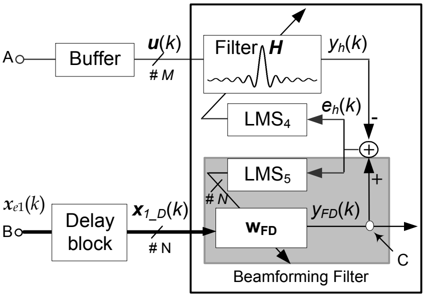


Fig. 10. FD-CMA Filter for frational time delay estimation and its corresponding path detection.

The following subsections present the FTDE filter developed and adopted in this work, and adaptive beamforming to estimate, the fractional delay and its corresponding path, respectively.

A- Fractional Time Delay Estimation

Once the first path is estimated by the MCMA filter, it is delayed by an estimated value $\hat{\tau}$ using the fractional time delay filter $\boldsymbol{H}$. This filtering is carried out by using the following equation using ideal fractional-delay filter with *sinc* function interpolation:

$$\begin{aligned} y_h(k) &\cong y_{MCMA}(k-\hat{\tau}) \\ &= \sum_{n=-\infty}^{\infty} sinc(n-\hat{\tau}) \cdot y_{MCMA}(k-n) \\ &= \sum_{n=-P}^{P} sinc(n-\hat{\tau}) \cdot y_{MCMA}(k-n), \end{aligned} \tag{41}$$

where the infinity sign in the summation is replaced by an integer P, which is chosen sufficiently large to minimize the truncation error. $\hat{\tau}$ is the instantaneous estimated time delay. If $\hat{\tau}$ is a fractional number, i.e. $0 < \hat{\tau} < 1$, the *sinc* interpolation impulse response has non-zero values for all n:

$$0 < \tau < 1 \rightarrow sinc(n-\tau) \neq 0, \quad \forall\, n \in \mathbb{Z}. \tag{42}$$

The delayed signal, $y_h(k)$, is the output of the FIR filter $\boldsymbol{H}$ whose coefficients are $sinc(n-\hat{\tau})$ and input is $y_{MCMA}(k)$. For this issue, a lookup table of the *sinc* function is constructed that consists of a matrix $\boldsymbol{H}$ of dimension K×(2P+1), with a generic element:

$$h_{ij} = sinc\left(\frac{i-1}{K} - j\right), \quad 1 \le i \le K, -P \le j \le P, \tag{43}$$

where K represents the inverse resolution over T_s of the estimated delay $\hat{\tau}$. The theoretical elements of the i-th row of the matrix $\boldsymbol{H}$ are therefore identical to the samples of the truncated *sinc* function with delay equal to:

$$\tau_i = \left(\frac{i-1}{K}\right), \quad 1 \le i \le K. \tag{44}$$

For the time delay estimation process, only the estimated time delay $\hat{\tau}_i(k)$ is adapted in our approach, and it is used as an index to obtain the vector $\boldsymbol{h}_i$ from a lookup table. As mentioned previously, this lookup table is a two-dimensional matrix called $\boldsymbol{H}$ of size $K\times(2P+1)$ that contains samples of the *sinc* function with delay ranging from 0 to $(K - 1)/K$. For a given vector τ with theoretically delayed value elements τ_i given by (44), the i-th row is computed as follows

$$i = \tau_i \cdot K + 1. \tag{45}$$

So, at each iteration, the integer part of $(\tau_i \cdot K + 1)$ is used to locate the i-th row of the matrix $\boldsymbol{H}$, i.e. $\boldsymbol{h}_i$, that is used to delay the signal $y_{MCMA}(k)$ using

$$y_h(k) = \boldsymbol{h}_i^H \cdot \boldsymbol{u}(k), \tag{46}$$

where $\boldsymbol{u}(k)$ is given by:

$$\boldsymbol{u}(k) = [y_{MCMA}(k), \dots, y_{MCMA}(k-(2P+1))]^T. \tag{47}$$

The estimated fractional time delay is obtained by using the gradient descent of the instantaneous squared error $|e_h|^2$ surface to locate the global minimum, i.e., using LMS (So et al., 1994). The estimated gradient is equal to the derivative of $|e_h|^2$ with respect to $\hat{\tau}$. The FTDE algorithm may be summarized as follows. The complex error signal, $e_h(k)$, is given by:

$$\begin{aligned} e_h(k) &= y_{FD}(k) - y_h(k) \\ &= y_{FD}(k) - \textstyle\sum_{n=-P}^{P} sinc(n-\hat{\tau}) \cdot y_{MCMA}(k-n), \end{aligned} \tag{48}$$

where

$$y_{FD}(k) = \boldsymbol{w}_{FD}^H(k) \cdot \boldsymbol{x}_{1\_D}(k), \tag{49}$$

$$\boldsymbol{x}_{1\_D}(k) = \boldsymbol{x}_1(k-(P+1)). \tag{50}$$

$x_{e1}(k)$ is delayed by $(P+1)$. T_s to be aligned with the output of the filter $\boldsymbol{H}$, i.e., $y_h(k)$, that has latency depending on its order value $M = 2P + 1$ as shown in Figure 10. The estimated time delay can be adapted by minimizing the cost function given by:

$$J(\tau, \boldsymbol{w}_{FD}) = E[|e_h(k)|^2] = E[|y_{FD}(k) - y_h(k)|^2]. \tag{51}$$

The constrained LMS algorithm becomes:

$$\hat{\tau}(k+1) = \hat{\tau}(k) - \mu_4 \cdot \nabla J(\tau, \boldsymbol{w}_{FD}), \tag{52}$$

where μ_4 is a small positive step size.

By differentiating the instantaneous error surface, $|e_h(k)|^2$, with respect to the estimated time delay, we have:

$$\frac{\partial\left(e_h^2(k)\right)}{\partial\hat{\tau}(k)} = \frac{\partial e_h(k)}{\partial\hat{\tau}(k)} \cdot e_h^*(k) + \frac{\partial e_h^*(k)}{\partial\hat{\tau}(k)} \cdot e_h(k)$$

$$= \sum_{n=-P}^{P} \frac{-\partial\left(sinc(n-\hat{\tau})\right)}{\partial\hat{\tau}(k)} y_{MCMA}(k-n)e_h^*(k) - \sum_{n=-P}^{P} \frac{\partial\left(sinc(n-\hat{\tau})\right)}{\partial\hat{\tau}(k)} y_{MCMA}^*(k-n)e_h(k)$$

$$= \sum_{n=-P}^{P} f(n-\hat{\tau})\, y_{MCMA}(k-n)e_h^*(k) + \sum_{n=-P}^{P} f(n-\hat{\tau})\, y_{MCMA}^*(k-n)e_h(k), \qquad (53)$$

where
$$f(v) = \frac{\cos(\pi v) - sinc(v)}{v}. \qquad (54)$$

Finally, the estimated time delay $\hat{\tau}$ is given by:

$$\hat{\tau}(k+1) = \hat{\tau}(k) - \mu_4 \cdot \left[\sum_{n=-P}^{P} f(n-\hat{\tau}) \cdot y_{MCMA}(k-n) \cdot e_h^*(k)\right.$$
$$\left. + \sum_{n=-P}^{P} f(n-\hat{\tau}) \cdot y_{MCMA}^*(k-n) \cdot e_h(k)\right]. \qquad (55)$$

In our implementation, lookup tables of *cos* and *sinc* functions are constructed for different values of v and used to calculate $f(n - \hat{\tau}(k))$. At each iteration, the integer part of $(\hat{\tau}_i(k) \cdot K + 1)$ is used to locate the i-th row of the matrix $\boldsymbol{H}$, i.e. $\boldsymbol{h}_i$ that is used to delay the signal $y_{MCMA}(k)$ by the estimated fractional delay $\hat{\tau}$ using (46).

B- Beamforming for fractional-delay path extraction

Now to extract the fractional-delay path, the weight vector of the FD-CMA filter is adapted using LMS by minimizing the cost function given in (51) as follows:

$$\boldsymbol{w}_{FD}(k+1) = \boldsymbol{w}_{FD}(k) + \mu_5 \cdot e_h^*(k) \cdot x_{1\_D}(k), \qquad (56)$$

where μ_5 is a small positive step size.

5. General SBB Approach

According to statistical modeling presented in (Boutin et al., 2008) of the studied underground channel, we were able to characterize, among many other channel parameters, the maximum number of paths at a given operation frequency and a given path resolution. Thus, we can assume for a given transmission rate and modulation type that the maximum number of paths arriving with delays that are a multiple integer of the sampling interval as well as the maximum number of paths arriving with fractional time delays are both predicted accurately. Consequently, we assume n paths causing ISI and p paths causing isi. In this general case of the presence of paths arriving with integer and fractional delay multiples of the sampling intervals, the two ID-CMA SBB and FD-CMA SBB proposed methods can be combined in a single approach named here as General Sequential Blind Beamforming (G-SBB) approach.

To simplify, the following study is performed using a three-path channel model for illustration purposes where the TPAs are given by $\tau_1 = 0$ (the strongest path), $\tau_2 = \tau < T_s$, and $\tau_3 = T_s$. Hence the received signal at the m-th antenna can be expressed by:

$$x_m(k) = A_{1m}s(k) + A_{2m}s(k-\tau) + A_{3m}s(k-T_s) + \eta_m(k). \qquad (57)$$

Figure 11 depicts the new approach using sequential blind spatial-domain path-diversity beamforming (SBB) to remedy both the ISI and isi problems using jointly CMA, LMS and adaptive FTDE filtering. This approach is designed to sequentially recover multipath rays by using multiple beamformings for received power maximization. First, the strongest path is extracted using the MCMA (AitFares et al., 2004; AitFares et al., 2006 a; AitFares et al., 2006 b; AitFares et al., 2008). Second, the path coming with delay that is multiple integer of the sampling interval is estimated using ID-CMA filter (i.e., y_{ID}) adapted using LMS with the CMA delayed output as a reference signal (AitFares et al., 2004). Finally, the path coming with fractional delay is estimated using FD-CMA filter (i.e., y_{FD}) (AitFares et al., 2006 a) adapted using LMS and FTDE. However, in order to ensure the estimated path arriving with the fractional delay, two ASC filters are used to extract the contribution of path $y_{MCMA}(k)$ and y_{ID} (k) from the received signal vector $x(k)$. As for the estimated path combination, we propose in the next section a combination based on MRC.

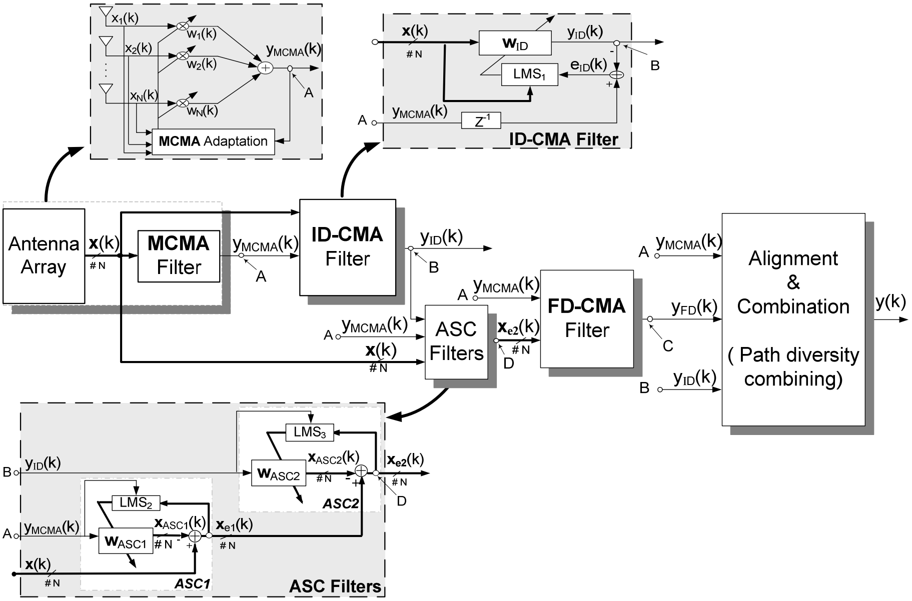


Fig. 11. Proposed G-SBB approach.

6. MRC Path Combination

The paths y_{MCMA}, y_{FD} and y_{ID}, estimated by the filters MCMA, FD-CMA and ID-CMA, respectively, possess a common phase ambiguity, since they are sequentially extracted using y_{MCMA} as a reference signal. As a result, a combination based on a simple addition of the estimated paths can only be constructive and it represents the output of a coherent Equal Gain Combiner (EGC) as illustrated in Figure 12(a). After appropriate delay alignments, the final estimated signal is given by EGC combining of the extracted paths as follows:

$$y(k) = y_{MCMA}(k) + y_{FD}(k) + y_{ID}(k+1). \tag{58}$$

For a Differential Binary Phase Shift Keying (DBPSK) modulation scheme, where the common phase ambiguity is actually a sign ambiguity, an EGC is equivalent to MRC.

However, for higher order modulations such as Differential Quadrature Phase Shift Keying (DQPSK), where the common phase ambiguity is an unknown angular rotation, more substantial improvement compared to EGC can be obtained by implementing coherent MRC with hard DFI as shown in Figure 12(b), which strives to force this common phase ambiguity to known quantized values that keep the constellation invariant by rotation (Affes & Mermelstein, 2003), thereby allowing coherent demodulation and MRC detection. In the first step, all paths y_{MCMA}, y_{FD} and y_{ID} are aligned by appropriate additional delays, and then scaled by an MRC weighting vector $\boldsymbol{g}(k)$. The summation of these scaled paths, $\tilde{s}(k)$, is given by

$$\tilde{s}(k) = \boldsymbol{g}^H(k) \cdot \boldsymbol{y}_d(k), \tag{59}$$

where

$$\boldsymbol{y_d}(k) = [y_{MCMA}(k), y_{FD}(k), y_{ID}(k+1)]^T, \tag{60}$$

$$\boldsymbol{g}(k) = [g_1(k), g_2(k), g_3(k)]^T. \tag{61}$$

In the next step, $\tilde{s}(k)$, is quantized by making a hard decision to match it to a tentative symbol $\hat{s}_h(k)$. This coherent-detection operation can be expressed as follows:

$$\hat{s}_h(k) = Hard\{\tilde{s}(k)\} = \arg_{a_k \,\epsilon\, A_M} min\{|\tilde{s}(k) - a_k|\}, \tag{62}$$

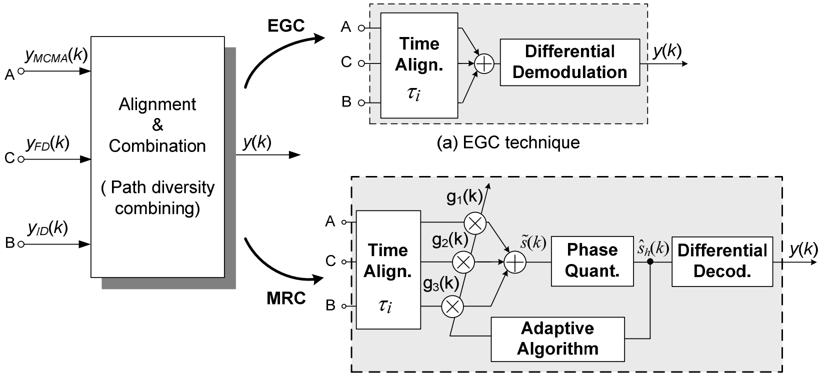
where A_M represents the MPSK modulation constellation defined by:

$$A_M = \{\dots, a_k, \dots\} = \left\{\dots, e^{\frac{j\pi((2k-1)-\delta(M-2))}{M}}, \dots\right\};\ k \,\epsilon\, \{1, \dots, M\}. \tag{63}$$

Since $\hat{s}_h(k)$ provides a selected estimate of the desired signal, it can be used as a feedback reference signal to update the weight vector $\boldsymbol{g}(k)$ using LMS-type adaptation referred to as Decision Feedback Identification (DFI):

$$\hat{\boldsymbol{g}}(k+1) = \hat{\boldsymbol{g}}(k) - \mu_6 \cdot \left(\boldsymbol{y}_d(k) - \hat{\boldsymbol{g}}(k) \cdot \hat{s}_h(k)\right) \cdot \hat{s}_h^*(k), \tag{64}$$

where μ_6 is a small positive step size.



(a) EGC technique

(b) Coherent Detection MRC with hard DFI technique

Fig. 12. Path diversity combining stage for the SBB using EGC or Coherent MRC with hard DFI.

It is this DFI procedure that enables coherent MRC detection by forcing the common phase ambiguity of the extracted paths to a value by which the constellation is invariant by rotation (Affes & Mermelstein, 2003; Aitfares et al., 2008). Finally the desired output signal $y(k)$ is estimated from $\hat{s}_h(k)$ by differential decoding, as shown in Figure 12(b), instead of differential demodulation needed previously with simple EGC. This final decoding step is expressed by:

$$y(k) = \hat{s}_h(k) \cdot s_h^*(k-1). \tag{65}$$

The proposed SBB technique enabling MRC path diversity combining (i.e., MRC-SBB) offers an *SNR* gain of about 2 *dB* gain compared to that using simple EGC implementation (i.e., EGC-SBB) (Affes & Mermelstein, 2003; Aitfares et al., 2008).

7. Computer simulation results

In this section, simulation results are presented to assess the performance of the proposed SBB method and to compare it with MCMA beamforming (Oh & Chin, 1995). A two-element array with half-wavelength spacing is considered. A desired signal is propagated along four multipaths to the antenna array while the interference and noise are simulated as additive white Gaussian noise. The first path is direct with a path arrival-time delay $\tau_1 = 0$. The second and third paths arrive, respectively, with delays τ_2 and τ_3 lower than the sampling interval, and the last path arrives with delay $\tau_4 = T_s$. Differential encoding is employed to overcome the phase ambiguity in the signal estimation. Performance study was carried out with two channel models and for two kinds of modulation (DBPSK and DQPSK). Type-A channel is Rayleigh fading with a Doppler shift $f_{d1} = 20$ *Hz*. Type-B channel is Rayleigh fading with a higher Doppler shift $f_{d2} = 35$ *Hz*. The use of these two Doppler frequencies reflects the typical range of the vehicle speed in underground environments[2]. The Bit Error Rate (*BER*) performance for different Doppler frequencies (f_{d1} and f_{d2}) was also studied. The figure of merit is the required *SNR* to achieve a *BER*[3] below 0.001. Table 1 summarizes the system parameters for the computer simulations.

---

[2] For operations at a carrier frequency $f_c = 2{:}4$ *GHz* and vehicle speeds $v_1 = 10km{=}h$, and $v_2 = 15km{=}h$, we found approximately that $f_{d1} = 20$ *Hz* and $f_{d2} = 35Hz$.

[3] The *BER* is calculated after steady-state convergence to avoid biasing the results.

Modulation	DBPSK or DQPSK.
Antenna array type	Linear uniform, with $\lambda/2$ element spacing.
Antenna array size	2 elements or 4 elements.
Max. Doppler frequency	f_{d1}=20Hz and f_{d2} =35Hz.
Channel model	Type-A: Rayleigh fading with f_{d1} Type-B: Rayleigh fading with f_{d2}
Adaptive algorithm	CMA & LMS
Carrier Frequency	f_c=2.4GHz
Noise	AWGN
Filter order	M=21
Path resolution	K =200, i.e. T_r=0.005 T_s
Step sizes	μ=0.009; μ_1= 0.008; μ_2=0.0095; μ_3=0.008; μ_4= 0.001; μ_5= 0.009 and μ_6= 0.001.
Number of symbol	10.000

Table 1. Simulation parameters.

Figs. 13 and 14 show the measured *BER* performance versus *SNR* of G-SBB and MCMA for Type-A and -B channels, with different values of τ_2 and τ_3 using a DBPSK modulated signal. As expected, it can be noted that for both algorithms, the *BER* performance decreases with increasing Doppler frequency values. Despite the speed increasing due to the Doppler effect, the proposed algorithm G-SSB provides significant gains and outperforms MCMA by approximately 5 *dB* for both channel environments (A and B).

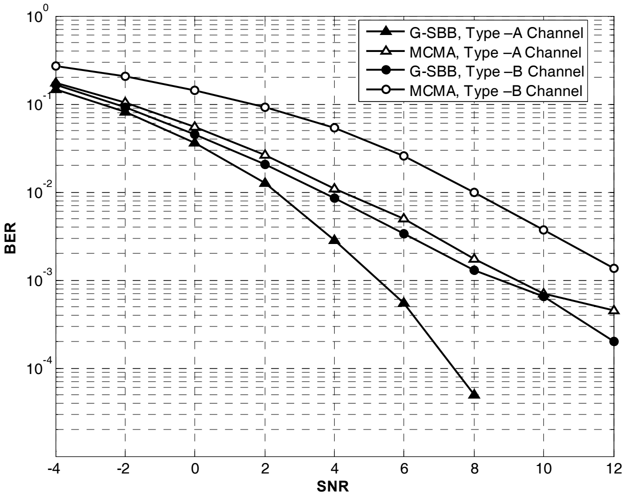


Fig. 13. *BER* performance versus *SNR* with τ_2=0.4T_s and τ_3= 0.8T_s for DBPSK modulation scheme using a 2-element antenna array.

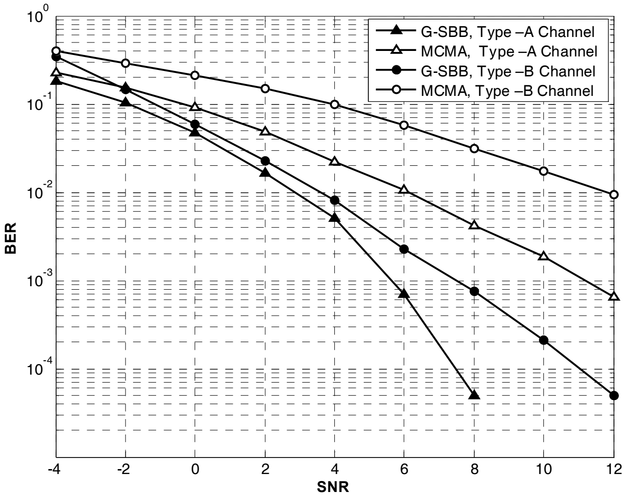


Fig. 14. BER performance versus SNR with τ_2=0.3T_s and τ_3= 0.7T_s for DBPSK modulation scheme using a 2-element antenna array.

Let us now study the convergence rate of the proposed G-SBB method compared to the MCMA algorithm for the Type-A channel with τ_2= 0.4 T_s and τ_3 = 0.8 T_s at 2.4 *GHz* and for *SNR* = 4 *dB*. Figure 15 illustrates the average *BER* in terms of the number of iterations for the first 8000 samples. A benchmark comparison with AAA using the LMS algorithm is also provided. From Figure 15, it can be seen that the LMS algorithm is the fastest one followed by the MCMA and than the G-SBB algorithms. However, the proposed G-SBB algorithm reaches a much lower steady-state *BER* after convergence within a shorter delay compared to AAA and MCMA.

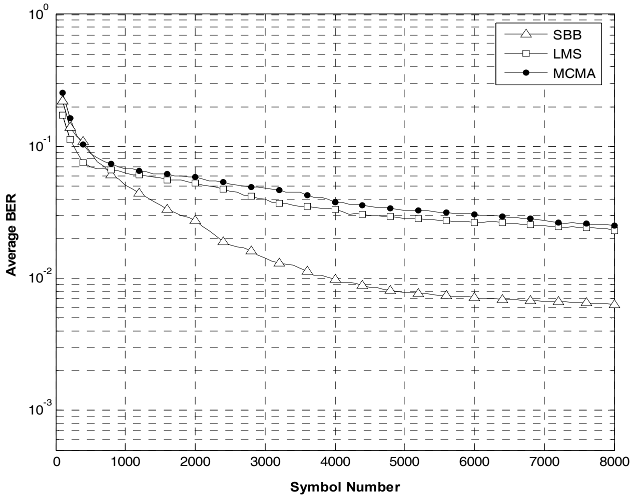


Fig. 15. The real-time performance of the proposed system compared with the MCMA and LMS algorithms at *SNR* = 4 *dB* for DBPSK modulation scheme using a 2-element antenna array.

Here we discuss the trade-off between the hardware complexity related to the delay resolution implementation and the *BER* performance. As mentioned above, *K*, given in equation (44), represents the number of the tap filter coefficients used to implement the fractional delay resolution. For instance, when K = 10, the delay resolution is equal to $T_r=1/(K.T_s)$ = 0.1 T_s. By increasing the value of *K*, we increase the FTDE resolution and consequently the FTDE filter will be able to estimate faithfully the fractional delay path which will in turn improve the *BER* performance. On the other hand, increasing *K* increases the hardware complexity needed to implement the FTDE. To find an optimal trade-off between resolution and hardware complexity, several simulations with different values of *K* in terms of *BER* performance were conducted.

Figure 16 illustrates the simulated *BER* performance versus *SNR* of the G-SBB for Type-A channel environment at different values of T_r. From this figure, it can be seen that the resolution of *K* impacts greatly the *BER* performance when *K* is less than 50. For *K* greater than 50, the optimal performance is attained and further increase of the *K* value is unnecessary.

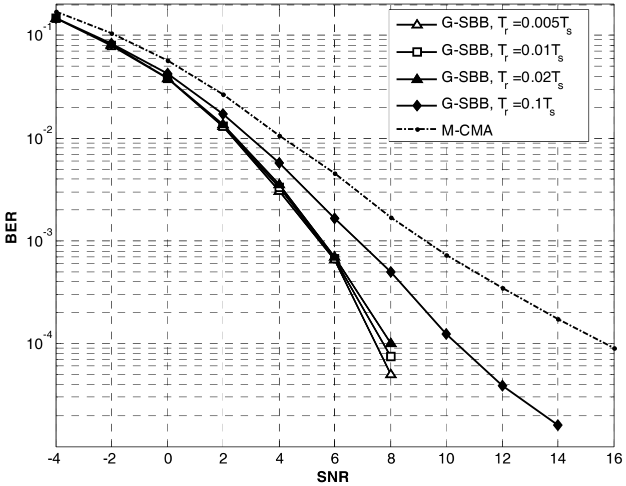


Fig. 16. *BER* performance versus *SNR* in Type -A Channel for τ_2= 0.4T_s and τ_3 = 0.8T_s when T_r is varied using a 2-element antenna array.

For high order modulation using DQPSK, Figs. 17 and 18 illustrate the *BER* performance versus *SNR* for G-SBB using MRC or EGC in the combining step for Type-A and –B channels with τ_2= 0.4 T_s and τ_3= 0.8 T_s, respectively, at 2.4 *GHz*. A benchmark comparison with AAA using MCMA is also provided. For the type-A channel, the results show that G-SBB with MRC provides a good enhancement and outperforms G-SBB with EGC and the AAA using MCMA by approximately 2 *dB* and up to 7 *dB* at a required *BER* =0.001, respectively (Figure 17). For the type- B channel with higher Doppler frequency, the measured results show that G-SBB with MRC maintains its advantage compared to G-SBB with EGC and to the AAA using MCMA where improvements of approximately 2 *dB* and up to 7 *dB* at a required *BER*=0.001 are obtained, respectively (Figure 18).

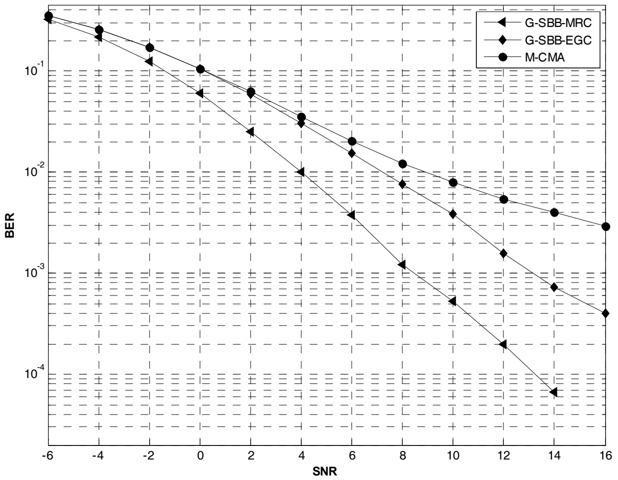


Fig. 17. *BER* performance versus *SNR* for Type -A Channel with τ_2=0.4T_s and τ_3= 0.8T_s for DQPSK modulation scheme using a 2-element antenna array.

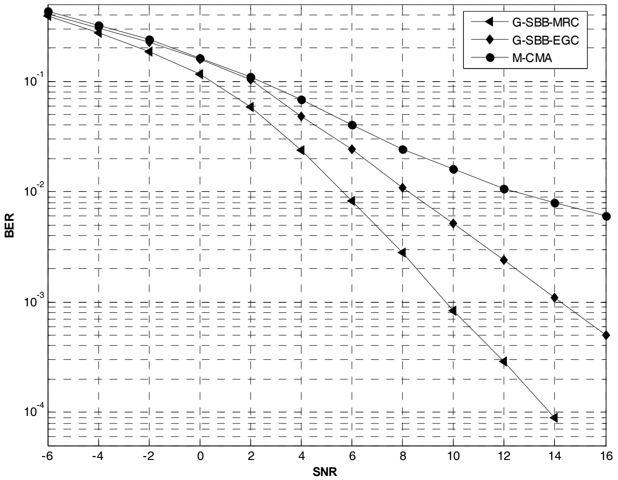


Fig. 18. *BER* performance versus SNR for Type -B Channel with τ_2=0.4T_s and τ_3= 0.8T_s for DQPSK modulation scheme using a 2-element antenna array.

Figure 19 shows the measured *BER* performance versus *SNR* for G-SBB using MRC or EGC in the combining step and with MCMA-AAA for Type-A channel using four antenna elements (*N* = 4). Again, it is clear that the G-SBB using the proposed MRC is more efficient than both previous G-SBB versions using EGC and the conventional MCMA algorithm.

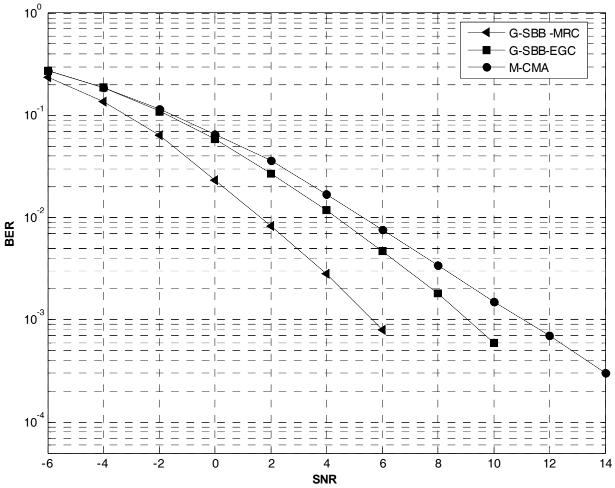


Fig. 19. *BER* performance versus *SNR* for Type -A Channel with τ_2=0.2T_s and τ_3= 0.8T_s for DQPSK modulation scheme using a 4-element antenna array.

8. Conclusion

In this Chapter, a new approach using sequential blind spatial-domain path-diversity beamforming (SBB) to remedy the ISI and isi problems has been presented. Using jointly CMA, LMS and adaptive FTDE filtering, this approach has been designed to sequentially recover multipath rays to maximize the received power by extracting all dominant multipaths. MCMA is used to estimate the strongest path while the integer path delay is estimated sequentially using adapted LMS with the first beamformer output as a reference signal. A new synchronization approach for multipath propagation, based on combining a CMA-AAA and adaptive fractional time delay estimation filtering, has been proposed to estimate the fractional path delay. It should be noted that the G-SBB architecture can be generalized for an arbitrary number of received paths causing ISI where several concurrent filters (ID-CMA and FD-CMA) can be implemented to resolve the different paths. Finally, to combine these extracted paths, an enabling MRC path diversity combiner with hard DFI has also been proposed. Simulation results show the effectiveness of the proposed SBB receiver especially at high *SNR*, where it is expected to operate in a typical underground wireless environment (Nerguizian et al., 2005).

9. References

AitFares, S.; Denidni, T. A. & Affes, S. (2004). Sequential blind beamforming algorithm using combined CMA/LMS for wireless underground communications, in *Proc. IEEE VTC'04*, vol. 5, pp. 3600-3604, Sept. 2004.

AitFares, S; Denidni, T. A.; Affes, S. & Despins, C. (2006). CMA/fractional delay sequential blind beamforming for wireless multipath communications, in *Proc. IEEE VTC'06*, vol. 6, pp. 2793-2797, May 2006.

AitFares, S; Denidni, T. A.; Affes, S. & Despins, C. (2006). Efficient sequential blind beamforming for wireless underground communications, in *Proc. IEEE VTC'06*, pp. 1-4, Sept. 2006.

AitFares, S; Denidni, T. A.; Affes, S. & Despins, C. (2008). Fractional-Delay Sequential Blind Beamforming for Wireless Multipath Communications in Confined Areas. IEEE *Transactions on Wireless Communications*, vol. 7, no. 1, pp. 1-10, January 2008.

Affes, S. & Mermelstein, P. (2003). Adaptive space-time processing for wireless CDMA, chapter 10, pp. 283-321, *in Adaptive Signal Processing: Application to Real-World Problems*, J. Benesty and A. H. Huang, eds. Berlin: Springer, 2003.

Amca, H.; Yenal, T. & Hacioglu, K. (1999). Adaptive equalization of frequency selective multipath fading channels based on sample selection, *Proc. IEE on Commun.*, vol. 146, no. 1, pp. 55-60, Feb. 1999.

Bellofiore, S.; Balanis, C. A.; Foutz, J. & Spanias, A. S. (2002). Smart antenna systems for mobile communication networks, part 1: overview and antenna design, *IEEE Antennas Propag. Mag.*, vol. 44, no. 3, pp. 145-154, June 2002.

Bellofiore, S.; Foutz, J.; Balanis, C. A. & Spanias, A. S. (2002). Smart-antenna systems for mobile communication networks, part 2: beamforming and network throughput, *IEEE Antennas Propagation Magazine*, vol. 44, no. 4, pp. 106-114, Aug. 2002.

Boutin, M. ; Benzakour, A; Despins, C & Affes, S. (2008). Radio Wave Characterization and Modeling in Underground Mine Tunnels, *IEEE Transaction on Antennas and Propagation*, vol. 56, no. 2, pp. 540-549, February 2008.

Chao, R. Y. & Chung, K. S. (1994). A low profile antenna array for underground mine communication, in *Proc. ICCS 1994*, vol. 2, pp. 705-709, 1994.

Cozzo, C. & Hughes, B. L. (2003). Space diversity in presence of discrete multipath fading channel, *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1629-1632, Oct. 2003.

Furukawa, H.; Kamio, Y. & Sasaoka, H. (1996). Co-Channel interference reduction method using CMA adaptive array antenna, *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 512-516, 1996.

Godara, L. C. (1997). Applications of antenna arrays to mobile communications, part I: performance improvement, feasibility, and system considerations, *Proc. IEEE*, vol. 85, no. 7, pp. 1031-1060, July 1997.

Lee, W. C. & Choi, S. (2005). Adaptive beamforming algorithm based on eigen-space method for smart antennas, *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 888-890, Oct. 2005.

McNeil, D.; Denidni, A. T. & Delisle, G. Y. (2001). Output power maximization algorithm performance of dual-antenna for personal communication handset applications, in *Proc. IEEE Antennas and Propagation Society International Symposium*, vol. 1, pp. 128-131, July 2001.

Nerguizian, C.; Despins, C; Affes, S. & Djadel, M. (2005). Radio-channel characterization of an underground mine at 2:4 *GHz* wireless communications, *IEEE Trans. Wireless Communication*, vol. 4, no. 5, pp. 2441-2453, Sept. 2005.

Oh, K. N. & Chin, Y. O. (1995). New blind equalization techniques based on constant modulus algorithm, in *Proc. Global Telecommunications Conference*, vol. 2, pp. 865-869, Nov. 1995.

Ogawa, Y.; Fujishima, K. & Ohgane, T. (1999). Weighting factors in spatial domain path-diversity using an adaptive antenna," in *Proc. IEEE VTC'99*, vol. 3, pp. 2184-2188, May 1999.

Sanada, Y. & Wang, Q. (1997). A co-channel interference cancellation technique using orthogonal convolutional codes on multipath Rayleigh fading channel, *IEEE Trans. Veh. Technol.*, vol. 46, no. 1, pp. 114-128, Feb. 1997.

Saunders, S. R. (1999). *Antenna and Propagation for Wireless Communication Systems.* Chichester, England: John Wiley & Sons, Ltd., 1999.

Slock, D. T. M. (1994). Blind joint equalization of multiple synchronous mobile users using over sampling and/or multiple antennas, in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1154-1158, 1994.

Stott, J. H. (2000). The how and why of COFDM, tutorial COFDM, BBC Research and Development, http://www.ebu.ch/en/technical/trev/trev 278-stott.pdf.

So, H. C.; Ching, P. C. & Chan, Y. T. (1994). New algorithm for explicit adaptation of time delay, *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1816-1820, July, 1994.

Tanabe, Y. et al. (2000), An adaptive antenna for spatial-domain path-diversity using a super-resolution technique, in *Proc. IEEE VTC'00*, vol. 1, pp. 1-5, May 2000.

Tanaka, T. (1994). A study on multipath propagation characteristics for RAKE receiving technique, in *Proc. 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communication*, vol. 2, pp. 711- 714, Sept. 1994.

Valimaki, V. & Laakso, T. I. (2000). Principles of fractional delay filters," in *Proc. IEEE ICASS*, vol. 6, pp. 3870-3873, June 2000.

Widrow, B.; Mantey, P. E.; Griffiths, L. J. & Goode, B. B. (1967). Adaptive antenna systems," *Proc. IEEE*, vol. 55, no. 12, pp. 2143-2159, Dec. 1967.

Youna, W. S. & Un, C. K. (1994). Robust adaptive beamforming based on the eigen-structure method, *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1543-1547, June 1994.

Yuan, J.T. & Tsai, K.D. (2005). Analysis of the multimodulus blind equalization algorithm in QAM communication systems, IEEE Transactions on Communications, vol. 53, no. 9, pp. 1427-1431, 2005.

6

Space-Time Diversity Techniques for WCDMA High Altitude Platform Systems

Abbas Mohammed
Blekinge Institute of Technology
Sweden

Tommy Hult
Lund University
Sweden

1. Introduction

Third generation mobile systems are gradually being deployed in many developed countries in hotspot areas. However, owing to the amount of new infrastructures required, it will still be some time before 3G is ubiquitous, especially in developing countries. One possible cost effective solution for deployments in these areas is to use High Altitude Platforms (HAPs) (Collela et al., 2000; Djuknic et al., 1997; Grace et al., 2001; 2005; Miura & Oodo, 2002; Park et al., 2002; Steele, 1992; Thornton et al., 2001; Tozer & Grace, 2001) for delivering 3G (WCDMA) communications services over a wide coverage area (Dovis et al., 2002; Falletti & Sellone, 2005; Foo et al., 2000; Masumura & Nakagawa, 2002; Vazquez et al., 2002). HAPs are either airships or planes that will operate in the stratosphere, 17-22 km above the ground. This unique position offers a significant link budget advantage compared with satellites and much wider coverage area than conventional terrestrial cellular systems. Such platforms will have a rapid roll-out capability and the ability to serve a large number of users, using considerably less communications infrastructure than required by a terrestrial network (Steele, 1992). In order to aid the eventual deployment of HAPs the ITU has allocated spectrum in the 3G bands for HAPs (ITU, 2000a), as well as in the mm-wave bands for broadband services at around 48 GHz worldwide (ITU, 2000b) and 31/28 GHz for certain Asian countries (Oodo et al., 2002).
Spectrum reuse is important in all wireless communications systems. Cellular solutions for HAPs have been examined in (El-Jabu, 2001; Thornton et al., 2003), specifically addressing the antenna beam characteristics required to produce an efficient cellular structure on the ground, and the effect of antenna sidelobe levels on channel reuse plans (Thornton et al., 2003). HAPs will have relatively loose station-keeping characteristics compared with satellites, and the effects of platform drift on a cellular structure and the resulting inter-cell handover requirements have been investigated (Thornton et al., 2005). Cellular resource management strategies have also been developed for HAP use (Grace et al., 2002).
Configurations of multiple HAPs can also reuse the spectrum. They can be used to deliver contiguous coverage and must take into account coexistence requirements (Falletti & Sellone, 2005; Foo et al., 2000). A technique not widely known is their ability to serve the same

coverage area reusing the spectrum to allow capacity enhancement. Such a technique has already been examined for TDMA/FDMA systems (Chen et al., 2005; Grace et al., 2005; Liu et al., 2005). In order to achieve the required reduction in interference needed to permit spectrum reuse, the highly directional user antenna is used to spatially discriminate between the HAPs. The degree of bandwidth reuse and resulting capacity gain is dependent on several factors, in particular the number of platforms and the user antenna sidelobe levels. An alternative method of enhancement is to apply space-time diversity techniques, such as Single-Input Multiple-Output (SIMO) receive diversity or Multiple-Input Multiple-Output (MIMO) diversity, to improve the spectrum reuse in the multiple HAP scenario.

In the case of many 3G systems the user antenna is either omni-directional or at best low gain, so in these cases it cannot be used to achieve the same effects. The purpose of this chapter is to examine how the unique properties of a WCDMA system can be exploited in multiple HAP uplink architectures to deliver both coverage and capacity enhancement (without the need for the user antenna gain).

In addition to the spectral reuse benefits, there are three main benefits for a multiple HAP architecture:

- The configuration also provides for incremental roll-out: initially only one HAP needs to be deployed. As more capacity is required, further HAPs can be brought into service, with new users served by the newly deployed HAPs.
- Multiple operators can be served from individual HAPs, without the need for complicated coexistence criteria since the individual HAPs could reuse the same spectrum.
- HAPs will be payload power, volume and weight constrained, limiting the overall capacity delivered by each platform. Capacity densities can be increased with more HAPs. Moreover, it may be more cost effective to use more lower capability HAPs (e.g., solar powered planes), rather than one big HAP (e.g., solar powered airship), when covering a large number of cells (Grace et al., 2006).

The chapter is organized as follows: in section 2 the multiple HAP scenario is explained. The interference analysis is presented in section 3. In section 4 we examine the completely overlapping coverage area case, different numbers of platforms, and simulation results showing the achievable capacity enhancement are presented. Finally, conclusions are presented in section 5.

2. Multiple HAP system setup

In this chapter we use a simple geometric positioning of the high altitude platforms to create signal environments that can easily be compared and analyzed. In each constellation, the HAPs are located with equal separation along a circular contour, as shown in figure 1.

The separation distance d_m along the line from the vertical projection of the HAP on the ground to the cell centre is varied from 70 km to zero (i.e., all the HAPs will be located on top of each other in the latter case). All HAPs are assumed to be flying in the stratosphere at an altitude of 20 km. The size of the coverage area assigned to each HAP is governed by the shape of the base station antenna pattern. If we assume that we only have one cell per HAP, then the coverage area is also synonymous with the total cell area of the HAP.

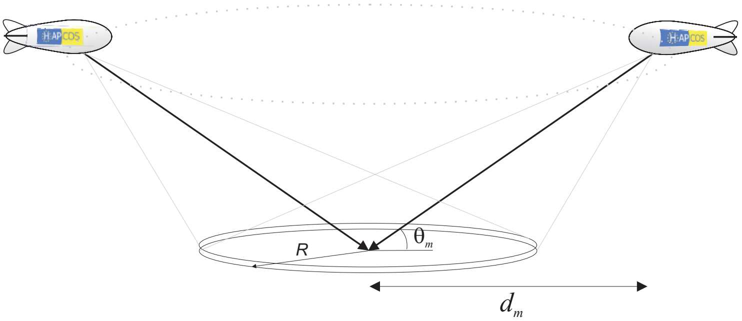


Fig. 1. An example of a system simulation setup with $N = 2$ HAPs with overlapping cells of radius R. d_m is the distance on the ground between the cell centre and the vertical projection of the HAP on the ground and θ_m is the elevation angle towards the HAP.

2.1 User Positioning Geometry

Each UE (User Equipment) is positioned inside the cell according to an independent uniform random distribution over the cell coverage area with radius R, as shown in figure 2. The position of each UE inside each cell is defined relative to the HAP base station that it is connected to, and also relative to every other HAP borne base station. This is necessary in order to evaluate the impact of interference between the different UE-HAP transmission paths.

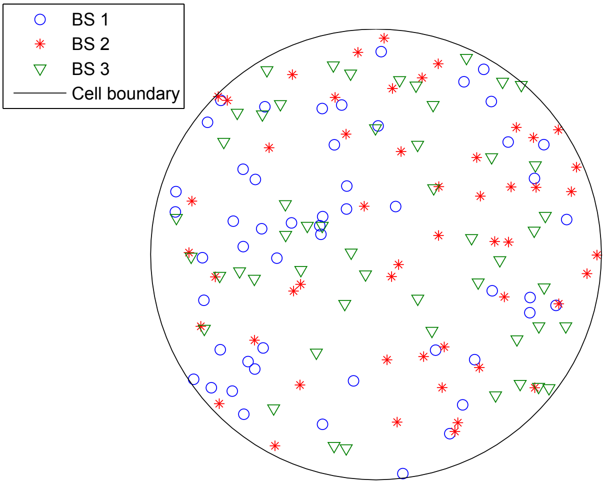


Fig. 2. A plot showing a sample distribution of 150 UE, where 50 UE are assigned to each of the three base stations (BS1, BS2 and BS3).

2.2 Base station antenna pattern

The base station antenna pattern for the simulations were chosen to be simple but detailed enough to show the effects of the main and side lobes, especially in the null directions, as illustrated in figure 3. A simple rotationally symmetric pattern based on a Bessel function is used for this purpose, and is defined by (Balanis, 1997)

$$G(\varphi) \approx 0.7 \cdot \left(\frac{2 \cdot J_1\left(\frac{70\pi}{\varphi_{3dB}} \sin(\varphi)\right)}{\sin(\varphi)} \right)^2, \tag{1}$$

where $J_1(\cdot)$ is a Bessel function of the first kind and order 1, φ_{3dB} is the 3 dB beamwidth in degrees of the main antenna lobe. The 3 dB beamwidth of the antenna is computed from the desired cell radius according to

$$\varphi_{3dB} = 2 \cdot \arctan\left(\frac{\text{cell radius}}{\text{HAP altitude}}\right). \tag{2}$$

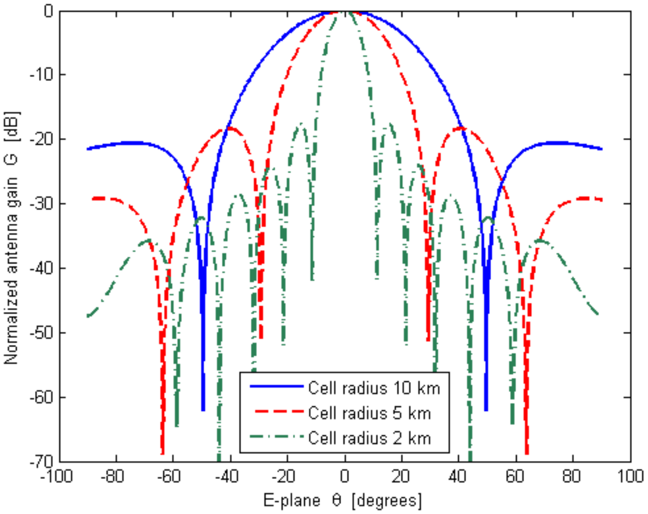


Fig. 3. HAP base station antenna patterns for different cell radii.

2.3 User equipment antenna pattern

In this analysis we assume that each UE employs a directive antenna and communicates with its corresponding HAP basestation. Using this assumption we only need to set the desired maximum gain of the UE antenna we want to use, as shown Table 1. The antenna pattern of the directive antennas is calculated according to equation (1), but with a fixed maximum gain instead of a fixed main beamwidth, the beamwidth is then $\varphi(G_{max})$.

User Equipment	Max. ant. Gain [dBi]
Mobile phone	0
Data terminal	2,4,12

Table 1. Antenna gains used in the simulation setup.

2.4 UE-HAP radio propagation channel model

In this chapter we use the Combined Empirical Fading Model (CEFM) together with the Free Space Loss (FSL) model. CEFM combines the results of the Empirical Roadside Shadowing (ERS) model (Goldhirsch & Vogel, 1992) for low elevation angles with the high elevation angle results from (Parks et al., 1993) for the L and S Bands. Using the FSL model the path loss from UE n to HAP base station m, is given by

$$l_{m,n}^{FSL} = \frac{(4\pi \cdot d_n^m)^2}{G_{m,n}^{tx} \cdot G_{m,n}^{rx} \cdot \lambda^2}, \tag{3}$$

where $d_{m,n}$ is the line of sight distance between the UE n and HAP m. The receiver $G_{m,n}^{rx}$ and transmitter $G_{m,n}^{tx}$ antenna gain patterns are calculated using equations (1) and (2), respectively. The carrier frequency f_c used in the simulation is 1.9 GHz which gives a wavelength λ of 0.1579 meters. The CEFM fading loss associated to HAP m is calculated as

$$L_f(\theta_m) = a \cdot \log_e(p) + b \quad [\text{dB}], \tag{4}$$

where p is the percentile outage probability, and the data fitting coefficients a and b are calculated according to (Goldhirsch & Vogel, 1992)

$$\begin{cases} a = 0.002 \cdot \theta_m^2 - 0.15 \cdot \theta_m - 0.7 - 0.2 \cdot f_c \\ b = 27.2 + 1.5 \cdot f_c - 0.33 \cdot \theta_m \end{cases}, \tag{5}$$

where θ_m is the elevation angle of HAP m. The total channel gain from UE n to HAP m is then given by

$$g_{m,n}(\theta_m) = \left(l_{m,n}^{FSL} \cdot 10^{\left(\frac{L_f(\theta_m)}{10}\right)} \right)^{-1}. \tag{6}$$

2.5 WCDMA Setup

The different service parameters used in this chapter are collected from the 3GPP standard (3GPP, 2005) and are summarized in Table 2. In order to account for the relative movement between the UE and the base stations, a fading propagation channel model based on equation (6) is simulated. This results in a Block Error Rate (BLER) requirement of 1% for the 12.2 kbps voice service and a BLER of 10% for 64, 144 and 384 kbps data packet services, respectively.

	Type of service			
Parameters	Voice	Data	Data	Data
Chip rate	3.84 Mcps			
Data rate	12 kbps	64 kbps	144 kbps	384 kbps
Req. E_b/N_0	11.9 dB	6.2 dB	5.4 dB	5.8 dB
Max. Tx. Power	125 mW	125 mW	125 mW	250 mW
Voice activity	0.67	1	1	1

Table 2. WCDMA service parameters employed in the simulation.

2.6 Space-Time Diversity Techniques

The spatial properties of wireless communication channels are extremely important in determining the performance of the systems. Thus, there has been great interest in employing space-time diversity schemes since they can offer a broad range of ways to improve wireless systems performance. For instance, receiver diversity techniques such as Single-Input Multiple-Output (SIMO) and Multiple-Input Multiple-Output (MIMO) can enhance link quality through diversity gain or increase the potential data rate or capacity through multiplexing gain. In this section, we apply these techniques to HAPs and in the next section we determine their impact on performance via simulations.

In this scenario, we assume that the link between the UE and the HAP BS is setup according to the previous sections in this chapter. The total spatio-temporal and polarization degrees of freedom is, in an Orthogonal User Multiple Access SIMO system, restricted by the number of users and the number of receiving antennas. If E_s is the average transmit energy per symbol, the received signal r is given by (Li & Wang., 2004)

$$r = \sqrt{E_s} \cdot \mathbf{w}^H \mathbf{h} s + \mathbf{w}^H \mathbf{n}, \tag{7}$$

where s is the transmitted signal, $\mathbf{h}$ is the channel response vector, $h_n = |h_n| e^{j\phi_n}, n = 1, 2, \cdots, N_{rx}$, for all receiving antennas, in which $|h_n|$ is defined as the inverse of the channel gain in equation (6) assuming that the separate channels are independent. The received noise vector $\mathbf{n}$ for all receiving antennas is assumed to be AWGN and $\mathbf{w}$ are the combining weights at the receiver. Choosing the combining weights $\mathbf{w}$ to be equal to the channel response vector $\mathbf{h}$ will result in the Maximum Ratio Combining (MRC) method, which can be represented as

$$r = \sqrt{E_s} \cdot ||\mathbf{h}||^2 s + \mathbf{h}^H \mathbf{n}. \tag{8}$$

The SNR for the received signal can now be written as

$$\text{SNR}_{MRC} = \frac{\left(\sqrt{E_s} \cdot ||\mathbf{h}||^2\right)^2}{\left(\mathbf{h}^H \mathbf{n}\right)^2} = \frac{s \cdot E_s}{\sigma_n^2} \cdot \mathcal{E}\left\{\frac{||\mathbf{h}||^4}{||\mathbf{h}||^2}\right\} = \text{SNR}_n \cdot ||\mathbf{h}||^2 = \text{SNR}_n \cdot N_{rx}, \tag{9}$$

where SNR_n is the signal to noise ratio in each receiving antenna and N_{rx} is the number of receiving antennas.

A similar combining method as in the SIMO receiver diversity is used in the MIMO diversity method. MIMO diversity utilize N_{tx} transmitting antennas and N_{rx} receiving antennas and assumes the channel response matrix $H_{nm} = |H_{nm}| e^{j\phi_{nm}}, n = 1, 2, \cdots, N_{rx}, m = 1, 2, \cdots, N_{tx}$. $|H_{nm}|$ is the inverse of the channel gain from equation (6), and provided that the separate channels are independent then $\mathbf{H}$ is a diagonal matrix. The noise is AWGN and the received signal from the MIMO diversity system can then be expressed as (Li & Wang., 2004)

$$r = \sqrt{E_s} \cdot \mathbf{w}_{rx}^H \mathbf{H} \mathbf{w}_{tx} s + \mathbf{w}_{rx}^H \mathbf{n}, \tag{10}$$

The SNR for the received signal is then given by

$$\text{SNR}_{MRC} = \frac{\left(\sqrt{E_s} \cdot ||\mathbf{H}||_F^2\right)^2}{\left(\mathbf{H}^H \mathbf{n}\right)^2} = \frac{s \cdot E_s}{\sigma_n^2} \cdot \mathcal{E}\left\{\frac{||\mathbf{H}||_F^4}{||\mathbf{H}||_F^2}\right\} = \text{SNR}_n \cdot N_{tx} \cdot N_{rx}, \tag{11}$$

where SNR_n is the signal to noise ratio in each receiving antenna and N_{rx} is the number of receiving antennas and N_{tx} is the number of transmitting antennas.

3. Interference analysis

Assuming that we have a setup of M different HAPs covering the same cell area and N users connected to each HAP, we can denote each UE position as $(x_{m,n}, y_{m,n})$, where $n = \{1, 2, \ldots, N\}$ and $m = \{1, 2, \ldots, M\}$. An example of a scenario setup with $N = 50$ and $M = 3$ is shown in figure 2. The maximum power $p_{m,n}^{tx}$ that the user in location $(x_{m,n}, y_{m,n})$ is transmitting dependent of the type of service used and can be obtained from Table 2. In WCDMA systems, power control is a powerful and essential method exerted in order to mitigate the near-far problem. The power received at base station (HAP) m from user n is

$$p_{m,n}^{rx}(\theta_m) = p_{m,n}^{tx} \cdot g_{m,n}(\theta_m), \tag{12}$$

where $g_{m,n}(\theta_m)$ is the total link gain, as defined in equation (6), between UE transmitter n and its own cell's BS receiver m. To be able to maintain a specific quality of service we need to assert that we maintain a good enough SINR (Signal to Interference plus Noise Ratio) level. From Table 2 we can see the required E_b/N_0 values for different services, and we can express the required SINR, $\gamma_{m,n}$ for user n at HAP base station m as

$$\gamma_{m,n}^{req} = \frac{R}{W} \cdot \left(\frac{E_b}{N_0}\right)_{req}, \tag{13}$$

where R is the data rate of the service and W is the *Chip-rate* of the system. The required SINR can then be expressed as

$$\gamma_{m,n}^{req} = \frac{p_{m,n}^{rx}}{I_{tot}} = \frac{p_{m,n}^{tx}}{\sum\limits_{m'=1}^{M} \sum\limits_{\substack{n'=1 \\ n' \neq n}}^{N} p_{m,n}^{tx} \cdot \frac{g_{m',n'}(\theta_{m'})}{g_{m,n}(\theta_m)} + \frac{p_w}{g_{m,n}(\theta_m)}}, \quad \begin{array}{l} m = \{1, 2, \ldots, M\} \\ n = \{1, 2, \ldots, N\} \end{array} \tag{14}$$

which can be formulated as

$$\gamma_i^{req} = \frac{p_i^{tx}}{\sum\limits_{\substack{k=1 \\ n' \neq n}}^{K} p_k^{tx} \cdot \frac{g_k(\theta_{m'})}{g_i(\theta_m)} + \frac{p_w}{g_i(\theta_m)}}, \quad \begin{array}{l} m = \{1, 2, \ldots, M\} \\ n = \{1, 2, \ldots, N\} \\ i = 1 + (n-1) + N(m-1) \end{array} \tag{15}$$

with $K = M \cdot N$ as the total number of users in all cells and p_w is the additive white Gaussian noise (AWGN) at the receiver, $\gamma_{m,n}^{req} \mapsto \gamma_i^{req}$, $g_{m',n'}(\theta_{m'}) \mapsto g_k(\theta_{m'})$, $g_{m,n}(\theta_m) \mapsto g_i(\theta_m)$, $p_{m',n'}^{tx} \mapsto p_k^{tx}$ are performed according to the index mapping rules in equation (15). To solve for the transmitter power p_k^{tx} of each of the K individual UE simultaneously, equation (14) can be reformulated into a matrix form as

$$\mathbf{p}_{tx} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}, \tag{16}$$

where the calculated vector $\mathbf{p}_{tx}$ contains the necessary transmitter power level assigned to each of the K UE to fulfil the SINR requirement and where matrix $[\mathbf{A}]_{K\times K}$ and vector $[\mathbf{b}]_{K\times 1}$ are defined as

$$\begin{aligned}
&[a_{ik}]_{K\times K} = \gamma_i^{req} \cdot \frac{g_k(\theta_{m'})}{g_i(\theta_m)} \quad \text{for} \quad n' \neq n \quad \text{and} \\
&[a_{ik}] = 0 \quad \text{for} \quad n' = n, \quad [b_i]_{K\times 1} = \gamma_i^{req} \cdot \frac{p_w}{g_i(\theta_m)}, \\
&m = \{1,2,\ldots,M\}, \quad n = \{1,2,\ldots,N\}, \quad i = 1 + (n-1) + N(m-1) \\
&m' = \{1,2,\ldots,M\}, \quad n' = \{1,2,\ldots,N\}, \quad k = 1 + (n'-1) + N(m'-1)
\end{aligned} \tag{17}$$

Using the $\mathbf{p}_{rx} = \mathbf{g} \odot \mathbf{p}_{tx}$, where $\odot$ denotes an elementwise multiplication and $\mathbf{g}$ is the total channel gain vector $[g_k]_{K\times 1}$ for all $k = \{1,2,\ldots,K\}$ users, then all elements in the vector $\mathbf{p}_{rx}$ for each block that contain the UE of each of the M cells are balanced. The total cell interference can then be calculated as

$$I_m^{own}(\theta_m) = \sum_{n=1}^{N} p_{m,n}^{rx}(\theta_m), \quad m = \{1,2,\ldots,M\} \tag{18}$$

$$I_m^{oth}(\theta_m) = \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{n=1}^{N} p_{m',n}^{rx}(\theta_{m'}) + p_w, \quad m = \{1,2,\ldots,M\} \tag{19}$$

where p_w is the thermal noise at the receiver, $I_m^{own}(\theta_m)$ is the interference from the UE within the own cell m and $I_m^{oth}(\theta_m)$ is the interference from the UE in the $M-1$ other cells where M is the total number of cells. We can now calculate $i_{UL}(\theta_m)$ which defines the other to own interference ratio for the uplink to HAP m and is given by

$$i_{UL}(\theta_m) = \frac{I_m^{oth}(\theta_m)}{I_m^{own}(\theta_m)}. \tag{20}$$

This is a performance measure of the simulated system capacity at a specific elevation angle θ_m towards the HAP (see figure 1). If $i_{UL}(\theta_m)$ is between zero and one there is possibility to have multiple HAP base stations covering the same coverage area. The actual number of users that can access the HAP base stations is also dependent of which data rate each user is using for transmission.

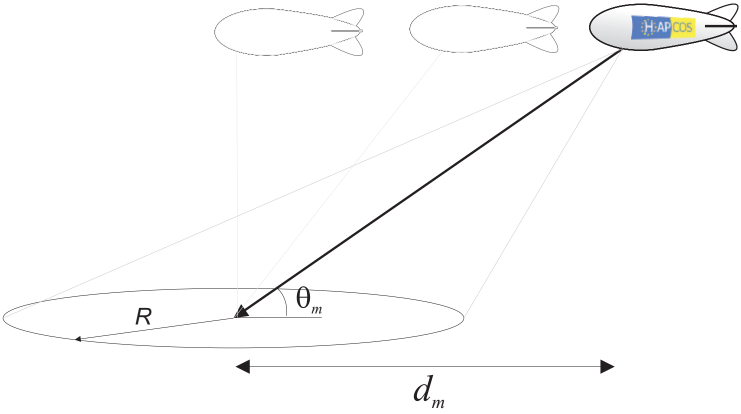


Fig. 4. A plot illustrating the change of HAP position d_m to create different elevation angles θ_m.

4. Simulation Results

In this simulation we assume M HAPs uniformly located along a circular boundary, with the centre of the circular boundary acting as the pointing direction of the HAPs base station antennas which simulate several overlapping cells, see figure 1. The beamwidth of these base station antennas are determined by the radius of the cell coverage area (see figures 1 and 3).

These results are acquired through running Monte Carlo simulations of the multiple HAP system. The aim of the simulation is to assess the effect of adding more HAPs on the system's capacity and of the impact of using space-time diversity techniques. The distance d_m between the cell centre and the vertical projection of the HAP on the earth's surface is denoted as "distance on the ground" and is varied from 0 to 70 km with a fixed cell position, as shown in figure 4. The distance to the cell centre is also changing the elevation angle θ_m towards the HAP base station m as seen from the user. The cell radius has been set to 10 km and 30 km, and the HAP altitude is 20 km. Each HAP base station serves 100 users within each corresponding cell.

From figure 5 it is clear that with the smaller cell radius (10 km) the worst case scenario will occur when all the HAPs are stacked on top of each other at 90 degrees elevation angle from the cell centre (i.e., at a distance d_m on the ground of 0 km). In the larger cell radius case (30 km) the worst case scenario happens approximately at 30 km which is at the edge of the cell.

Comparing the bottom diagram in figure 5 with the two diagrams in figure 6, we can see that if we utilize a maximum allowed other-to-own interference ratio equal to one, then as the service data rate decreases, the number of possible HAP base stations covering the same area can increase from 2-4 HAPs (depending on the distance d_m between the cell centre and the vertical projection of the HAP on the ground) for the combined service (12 kbps and 384 kbps) to 6 HAPs with the same service (12 kbps on all HAPs).

Next, we analyze the impact of different space-time diversity techniques (SIMO and MIMO) on the possible number of HAPs that can coexist within the same cell area and compare them to a single-input single-output (SISO) system. From figure 7 it is obvious that using a space-

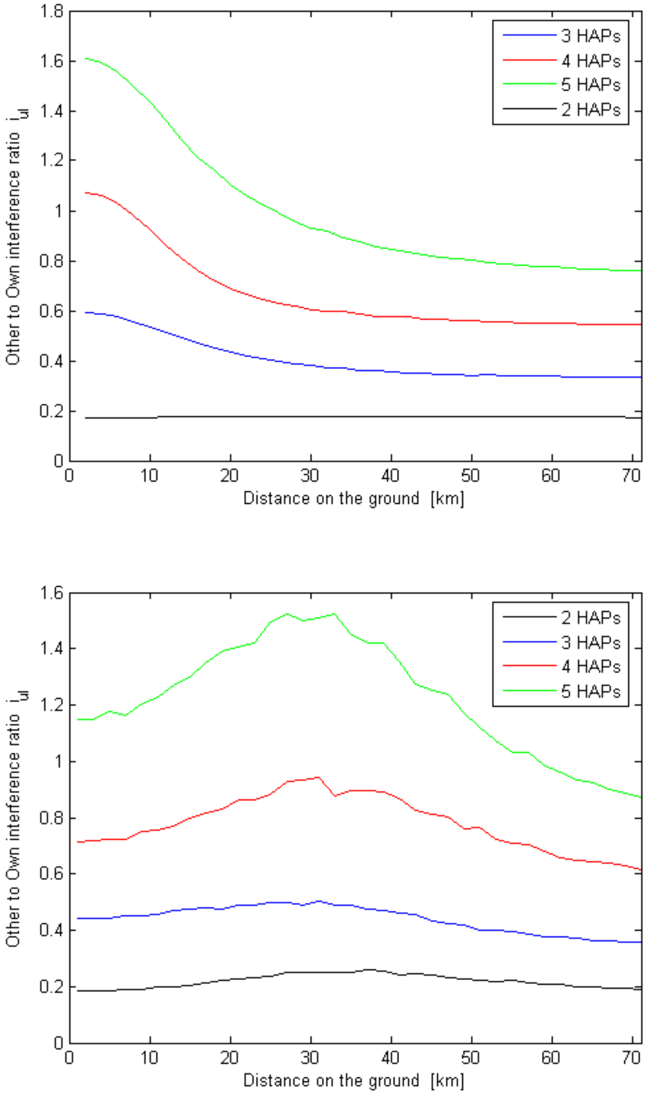


Fig. 5. The performance of the voice service (12 kbps) from one HAP in combination with the data service (384 kbps) on the remaining HAPs for cell radius of 10 km (top) and 30 km (bottom). The distance on the ground dm is varied from 0 to 70 km.

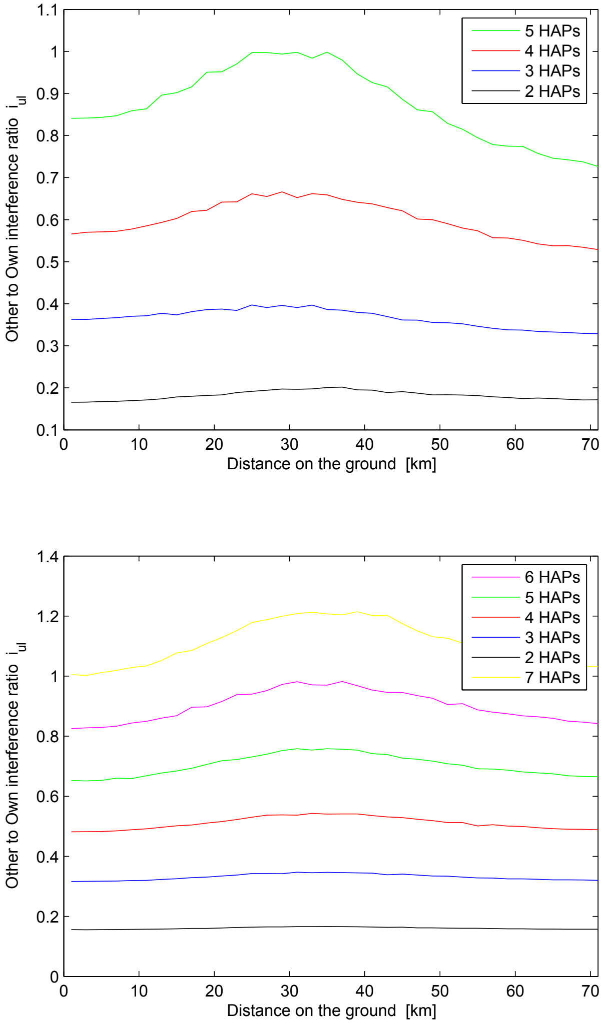


Fig. 6. The other to own interference ratio obtained for a 30 km cell radius for: (top) the performance of the voice service (12 kbps) from one HAP in combination with the data service (144 kbps) on the remaining HAPs and (bottom) the performance when we have voice services (12 kbps) on all HAPs. The distance on the ground d_m is varied from 0 to 70 km.

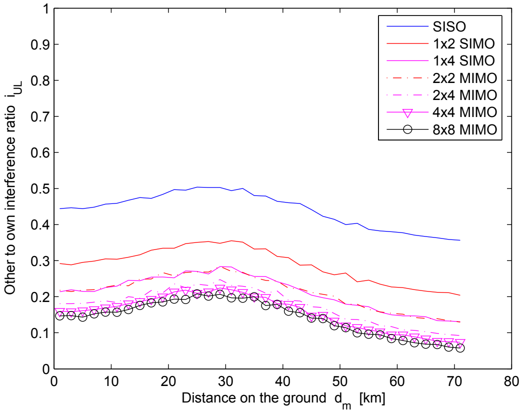


Fig. 7. The other to own interference ratio obtained for a 30 km cell radius for the performance of the voice service (12 kbps) from one HAP in combination with the data service (384 kbps) on the remaining two HAPs and utilizing different SISO, SIMO and MIMO space-time diversity systems. The distance on the ground d_m is varied from 0 to 70 km.

time diversity technique will enhance the interference mitigating capability and improve the overall performance of the multiple HAP system. This interference mitigation technique can also be interpreted as a capacity improvement, which is clearly seen in figure 7 for a three HAP system and in figure 8 for a seven HAP system. In both of these figures we can observe a decrease in the other-to-own interference ratio as we use an increasing number of antennas at the transmitter and receiver, which in turn will allow more HAPs to provide wireless service to more users by utilizing the remaining degrees of freedom of the system.

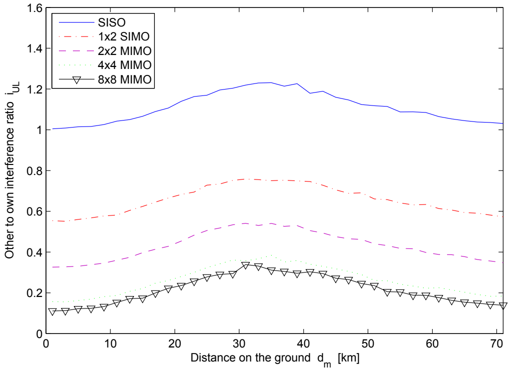


Fig. 8. The other to own interference ratio obtained for a 30 km cell radius for the performance of the voice service (12 kbps) from one HAP in combination with the data service (384 kbps) on the remaining six HAPs and utilizing different SISO, SIMO and MIMO space-time diversity systems. The distance on the ground d_m is varied from 0 to 70 km.

Comparing the graphs in figure 8, we can observe that a seven HAP system using SISO would not be possible due to the interference. However, a SIMO diversity system (utilizing two receiving antennas at the HAP base station) would make a seven HAP system possible. Adding more antennas at the receiver and transmitter respectively will increase the number of possible HAPs that can be used in the multiple HAP system. However, the benefit of the diversity system will diminish even with increasing the number of antennas beyond a certain limit. From figure 7 and figure 8 it is obvious that this limit is obtained at approximately a 4x4 MIMO system, beyond which diversity gain is negligible as is evident from the graph of the 8x8 MIMO system.

It is also clear from figure 6 that the worst case distance (highest interference level) is at approximately 30 km, and consequently a worst case elevation angle of 34 degrees. This maximum interference level depends on the cell radius chosen for the HAP base station as shown in figure 9. Simulation results show that for cell radii larger than 10 km the maximum interference level will occur at the cell boundary.

5. Conclusions

In this chapter we have investigated the possibility of multiple HAP coverage of a common cell area in WCDMA systems with and without space-time diversity techniques. Simulation results have shown that as the service data rate decreases, the number of possible HAP base stations that can be deployed to cover the same geographical area increases. It has further been shown that this increment in number of HAP base stations can be enhanced to some extent by using space-time diversity techniques. We have also shown that the worst case position of the HAPs is in the centre of the cell if the cell radius is small ($\leq$ 20 km) and at the cell boundary for large cells ($\geq$ 20 km). We can conclude that there is a possibility of deploying 3-5 (SISO), or 5-8 (1x2 SIMO, 2x2 MIMO and 4x4 MIMO) HAPs covering the same cell area in response to an increase in traffic demands, depending on the type of service used. There also appear to be

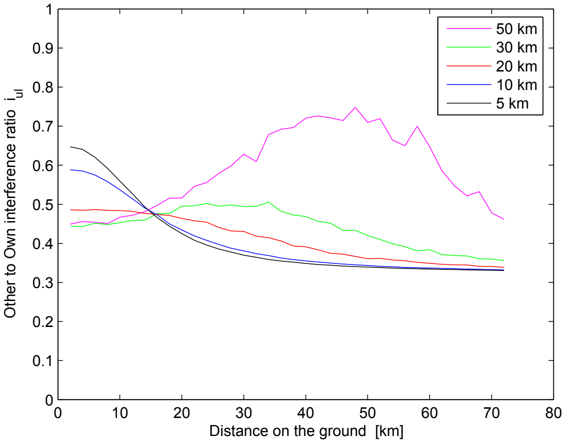


Fig. 9. Illustrating the effect of HAP base station cell radius on interference levels. A system of 3 HAPs is utilized here and a voice service (12 kbps) from one HAP in combination with the data service (384 kbps) on the other HAPs. The distance on the ground d_m is varied from 0 to 70 km.

a limit on the number of HAPs that could be deployed using space-time diversity techniques. Simulation results have shown that the maximum number of HAPs that could be sustained is approximately eight when using the voice services with 4x4 MIMO on all HAPs and users.

6. References

3GPP (2005). http://www.3gpp.org/specs/specs.htm. Base station (BS) radio transmission and reception. *3GPP* TS 25.104, 2005.

Balanis, C. (1997). *Antenna theory: analysis and design*, Chapter 6, John Wiley, 1997.

Chen, G.; Grace, D. & Tozer, T. (2005). Performance of multiple HAPs using directive HAP and user antennas. International Journal of Wireless Personal Communications - Special issue on High Altitude Platforms, Vol. 32, No. 3-4, February 2005, 275 -299.

Collela, N; Martin, J. & Akyildiz, I. (2000). The HALO network. *IEEE Communications Magazine*, Vol. 38, No. 6, June 2000, 142-148.

Djuknic, G.; Freidenfelds, J. & Okunev, Y. (1997). Establishing wireless communications services via high-altitude aeronautical platforms: a concept whose time has come?. *IEEE Communications Magazine*, Vol. 35, No. 9, September 1997, 128-135.

Dovis, F.; Fantini, R.; Mondin, M. & Savi, P. (2002). Small-scale fading for high-altitude platform (HAP) propagation channels. *IEEE Journal on Selected Areas in Communications*, Vol. 20, No. 3, April 2002, 641-647.

El-Jabu, B. & Steele, R. (2001). Cellular communications using aerial platforms. *IEEE Transactions on Vehicular Technology*, Vol. 50, May 2001, 686-700.

Falletti, E.; Mondin, M.; Dovis, F. & Grace, D. (2003). Integration of a HAP within a terrestrial UMTS network: interference analysis and cell dimensioning. *International Journal*

Wireless Personal Communications - Special Issue on Broadband Mobile Terrestrial-Satellite Integrated Systems, Vol. 24, No. 2, February 2003, 291-325.

Falletti, E. & Sellone, F. (2005). A multi-antenna channel simulatorfor transmit and receive smart antennas systems. *IEEE Transactions on Vehicular Technology*, June 2005.

Foo, Y.; Lim, W. & Tafazolli, R. (2000). Performance of high altitude platform station (HAPS) in delivery of IMT-2000 W-CDMA. *Stratospheric Platform Systems Workshop*, Tokyo, Japan, September 2000.

Grace, D; Daly, N.; Tozer, T.; Burr, A. & Pearce, D. (2001). Providing multimedia communications from high altitude platforms, *International Journal of Satellite communications*, Vol. 19, No. 6, November 2001, pp. 559-580.

Grace, D.; Spillard, C.; Thornton, J. & Tozer, T. (2002). Channel assignment strategies for a high altitude platform spot-beam architecture. *IEEE PIMRC 2002*, Lisbon, Portugal, September 2002.

Grace, D.; Mohorcic, M.; Capstick, M.; Pallavicini, B. & Fitch, M. (2005). Integrating users into the wider broadband network via high altitude platforms. *IEEE Wireless Communications*, Vol. 12, No. 5, October 2005, 98-105.

Grace, D.; Thornton, J.; Chen, G.; White, G. & Tozer, T. (2005). Improving the system capacity of broadband services using multiple high altitude platforms, *IEEE Transactions on Wireless Communications*, Vol. 4, No. 2, March 2005, 700-709.

Grace, D. & Likitthanasate, P. (2006). A business modelling approach for broadband services from high altitude platforms. *ICT'06*, Madeira, Portugal, May 2006.

Goldhirsch, J. & Vogel, W. (1992). Propagation effects for land and mobile satellite systems: overview of experimental and modelling results. *NASA Ref. Publication 1274*, February 1992.

ITUa (2000). Recommendation ITU-R M.1456. Minimum Performance characteristics and operational conditions for high altitude platform stations providing IMT-2000 in the bands 1885-1980 MHz, 2010-2025 MHz and 2110-2170 MHz in the Regions 1 and 3 and 1885-1980 MHz and 2110-2160 MHz in Region 2. *International Telecommunications Union*, 2000.

ITUb (2000). Recommendation ITU-R F.1500. Preferred characteristics of systems in the fixed service using high-altitude platform stations operating in the Bands 47.2-47.5 GHz and 47.9-48.2 GHz. *International Telecommunications Union*, 2000.

Liu, Y.; Grace, D. & Mitchell, P. (2005) Effective system spectral efficiency applied to a multiple high altitude platform system, *IEE Proceedings - Communications*, Vol. 152, No. 6, December 2005, 855-860.

Li, C. & Wang, X. (2004). Performance Comparisons of MIMO Techniques with Application to WCDMA Systems. *EURASIP Journal on Applied Signal Processing*, Vol. 2004, No. 5, 2004, 49-661.

Masumura, S. & Nakagawa, M. (2002). Joint system of terrestrial and high altitude platform stations (HAPS) cellular for W-CDMA mobile communications, *IEICE Transactions on Communications*, Vol.E85-B, No. 10, October 2002, 2051-2058.

Miura, R. & Oodo, M. (2001). Wireless communications system using stratospheric platforms. *Journal of the Communication Research Laboratory*, Vol. 48, No.4, 2001, 33-48.

Oodo, M.; Miura, R.; Hori, T.; Morisaki, T.; Kashiki, K. & Suzuki, M. (2002). Sharing and compatability study between fixed service using high altitude platform stations (HAPs) and other services in 31/28 GHz bands, *Wireless Personal Communications*, Vol. 23, 2002, 3-14.

Park, J.; Ku, B.; Kim, Y. & Ahn, D. (2002). Technology development for wireless communications system using stratospheric platform in Korea. *IEEE PIMRC 2002*, pp. 1577-1581, Lisbon, Portugal, Sept 2002.

Parks, M.; Butt, G.; Evans, B. & Richharia, R. (1993). Results of multiband (L, S, Ku Band) propagation measurements and model for high elevation angle land mobile satellite channel. *Proceedings of XVII NAPEX Conference*, pp. 193-202, Pasadena, California, USA, June 1993.

Steele, R. (1992). Guest Editorial: An update on personal communications, *IEEE Communications Magazine*, December 1992, 30-31.

Thornton, J.; Grace, D.; Spillard, C.; Konefal, T. & Tozer, T. (2001). Broadband Communications from a High Altitude Platform - The European HeliNet Programme. *IEE Electronics and Communications Engineering Journal*, Vol. 13, No.3, June 2001. 138-144.

Thornton, J.; Grace, D.; Capstick, M. & Tozer, T. (2003). Optimising an Array of antennas for cellular coverage from a high altitude platform, *IEEE Transactions on Wireless communications*, Vol. 2, No. 3, May 2003, 484-492.

Thornton, J. & Grace, D. (2005). Effect of lateral displacement of a high altitude platform on cellular interference and handover. *IEEE Transactions on Wireless Communications*, Vol. 4, No. 4, July 2005, 1483-1490.

Tozer, T. & Grace, D. (2001). High-altitude platforms for wireless communications, *IEE Electronics and Communications Engineering Journal*, June 2001, Vol. 13, No. 3, 127-137.

Vazquez-Castro, M.; Belay-Zelek, D. & Curieses-Guerrero, A. (2002). Availability of systems based on satellites with spatial diversity and HAPS, *Electronics Letters*, Vol. 38, No. 6, 286-287.

7

High-Rate, Reliable Communications with Hybrid Space-Time Codes

Joaquín Cortez[1] and Miguel Bazdresch[2]
[1]*Instituto Tecnológico de Sonora*
[2]*Instituto Tecnológico de Estudios Superiores de Occidente*
México

1. Introduction

Current wireless services and applications, such as third-generation (3G) cellular systems and Wi-Fi networks, offer capabilities far beyond what was previously available. With data rates on the order of 100kbit/s for mobile cellular users and up to 54Mbit/s on fixed WLANs, these systems provide attractive services such as internet access and video telephony.

In the near- and medium-term, however, it is expected that the capabilities of wireless networks will grow exponentially. The future of wireless applications and services will require high spectral efficiency, data rates on the order of 1Gbit/s, WLAN and WMAN integration and seamless connectivity, for devices ranging from a cell phone to a full-fledged desktop computer. Examples of services that will be available to users are Multimedia Messaging Service (MMS), HDTV-quality digital video, mobile TV, and Quality of Service guarantees.

The fulfillment of these promises hinges on several key telecommunications technologies. OFDM and related modulation techniques promise high spectral efficiency on wideband channels. For example, adaptive radio interfaces and cognitive radio will allow efficient spectrum use and smooth handoff between disparate networks. Software-defined radio and advanced circuit design techniques are needed to support all required functionality while meeting size, weight and power consumption requirements. All of these areas present heavy research activity.

Another key technology is known as multiple-input, multiple-output (MIMO) systems. These communications systems use multiple antennas at the transmitter and receiver, and have powerful capabilities in two respects: they can improve link reliability, and/or they can increase the data rate, without requiring extra power or bandwidth. Compared to more conventional systems, with only one antenna at the transmitter end (single-input multiple-output, SIMO), at the receiver end (multiple-input single-output, MISO), or at both ends (single-input single-output, SISO), MIMO systems offer additional (spatial) degrees of freedom (Tse & Viswanath, 2005), (Biglieri et al., 2007). While information-theoretic capacity analyses support the potential gains (and illustrate the limitations) offered by MIMO systems (Telatar, 1999), (Foschini & Gans, 1998), practical coding strategies that take advantage of them must be devised.

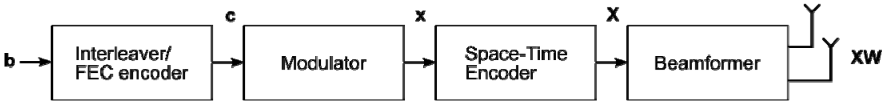


Fig. 1. MIMO system transmit chain.

Assuming a narrowband channel and adequate antenna separation, MIMO systems allow signal coding over time (that is, over multiple symbol periods) and over space (using all the available antennas). A space-time code is a mapping from modulated symbols to n_t spatial data streams, each of which is transmitted by a different antenna. This process is illustrated in Figure 1. A data stream **b** is interleaved and coded with a conventional FEC coder. The interleaved/coded stream **c** is modulated, and the resulting stream **x** is space-time coded. The space-time encoder takes R_sT symbols from **x** at a time, and (linearly) maps it to space-time code matrix **X**. Each column of **X** is transmitted during a symbol period. If the transmitter has any channel-state information (CSI), then a beamformer may be used to allocate power in an optimal way among the transmitter antennas. We will assume no transmitter CSI, so that beamformer matrix **W** is equal to the identity matrix. Matrix **X** has dimensions $n_t \times T$, so that it takes T symbol periods to transmit R_sT symbols and the code rate is R_s. The set of all possible code matrices is the space-time code, and the design problem consists in finding a set that meets given performance criteria.

Assuming that the channel presents quasi-static Rayleigh fading (the channel remains constant during T symbol periods), and assuming there are n_r receiver antennas, then the channel may be modeled as a matrix **H** of dimensions $n_r \times n_t$, where each element $h_{i,j}$ is a complex Gaussian random variable with 0 mean and variance 1, and represents the channel coefficient from transmitter antenna j to receiver antenna i. Assuming perfect CSI at the receiver, the received matrix **Y** may be written as

$$\mathrm{Y} = \mathrm{HX} + \mathrm{Z}, \tag{1}$$

where matrix **Z** corresponds to additive Gaussian white noise. Its entries are complex Gaussian random variables with 0 mean and variance N_0. If E_S is the signal energy transmitted for each antenna during each symbol period, then the signal-to-noise ratio (SNR) at the receiver is defined as

$$SNR = \frac{n_t E_S}{N_0}. \tag{2}$$

Under these conditions, we may identify three important code performance measures that characterize a given space-time code.

Multiplexing gain. Channel capacity C, or the achievable data rate assuming optimum coding and decoding, scales with $\min(n_T, n_R)$:

$$C(n_T, n_R, SNR) \sim \min(n_T, n_R)\log(SNR). \tag{3}$$

This means that increasing the number of antennas at either side allows an increase in data rate without increasing error probability. This may be interpreted as the separation of channel **H** into a number of parallel, independent channels that do not interfere with each other. Equation (3) suggests that this separation is possible at the receiver side, so that, for a fixed error probability, the data rate may be increased by up to $\min(n_T, n_R)$ bps/Hz for each 3dB increase in SNR. The increase in data rate afforded by a particular code is known as its multiplexing gain.

Diversity gain. Assuming code matrix **C** is transmitted, an upper bound on the probability that the receiver instead decides in favor of code matrix **D** has the form

$$K^{-n_R} J^{-rn_R}, \tag{4}$$

where $\boldsymbol{r}$, $\boldsymbol{K}$ and $\boldsymbol{J}$ are real numbers that depend on the particular properties of the space-time code. In particular, r and K depend on the code matrices, while $\boldsymbol{J}$ depends on $\boldsymbol{E_S}$ and $\boldsymbol{N_0}$. The number $\boldsymbol{rn_R}$ is known as the diversity gain of the code, since it is the asymptotic negative slope of the error probability, in a log-log scale, as a function of SNR. The maximum achievable diversity gain is $n_R n_T$. This means that for a 3dB increase in SNR, the error probability decreases by $2^{-n_R n_T}$.

Coding gain. In Equation (4), the number $\boldsymbol{K}$ is known as the coding gain. It is independent of SNR, so it has no effect on the slope of the error probability. However, it is important to make it as large as possible to improve code performance.

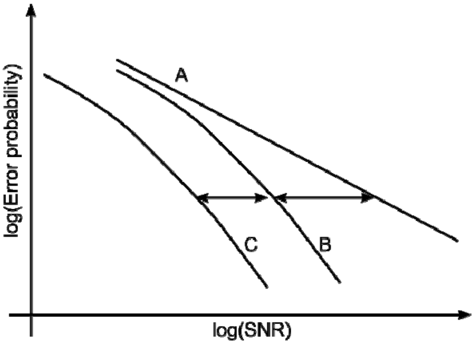


Fig. 2. Coding gain and diversity gain.

Coding and diversity gains are illustrated in Figure (2). Consider three space-time codes, A, B, and C, with error performance as shown in Figure (2). Code B has a diversity gain over Code A; this can be seen as an improvement on the slope of its error probability as SNR increases. On the other hand, Codes B and C provide the same diversity; however, code C has a coding gain advantage over Code B.

It is not possible to maximize all gains at the same time (Zheng & Tse, 2003). In particular, there is a trade-off that must be made between multiplexing and diversity gains. If a space-time code has multiplexing gain $\boldsymbol{R}$, then the maximum diversity gain it may achieve is given by

$$d(R) = (n_R - R)(n_T - R)\,. \tag{5}$$

There has been a large research effort to design space-time codes that achieve the gains promised by information theory. In the rest of this chapter, we summarize some of these efforts, focusing on so-called hybrid codes.

2. Design of space-time codes

2.1 Early developments

Two of the first space-time codes are V-BLAST (Golden et al., 1999) and the Alamouti Scheme (Alamouti, 1998). V-BLAST is a purely spatial multiplexing code; the space-time coder is simply a serial-to-parallel converter. It maximizes data rate while offering relatively poor link reliability. In contrast, the Alamouti Scheme is a pure diversity scheme; its data rate is the same as a SISO system but it offers maximum diversity gain. While V-BLAST's receiver is more complex, both are practical in the sense that receiver design and implementation is feasible.

Space-time trellis codes (STTCs) were proposed shortly afterwards; however, most research has focused on space-time block codes (STBCs), first introduced in (Tarokh et al., 1999). The idea is to extend the Alamouti scheme to systems with larger numbers of antennas (and to systems where $n_T \neq n_R$); maximum likelihood (ML) decoding is feasible because the code matrices have a particular, orthogonal structure. It was soon found, however, that this approach is incapable of producing high-rate, high-diversity codes see (Gesbert et al., 2003) and (Paulraj et al., 2004) for good overviews of the development of space-time codes).

Broadly speaking, it could be said that from this point, research into code design followed two different routes, one based on algebraic construction of optimal codes, and the other based on a more practical search for possibly sub-optimal, but still useful codes.

2.2 Algebraic design of space-time codes

Algebraic methods have been used to design codes since the very earliest stages of the development of information theory. It is small surprise, then, that these methods have been extended to the design of space-time codes. In particular, cyclic division algebras are very useful to derive families of space-time codes with full diversity and high rate. In particular, the design objective is to find codes that achieve the diversity-multiplexing tradeoff and have maximum data rate. The constellation shape (a lattice) is included in code design, and decoding is reformulated as lattice decoding (Oggier et al., 2007). One of the main results of this work has been the definition of Perfect space-time codes; these codes are the best possible square block codes, and satisfy these criteria: full diversity, maximum coding gain, optimum diversity-multiplexing trade-off, optimum constellation shape, and uniform average energy emitted per antenna. In a very important result (Berhuy & Oggier, 2009), it was shown that Perfect codes exist only for MIMO systems with six or less antennas.

2.3 A simplified approach

In contrast with the theoretical, rigorous development of algebraic codes, hybrid codes are developed in a more ad hoc manner. The objective is the design of codes offering a good tradeoff between multiplexing and diversity gains; however, the design consists of a straightforward mixture of pure diversity and pure multiplexing codes. A hybrid space-time encoder distributes the modulated data symbols among a spatial-multiplexing encoder and possibly several full-diversity encoders, as shown in Figure 3.

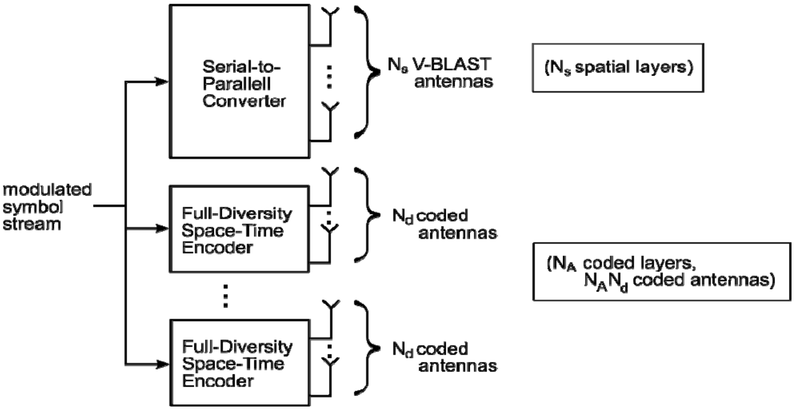


Fig. 3. A general hybrid space-time encoder.

In this general scheme, different space-time codes are simply stacked together and transmitted simultaneously. Conceptually, the transmitted data is divided into several layers; one spatial layer for each V-BLAST antenna, and N_A coded layers, one for each diversity encoder. Hybrid code design consists mainly in the selection of number of transmit and receive antennas, and in choosing appropriate encoders. In order for a hybrid code to be practical, a decoding algorithm must be devised that is able to take advantage of the hybrid code properties and that has low complexity. Hybrid codes are explored in more detail in the following section.

3. Hybrid Space-Time Codes

A large variety of hybrid space-time codes have been proposed in recent years. They differ mainly in the selection of coding layers and in decoding strategies. In this section, we identify and categorize some of the main results in hybrid code design.

3.1 Quasi-Orthogonal Space-Time Codes

Space-time codes designed from orthogonal matrices can be linearly decoded in a very simple way. However, as mentioned before, full-diversity, full-rate orthogonal codes don't exist for more than two transmit antennas. Quasi-orthogonal codes (Jafarkhani, 2001), (Tirkkonen et al., 2000) use almost-orthogonal matrices to achieve full rate, while providing only partial diversity. As an example, ABBA codes have code matrices of the form

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix}, \tag{6}$$

where matrix blocks A and B are **2x2** Alamouti matrices and each one transmits a different pair of symbols. A generalization of ABBA to any number of antennas is presented in (Dai et al., 2007).

In quasi-orthogonal codes, some transmitted symbols interfering each other (in contrast, in pure spatial multiplexing schemes such as V-BLAST, all symbols interfere with each other). The decoding algorithm consists in a combination of ML decoding where there is no interference, combined with interference suppression (possibly iterative) where needed. Joint (ML) detection of all symbols simultaneously is also possible, but has increased complexity.

3.2 Double Space-Time Transmit Diversity

This scheme, known as DSTTD, uses two coded layers, where each layer is simply an Alamouti encoder. It is a variant of ABBA codes, where only the first column of the ABBA matrix is used, resulting in an increased data rate at the cost of less diversity. A generalization to **2Nx2N** antennas, along with an efficient detector algorithm, was proposed in (Arar & Yongacoglu, 2006). In (Kwak et al., 2005) a theoretical analysis of DSTTD proved that, for a range of multiplexing gains, DSTTD offers a multiplexing-diversity tradeoff that is superior to both V-BLAST and quasi-orthogonal codes. Since DSTTD may be seen as a special case of quasi-orthogonal codes, the same decoding strategies are applicable.

3.3 Hybrid Codes with Multi-User Interference Suppression

A different decoding strategy is proposed in (Freitas et al., 2006). A number of different combinations of space-time encoders are proposed. In the receiver, each hybrid code layer is considered to be a different user in a multi-user system. Traditional multi-user access techniques are used to separate the symbols from each layer. In particular, the decoding process is divided in two steps, one for interference cancellation, and another for space-time decoding. In the interference cancellation step, a MIMO-MMSE spatial filter is used to eliminate the interference from undesired blocks. This filter is a traditional adaptive filter that is calculated minimizing a cost function. This technique is interesting since it leverages the large body of knowledge concerned with multi-user environments. As the number of layers increases, though, the interference is harder to eliminate; simulation results suggest that this scheme does not effectively extract diversity from the code and works best in high-SNR scenarios.

3.4 STBC-VBLAST

In (Mao & Motani, 2005) another decoding algorithm is proposed. In a sense, the V-BLAST iterative, successive interference cancellation (SIC) algorithm is extended to hybrid codes. It is shown that it is advantageous to decode the diversity layers first; each layer is decoded and then its interference is suppressed from the rest of the layers. This results in an increase in diversity, at the cost of some spectral efficiency. A similar detection technique is known as QR-Group Receiver (Zhao & Dubey, 2005). This scheme uses any number of stacked Alamouti encoders, and uses spatial filtering and successive interference cancellation to eliminate inter-block interference. It achieves a performance in between that of V-BLAST and Alamouti codes.

3.5 Linear Dispersion Codes

Linear dispersion codes (LDCs) (Hassibi & Hochwald, 2002) are an extremely flexible framework for designing space-time codes with arbitrary number of antennas. The time- and space-spreading of each transmitted symbol is specified in a dispersion matrix; this method of specifying the space-time code is so powerful that it describes a large collection of codes, from V-BLAST to perfect codes. The main disadvantage of this technique is that it is not always clear how to specify the dispersion matrices; in general, only unproven heuristics are provided.

An important property of LDCs is that they may be transformed into purely spatial codes, which means they may be decoded using a variety of algorithms, such as ordered SIC or ML decoding. It was recently shown (Longoria, 2007) that many hybrid space-time codes may be transformed into LDCs and subsequently into spatial codes. This result generalizes all the different decoding algorithms presented in this section. In the following section, two hybrid codes that build on this result are presented in detail.

4. Two Hybrid Space-Time Code Architectures

In this section, we present in some detail two hybrid space-time codes. Significant features of these codes are:

- Both mix spatial layers and diversity layers, allowing any number of each.
- The number of receiver antennas is also flexible, with just a minimum number specified.
- One of the codes uses Alamouti space-time encoders, while the other uses ABBA codes with either 3 or 4 transmitter antennas each.
- The receiver architecture is based on a code transformation to an LDC code, and subsequently to a purely spatial code.
- As will be seen, transformation of a hybrid code to a spatial code entails an increase in the size of the virtual channel matrix. We present a low-complexity decoder that avoids this problem.

Simulation results are presented for both architectures, and it is shown that their performance compares very favorably to the codes presented in section 3.

4.1 ZF-SQRD LDSTBC Scheme

This scheme is based on the hybrid architecture with an arrangement of Alamouti-STBC modules and V-BLAST layers. We show how to transform the hybrid system equation as a particular linear dispersion code. We denote by n_B and n_S the number of Alamouti blocks and number of V-BLAST antennas, respectively; at least $n_R = n_S + n_B$ receive antennas are required. We refer to this architecture as ZF-SQRD LDSTBC. We will show how to arrive at a symbol by symbol (OSIC) decoder. This code meets the first constraint of the design method established in (Hassibi & Hochwald, 2003).

We exploit the matrices' structure to obtain a low-complexity receiver algorithm based on the sorted QR decomposition (Wubben et al., 2001). BER performance is further increased by allocating the same energy to each transmitted symbol, in contrast to other recent proposals where equal power is allocated to each antenna.

A simplified block diagram of the ZF-SQRD LDSTBC system, based on the scheme proposed in (Mao & Motani, 2005), is depicted in Figure 4. A single data stream is demultiplexed into n_L spatial layers, and each of them is mapped to the constellation chosen. The modulated stream feeds two different kinds of transmitters: n_S V-BLAST layers and n_B STBC encoders with $n_A = 2$ antennas each one; therefore, $n_T = n_S + n_A n_B$.

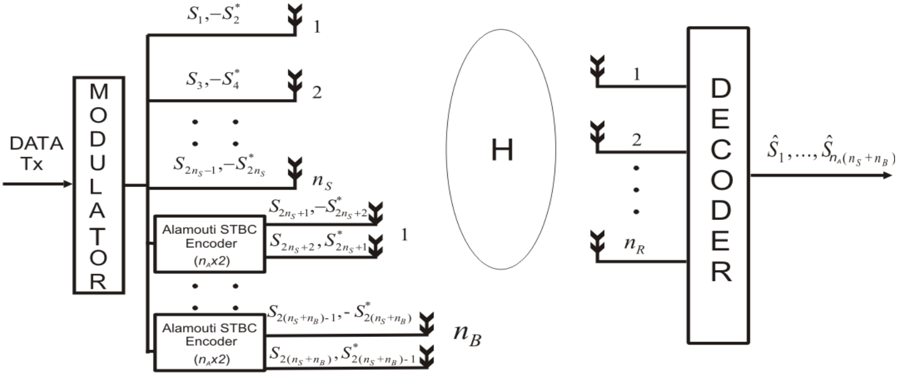


Fig. 4. ZF-SQRD LDSTBC Transmitter/Receiver Architecture

It is assumed that $n_R \geq n_S + n_B$. In the transmission of one block, the symbol sequence $s_1, s_2, \ldots, s_{nsym}$, where $nsym = n_A(n_S + n_B)$ is transmitted. The mapping of symbols to antennas is shown in Table I. The fraction of power allocated to the VBLAST layers is given by:

$$P_S = \frac{2}{n_A(n_S + n_B)}, \tag{7}$$

and the fraction allocated to the Alamouti encoder is given by

$$P_A = \frac{P_S}{n_A n_B}. \tag{8}$$

	Spatial Antennas	STBC Blocks $B=1,2,\ldots,n_B$	
Time	Antenna $V=1,2,\ldots,n_S$	Antenna 1	Antenna 2
t	s_V	s_{kn_A+1}	s_{kn_A+2}
$T+T$	$-s_{V+1}^*$	$-s_{kn_A+2}^*$	$s_{kn_A+1}^*$

Table 1. ZF-SQRD-LDSTBC Symbol to Antenna Mapping with $k = B - 1 + n_S$

This allocation results in all symbols being transmitted with the same power. Under the assumption of perfect CSI at the receiver, the detection and decoding process of the transmitted signal vector S, at the m^{th} time block slice, where $m = 1, 2$ and $n_A = 2$, the receive signal can be written as:

$$\begin{bmatrix} y_1^{(1)} & y_1^{(2)} \\ \vdots & \vdots \\ y_{n_R}^{(1)} & y_{n_R}^{(2)} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R,1} & h_{n_R,2} & & h_{n_R,n_T} \end{bmatrix} \begin{bmatrix} S_{spa} \\ S_A \end{bmatrix} + \begin{bmatrix} n_1^{(1)} & n_1^{(2)} \\ \vdots & \vdots \\ n_{n_R}^{(1)} & n_{n_R}^{(2)} \end{bmatrix}, \tag{9}$$

or equivalently,

$$Y = HS + N. \tag{10}$$

In equation (9) and the following, the sub indices indicate the receiver antenna, and the super indices indicate the block emission time. Matrix $Y \in C^{n_R \times 2}$ represents the symbols received in a block. Matrix H is the channel matrix defined above. Matrix $N \in C^{n_R \times 2}$ represents the noise added to each received symbol. Matrix S is composed for two blocks: The matrix S_{spa} corresponds to the symbols transmitted by VBLAST layers and the matrix S_A to the symbols transmitted by STBC encoders. The spatial multiplexing block S_{spa} is defined as:

$$S_{spa} = \begin{bmatrix} s_1 & -s_2^* \\ s_3 & -s_4^* \\ \vdots & \vdots \\ s_{2n_S-1} & -s_{2n_S}^* \end{bmatrix} = \begin{bmatrix} s_1^{(1)} & s_1^{(2)} \\ s_2^{(1)} & s_2^{(2)} \\ \vdots & \vdots \\ s_{n_S}^{(1)} & s_{n_S}^{(2)} \end{bmatrix}, \tag{11}$$

and the STBC block is defined as:

$$S_A = \begin{bmatrix} S_1^A \\ \vdots \\ S_{n_B}^A \end{bmatrix} = \begin{bmatrix} S_1^{(1)A} & S_1^{(2)A} \\ \vdots & \vdots \\ S_{n_B}^{(1)A} & S_{n_B}^{(2)A} \end{bmatrix}, \tag{12}$$

where each element of equation (4) is given by:

$$\begin{bmatrix} S_B^{(1)A} & S_B^{(2)A} \end{bmatrix} = \begin{bmatrix} s_{kn_A+1} & -s_{kn_A+2}^* \\ s_{kn_A+2} & s_{kn_A+1}^* \end{bmatrix}, \tag{13}$$

with $B = 1,2,\ldots,n_B$ and $k = B-1+n_S$. In equations (11) and (12), the matrix on the left is a direct mapping from Table 1; the notation of the matrix on the right, where antenna number and symbol period are made explicit, is adopted to simplify the explanation of the receiver algorithm. Reformulating the system equation (9) as a linear dispersion code (Longoria et al., 2007), we have:

$$\left[y_1^{(1)}\ y_1^{(2)*}\cdots y_{n_R}^{(1)}\ y_{n_R}^{(2)*}\right]^T = \left[H_{spa}\ H_A\right]S_{LD} + \left[n_1^{(1)}\ n_1^{(2)*}\cdots n_{n_R}^{(1)}\ n_{n_R}^{(2)*}\right]^T. \tag{14}$$

Equation (14) can be expressed in compact form as:

$$Y_{LD} = H_{LD}S_{LD} + N_{LD}, \tag{15}$$

where H_{LD} is composed of two blocks is named Linear Dispersion Matrix, one corresponding to the V-BLAST layers and another to the STBC layers. The V-BLAST block H_{spa} is given by:

$$H_{spa} = \begin{bmatrix} H_{1,1}^{spa} & H_{1,2}^{spa} & \cdots & H_{1,n_S}^{spa} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n_R,1}^{spa} & H_{n_R,2}^{spa} & & H_{n_R,n_S}^{spa} \end{bmatrix}, \tag{16}$$

where

$$H_{i,j}^{spa} = \begin{bmatrix} h_{i,j} & 0 \\ 0 & -h_{i,j}^* \end{bmatrix}, \tag{17}$$

for $i = 1,2,\cdots,n_R$ and $j = 1,2,\cdots,n_S$. The STBC block H_A is itself a block matrix; it is given by:

$$H_A = \begin{bmatrix} H_{1,1}^A & H_{1,2}^A & \cdots & H_{1,n_S}^A \\ \vdots & \vdots & \ddots & \vdots \\ H_{n_R,1}^A & H_{n_R,2}^A & & H_{n_R,n_S}^A \end{bmatrix}, \tag{18}$$

where each element of equation (18) is given by:

$$H_{k,B}^A = \begin{bmatrix} h_{k,n_A B+n_S-1} & h_{k,n_A B+n_S} \\ h_{k,n_A B+n_S}^* & h_{k,n_A B+n_S-1}^* \end{bmatrix}, \tag{19}$$

for $k = 1,2,\cdots,n_R$ and $B = 1,2,\cdots,n_B$. The matrix $H_{i,j}^{spa}$ links the j^{th} spatial antenna with the i^{th} receiver antenna. Likewise, $H_{k,B}^{A}$ links the B^{th} STBC block to the k^{th} receiver antenna. To complete the reformulation of system equation (9), it remains to rearrange matrix S. We define S_{LD} as:

$$S_{LD} = \begin{bmatrix} S_{LD}^{spa} \\ S_{LD}^{A} \end{bmatrix} \tag{20}$$

where

$$S_{LD}^{spa} = \left[s_1^{(1)} \; s_1^{(2)} \cdots s_{n_S}^{(1)} \; s_{n_S}^{(2)} \right]^T \tag{21}$$

and

$$S_{LD}^{A} = \left[S_1^{(1)A} \; S_2^{(1)A} \cdots S_{n_B}^{(1)A} \right]^T . \tag{22}$$

The reformulation of equation (9) as equation (14) transforms the hybrid code to a simpler, equivalent purely spatial system with $N_T = n_A(n_S + n_B)$ transmit antennas and without distinction between the STBC and VBLAST layers. This simpler system is shown in Figure 5.

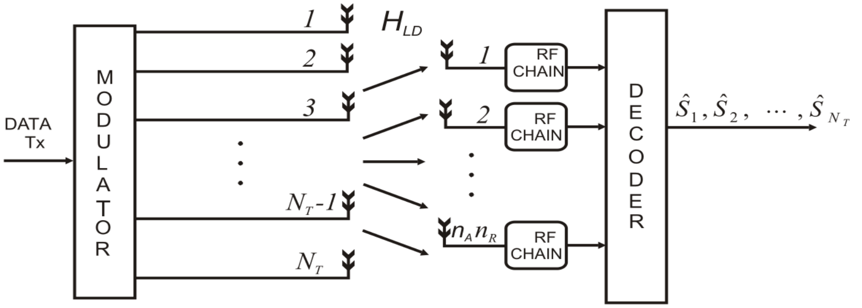


Fig. 5. ZF-SQRD LDSTBC Architecture Transmitter/Receiver as Linear Dispersion Code

4.2 ZF-SQRD LQOSTBC Scheme

The hybrid code called ZF-QR-SIC-LQOSTBC (Cortez et al., 2008) allows the use of any number of V-BLAST antennas, and any number of ABBA encoders of 3 or 4 antennas each. A block diagram of the ZF-SQRD LQOSTBC is shown in Figure 6.

Of the n_T transmit antennas, n_S are spatially multiplexed. There are n_B ABBA encoders, of $n_A \in \{3,4\}$ antennas each, so that $n_T = n_S + n_B \cdot n_A$. It is assumed that $n_R \geq n_S + n_B$. We will assume $n_A = 4$; it is straightforward to extend the results to the case $n_A = 3$. In the transmission of one block, the symbol sequence $\{s_i\}_{i=1}^{4(n_S+n_B)}$ is transmitted. The mapping of symbols to antennas is shown in Table 2. Using this code, $n_S + n_B$ symbols are transmitted per channel use, for a code rate equal to

$$\frac{n_S + n_B}{n_T}. \tag{23}$$

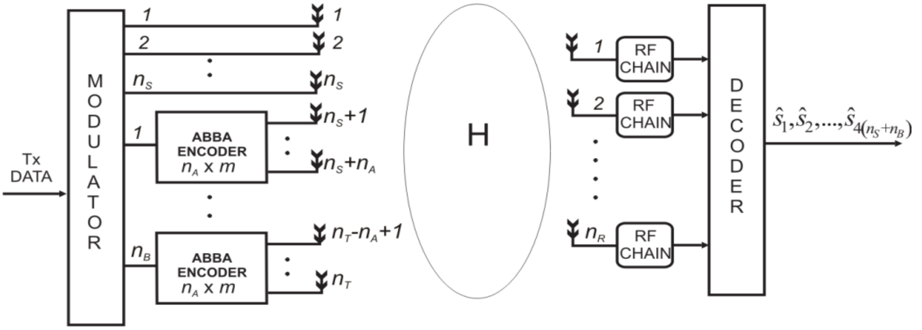


Fig. 6. ZF-SQRD LQOSTBC Architecture Transmitter/Receiver

ANTENNA	Symbol Period			
	1	2	3	4
(VBLAST)				
1	s_1	$-s_2^*$	s_3	$-s_4^*$
2	s_5	$-s_6^*$	s_7	$-s_8^*$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_S	s_{k-3}	$-s_{k-2}^*$	s_{k-1}	$-s_k^*$
(ABBA)				
n_S+1	s_{k+1}	$-s_{k+2}^*$	s_{k+3}	$-s_{k+4}^*$
n_S+2	s_{k+2}	s_{k+1}^*	s_{k+4}	s_{k+3}^*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_T	s_{k+4n_B}	$s_{k+4n_B-1}^*$	s_{k+4n_B-2}	$s_{k+4n_B-3}^*$

Table 2. ZF-SQRD LQOSTBC Symbol to Antenna Mapping with $k = 4n_S$

Since the transmitter has no knowledge of the channel, all symbols must be transmitted with equal energy. In the ABBA layers, each symbol's transmission is spread across multiple time intervals; in consequence, the signal constellations must be scaled accordingly. If E_v is the average energy of the signal constellation employed by each antenna in the VBLAST layers, then the average constellation energy E_a of the ABBA layers is given by $E_a=E_v/n_a$. It should be noted that the coding schemes referenced above use a single constellation, resulting in unequal symbol energy and suboptimal BER performance.

The system equation for ZF-SQRD LQOSTBC over four symbol periods may be written as follows, where subindices indicate antenna number, and superindices indicate symbol period within a block:

$$\begin{bmatrix} y_1^{(1)} \cdots y_1^{(4)} \\ \vdots \ \ddots \ \vdots \\ y_{n_R}^{(1)} \cdots y_{n_R}^{(4)} \end{bmatrix} = \begin{bmatrix} h_{1,1} \ h_{1,2} \ \cdots \ h_{1,n_T} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ h_{n_R,1} \ h_{n_R,2} \ h_{n_R,n_T} \end{bmatrix} \begin{bmatrix} S_{abba} \\ S_A \end{bmatrix} + \begin{bmatrix} n_1^{(1)} \cdots n_1^{(4)} \\ \vdots \ \ddots \ \vdots \\ n_{n_R}^{(1)} \cdots n_{n_R}^{(4)} \end{bmatrix}, \tag{24}$$

or equivalently,

$$Y = HS + N \,. \tag{25}$$

Matrix $Y \in C^{n_R \times 4}$ represents the symbols received in a block. Matrix H is the channel matrix defined above. Matrix $N \in C^{n_R \times 4}$ represents the noise added to each received symbol. Matrix S is composed of a spatial multiplexing block and n_B ABBA blocks. The spatial multiplexing block S_{spa} is defined as:

$$S_{spa} = \begin{bmatrix} s_1 & -s_2^* & s_3 & -s_4^* \\ \vdots & \vdots & \vdots & \vdots \\ s_{4n_S-3} & -s_{4n_S-2}^* & s_{4n_S-1} & -s_{4n_S}^* \end{bmatrix} = \begin{bmatrix} s_1^{(1)} \cdots s_1^{(4)} \\ \vdots \quad\ \ \vdots \\ s_{n_S}^{(1)} \cdots s_{n_S}^{(4)} \end{bmatrix}, \tag{26}$$

which corresponds to the VBLAST layer mapping in Table 2. The ABBA block is defined as:

$$S_{abba} = \left[S_1^{abba} \ S_2^{abba} \ \cdots S_{n_B}^{abba} \right], \tag{27}$$

where every element of equation (27) is given by:

$$S_B^{abba} = \begin{bmatrix} s_{k+1} & -s_{k+2}^* & s_{k+3} & -s_{k+4}^* \\ s_{k+2} & s_{k+1}^* & s_{k+4} & s_{k+3}^* \\ s_{k+3} & -s_{k+4}^* & s_{k+1} & -s_{k+2}^* \\ s_{k+4} & s_{k+3}^* & s_{k+2} & s_{k+1}^* \end{bmatrix} = \begin{bmatrix} s_{l-3}^{(1)} \cdots s_{l-3}^{(4)} \\ s_{l-2}^{(1)} \cdots s_{l-2}^{(4)} \\ s_{l-1}^{(1)} \cdots s_{l-1}^{(4)} \\ s_l^{(1)} \cdots s_l^{(4)} \end{bmatrix}, \tag{28}$$

with $l = n_S + 4B$, $k = 4(B - 1 + n_S)$ and $B = 1, 2, \ldots, n_B$. Rewriting the system equation (24) as a linear dispersion code, we have:

$$\begin{bmatrix} y_1^{(1)} \\ y_1^{(2)*} \\ y_1^{(3)} \\ y_1^{(4)*} \\ \vdots \\ y_{n_R}^{(1)} \\ y_{n_R}^{(2)*} \\ y_{n_R}^{(3)} \\ y_{n_R}^{(4)*} \end{bmatrix} = \begin{bmatrix} H_{spa} \; H_{abba} \end{bmatrix} S_{LD} + \begin{bmatrix} n_1^{(1)} \\ n_1^{(2)*} \\ n_1^{(3)} \\ n_1^{(4)*} \\ \vdots \\ n_{n_R}^{(1)} \\ n_{n_R}^{(2)*} \\ n_{n_R}^{(3)} \\ n_{n_R}^{(4)*} \end{bmatrix}, \tag{29}$$

expressed in compact form as:

$$Y_{LD} = H_{LD} S_{LD} + N_{LD}, \tag{30}$$

where H_{LD} is a linear dispersion matrix with two blocks, one corresponding to the V-BLAST layers and another to the ABBA layers. The V-BLAST block H_{spa} is given by:

$$H_{spa} = \begin{bmatrix} H_{1,1}^{spa} & H_{1,2}^{spa} & \cdots & H_{1,n_S}^{spa} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n_R,1}^{spa} & H_{n_R,2}^{spa} & & H_{n_R,n_S}^{spa} \end{bmatrix}, \tag{31}$$

where

$$H_{i,j}^{spa} = \begin{bmatrix} h_{i,j} & 0 & 0 & 0 \\ 0 & -h_{i,j}^* & 0 & 0 \\ 0 & 0 & h_{i,j} & 0 \\ 0 & 0 & 0 & -h_{i,j}^* \end{bmatrix}, \tag{32}$$

for $i = 1,2,\cdots,n_R$ and $j = 1,2,\cdots,n_S$. The ABBA block H_{abba} is itself a block matrix; it is given by:

$$H_{abba} = \begin{bmatrix} H_{1,1}^{abba} & H_{1,2}^{abba} & \cdots & H_{1,n_S}^{abba} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n_R,1}^{abba} & H_{n_R,2}^{abba} & & H_{n_R,n_S}^{abba} \end{bmatrix}, \quad (33)$$

where every element of equation (33) is given by:

$$H_{i,k}^{abba} = \begin{bmatrix} h_{i,l-3} & h_{i,l-2} & h_{i,l-1} & h_{i,l} \\ h_{i,l-2}^{*} & -h_{i,l-3}^{*} & h_{i,l}^{*} & -h_{i,l-1}^{*} \\ h_{i,l-1} & h_{i,l} & h_{i,l-3} & h_{i,l-2} \\ h_{i,l}^{*} & -h_{i,l-1}^{*} & h_{i,l-2}^{*} & -h_{i,l-3}^{*} \end{bmatrix}, \quad (34)$$

for $i = 1,2,\cdots,n_R$, $k = 1,2,\cdots,n_B$ and $l = n_S + 4B$. The matrix $H_{i,j}^{spa}$ of H_{LD} that links the j^{th} spatial antenna with the i^{th} receiver antenna. Likewise, $H_{i,k}^{abba}$ links the k^{th} ABBA block to the i^{th} receiver antenna. To complete the reformulation of system equation (24), it remains to rearrange matrix S. We define S_{LD} as:

$$S_{LD} = \left[S_{LD}^{spa} \quad S_{LD}^{abba} \right]^T \quad (35)$$

where

$$S_{LD}^{spa} = \left[s_1^{(1)} \; s_1^{(2)} \; s_1^{(3)} \; s_1^{(4)} \cdots s_{n_S}^{(1)} \; s_{n_S}^{(2)} \; s_{n_S}^{(3)} \; s_{n_S}^{(4)} \right]^T \quad (36)$$

and the ABBA block for $n_A = 4$ is, then, given by:

$$S_{LD}^{abba} = \left[s_{n_S+1}^{(1)} \; s_{n_S+2}^{(1)} \; s_{n_S+3}^{(1)} \; s_{n_S+4}^{(1)} \cdots s_{l-3}^{(1)} \; s_{l-2}^{(1)} \; s_{l-1}^{(1)} \; s_{l}^{(1)} \right]^T. \quad (37)$$

We have rewritten the hybrid space-time matrices as linear dispersion code matrices. Now we can substitute the original V-BLAST plus ABBA hybrid transceiver with a simpler, purely spatial system with $N_T = 4n_S + n_A n_B$ transmit antennas like is depicted in Figure 7 and without distinction between the ABBA and VBLAST layers.

5. Receiver Architectures for Hybrid Space-Time Codes

Since schemes ZF-SQRD LSTBC and ZF-SQRD LQOSTBC are equivalent to a purely spatial system with N_T transmit antennas, it is possible to propose a linear detector based on the sorted QR decomposition and OSIC, that takes advantage of the structure of the linear dispersion matrices to achieve low complexity and high performance.

5.1 OSIC Detection for Hybrid Schemes Proposed

We first calculate *HC Sorted QR* (Hybrid Coding Sorted QR) of the matrix H_{LD}=$Q_{LD}R_{LD}$ where Q_{LD} is a unitary matrix and R_{LD} is an upper triangular matrix. By multiplying the received signal equations (10) and (35) by Q_{LD}^H, the modified received vector is:

$$\tilde{Y}_{LD} = Q_{LD}^H Y_{LD} = R_{LD} S_{LD} + \tilde{N}_{LD}, \tag{38}$$

if vector S_{LD} is transmitted. Note that the statistical properties of the noise term $\tilde{N}_{LD}$ remain unchanged. Due to the upper triangular structure of R_{LD}, the k^{th} element of $\tilde{Y}_{LD}$ is:

$$\tilde{y}_k = r_{k,k} s_k + \sum_{i=k+1}^{N_T} r_{k,i} s_i + \tilde{n}_k . \tag{39}$$

Symbols are estimated in sequence, from lower stream to higher stream, using OSIC; assuming that all previous decisions are correct; the interference can be perfectly cancelled in each step except for the additive noise. The estimated symbol s_k is given by:

$$s_k = \mathrm{D}\left[\frac{y_k - \sum_{i=k+1}^{N_T} r_{k,i} \hat{s}_i}{r_{k,k}} \right], \tag{40}$$

where $\hat{s}_k$ is the estimate of s_k and **D[.]** is a decision device that maps its argument to the closest constellation point. Therefore the receiver requires calculating the QR decomposition for the linear dispersion matrix H_{LD}; the main challenge lies in finding the most efficient way to obtain this decomposition.

We use the permutation vector order provided by *HC Sorted QR* algorithm to reorder the received symbols; the QR decomposition is obtained using the modified Gram-Schmidt (MGS) algorithm.

5.2 HC Sorted QR Decomposition

Matrix H_{LD} is $N_T \times n_A n_R$ for both hybrid schemes. A direct application of MGS on it would result in unacceptable complexity. However, taking advantage of the structure imposed on H_{LD} by the proposed code, we can decrease this complexity significantly. We now explain how this simplification is obtained.

From the equations (15) and (30) we can see that the structure presented for the H_{LD} matrix allows us to reduce the computational complexity that is required for to calculate the *HC*

Sorted QR decomposition, since many of the elements of each matrix are equal, and their locations in each matrix are fixed and can be calculated in advance. This method involves obtaining the QR decomposition of the H_{LD} matrix in two stages: first we obtain the QR decomposition corresponding to the spatial layers of the hybrid system; in the second stage we calculate the QR decomposition for the diversity layers.
As a first step, we calculate the QR decomposition $H = Q_m R_m$ using the Sorted QR algorithm; in this process, we also produce vector *order* which specifies the detection order of the spatial layers. Then, using Q_m and R_m, and non normalized columns of the matrix H, we build the matrices H_{div}^{ala} or H_{div}^{abba} . The next step is analogous to MGS: column $k + 1$ is normalized and used to fill column $k+2$ of each block (Alamouti/ABBA); the process is repeated for the remaining columns for each block of H_{div}^{ala} or H_{div}^{abba} . In the process, matrix R_{LD} is also calculated. A block diagram of the process is shown in the Figure 7.

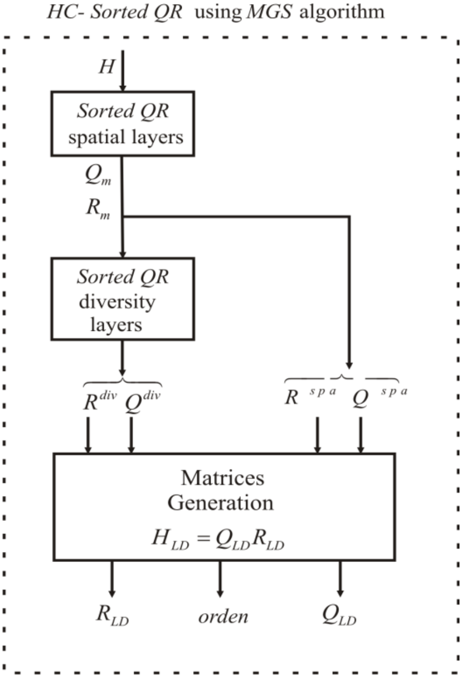


Fig. 7. *HC Sorted QR* Process Using MGS Algorithm

The structure of matrices Q_{LD} and R_{LD}, and their relation to Q_m and R_m, has been detailed in (Cortez et al., 2007), (Kim et al., 2006), (Le et al., 2005). The complete process is presented in two stages. In the first stage the algorithm 1 takes the channel matrix H and outputs the intermediate matrices Q_m, R_m and vector *order*.

Algorithm 1. *HC Sorted QR* of the Spatial Layers

```
INPUT: $H^{n_R \times n_T}$ , $L$, $nsym$, $n_S$
OUPUT: $Q_m$, $R_m$, $order$
$Q_m = H$ , $R_m^{n_T \times n_T} = 0$, $orden = [1:1:nsym]$
for $i = 1$ to $n_S$ do
    $k = \arg\min_{j=i:n_S} \|Q_m(:,j)\|$
    Exchange columns i and k of $Q_m$ and $R_m$
    Exchange columns $Li:-1:i(L-1)+1$ and $Lk:-1:k(L-1)+1$ of $order$
    $R_m(i,i) = \|Q_m(:,i)\|_2$
    $Q_m(:,i) = Q_m(:,i)/R_m(i,i)$
    for $j = i+1$ to $n_T$ do
        $R_m(i,j) = Q_m(:,i)^H Q_m(:,j)$
        $Q_m(:,j) = Q_m(:,j) - Q_m(:,i)R_m(i,j)$
    endfor
endfor
```

The structure for the matrices Q_m and R_m are:

$$Q_m = \begin{bmatrix} q_{1,1} \cdots & q_{1,n_S} & \underline{h}_{1,n_S+1} & \cdots \underline{h}_{1,n_T} \\ q_{2,1} \cdots & q_{2,n_S} & \underline{h}_{2,n_S+1} & \cdots \underline{h}_{2,n_T} \\ \vdots & \vdots & \vdots & \ddots \ \vdots \\ q_{n_R,1} \cdots & q_{n_R,n_S} & \underline{h}_{n_R,n_S+1} & \cdots \underline{h}_{n_R,n_T} \end{bmatrix}, \tag{41}$$

$$R_m = \begin{bmatrix} r_{1,1} \cdots & r_{1,n_S} & r_{1,n_S+1} & \cdots r_{1,n_T} \\ 0 \ \cdots & r_{2,n_S} & r_{2,n_S+1} & \cdots r_{2,n_T} \\ \vdots & \vdots \ \ddots & \vdots & \ddots \ \vdots \\ 0 & 0 \cdots r_{n_s,n_S} & \cdots & r_{n_S,n_T} \end{bmatrix}. \tag{42}$$

We choose the first n_S columns of Q_m and the first n_S rows of R_m to built the matrices Q^{spa} and R^{spa} with the next structure:

$$Q^{spa} = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,n_S} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,n_S} \\ \vdots & \vdots & & \vdots \\ q_{n_R,1} & q_{n_R,2} & \cdots & q_{n_R,n_S} \end{bmatrix}, \tag{43}$$

$$R^{spa} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n_S} & r_{1,n_S+1} & \cdots & r_{1,n_T} \\ 0 & \cdots & r_{2,n_S} & r_{2,n_S+1} & \cdots & r_{2,n_T} \\ \vdots & & \vdots \ddots & \vdots & \ddots & \vdots \\ 0 & & 0 \cdots r_{n_s,n_S} & & \cdots & r_{n_S,n_T} \end{bmatrix}. \tag{44}$$

The matrices Q^{spa} and R^{spa} represent the contribution of the spatial layers in both hybrid schemes. The columns with elements $\underline{h}_{i,j}$ in equation (41) are non normalized columns that we used to build the matrices H_{div}^{ala} and H_{div}^{abba} that are required in the second stage of the QR decomposition for the diversity layers. In the case of ZF-SQRD LDSTBC the matrix H_{div}^{ala} has the following structure:

$$H_{div}^{ala} = \begin{bmatrix} \underline{h}_{1,n_S+1} & \underline{h}_{1,n_S+2} & \cdots & \underline{h}_{1,n_T-1} & \underline{h}_{1,n_T} \\ \underline{h}^*_{1,n_S+2} & -\underline{h}^*_{1,n_S+1} & \cdots & \underline{h}^*_{1,n_T} & -\underline{h}^*_{1,n_T-1} \\ \vdots & \vdots & & \vdots & \vdots \\ \underline{h}^*_{n_R,n_S+2} & -\underline{h}^*_{n_R,n_S+1} & \cdots & \underline{h}^*_{n_R,n_T} & -\underline{h}^*_{n_R,n_T-1} \end{bmatrix}. \tag{45}$$

For the case of ZF-SQRD LQOSTBC the matrix H_{div}^{abba} with $n_A = 4$ has the structure:

$$H_{div}^{abba} = \begin{bmatrix} \underline{h}_{1,n_S+1} & \underline{h}_{1,n_S+3} & \cdots & \underline{h}_{1,n_T-3} & \underline{h}_{1,n_T-1} \\ \underline{h}^*_{1,n_S+2} & \underline{h}^*_{1,n_S+4} & \cdots & \underline{h}^*_{1,n_T-2} & \underline{h}^*_{1,n_T} \\ \vdots & \vdots & & \vdots & \vdots \\ \underline{h}^*_{n_R,n_S+4} & \underline{h}^*_{n_R,n_S+2} & \cdots & \underline{h}^*_{n_R,n_T} & \underline{h}^*_{n_R,n_T-2} \end{bmatrix}. \tag{46}$$

Once the matrix $H_{div}^{ala\ /\ abba}$ is found, the next step is to apply the Sorted QR decomposition on it. This calculation may be carried out using Algorithm 2 below; the ordering among elements of the matrix $H_{div}^{ala\ /\ abba}$ is by block and not by column. It is only necessary to calculate the odd rows and columns of matrices Q^{div} and R^{div}. It can be seen as H_{div}^{ala} and H_{div}^{abba} have the same structure, the matrix H_{div}^{abba} can be seen as a particular case of H_{div}^{ala} with two Alamouti coders in each block ABBA. The above consideration allows us to use the same algorithm to calculate the *HC Sorted QR* decomposition for both schemes. The matrices Q^{div} and R^{div} generated in this part of the process represent the contribution of diversity layers in the *HC Sorted QR* decomposition.

Algorithm 2. *HC Sorted QR* of the Diversity Layers

```
INPUT: $H_{div}^{ala/abba}$, $L$, $nsym$, $n_B, n_S, order$
OUPUT: $Q^{div}$, $R^{div}$, $order$
$Q^{div} = H_{div}^{ala/abba}$, $R^{div} = 0$, $m = Ln_S$
for $i = 1:2$ to $2n_B$ do
    $k = \arg\min_{j=i:2:2n_B} \left\| Q^{div}(:,j) \right\|$
    Exchange columns i and i+1 for k and k+1 of $Q^{div}$ and $R^{div}$
    Exchange columns $m+Li:-1:m+i(L-1)+1$ and
        $m+Lk:-1:m+k(L-1)+1$ of $order$
    $R^{div}(i,i) = \left\| Q^{div}(:,i) \right\|_2$
    $Q^{div}(:,i) = Q^{div}(:,i) / R^{div}(i,i)$
    $R^{div}(i+1,i+1) = -R^{div}(i,i)$
    $Q^{div}(1:2:Ln_R-1,i+1) = -Q^{div}(2:2:Ln_R,i)^*$
    $Q^{div}(2:2:Ln_R,i+1) = Q^{div}(1:2:Ln_R-1,i)^*$
    for $j = i+1$ to $2n_B$ do
        $R^{div}(i,j) = Q^{div}(:,i)^H Q^{div}(:,j)$
    endfor
    $R^{div}(i+1,i+2:2:Ln_B-1) = R^{div}(i,i+3:2:Ln_B)^*$
    $R^{div}(i+1,i+3:2:Ln_B) = -R^{div}(i,i+2:2:Ln_B-1)^*$
    for $j = i+1$ to $2n_B$ do
        $Q^{div}(:,j) = Q^{div}(:,j) - Q^{div}(:,i)R^{div}(i,j)$
        $Q^{div}(:,j) = Q^{div}(:,j) - Q^{div}(:,i+1)R^{div}(i+1,j)$
    endfor
endfor
```

The matrices Q_{LD} and R_{LD} are generated from the matrices Q^{spa}, R^{spa}, Q^{div} and R^{div}. The construction process is described in Algorithms 3 and 4. Once the matrices Q_{LD} and R_{LD} are generated the detection of the received symbols was carried out according to the procedure described in section 5.1.

Algorithm 3. Generation of matrices Q_{LD} and R_{LD} for the scheme ZF-SQRD LDSTBC

```
INPUT: $Q^{spa}$, $R^{spa}$, $Q^{div}$, $R^{div}$, $n_B$, $n_S$
OUPUT: $Q_{LD}$, $R_{LD}$
$col = 1$
  for $i = 1$ to $n_S$ do
      $Q_{LD}^{spa}(1:2:2n_R - 1, col) = Q^{spa}(:,k)$
      $col = col + 1$
      $Q_{LD}^{spa}(2:2:2n_R, col) = -Q^{spa}(:,k)^*$
      $col = col + 1$
  endfor
  $row = 1$
  for $i = 1$ to $n_S$ do
      $R_{LD}^{spa}(row,1:2:2n_S - 1) = R^{spa}(k,1:n_S)$
      $R_{LD}^{spa}(row,1:2:2n_S - 1) = R^{spa}(k,1:n_S)$
      $row = row + 2$
  endfor
  $row = 1, col = 2n_S + 1$
  for $i = 1$ to $n_S + n_B$ do
      for $j = 1$ to $n_B$ do
          $R_{LD}^{spa}(row, col) = R^{spa}(i, 2j - 1 + n_S)$
          $R_{LD}^{spa}(row, col + 1) = R^{spa}(i, 2j + n_S)$
          $R_{LD}^{spa}(row + 1, col) = R^{spa}(i, 2j + n_S)^*$
          $R_{LD}^{spa}(row + 1, col + 1) = -R^{spa}(i, 2j - 1 + n_S)^*$
          $col = col + 2$
      endfor
      $col = 2n_S + 1$
      $row = row + 2$
  endfor
```

```
$R_{LD}^{ala}(2n_S+1:2(n_S+n_B),2n_S+1:2(n_S+n_B))=R^{div}$
$R_{LD}=\begin{bmatrix}R_{LD}^{spa}\\ R_{LD}^{ala}\end{bmatrix}$, $Q_{LD}=\begin{bmatrix}Q_{LD}^{spa} & Q^{div}\end{bmatrix}$
```

Algorithm 4. Generation of matrices Q_{LD} and R_{LD} for the scheme ZF-SQRD LQOSTBC

```
INPUT: $Q^{spa}$, $R^{spa}$, $Q^{div}$, $R^{div}$, $n_B$, $n_S$
OUPUT: $Q_{LD}$, $R_{LD}$
$row=1;\ k=4n_S$
for $i=1$ to $n_S$ do
    $R_{LD}^{spa}(row,1:4:k)=R^{spa}(i,1:n_S)$
    $R_{LD}^{spa}(row+1,2:4:k)=-R^{spa}(i,1:n_S)^*$
    $R_{LD}^{spa}(row+2,3:4:k)=R^{spa}(i,1:n_S)$
    $R_{LD}^{spa}(row+3,4:4:k)=-R^{spa}(i,1:n_S)^*$
    $row=row+4$
endfor
$col=1$
for $i=1$ to $n_S$ do
    $Q_{LD}^{spa}(1:4:4n_R,col)=Q^{spa}(:,i)$
    $Q_{LD}^{spa}(2:4:4n_R,col+1)=Q^{spa}(:,i)^*$
    $Q_{LD}^{spa}(3:4:4n_R,col+2)=Q^{spa}(:,i)$
    $Q_{LD}^{spa}(4:4:4n_R,col+3)=Q^{spa}(:,i)^*$
    $col=col+4$
endfor
$row=1,col=4n_S+1,col2=n_S+1$
for $i=1$ to $n_S$ do
    for $j=1$ to $n_B$ do
        $R_{LD}^{abba}(row,col)=R^{spa}(i,col2)$
        $R_{LD}^{abba}(row+1,col)=R^{spa}(i,col2+1)^*$
        $R_{LD}^{abba}(row+2,col)=R^{spa}(i,col2+2)$
        $R_{LD}^{abba}(row+3,col)=R^{spa}(i,col2+3)^*$
```

```
            $R_{LD}^{abba}(row, col+2) = R^{spa}(i, col2)$
            $R_{LD}^{abba}(row+1, col+2) = R^{spa}(i, col2+1)^*$
            $R_{LD}^{abba}(row+2, col+2) = R^{spa}(i, col2+2)$
            $R_{LD}^{abba}(row+3, col+2) = R^{spa}(i, col2+3)^*$
            $R_{LD}^{abba}(row, col+1) = R^{spa}(i, col2+1)$
            $R_{LD}^{abba}(row+1, col+1) = -R^{spa}(i, col2)^*$
            $R_{LD}^{abba}(row+2, col+1) = R^{spa}(i, col2+3)$
            $R_{LD}^{abba}(row+3, col+1) = -R^{spa}(i, col2+2)^*$
            $R_{LD}^{abba}(row, col+3) = R^{spa}(i, col2+1)$
            $R_{LD}^{abba}(row+1, col+3) = -R^{spa}(i, col2)^*$
            $R_{LD}^{abba}(row+2, col+3) = R^{spa}(i, col2+3)$
            $R_{LD}^{abba}(row+3, col+3) = -R^{spa}(i, col2+2)^*$
            $col = col+4, col2 = col2+4$
        endfor
        $col = 4n_S + 1$, $col2 = n_S + 1$, $row = row + 4$
    endfor
$R_{LD} = \begin{bmatrix} R_{LD}^{spa} & R^{div} \end{bmatrix}$, $Q_{LD} = \begin{bmatrix} Q_{LD}^{spa} & Q^{div} \end{bmatrix}$
```

6. Simulation results

To demonstrate the advantages of the codes presented in section 5 we compare the bit error rate (BER) performance of recent hybrid codes, employing 16-QAM modulation. In all cases, we have fixed the code rate to 3 symbols per channel use. The block length is fixed to $L = 4$. We remark that, besides better BER performance, the codes we have presented have lower receiver complexity, requiring between 4% and 12% fewer multiplications.

In Figure 7, we show the BER performance comparison between QR Group Receiver 6 × 6 (Zhao & Dubey, 2005), STBC-VBLAST algorithm 6 × 6 (2, 2, 3) (Mao et al., 2005), and ZF-SQRD LQOSTBC with $n_R = 6$, $n_T = 6$ and $n_A = 4$.

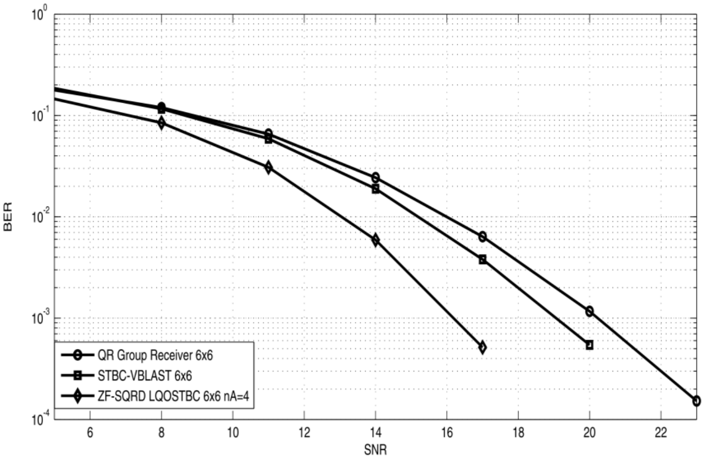


Fig. 7. BER vs. SNR of ZF-SQRD LQOSTBC, STBC-VBLAST 6 × 6 (2, 2, 3) and QR Group Receiver 6 × 6.

Regarding ZF-SQRD LDSTBC, the block length is fixed to $L = 2$. In Figure 8, we show the BER performance comparison between QR Group Receiver 6 × 6, STBC-VBLAST 6 × 6 (2, 2, 3), and ZF-SQRD LDSTBC with $n_R = 6$, $n_T = 6$ and $n_B = 3$.

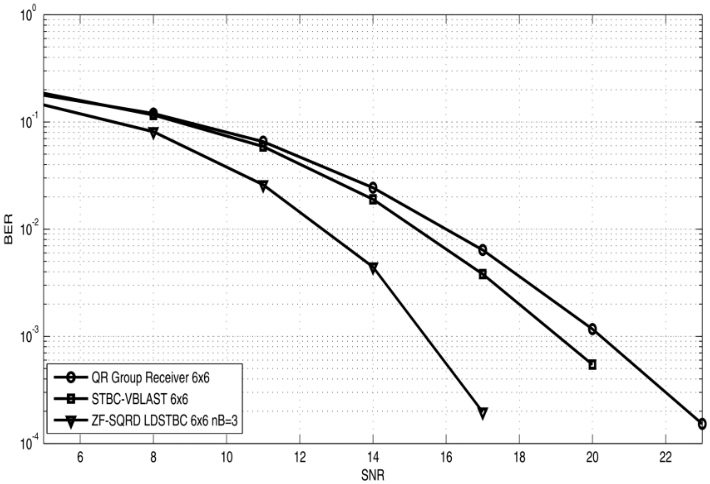


Fig. 8. BER vs. SNR of ZF-SQRD LDSTBC, STBC-VBLAST 6 × 6 (2, 2, 3) and QR Group Receiver 6 × 6.

7. Conclusions

We have presented an overview of space-time block codes, with a focus on hybrid codes, and analyzed in some depth two hybrid MIMO space-time codes with arbitrary number of

STBC/ABBA and spatial layers, and a receiver algorithm with very low complexity. We have used the theory of linear dispersion codes to transform the original MIMO system to an equivalent system where the OSIC method for nulling and cancellation of the interference among layers can be applied.

8. Future Research

There are many open, interesting lines of research on hybrid codes. One is to carry out a theoretical analysis of their spectral efficiency versus diversity trade-off. It would also be interesting to explore a hardware implementation of a hybrid MIMO transceiver. Finally, as the performance limits of perfect space-time codes become well understood, hybrid codes remain an attractive alternative, although more rigorous and powerful construction and analysis techniques are required.

9. Acknowledgments

We thank the CONACYT research (grant 51332-Y) and Intel research grant INTEL-CERMIMO2008 and PROMEP for supporting this work.

10. References

Alamouti S. M. (1998). A Simple Transmit Diversity Technique for Wireless Communications, *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, October 1998, pp. 1451-1458, ISSN:0733-8716

Arar, M. & Yongacoglu A. (2006) Efficient Detection Algorithm for 2N×2N MIMO Systems Using Alamouti Code and QR Decomposition, *IEEE Communication Letters*, vol. 10, no. 12, December 2006, pp. 819-821, ISSN:1089-7798

Berhuy G. & Oggier F. (2009). On the Existence of Perfect Space-Time Codes, *IEEE Transactions on Information Theory*, Vol. 55, No. 5, May 2009, pp. 2078-2082, ISSN: 0018-9448

Biglieri E., Calderbank R., Constantinides A., Goldsmith A., Paulraj A., & Poor H. V. (2007). *MIMO Wireless Communications*, Cambridge University Press, ISBN: 978-0-521-87328-4, United Kingdom

Cortez J., Bazdresch M., Torres D. & Parra-Michel R. (2007). An Efficient Detector for Non-Orthogonal Space-Time Block Codes with Receiver Antenna Selection, *Proceedings of the 18th IEEE International Symposium on Personal Indoor and Mobile Wireless Communications*, pp. 1-5, ISBN: 978-1-4244-1144-3, Athens, Greece, September 2007

Cortez J., Bazdresch M., Torres D. & Parra-Michel R. (2008). Generalized ABBA-VBLAST Hybrid Space-Time Code for MIMO Wireless Communications, *Proceedings of the 9th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, pp. 481-485, ISBN:978-1-4244-2045-2, Recife, Brazil, July 2008

Dai L., Sfar S. & Letaief K. (2007). A Quasi-Orthogonal Group Space-Time Architecture to Achieve a Better Diversity-Multiplexing Tradeoff, *IEEE Transactions on Wireless Communications*, Vol. 6, No. 4, pp. 1295-1307, April 2007, ISSN: 1536-1284

Foschini G. J. & Gans M. J. (1998). On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas. Wireless Personal Communications, Vol. 6, No. 3, March 1998, pp. 311-335, ISSN: 0929-6212

Freitas W., Cavalcanti F. & Lopes R. (2006). Hybrid MIMO Transceiver Scheme with Antenna Allocation and Partial CSI at Transmitter, *Proceedings of the 17th Annual International Symposium on Personal Indoor and Mobile Radio Communications*, pp. 1-5, ISBN:1-4244-0329-4, Helsinki, Finland, September 2006

Gesbert D., Shafi M., Shiu D.-S., Smith P. J. & Naguib A. (2003). From Theory to Practice: an Overview of MIMO Space-Time Coded Wireless Systems, *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 3, April 2003, pp. 281-302, ISSN: 0733-8716

Golden G., Foschini G., Valenzuela R. & Wolniansky P. (1999). Detection Algorithm and Initial Laboratory Results Using VBLAST, *Electronic Letters*, Vol. 35, No. 1, January 1999, pp. 14-16, ISSN:0013-5194

Hassibi B. & Hochwald B. (2002) High-rate Codes that are Linear in Space and Time, *IEEE Transactions on Information Theory*, vol. 48, no. 7, July 2002, pp. 1804-1824, ISSN:0018-9448

Jafarkhani H. (2001). A Quasi-Orthogonal Space-Time Block Code. IEEE Transactions on Communications Vol. 49, No. 1, January 2001, pp. 1-4, ISSN: 0090-6778

Kim H., Park H., Kim T. & Eo I., Performance Analysis of DSTTD System with Decision Feedback Detection, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 14-18, ISBN: 1-4244-0728-1, May 2006, Tolouse

Kwak K., Kim J., Park B. & Hong D. (2005). Performance Analysis of DSTTD Based on Diversity-Multiplexing Trade-off, *61st IEEE Vehicular Technology Conference*, 2005, Vol. 2, pp. 1106-1109, 30 May-1 June 2005

Longoria O., Sanchez A., Cortez J., Bazdresch M. & Parra-Michel R. (2007) Linear Dispersion Codes Generation from Hybrid STBC-VBLAST Architectures, *Proceedings of IEEE International Conference on Electrical and Electronics Engineering*, pp. 142-145, ISBN: 1-4244-1166-1, Septemer 2007, Mexico

Mao T. & Motani M. (2005) STBC-VBLAST for MIMO Wireless Communication Systems, *Proceedings of IEEE International Conference on Communications*, pp. 2266-2270, ISBN: 0-7803-8938-7, May 2005, Seuol

Oggier F., Belfiore J.-C. & Viterbo E. (2007). Cyclic Division Algebras: A Tool for Space-Time Coding. *Foundations and Trends in Communications and Information Theory*, Vol. 4, No. 1, pp. 1-95, 2007, ISBN: 978-1-60198-050-2

Paulraj A., Gore D. A., Nabar R. U. & Bölcskei H. (2004). An Overview of MIMO Communicatios-A Key to Gibabit Wireless. *Proceedings of the IEEE*, Vol. 92, No. 2, February 2004, pp. 198-218, ISSN: 0018-9219

Pe M., Pham V., Mai L. & Yoon G. (2005). Low-complexity maximum-likelihood decoder for four-transmit-antenna quasi-orthogonal space-time block code, *IEEE Transactions on Communications*, vol. 53, no. 11, November 2005, pp. 1817-1821, ISSN: 0090-6778

Tarokh V., Jafarkhani H. & Calderbank A. (1999). Space-Time Block Codes from Orthogonal Designs, *IEEE Transactions on Information Theory*, vol. 45, no. 5, July 1999, pp. 1456-1467, ISSN: 0018-9448

Telatar I. (1999). Capacity of Multiple-Antenna Gaussian Channels. *European Transactions on Telecommunications*, Vol. 10, No. 6, February 1999, pp. 585-595, ISSN: 1120-3862.

Tirkkonen O., Boariu A. & Hottinen A. (2000). Minimal Non-Orthogonality Rate 1 Space-Time Block Code for 3+ Tx Antennas, *Proceedings of the 6th International Symposium on Spread-Spectrum Technology and Applications*, New Jersey, USA, Septermber 6-8, 2000

Tse D. & Viswanath D. (2005), *Fundamentals of Wireless Communications*, Cambridge University Press, ISBN: 0521845270, England.

Wubben D., Bohnke R., Rinas J., Kuhn V. & Kammeyer K. (2001). Efficient algorithm for decoding layered space-time codes, *IEE Electronic Letters*, vol. 37, no. 22, October 2001, pp. 1348-1350, ISSN: 0013-5194.

Zhao L. & Dubey V. (2005). Detection Schemes for Space-Time Block Code and Spatial Multiplexing Combined System, *IEEE Communication Letters*, vol. 9, no. 1, January 2005, pp. 49-51, ISSN: 1089-7798

Zheng L. & Tse D. (2003) Diversity and Multiplexing: Fundamental TradeOff in Multiple Antenna Channels, *IEEE Transactions on Information Theory*, vol. 49, no. 5, May 2003, pp. 1073-1096, ISSN: 0018-9448

8

MIMO Channel Characteristics in Line-of-Sight Environments

Leilei Liu, Wei Hong, Nianzu Zhang, Haiming Wang and Guangqi Yang
State Key Lab. of Millimeter Waves, School of Information Science and Engineering,
Southeast University
China

1. Introduction

It is known that the performance of Multiple-Input-Multiple-Output (MIMO) system is highly dependent on the channel characteristics, which determined by antenna configuration and richness of scattering. In this chapter, we address the utilization issue of MIMO communication in strong line-of-sight (LOS) component propagation. It will be focused on the characteristics of the MIMO channel matrix, the channel capacity and the condition number of the matrix. Two typical scenarios will be discussed: the pure LOS environment and the LOS environment with a scatterer. Our previous researches (Liu et al., 2007 & 2009) formed the basis of this chapter.

For the first case, the design constraint for antenna arrangement as a function of frequency and distance is discussed for the LOS MIMO communication. Then is can be seen how this constraint works and how could this constraint be weaken by smart geometrical arrangement and multi-polarization.

For the second case, the effects of scatterer on the MIMO channel characteristics in LOS environment are described. The MIMO channel matrix is expressed analytically for typical scatterer and the microstrip antennas are considered in this case. Some suggestions for practical MIMO system design will be presented in the end.

2. MIMO channel model and channel characteristics

2.1 MIMO channel matrix H

We consider a MIMO channel model with n_t transmit antennas and n_r receive antennas [illustrated in Fig. 1]. The channel impulse response between the *i* transmit antenna and the *j* receive antenna is denoted as $h_{j,i}$. Given that the signal $x_i(t)$ is launched from the *i* transmit antenna, the signal received at the *j* receive antenna is given by

$$y_j(t) = \sum_{i=1}^{n_t} h_{j,i}(t) * x_i(t) + \eta_j(t) \quad i = 1,2,...,n_t;\ j = 1,2,...,n_r \tag{1}$$

where $*$ denotes the convolution operation, $\eta_j(t)$ is the additive noise in the receiver.

The channel is assumed to be frequency-flat over the band of interest, then (1) is rewritten as

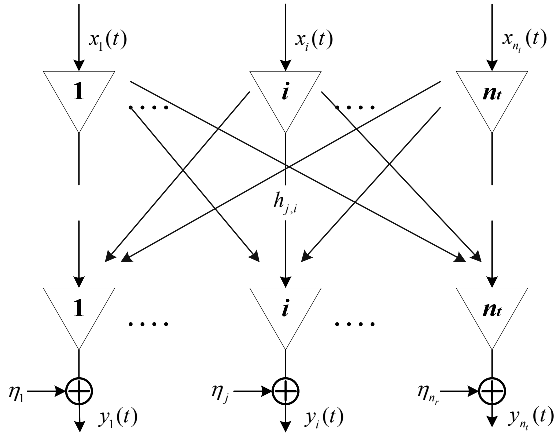


Fig. 1. MIMO channel model with n_t transmit antennas and n_r receive antennas

$$y_j(t) = \sum_{i=1}^{n_t} h_{j,i}(t)x_i(t) + \eta_j(t) \tag{2}$$

It can be described by matrix form

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \boldsymbol{\eta}(t) \tag{3}$$

where

$$\mathbf{x}(t) = (x_1(t), x_2(t), ..., x_{n_t}(t))^{\mathrm{T}} \in \mathrm{C}^{n_t \times 1}$$
$$\mathbf{y}(t) = (y_1(t), y_2(t), ..., y_{n_r}(t))^{\mathrm{T}} \in \mathrm{C}^{n_r \times 1}$$
$$\boldsymbol{\eta}(t) = (\eta_1(t), \eta_2(t), ..., \eta_{n_r}(t))^{\mathrm{T}} \in \mathrm{C}^{n_r \times 1}$$

are the transmitted signal vector, the received signal vector and the zero-mean complex Gaussian noise vector respectively.

The composite MIMO channel response is given by the matrix **H** with

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n_t} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n_t} \\ \vdots & \vdots & & \vdots \\ h_{n_r,1} & h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix} \in \mathrm{C}^{n_r \times n_t} \tag{4}$$

The discrete signal model is obtained by sampling as symbol time $\mathrm{T_s}$

$$\mathbf{y}(k) = \sqrt{E_s/n_t}\,\mathbf{H}\mathbf{x}(k) + \boldsymbol{\eta}(k) \tag{5}$$

where E_s is the total average energy available at the transmitter over a symbol period, the normalized transmitted energy on every transmit antenna is $\sqrt{E_s/n_t}$ during a symbol period.

2.2 LOS component and NLOS component

The Ricean MIMO channel model decomposes the channel into a deterministic LOS component and a stochastic NLOS (non-LOS) component for the scattered multi-path signal (Erceg et al., 2002), where the Ricean K-factor is defined as the ratio between the power of the two (Tepedelenlioglu et al., 2003), and the common path loss is moved out of the matrix **H** as being normalized.

$$\mathbf{H} = \sqrt{\frac{K}{1+K}} \cdot \mathbf{H}_{LOS} + \sqrt{\frac{1}{1+K}} \cdot \mathbf{H}_{NLOS} \tag{6}$$

Note that when $K = \infty$ the matrix becomes a pure LOS matrix and when $K = 0$ it corresponds to the case of pure Rayleigh fading. As the environment we concerned is related to the microwave relay system, which is a pure LOS channel generally, our discussion will be focused on the LOS scenario, which implies $K = \infty$.

2.3 Deterministic MIMO Channel Capacity

Chanel capacity evaluates the performance of MIMO channels by quantifying the maximum information able to be transmitted by the propagation channel without error (Paulraj et al., 2004). We assume that the channel H is perfectly known to the receiver. The capacity expressed in Bit/s/Hz of the MIMO channel is given by (Telatar, 1999)

$$C = \max_{\mathbf{R_{xx}}, tr(\mathbf{R_{xx}})=n_t} \log_2 \left[\det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{H} \mathbf{R_{xx}} \mathbf{H}^{\mathrm{H}} \right) \right] \tag{7}$$

where E_s is the total average energy available at the transmitter over a symbol period, N_0 is additive temporally white complex Gaussian noise. $(\cdot)^H$ stands for complex conjugate transpose, $tr(\cdot)$ stands for trace.

$\mathbf{R}_{xx}$ is the covariance matrix of transmitted signal *x*. The maximization is performed over all possible input covariance matrices satisfying $tr(\mathbf{R}_{xx}) = n_t$.

Given a bandwidth of *W*Hz, the maximum asymptotically error-free data rate supported by the MIMO channel is simply *WC*bit/s.

Assume that CSI (Channel State Information) is known only at the receiver, and then the covariance matrix should be

$$\mathbf{R_{xx}} = \mathbf{I}_{n_t} \tag{8}$$

This implies that the signals transmitted from the individual antennas are independent and equi-powered. With (8)submitted to (7), the channel capacity is given as (Foschiniet al. , 1998)

$$C = \log_2 \left[\det \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \mathbf{H} \mathbf{H}^{\mathrm{H}} \right) \right] \tag{9}$$

which may be decomposed as

$$C = \sum_{i=1}^{m} \log_2 \left(\mathbf{I}_{n_r} + \frac{E_s}{n_t N_0} \lambda_i \right) \tag{10}$$

where $m = \min\{n_t, n_r\}$, λ_i denotes the positive eigenvalues of W, or the singular value of the matrix $\mathbf{H}$.

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^{\mathrm{H}} & n_r \le n_t \\ \mathbf{H}^{\mathrm{H}}\mathbf{H} & n_r > n_t \end{cases}$$

Equation (10) expresses the spectral efficiency of the MIMO channel as the sum of the capacities of m SISO channels with corresponding channel gains $\sqrt{\lambda_i}\,(i = 1,2,...m)$ and transmit energy E_s/n_t (Paulraj et al., 2004).

2.3 Condition number of the channel matrix

The condition number of the channel matrix is the second important characteristic parameter to evaluate the environmental modelling impact on MIMO propagation. It is known that low-rank matrix brings correlations between MIMO channels and hence is incapable of supporting multiple parallel data streams. Since a channel matrix of full rank but with a large condition number will still bring high symbol error rate, condition number is preferred to rank as the criterion.

The condition number is defined as the ratio of the maximum and minimum singular value of the matrix **H**.

$$cond(\mathbf{H}) = \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \tag{11}$$

The closer the condition number gets to one, the better MIMO channel quality is achieved. As a multiplication factor in the process of channel estimation, small condition number decreases the error probability in the receiver.

3. MIMO technique utilized in LOS propagation

As discussed above, the high speed data transmission promised by the MIMO technique is highly dependent on the wireless MIMO channel characteristics. The channel characteristics are determined by antenna configuration and richness of scattering. In a pure LOS component propagation, low-rank channel matrix is caused by deficiency of scattering (Hansen et al., 2004).

Low-rank matrix brings correlations between MIMO channels and hence is incapable of supporting multiple parallel data streams. But some propagation environments, such as microwave relay in long range communication and WLAN system in short range communication, are almost a pure LOS propagation without multipath environment. However, by proper design of the antenna configuration, the pure LOS channel matrix could also be made high rank. It is interesting to investigate how to make MIMO technique utilized in LOS propagation.

3.1 The design constraint

We firstly consider a symmetrical 4×4 MIMO scheme with narrow beam antennas. The practical geometric approach is illustrated in Fig. 2, this geometrical arrangement can extend the antenna spacing and hence reduce the impact of MIMO channel correlation. On each side, the four antennas numbered clockwise are distributed on the corners of a square with the antenna spacing *d*. R represents the distance between the transmitter and the receiver.

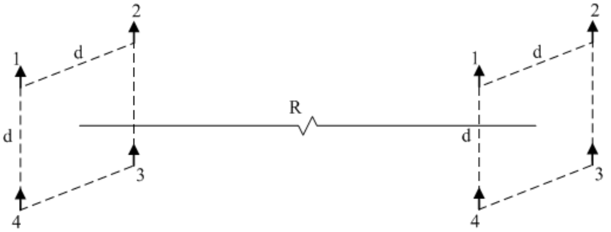


Fig. 2. Arrangement of 4 Rx and 4 Tx antennas model

We assume the distance *R* is much larger than the antenna spacing *d*. This assumption results in a plane wave from the transmitter to the receiver. In addition, the effect of path loss differences among antennas can be ignored, only the phase differences will be considered.

From the geometrical antenna arrangement, we have the different path lengths $r_{m,n}$ from transmitting antenna *n* to receive antenna *m*:

$$r_{1,1} = R$$

$$r_{2,1} = r_{4,1} = \sqrt{R^2 + d^2} \approx R + d^2/(2R)$$

$$r_{3,1} = \sqrt{R^2 + 2d^2} \approx R + d^2/R, \ldots$$

All the approximations above are made use of first order Taylor series expansion, which becomes applicable when the distance is much larger than antenna spacing.

Denoting the received vector from transmitting antenna *n* as

$$\mathbf{h}_n = [\exp(-j\frac{2\pi}{\lambda}r_{1,n}), \ldots, \exp(-j\frac{2\pi}{\lambda}r_{4,n})]^T,\ n = 1\ldots4 \tag{12}$$

where λ is the wavelength and $(\cdot)^T$ denotes the vector transpose. Thus the channel matrix is given as

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4] \tag{13}$$

The best situation for the channel matrix is that its condition number (11) equals to one. It is satisfied when **H** is the full orthogonality matrix which means all the columns (or rows) are orthogonal.

Orthogonality between different columns in (13) is obtained if the inner product of two received vectors from the adjacent transmitting antennas equals to zero:

$$< h_k, h_{k+1} > = 2\exp(-j\frac{2\pi}{\lambda}(2R + \frac{d^2}{2R}))[1 + \exp(-j\frac{2\pi}{\lambda}\cdot\frac{d^2}{R})] = 0 \tag{14}$$

which results in

$$d^2 = \frac{\lambda R}{2}(2k+1) \qquad k = 0,1\ldots \tag{15}$$

To get practical values of d, we choose $k = 0$ to update (15). The optimal design constraint therefore becomes

$$d^2 = \frac{\lambda R}{2} \tag{16}$$

From (16) we can see the optimal antenna spacing is a function of carrier frequency and propagation distance as well as the geometrical arrangement. As $\lambda = c/f$ shows, higher frequency results in smaller antenna spacing requirement, but longer distance increases it. Here c denotes the velocity of light. According to the deriving process above, it is obvious that when constraint (16) is satisfied, the channel matrix **H** in (16) is a full-rank matrix. This optimal design constraint is also determined by the antenna array arrangement, since different path lengths $r_{m,n}$ gives different channel matrix.

The derivation foundation of (16) is the condition number of **H** in (11) equals to one, which also satisfies all singular values of H are equal, that is

$$\lambda_1(\mathbf{H}) = \cdots = \lambda_4(\mathbf{H}) \tag{17}$$

From (17) and the method of Lagrange multipliers, the highest channel capacity in (10) is obtained (Bohagen et al. 2005). Therefore, maximum capacity and best condition number agree well.

The relation between the condition number and the capacity in pure LOS propagation is depicted in Fig. 3. The capacity is in linear inverse proportion to condition number, i.e., the closer the condition number is to one, higher the capacity is. It achieves the maximum capacity of 17.6 bit/s/Hz at 3km transmission distance with 30GHz frequency and 20dB SNR (Signal to Noise Ratio).

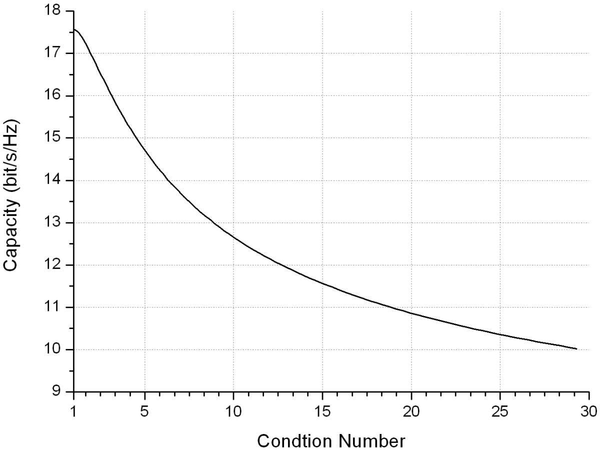


Fig. 3. Capacity as a function of condition number in pure LOS propagation, for the case that SNR 20dB, optimal frequency 30GHz, and optimal transmit distance 3km.

3.2 MIMO channel characteristics analysis and suggestions

Antenna spacing larger than half wavelength is usually required for achieving uncorrelated subchannels in dense scattering environment (Foschini, 1996). The design constraint in (16) shows half wavelength is no longer enough for pure LOS propagation when distance between transmitter and receiver is large. In the following, we will discuss how to construct a feasible LOS MIMO channel in accordance with this design constraint. It is also interesting to explore what is the acceptable situation and how it affects the practical design.

Based on the 4×4 MIMO ray tracing model above, the relation between condition number and antenna spacing is investigated and shown in Fig. 4. It confirms that larger distance requires larger antenna spacing while higher frequency requires smaller antenna spacing. The optimal condition number can be achieved at many points because of the periodicity of traveling wave phase.

The design constraint in (16) obtains the optimal channel quality, but the large antenna spacing is difficult to achieve in practice. However, practically, the condition number around 10 is allowed from the view of link quality. For example, if carrier frequency and transmission distance are 30GHz and 2km respectively, instead of 3.2m antenna spacing for the best case, 2m antenna spacing also performs well as condition number equals to 10.

It is noted that the antenna spacing is fairly large, but potentially acceptable for microwave relay or mobile telephone towers. In addition, NLOS elements also exist in actual situation, such as weak scattering elements, rain event, etc. These causes will increase the independence of MIMO channels and therefore improve the condition number.

Some position errors will exist in practical setting, and Fig. 5 investigates how sensitive the performance of channel matrix is to the distance between two relay stations, with different or frequency 30GHz or 40GHz, and different antenna spacing 2m or 4m respectively. It shows that these four scenarios have the same degradation rate of channel quality with the

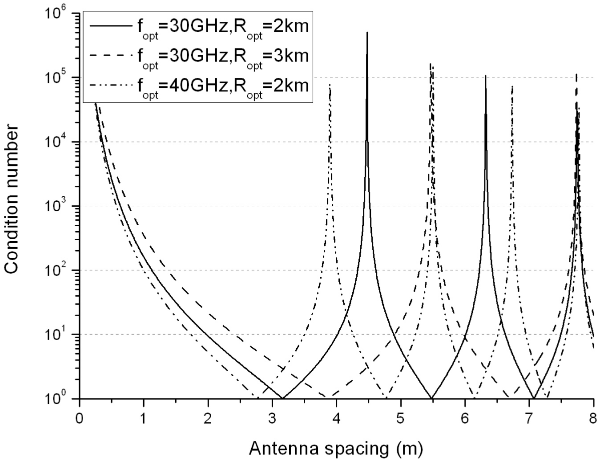


Fig. 4. Antenna spacing deviation impact on condition number with fixed frequency (30GHz, 40GHz) and fixed distance (2km, 3km). Condition number below 10 can be accepted in practice.

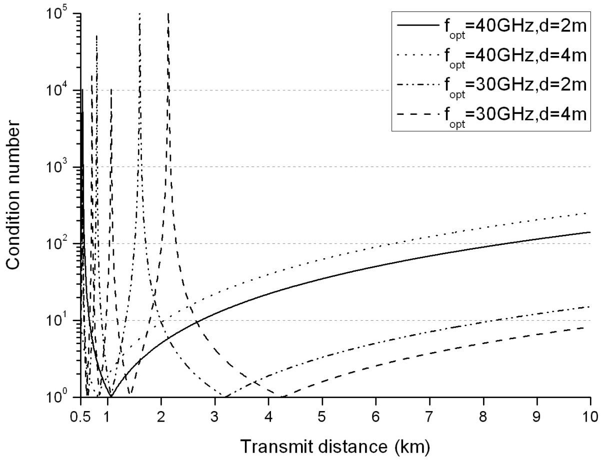


Fig. 5. Transmit distance deviation impact on condition number with fixed frequency (30GHz, 40GHz) and optimal antenna spacing (2m, 4m). 1000m location deviation yields slight performance degradation.

distance offsets. This figure also indicates that even 1000 meters location deviation yields slight performance degradation.

3.3 Effects of multi-polarization

As the design constraint shows in (16) , considerable antenna spacing is needed to introduce phase differences among antennas when operating MIMO system in LOS environment. To increase the independence among MIMO channels, multi-polarized antennas can be applied. Using the same geometry depicted in Fig. 2, we assume that α_i ($i=1,2,3,4$) denotes the offset angle of the polarization of *i*th transmitting antenna with respect to vertical polarization, while β_j ($j=1,2,3,4$) denotes the offset angle of the polarization of *j*th receiving antenna. For simplicity, we neglect the effect of cross polarization. Then the channel matrix (12) in multi-polarized LOS MIMO scenario is updated to

$$\mathbf{h}_n = [\cos(\beta_1-\alpha_n)\cdot\exp(-j\frac{2\pi}{\lambda}r_{1,n}),\ ...,\cos(\beta_4-\alpha_n)\cdot\exp(-j\frac{2\pi}{\lambda}r_{4,n})]^T,\quad n=1,...,4 \qquad (18)$$

where $\cos(\beta_j-\alpha_i)$ is the square root of normalized signal power on *j*th receiving antenna relative to *i*th transmitting antenna. With regard to this new channel matrix, we will see the improvements of channel matrix characteristics brought by multi-polarization.

Fig. 6 illustrates how the multi-polarization impacts on the MIMO channel characteristic. Three typical polarized cases are plotted compared with the uni-polarized case. By searching all the values of polarization degree in $[0^\circ,90^\circ]$, some points can be concluded: the use of multi-polarized antennas is an effective way to decrease the antenna spacing.

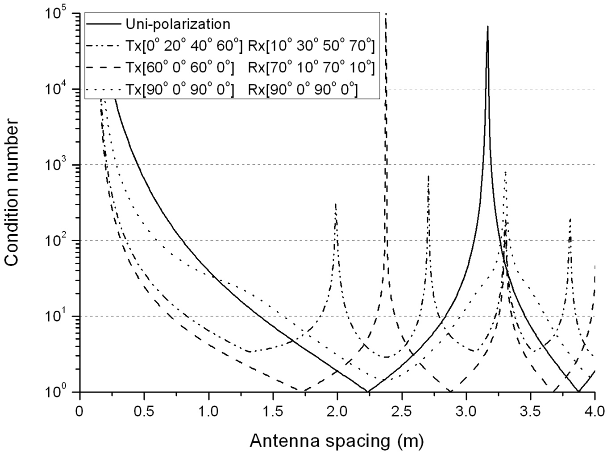


Fig. 6. Condition number as a function of antenna spacing with three polarized cases compared to the uni-polarized case. Degree of polarized antennas on transmitter side (Tx) and receiver side (Rx) follows: Case1: Tx $0^\circ,20^\circ,40^\circ,60^\circ$, Rx $10^\circ,30^\circ,50^\circ,70^\circ$; Case2: Tx $60^\circ,0^\circ,60^\circ,0^\circ$, Rx $70^\circ,10^\circ,70^\circ,10^\circ$; Case3: Tx $90^\circ,0^\circ,90^\circ,0^\circ$, Rx $90^\circ,0^\circ,90^\circ,0^\circ$. The use of multi-polarization appears as a space- and cost-effective alternative.

For instance, case 2 saves 0.8m antenna spacing to achieve condition number 10 relative to the uni-polarized case. Moreover, dual-polarization on each side leads to better channel matrix characteristic than four-polarization. Furthermore, the minimal antenna spacing we get is the orthogonal polarization $0^\circ/90^\circ$ on each side. But this is not the best choice for system performance, because the improvement of channel quality is based on sacrificing the transmitting power and the receiving diversity gain.

4. Effects of scatterer on the LOS MIMO channel

As we discuss above, the implement of MIMO technique to pure LOS propagation enviroment is restricted by a constraint which is a function of antenna arrangement, frequency and transmission distance. In actual outdoor radio channels, the existence of scattering will improve the MIMO channel performance effectively (Gesbert et al., 2002). Start from the electromagnetic knowleges, we will give the theoretical explanation on how the channel performance improved and how much it will be improved by a typical scatter.

4.1 A 2D MIMO channel model in outdoor propagation

We focused on the outdoor LOS environment but with a scatterer. It is an abstract model for the propagation enviroment of microwave relay or mobile telephone towers. Analytical method will be adopted in this channel model combining the electromagnetic theory and antenna theory.

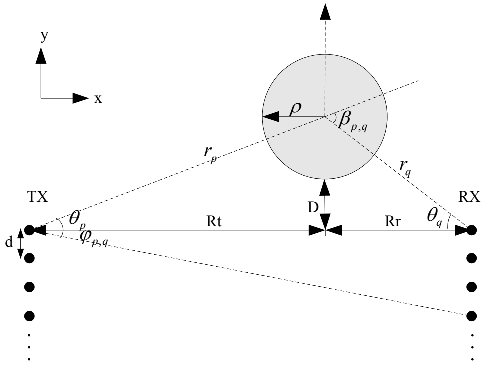


Fig. 7. A 2D MIMO channel model in outdoor propagation

A 2D MIMO channel model in outdoor LOS propagation is shown Fig. 7. Combining with the practical applications, microstrip patch array antennas are used in this model. Every rectangular patch antenna on each side is arranged along z-axis. To simplify the MIMO system is projected to x-y plane, it has *P* transmitter and *Q* receiver. The propagation is considered as a LOS situation, a cylindrical scatter is on the side of the direct path. The cylinder is the simplified model of the actual architecture in outdoor environments. Only transverse magnetic wave (vertical polarization) is considered in this electromagnetic scattering problem.

4.2 Radiation patterns of microstrip antennas

The geometry for far-field pattern of rectangular microstrip patch is shown in Fig. 8. The far-field radiation pattern of such a rectangular microstrip patch operation in the TM_{10} mode is broad in both the E and H planes. The pattern of a patch over a large ground plane may be calculated by modelling the radiator as two parallel uniform magnetic line sources of length *a*, separated by distance *b*. If the slot voltage across either radiating edge is taken as V_0, the calculated fields are (Carver et. al, 1981)

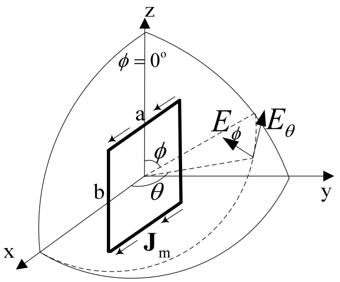


Fig. 8. Geometry for far-field pattern of rectangular microstrip patch

$$E_\theta = \frac{jV_0 k_0 a e^{-jk_0 r}}{\pi r}[\cos(kh\cos\theta)] \cdot \left[\frac{\sin[k_0 \frac{a}{2}\sin\theta\sin\phi]}{k_0 \frac{a}{2}\sin\theta\sin\phi} \right] \cdot \left[\cos(k_0 \frac{b}{2}\sin\theta\cos\phi) \right]\cos\phi, \qquad \left(0 \le \theta \le \frac{\pi}{2}\right) \tag{19}$$

$$E_\phi = \frac{jV_0 k_0 a e^{-jk_0 r}}{\pi r}[\cos(kh\cos\theta)] \cdot \left[\frac{\sin[k_0 \frac{a}{2}\sin\theta\sin\phi]}{k_0 \frac{a}{2}\sin\theta\sin\phi} \right] \cdot \left[\cos(k_0 \frac{b}{2}\sin\theta\cos\phi) \right]\cos\theta\sin\phi, \qquad \left(0 \le \theta \le \frac{\pi}{2}\right), \tag{20}$$

where h is the substrate thickness, $k = k_0\sqrt{\varepsilon_r}$, k_0 is the wave number in vacuum, ε_r is the dielectric constant , *r* is the radiation distance.

4.3 Cylindrical scattering

To obtain the analytical expression, we suppose the cylindrical scatter in Fig. 7 is a conducting cylinder with the radius ρ . The plane wave incident upon this cylinder is considered since the propagation distance from the transmitter to the cylinder is long enough.

Take the incident wave to be z-polarized, that is (Harrington, 2001)

$$E_z^i = E_0 e^{-jkr\cos\beta} \tag{21}$$

where E_0 is the far-field from transmitter to cylinder, *r* is the propagation distance and β is the scattering angle in Fig. 7.

Using the wave transformation, we can express the incident field as

$$E_z^i = E_0 \sum_{n=-\infty}^{\infty} j^{-n} J_n(kr) e^{jn\beta} \tag{22}$$

where J_n is the first kind Bessel function.

The total field with the conducting cylinder present is the sum of the incident and scattered fields, that is

$$E_z = E_z^i + E_z^s \tag{23}$$

To presnet outward-traveling waves, the scattered field must be of the form

$$E_z^s = E_0 \sum_{n=-\infty}^{\infty} j^{-n} a_n H_n^{(2)}(kr) e^{jn\beta} \tag{24}$$

where $H_n^{(2)}$ is the second kind Hankel function.

Hence the total field is

$$E_z = E_0 \sum_{n=-\infty}^{\infty} j^{-n} \left[J_n(kr) + a_n H_n^{(2)}(kr) \right] e^{jn\beta} \tag{25}$$

At the cylinder the boundary condition $E_z = 0$ at $r = \rho$ must be met. It is evident from the above equation that this condition is met if

$$a_n = \frac{-J_n(k\rho)}{H_n^{(2)}(k\rho)} \tag{26}$$

Which completes the solution.

4.4 Analytical mathematical expression of channel matrix H

According to 2D model in Fig. 7, $\phi = 90^\circ$ is adopted in the far-field radiation pattern of patch antenna. Thus, $E_\theta = 0$, and the incident wave from each microstrip antenna of the transmitter to the cylinder is

$$E_z^i = \frac{jV_0 k_0 a e^{-jk_0 r}}{\pi r}[\cos(kh\cos\theta)]\left[\frac{\sin[k_0 \frac{a}{2}\sin\theta]}{k_0 \frac{a}{2}\sin\theta}\right]\cos\theta, \quad \left(0 \le \theta \le \frac{\pi}{2}\right) \tag{27}$$

In accordance with the Geometric Relationship between the cylinder and each antenna shown in Fig. 7, the scattered field from the p antenna in transmitter to the q antenna in receiver affected by the cylinder is calculated by (24) and (27)

$$\begin{aligned} E_{p,q}^s &= \frac{jV_0 k_0 a}{\pi r_p}\exp(-jk_0 r_p)\cdot[\cos(kh\cos\theta_p)]\cdot\left[\frac{\sin[k_0(a/2)\sin\theta_p]}{k_0(a/2)\sin\theta_p}\right]\cdot\cos\theta_p \\ &\cdot\sum_{n=-N}^{N} j^{-n}\frac{-J_n(k_0\rho)}{H_n^{(2)}(k_0\rho)}H_n^{(2)}(k_0 r_q)\exp(jn\beta_{p,q}) \qquad p = 1...P, q = 1...Q \end{aligned} \tag{28}$$

where r_p is the distance from the *p* transmit antenna to the cylinder, r_q is the distance from the cylinder to the *q* receive antenna. θ_p is the angle between the LOS path and the ray path from the *p* transmit antenna to the cylinder, θ_q is the angle between the LOS path and the ray path from the cylinder to the *q* receive antenna. $\beta_{p,q}$ is the scattering angle in Fig. 7, which $\beta_{p,q} = \theta_p + \theta_q$. The value of *N* order is determined by the convergence of the Bessel function and the Hankel fuction.

The incident wave at the receiver from the LOS path is

$$\begin{aligned} E_{p,q}^i &= \frac{jV_0 k_0 a}{\pi R_{p,q}}\exp(-jk_0 R_{p,q})\cdot[\cos(kh\cos\varphi_{p,q})] \\ &\cdot\left[\frac{\sin[k_0(a/2)\sin\varphi_{p,q}]}{k_0(a/2)\sin\varphi_{p,q}}\right]\cos\varphi_{p,q}, \qquad p = 1...P, q = 1...Q \end{aligned} \tag{29}$$

where $R_{p,q}$ and $\varphi_{p,q}$ are the distance and the angle between the *p* transmit antenna and the *q* receive antenna respectively.

The antenna directivity of the receiving antenna is given by the rectangle microstrip antenna pattern

$$D(\theta)=[\cos(kh\cos\theta)]\cdot\left[\frac{\sin[k_0(a/2)\sin\theta]}{k_0(a/2)\sin\theta}\right]\cdot\cos\theta \tag{30}$$

The total field is the sum of the incident and scattered fields, that is

$$\begin{aligned}E_{p,q}&=E^{i}_{p,q}\cdot D(\varphi_{pq})+E^{s}_{p,q}\cdot D(\theta_q)\\&=\frac{jV_0k_0a}{\pi R_{p,q}}\exp(-jk_0R_{p,q})[\cos(kh\cos\varphi_{p,q})]^2\left[\frac{\sin[k_0(a/2)\sin\varphi_{p,q}]}{k_0(a/2)\sin\varphi_{p,q}}\right]^2(\cos\varphi_{p,q})^2\\&+\frac{jV_0k_0a}{\pi r_p}\exp(-jk_0r_p)[\cos(kh\cos\theta_p)]\left[\frac{\sin[k_0(a/2)\sin\theta_p]}{k_0(a/2)\sin\theta_p}\right]\cos\theta_p\cos\theta_q\\&\cdot\sum_{n=-N}^{N}j^{-n}\frac{-J_n(k_0\rho)}{H_n^{(2)}(k_0\rho)}H_n^{(2)}(k_0r_q)\exp(jn\beta_{p,q})[\cos(kh\cos\theta_q)]\left[\frac{\sin[k_0(a/2)\sin\theta_q]}{k_0(a/2)\sin\theta_q}\right]\end{aligned} \tag{31}$$

This can be also considered as the sum energy of the LOS element and the NLOS element at the receiver. This electromagnetic interpretation agrees well with the Ricean model in (6).

Define E_p^{TX} as the transmitted field, thus the channel matrix element follows

$$h_{p,q}=E_{p,q}\big/E_p^{TX} \tag{32}$$

Hence, the MIMO channel matrix is composed

$$\mathbf{H}=\begin{bmatrix}h_{1,1}&\cdots&h_{1,P}\\\cdots&\cdots&\cdots\\h_{1,Q}&\cdots&h_{P,Q}\end{bmatrix}\in C^{P\times Q} \tag{33}$$

4.5 Numerical Evaluation

Our simulation is based on a 4×4 MIMO system with working frequency at 3GHz. The antennas are excited by voltage 1V. The dielectric constant of the microstrip antennas substrate is 2.5, and its thickness is 0.03λ which determined by the working wavelength. For a matched antenna, the size could be referenced to (Bahl, et al., 1982).

The simulation parameters are initialized as follows: the spacing between antenna elements is $d=0.4\text{m}$; the radius of cylinder is $\rho=50\text{m}$ as the actual size of buildings; the propagation distance between transmitter and the receiver is $R=1\text{km}$; the projected distance from the cylinder to the transmitter and to the receiver are $Rt=800\text{m}$ and $Rr=200\text{m}$ respectively; the distance between the cylinder to the LOS path is D.

We mentioned above the order N in the scattering field expression (28) is determined by the convergence of Bessel function and Hankel function. Because the practical scatter is relatively big, large N is needed. Hence, we need to investigate the convergence of these functions first, in order to reduce the calculation complexity.

We redefine the determinative part in (28) as

$$f(n)=\sum_{n=-N}^{N}j^{-n}\frac{-J_n(k_0\rho)}{H_n^{(2)}(k_0\rho)}H_n^{(2)}(k_0r_q)\exp(jn\beta_{p,q}) \tag{34}$$

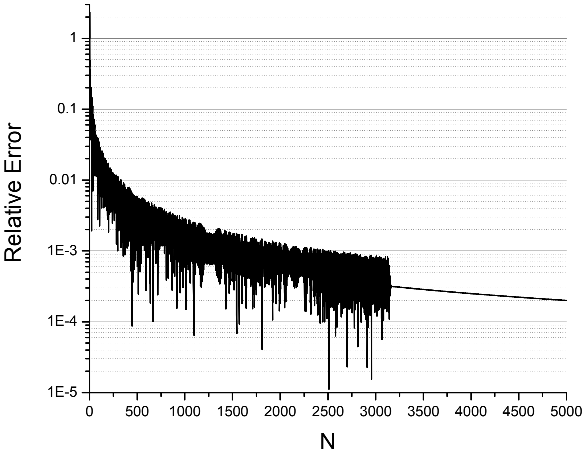


Fig. 9. The relative error function $\varepsilon(n)$, at $f = 3\text{GHz}$, $\rho = 50\text{m}$, $Rr = 200\text{m}$, and $D = 100\text{m}$

the relative error function is

$$\varepsilon(n) = \frac{|f(n+1) - f(n)|}{|f(n)|} \tag{35}$$

The relation between the error function $\varepsilon(n)$ and the order *N* is shown in Fig. 9. This curve corresponds to the MIMO system with working frequency $f = 3\text{GHz}$, and geometric parameters $\rho = 50\text{m}$, $Rr = 200\text{m}$ and $D = 100\text{m}$. It shows that when *N* is larger than 2500, it has $\varepsilon < 10^{-3}$. The $\varepsilon(n)$ curve starts smoothly when *N* is larger than 3170. The simulation shows that the value of *N* makes a strong effect on the accuracy. With accordance to (34), a higher frequency, a larger scatter or a longer propagation distance needs a larger *N* to meet the same accuracy.

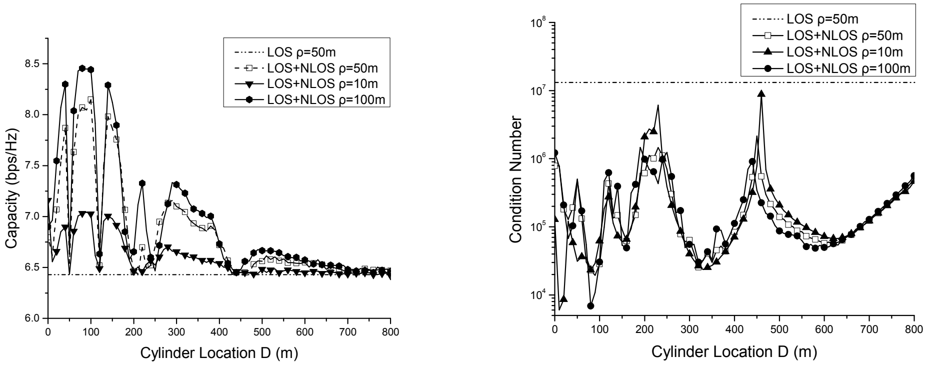
Suppose the transmitted power doesn't depend on the system frequency and propagation distance. The SNR is defined as a variable which depend on the system parameters and the actual propagation environment. Set SNR_0=10dB at 3GHz system frequency and 2km propagation distance. Then the SNR can be calculated by the transmission loss L_{bf} in free space:

$$SNR = SNR_0 + (L_{bf0} - L_{bf}) \tag{36}$$

where $L_{bf} = 20\lg(4\pi R/\lambda)$ (dB)

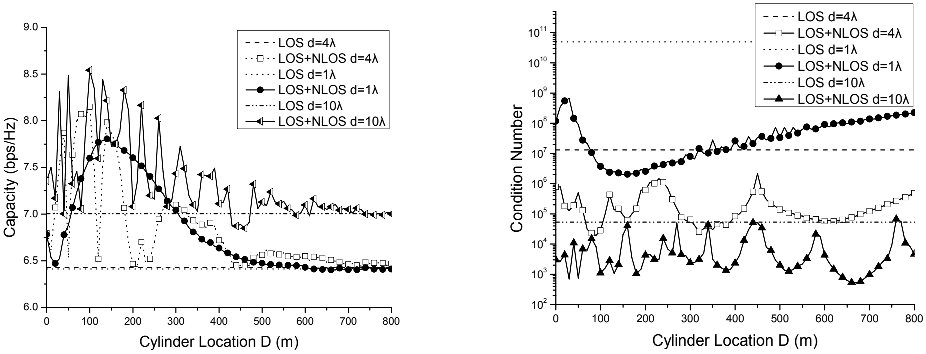
4.6 MIMO channel characteristics analysis and suggestions

Fig. 10 shows the effects of difference cylinder size on the MIMO channel performance. The cylinder distance to the LOS path steps by 10m in the simulation. Compared with the pure LOS case, the scattering in MIMO propagation improves the channel performance significantly. The larger cylinder, the higher channel capacity achieves. If the cylinder radius is 100m, the channel capacity improves more than 2bps/Hz. Fig. 10(b) shows the correlation of the MIMO sub-channels from the condition number of channel matrix. When the distance from cylinder to LOS path is smaller than 200m, the condition number reduces from 1E7 to 1E5 because of the cylindrical scattering.



(a) (b)

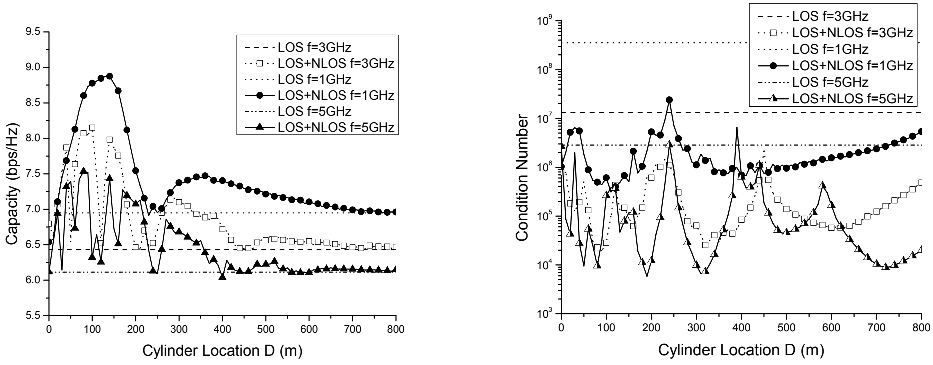
Fig. 10. The channel capacity (a) and the condition number (b) vary with the cylinder location for the different cylinder size



(a) (b)

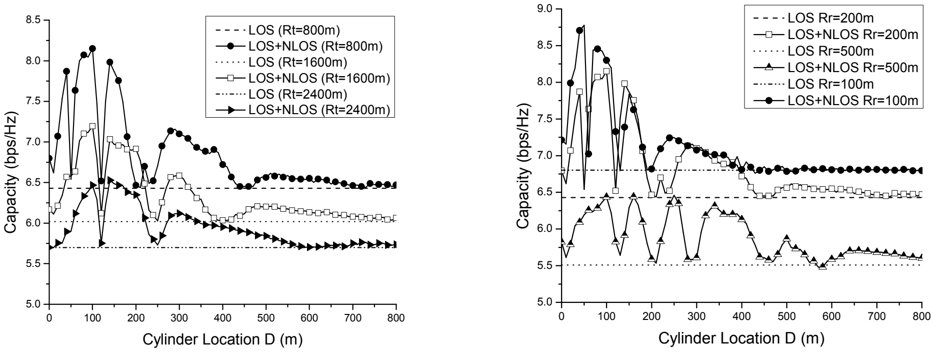
Fig. 11. The channel capacity (a) and the condition number (b) vary with the cylinder location for the different antenna spacing

From the comparison about Fig. 10(a) and Fig. 10(b), the capacity is inversely related to the condition number. Namely, the improvement on the capacity by the cylindrical scattering is the same as the condition number. Increase with the distance from the cylinder to the LOS path, the channel performance improves gradually. It behaves as the slow fading channel. This improvement is comparatively small when the cylinder is farther away than 600m. However, big fluctuation appears when the cylinder is near LOS path (*D*<200m). This is because of the fast fading caused by the superposition of the random phases in different multipath.



(a) (b)

Fig. 12. The channel capacity (a) and the condition number (b) vary with the cylinder location for the different frequency



(a) (b)

Fig. 13. The channel capacity (a) and the condition number (b) vary with the cylinder location for the different distance Rt

Fig. 11 shows that the antenna spacing increase, the correlation the subchannels becomes lower. Hence, larger capacity and smaller condition number are obtained. Besides, the channel performance is improved significantly. In addition, larger antenna spacing leads to bigger fluctuation of the capacity curves and the condition number curves.

Fig. 12 shows the different frequencies effect on the channel performance. It demonstrates that lower frequency results to less path loss and less sensitivity of fast fading. Namely, high frequency lets the curves fluctuate greatly.

Fig. 13 shows the different distances to the cylinder effect on the channel capacity. Rt is the projective distance from the transmitter to the cylinder, and Rr is the projective distance from the cylinder to the receiver. The further cylinder away from the antennas, the less improvement by the scattering could be achieved. Consequently, the channel capacity becomes lower. In Fig. 13(b) the slope of the curve $Rr = 100\text{m}$ is larger than the $Rr = 500\text{m}$ case. This is because of the received antenna orientation. With the increase distance, the angle among the received antenna and the cylinder becomes larger, which results in less received energy in former case.

There are several singular points in the curves from Fig. 10 to Fig. 13. At these points, few improvements of channel performance could be achieved. The reason for this is that the cylinder scattering causes plenty of multipath element, and the phases of these multipath superpose randomly. In the net-establishing of the base station, if the main scatters are known, the singular points could be avoided by simulated prediction.

5. Conclusion

In this chapter, we discussed the MIMO channel performance in the LOS environment, classified into two cases: the pure LOS propagation and the LOS propagation with a typical scatter.

We first deduced a useful constraint for the MIMO system applied in the pure LOS propagation. This constraint is a function of the antenna setup, the frequency and the transmission distance. Then we give the analytical mathematical expression of the MIMO channel matrix for the second environment case. From the electromagnetic explanation, we know how this scatter effects on the MIMO channel.

The MIMO channel capacity and the condition number of the matrix were investigated. We discussed how to construct such a LOS MIMO channel and how much a typical scatter could improve on the channel performance.

6. References

Bahl, I.; Bhartia, P. & Stuchly, S. (1982). Design of microstrip antennas covered with a dielectric layer. *IEEE Trans. Antennas and Propag.* Vol. 30, No. 2, pp. 314-318

Bohagen, F.; Orten, P. & Oien, G.E. (2005). Construction and capacity analysis of high-rank line-of-sight MIMO channels, *Proceedings of Wireless Communications and Networking Conference,* Vol. 1, pp. 432–437

Carver, K.; Mink, J. (1981). Microstrip antenna technology. *IEEE Trans. Antennas and Propag.* Vol. 29, No. 1, pp. 2-24

Erceg, V.; Soma, P.; Baum, D. S. & Paulraj, A. J. (2002). Capacity obtained from multiple-input multiple-output channel measurements in fixed wireless environments at 2.5 GHz, *Proceedings of IEEE International Conference on Communications,* Vol.1, pp. 396–400, April 2002

Foschini, G. J. (1996). Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas, *Bell Labs Tech. J.,* Vol. 1, No. 2, pp. 41-59.

Foschini, G. J. & Gans, M. J. (1998). On limits of wireless communications in a fading environment when using multiple antennas, *Wireless Pers. Commun.,* Vol. 6, No. 3, pp. 311–335

Gesbert, D.; Bolcskei, H.; et al. (2002). Outdoor MIMO wireless channels: models and performance prediction. *IEEE Trans. Commun.*, Vol. 50, No. 12, pp. 1926-1934.

Hansen, J. & Bölcskei, H. (2004). A geometrical investigation of the rank-1 Ricean MIMO channel at high SNR, *Proceeding of International Symposium on Information Theory*. pp. 64

Harrington, R. F. (2001). Time-harmonic Electromagnetic Fields. *IEEE Press Series on Electromagnetic Wave Theory*.

Liu, L.; Hong W.; et al. (2007). Characterization of Line-of-Sight MIMO Channel for Fixed Wireless Communications, IEEE *Antenna and Wireless Propagation Letters*,Vol.6, pp. 36-39

Liu, L.; Hong W.; (2009). Investigations on the Effects of Scatterers on the MIMO Channel Characteristics in LOS Environment, *Journal on Communications*,Vol. 30, No.2, pp: 65-70

Paulraj, A.J.; GORE, D.A.; NABAR, R.U. & BOLCSKEI, H. (2004). An overview of MIMO communications - a key to gigabit wireless, *Proceedings of the IEEE,*Vol. 92, No. 2, pp: 198 - 218

Telatar, E. (1999). Capacity of Multi-Antenna Gaussian Channels. *European Transactions on Telecommunications,* Vol. 10, No. 6, pp. 585-595

Tepedelenlioglu, C.; Abdi, A. & Giannakis, G. B. (2003). The Ricean K Factor: Estimation and Performance Analysis, *IEEE Trans. Wireless Commun.,* Vol. 2, No. 4, pp. 799–810

9

Iterative Joint Optimization of Transmit/Receive Frequency-Domain Equalization in Single Carrier Wireless Communication Systems

Xiaogeng Yuan, Osamu Muta and Yoshihiko Akaiwa
Kyushu University
Japan

1. Introduction

In recent years, high data-rate wireless transmission system such as IEEE802.16e has received increased attention. In such a system, inter-symbol interference (ISI) caused by frequency-selective fading severely degrades bit error rate (BER) performance. As a solution to overcome the effect of frequency-selective fading, Orthogonal Frequency Division Multiplexing (OFDM) has been considered. However, OFDM system has a problem that the Peak to Average Power Ratio (PAPR) becomes high as the number of sub-carriers increases.

As an alternative method to solve the above problems, Single Carrier transmission with Frequency-Domain Equalization (SC-FDE) has been investigated. In this system, a low PAPR is achieved in contrast to the OFDM system and the received signal passing through frequency-selective fading channel is equalized with frequency-domain processing based on minimum mean square error (MMSE) criterion. Hence, BER performance of SC-FDE system without error correction coding is improved by obtaining the diversity effect.

In this study, as an alternative method to improve BER performance of SC-FDE system, we propose an iterative optimization method of transmit/receive frequency domain equalization (TR-FDE) based on MMSE criterion, where both transmit and receive FDE weights are iteratively determined with a recursive algorithm so as to minimize the mean square error at a virtual receiver.

2. Proposed Iterative Optimization Method for TR-FDE

Figure 1 shows a SC-FDE system considered in this study, where a frequency-domain equalizer is equipped at both transmitter and receiver. In the proposed method, optimum weight vectors for transmit and receive equalizers are obtained by minimizing mean square error including ISI and noise at the receiver side, i.e., the error signal is defined as the difference between transmit and receive signal vectors under the constraint of the constant average transmit power. This means that error signal at the receiver side should be estimated at the transmitter side. For this purpose, a virtual channel and receiver are equipped at the transmitter, where it is assumed that channel state information (CSI) is known to the transmitter. In this system, the transmitter needs to know channel state information (CSI) measured at the receiver side.

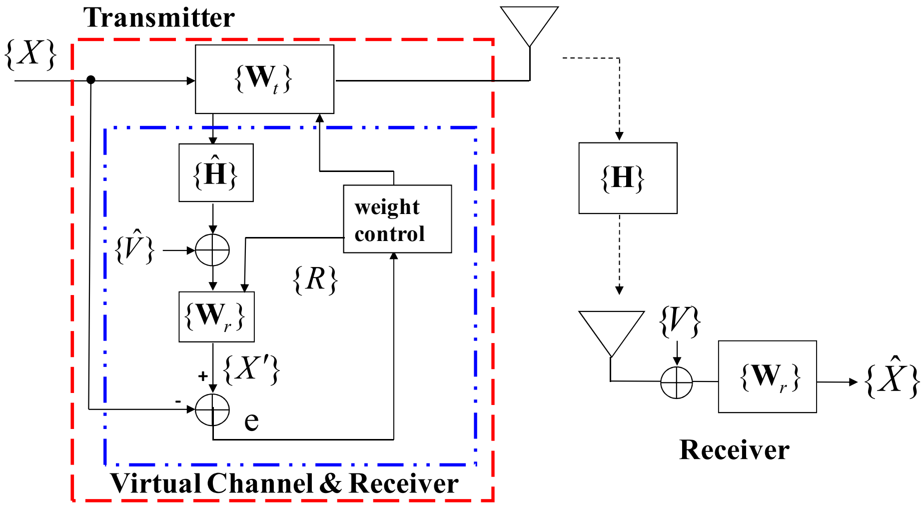


Fig. 1. Block diagram of the proposed system with transmit/receive frequency-domain equalization, where virtual channel and virtual receiver are equipped at the transmitter.

For simplicity of explanation, frequency-domain signal expressions are used in the following discussion. As shown in Figure 1, the received signal vector at the virtual receiver is expressed as

$$\hat{\mathbf{R}} = \hat{\mathbf{H}}^{\mathrm{d}}\mathbf{W}_{\mathrm{t}}^{\mathrm{d}}\mathbf{X} + \hat{\mathbf{V}} \tag{1}$$

where suffix d denotes diagonal matrix; $\hat{\mathbf{H}}^{\mathrm{d}} = \mathrm{diag}\{\hat{\mathbf{H}}\}$ and $\mathbf{W}_{\mathrm{t}}^{\mathrm{d}} = \mathrm{diag}\{\mathbf{W}_{\mathrm{t}}\}$. $\hat{\mathbf{H}} = \left(\hat{\mathrm{H}}_1,\cdots,\hat{\mathrm{H}}_{\mathrm{k}},\cdots,\hat{\mathrm{H}}_{\mathrm{N}}\right)^{\mathrm{T}}$ is the estimated channel transfer function vector.
$\mathbf{X}^{\mathrm{T}}$ denotes the transpose of vector $\mathbf{X}$. $\mathbf{X} = \left(\mathrm{X}_1,\cdots,\mathrm{X}_{\mathrm{k}},\cdots,\mathrm{X}_{\mathrm{N}}\right)^{\mathrm{T}}$, $\mathbf{W}_{\mathrm{t}} = \left(\mathrm{W}_{\mathrm{t1}},\cdots,\mathrm{W}_{\mathrm{tk}},\cdots,\mathrm{W}_{\mathrm{tN}}\right)^{\mathrm{T}}$, and $\hat{\mathbf{V}} = \left(\hat{\mathrm{V}}_1,\cdots,\hat{\mathrm{V}}_{\mathrm{k}},\cdots,\hat{\mathrm{V}}_{\mathrm{N}}\right)^{\mathrm{T}}$ are transmit signal vector, transmit equalizer weight vector, and virtual noise vector, respectively. At the virtual receiver, the signal passing through the receive equalizer is given as

$$\hat{\mathbf{Z}} = \hat{\mathbf{W}}_{\mathrm{r}}^{\mathrm{d}}\hat{\mathbf{R}} = \hat{\mathbf{W}}_{\mathrm{r}}^{\mathrm{d}}\left(\hat{\mathbf{H}}^{\mathrm{d}}\mathbf{W}_{\mathrm{t}}^{\mathrm{d}}\mathbf{X} + \hat{\mathbf{V}}\right) \tag{2}$$

where $\hat{\mathbf{Z}} = \left(\hat{\mathrm{Z}}_1,\cdots,\hat{\mathrm{Z}}_{\mathrm{k}},\cdots,\hat{\mathrm{Z}}_{\mathrm{N}}\right)^{\mathrm{T}}$ and $\hat{\mathbf{W}}_{\mathrm{d}}^{\mathrm{r}} = \mathrm{diag}\{\hat{\mathbf{W}}\}$. $\hat{\mathbf{W}}_{\mathrm{r}} = \left(\hat{\mathrm{W}}_{\mathrm{r1}},\cdots,\hat{\mathrm{W}}_{\mathrm{rk}},\cdots,\hat{\mathrm{W}}_{\mathrm{rN}}\right)^{\mathrm{T}}$ denotes virtual receive equalizer weight vector. The error function is defined as

$$\mathrm{E} = \sum_{\mathrm{k}=1}^{\mathrm{N}}\left|\mathrm{e}_{\mathrm{k}}\right|^2 = \sum_{\mathrm{k}=1}^{\mathrm{N}}\left|\mathrm{X}_{\mathrm{k}} - \hat{\mathrm{Z}}_{\mathrm{k}}\right|^2 \tag{3}$$

In Eq.(3), the error signal corresponding to the k-th frequency component is defined as

$$\mathrm{e}_{\mathrm{k}} = \mathrm{X}_{\mathrm{k}} - \hat{\mathrm{Z}}_{\mathrm{k}} = \mathrm{X}_{\mathrm{k}} - \hat{\mathrm{W}}_{\mathrm{rk}}\left(\hat{\mathrm{H}}_{\mathrm{k}}\mathrm{W}_{\mathrm{rk}}\mathrm{X}_{\mathrm{k}} + \hat{\mathrm{V}}_{\mathrm{k}}\right) \tag{4}$$

where X_k and $\hat{V}_k$ denote the k-th frequency component of transmit signal vector and virtual noise vector, respectively. H_k is the k-th component of frequency transfer function. W_{tk} and $\hat{W}_{rk}$ are the k-th weight of transmit equalizer and virtual receive equalizer, respectively.

The objective is to minimize the error function E in Eq.(3). From Eq.(4), frequency-domain square error signal is given as $|e_k|^2$. In order to minimize E with respect to both W_{tk} and $\hat{W}_{rk}$, derivations of $|e_k|^2$ with respect to both W_{tk} and $\hat{W}_{rk}$ are set to be zero. To solve this problem, we employ the following iterative algorithm which simultaneously updates two weight vector $\mathbf{W}_t$ and $\hat{\mathbf{W}}_r$, respectively.

First, we try to find the transmit equalizer weight W_{tk} by minimizing the square error of $|e_k|^2$ in Eq.(4) with respect to W_{tk}. The square error signal e_k corresponding to the k-th frequency component is given as

$$\left|e_k\right|^2 = \left|X_k - W_{rk}(H_k W_{tk} X_k + V_k)\right|^2 \tag{5}$$

The derivative of Eq.(4) with respect to W_{tk} is given as

$$\nabla_{W_{tk}}\left|e_k\right|^2 = -2(\hat{W}_{rk}\hat{H}_k X_k)^* e_k \tag{6}$$

where $\nabla_w = \frac{\partial}{\partial w_x} - j\frac{\partial}{\partial w_y}$, ($w=w_x+jw_y$). Thus, the recursive equation for iterative algorithm is obtained as

$$W_{tk}(n+1) = W_{tk}(n) + 2\mu(\hat{W}_{rk}\hat{H}_k X_k)^* E \tag{7}$$

where μ is step size to adjust convergence speed. By extending the above discussion to the vector expression, we can obtain

$$\mathbf{W}_t(n+1) = \mathbf{W}_t(n) + 2\mu(\hat{\mathbf{W}}_r^d \hat{\mathbf{H}}^d \mathbf{X})^* E \tag{8}$$

where the norm of the transmit equalizer weight vector is normalized to a constant value in order to keep the total average transmit power to be constant. Similarly, the receive equalizer weight vector is determined by minimizing the square error of $\mathbf{e}$ in Eq.(4) with respect to $\hat{\mathbf{W}}_r$. The error signal vector is defined as

$$\mathbf{e} = \mathbf{X} - \hat{\mathbf{W}}_r^d \hat{\mathbf{R}} \tag{9}$$

where $\hat{\mathbf{R}} = \left(\hat{R}_1,\cdots,\hat{R}_k,\cdots,\hat{R}_N\right)^T$ is the receive signal vector at the virtual receiver. Thus, the square error of the k-th component of the error vector is given as

$$\left|e_k\right|^2 = \left|X_k - \hat{W}_{rk}\hat{R}_k\right|^2 \tag{10}$$

By differentiating Eq.(10) with respect to the k-th weight $\hat{W}_{rk}$, we can obtain

$$\nabla_{\hat{W}_{rk}}\left|e_k\right|^2 = -2\hat{R}_k^* e_k \tag{11}$$

Thus, the updating equation for the receive weight vector $\hat{\mathbf{W}}_r$ at the n-th iteration is given as

$$\mathbf{W}_r(n+1) = \mathbf{W}_r(n) + 2\mu\left(\hat{\mathbf{R}}\right)^* E \tag{12}$$

With the proposed iterative algorithm, transmit and receive equalizer weights $\mathbf{W}_t$ and $\hat{\mathbf{W}}_r$ are determined based on recursive equations in Eqs.(8) and (12) until weight vectors are optimized, i.e., the square error of $\mathbf{e}$ is minimized. After optimum weights are determined, the signal weighted by $\mathbf{W}_t$ is transmitted to radio channel. At the receiver, the receive equalizer weight vector $\hat{\mathbf{W}}_r$ is determined by observing a received pilot sequence.

2.1 Extension to decision-feedback equalization case

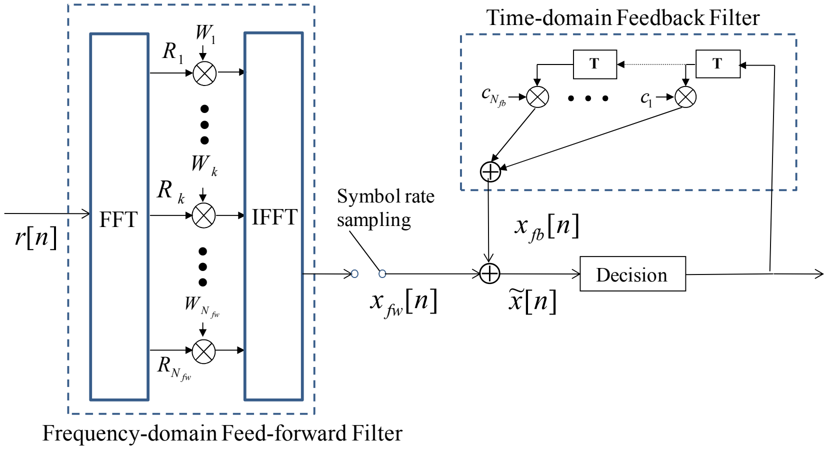


Fig. 2. Block diagram of the proposed system with decision feedback equalizer, where virtual channel and virtual receiver are equipped at the transmitter.

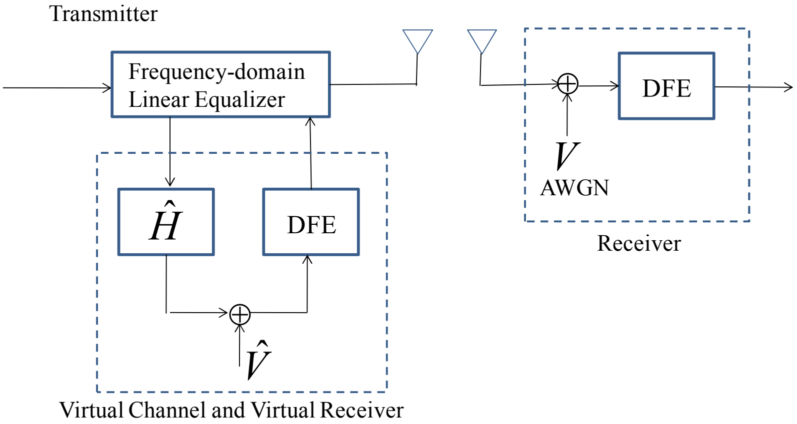


Fig. 3. Decision feedback equalizer with frequency-domain feedforward filter.

The proposed method is extendable to nonlinear equalizers such as decision feedback equalizer (DFE) which offers better equalization performance than linear ones. Block diagram of the system is the same as that in Figure 1 except that DFE is used at both virtual and real receivers. The detailed block diagram of DFE with frequency-domain feed-forward filter and time-domain feedback filter is shown in Figure 3, where N_{fw} and N_{fb} denote the number of taps in frequency-domain feed-forward filter and time-domain feedback filter, respectively. $x_{fw}[n]$ and $x_{fb}[n]$ denote the output signals of feed-forward and feedback filters at n-th time instant. c_l is the *l*-th tap coefficient of feedback filter. In Figure 3, DFE output signal for the n-th symbol is given as

$$\tilde{x}[n] = x_{fw}[n] + x_{fb}[n] \tag{13}$$

where feedback filter output $x_{fb}[n]$ is expressed as

$$x_{fb}[n] = \sum_{l=1}^{N_{fb}} c_l \hat{x}_{(n-1)}, \qquad (n > N_{fb}) \tag{14}$$

where $\hat{x}_{(n-1)}$ denotes the n-l th decided symbol. If a known sequence is used to determine tap coefficients, the decided symbol $\hat{x}_{(n-1)}$ in Eq.(14) is replaced with known symbol $x_{(n-l)}$. Hence, assuming a known training sequence, feedback filter output $x_{fb}[n]$ is expressed as

$$x_{fb}[n] = \sum_{l=1}^{N_{fb}} c_l x_{(n-1)}, \qquad (n > N_{fb}) \tag{15}$$

By taking Fourier transform of Eq.(15), we can obtain frequency-domain expression of the feedback filter output; the k-th frequency component of feedback signal is expressed as

$$x_{fb}[n] = \sum_{l=1}^{N_{fb}} c_l X_k e^{-j2\frac{\pi kl}{N}} = \sum_{l=1}^{N_{fb}} c_l Q_{k,l}, \qquad (n > N_{fb}) \tag{16}$$

where $Q_{k,l} \equiv X_k e^{-j2\frac{\pi kl}{N}}$ denote the k-th frequency component of lT time-delayed signal in feedback filter.

To determine optimum weight vectors of transmit and receive equalizers, we calculate the error signal at the virtual receiver. Similarly to the previous discussion, the square error signal corresponding to the k-th frequency component at the virtual receiver is defined as

$$|e_k|^2 = \left| X_k - \hat{X}_{fw,k} - \hat{X}_{fb,k} \right|^2 \tag{17}$$

where the k-th frequency component of the virtual feed-forward filter output $\hat{X}_{fw,k}$ is expressed as

$$\hat{X}_{fw,k} = \hat{W}_{rk} \hat{R}_k = \hat{W}_{rk} \left(\hat{H}_k W_{tk} X_k + \hat{V}_k \right) \tag{18}$$

where $\hat{R}_k = \hat{H}_k W_{tk} X_k + \hat{V}_k$ is the k-th frequency component of the received signal at the virtual receiver. The derivative of Eq.(18) with respect to W_{tk} is given as

$$\nabla_{W_{tk}} |e_k|^2 = 2 \left(\hat{W}_{rk} \hat{H}_k X_k \right)^* e_k \tag{19}$$

Thus, the recursive equation for iterative algorithm is given as

$$W_{tk}(n+1) = W_{tk}(n) + 2\mu \left(\hat{W}_{rk} \hat{H}_k X_k \right)^* E \tag{20}$$

Similarly, the recursive equations for $\hat{W}_{rk}$ and $\hat{c}_l$ are obtained by taking derivative of Eq.(17) with respect to $\hat{W}_{rk}$ and $\hat{c}_l$, respectively;

$$\hat{W}_{rk}(n+1) = \hat{W}_{rk}(n) + 2\mu \left(\hat{R}_k \right)^* E \tag{21}$$

$$\hat{c}_l(n+1) = \hat{c}_l(n) + 2\mu \left(\hat{\mathbf{Q}}_l \right)^* \mathbf{e} \tag{22}$$

where $\hat{\mathbf{Q}}_l = \left(\hat{Q}_{l1},\cdots,\hat{Q}_{lk},\cdots,\hat{Q}_{lN}\right)$ is the frequency-domain vector expression of lT time-delayed feedback signal, $\mathbf{e} = \left(e_1,\cdots,e_k,\cdots,e_N\right)^T$ is the error signal vector, and μ is a step size. By extending the above equations to the vector expression, we can obtain

$$\mathbf{W}_t(n+1) = \mathbf{W}_t(n) + 2\mu(\hat{\mathbf{W}}_r^d \hat{\mathbf{H}}^d \mathbf{X})^* E \tag{23}$$

$$\mathbf{W}_r(n+1) = \mathbf{W}_r(n) + 2\mu\left(\hat{\mathbf{R}}\right)^* E \tag{24}$$

$$\hat{\mathbf{c}}(n+1) = \hat{\mathbf{c}}(n) + 2\mu\left(\hat{\mathbf{Q}}\right)^* \mathbf{e} \tag{25}$$

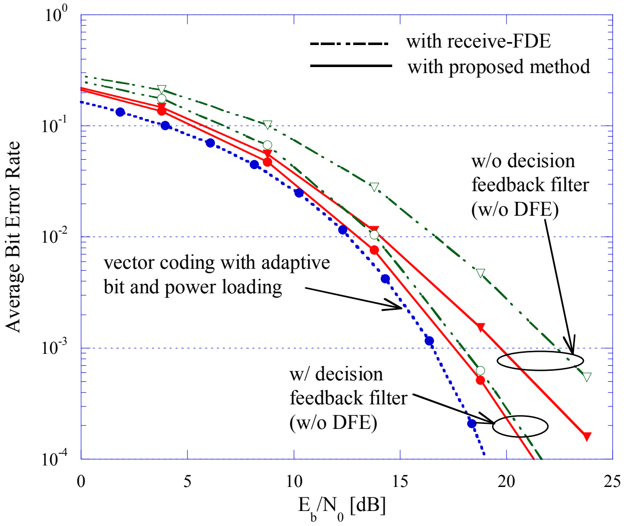
where $\hat{\mathbf{c}} = \left(\hat{c}_1,\cdots,\hat{c}_k,\cdots,\hat{c}_{N_{fb}}\right)$ denotes the feedback tap vector of virtual DFE and $\hat{\mathbf{Q}} = \left(\hat{\mathbf{Q}}_1,\cdots,\hat{\mathbf{Q}}_k,\cdots,\hat{\mathbf{Q}}_{N_{fb}}\right)^T$ denotes the feedback signal matrix of virtual DFE.

3. Performance evaluation

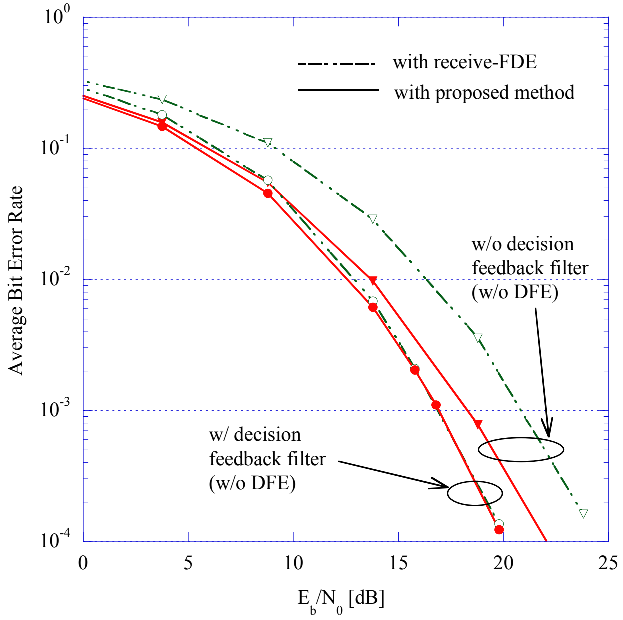
Performance of a SC system using transmit/receive equalization is evaluated by computer simulation. System block diagram is the same as that in Figure 1. QPSK modulation is adopted. A square root of raised cosine filtering with a roll-off factor of α=0.2 is employed. Propagation model is attenuated 6-path quasistatic Rayleigh fading. Block length for FDE is set to 128 symbols. Guard interval whose length is 16 symbols is inserted into every blocks to eliminate inter-block interference. Additive white Gaussian noise (AWGN) is added at the receiver. For simplicity, it is assumed that frequency channel transfer function is known to both transmitter and the receiver. Transmit/receive equalizer weights are determined with least mean square (LMS) algorithm, where sufficient number of training symbols is assumed for simplicity.

In this study, we also evaluate BER performance of *vector coding* (VC) transmission in SISO channel. The basic concept of VC is the same as that of E-SDM in MIMO system; eigenvectors of channel autocorrelation matrix is used for weight matrices of transmit and receive filters. Therefore, data streams are transmitted through multiple eigenpath channels between transmit and receive filters. To minimize the average BER in VC system, adaptive bit and power loading based on BER minimization criterion is adopted; the bit allocation pattern which minimize the average BER is selected among possible bit allocation patterns under constraint of a constant transmit power and a constant data rate, where modulation scheme is selected among QPSK, 16QAM, and 64QAM according to each eigenpath channel condition. Consequently, provided that CSI is known to the transmitter, the minimum average BER in SISO channel is achieved by VC transmission with adaptive bit and power loading.

Figure 4 shows BER performance of the SC system using the proposed method in attenuated 6-path quasistatic Rayleigh fading, where normalized delay spread values of τ/T are τ/T=0.769 and 2.69 for Figs(a) and (b), respectively. T is symbol duration. DFE is employed for both the proposed and conventional systems, where the number of feedback taps in DFE is set to 3. For comparison purpose, BER performance of the SC system using the conventional receive FDE with and without decision-feedback filter is also shown. BER performance of VC with adaptive bit and power loading is also shown. In Figure 4, in case of linear transmit/receive equalization (i.e., without decision-feedback filter), BER performance of the SC systems using the proposed method is improved by about 2.7dB at BER=10^{-3} compared to the case of the conventional receive FDE.



(a) τ/T=0.769



(b) τ/T=2.69

Fig. 4. BER performance of the SC system using the proposed method as a function of E_b/N_0, where normalized delay spread values in figures (a) and (b) are τ/T=0.769 and τ/T=2.69.

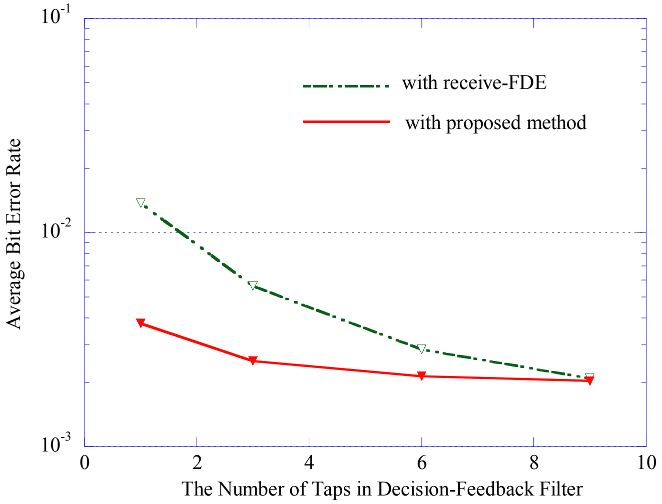


Fig. 5. BER performance of the SC system using the proposed method as a function of the number of feedback taps in decision-feedback filter, where E_b/N_0=15.8dB and normalized delay spread is τ/T=2.69.

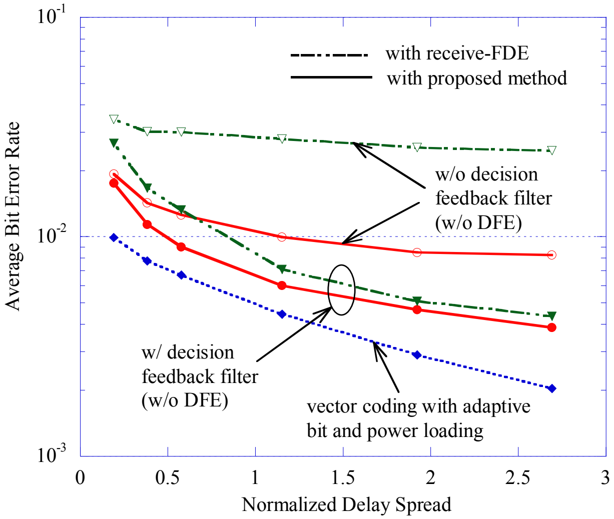


Fig. 6. BER performance of the SC system using the proposed method as a function of normalized delay spread, where E_b/N_0 is set to 13.8dB.

When decision-feedback filter is adopted in both systems, the proposed system achieves better BER performance than case using the conventional one in lower E_b/N_0 region. On the other hand, in higher E_b/N_0 region, it can be seen that BER performance of the proposed system becomes close to that of the conventional one as E_b/N_0 increases. In addition, it can be seen that difference between the proposed method and VC with adaptive bit and power loading in BER performance is about 2.3dB at BER=10^{-4}.

Figure 5 shows BER performance of the proposed and conventional SC systems with decision-feedback filter as a function of the number of feedback taps N_{fb}, where E_b/N_0=15.8dB and normalized delay spread τ/T is 2.69. The maximum delay time difference between the first path and last path is set to 8.75T. In general, the required number of taps in decision feedback filter is 9 to suppress intersymbol interference. In Figure 5, both the proposed and conventional systems achieves almost the same BER performance when N_{fb} is set to 9, because the number of feedback taps N_{fb}=9 is sufficient for suppressing ISI. From this figure, when the number of feedback taps N_{fb} is less than 9taps, it can be seen that the proposed system achieves better BER performance than the conventional system using the receive FDE. This result means that the required number of feedback taps in the proposed system is less than that in the conventional one.

Figure 6 shows BER performance of the SC systems using the proposed method as a function of normalized delay spread, where E_b/N_0=13.8dB is assumed. BER performance of the SC systems using the conventional receive equalization with and without decision-feedback filter is also shown. For case with DFE, the sufficient number of feedback taps is used for various delay spread channel. From this figure, it can be seen that the SC system with transmit/receive DFE with decision-feedback filter using the proposed algorithm achieves better BER performance than those using the receive FDE for various delay spread conditions.

4. Conclusion

An iterative optimization method of transmit/receive frequency domain equalization (FDE) was proposed for single carrier transmission systems, where both transmit and receive FDE weights are iteratively determined with a recursive algorithm so as to minimize the error signal at a virtual receiver. With computer simulation, it is confirmed that the proposed transmit/receive equalization method achieves better BER performance than that of the system using the conventional ones.

5. References

D. Falconer; S. L. Ariyavisitakul; A.Benyamin-Seeyar & B. Eidson (2002). *Frequency Domain Equalization for Single Carrier Broadband Wireless Systems*, IEEE Commun. Magazine, pp. 58-66.

F. Adachi; D. Garg; S. Takaoka & K. Takeda (2005). *Broadband CDMA Technique, IEEE Wireless Communications*, no. 4, pp. 8-18.

IEEE Std 802.16e/D9 (2005). *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*.

J. Chuang and N. Sollenberger, *Beyond 3G (2000): wideband wireless data access based on OFDM and dynamic packet assignment*, IEEE Commun. Mag., vol. 38, no. 7, pp. 78-87.

K. Ban, et al. (2000), *Joint optimization of transmitter/receiver with multiple transmit/receive antennas in band-limited channels*, IEICE Trans. Commun., vol. E83-B, no. 8, pp. 1697-1703.

S. Kasturia (1990), *Vector coding for partial response channels*, IEEE Trans. Infor. Theory, vol. 36, no. 4.

R. van Nee & R. Prasad, OFDM *for wireless multimedia communications*, Artech House

Y. Akaiwa, *Introduction to digital mobile communication*, John Wiley & Sons, Inc.

An Enhanced Iterative Flipping PTS Technique for PAPR Reduction of OFDM Signals

Byung Moo Lee[1] and Rui J. P. de Figueiredo[2]
[1]*Central R&D Laboratory, Korea Telecom (KT),*
Seoul, 137-792, Korea,
Email:blee@kt.com
[2]*Laboratory for Intelligent Signal Processing and Communications,*
Department of Electrical Engineering and Computer Science,
University of California, Irvine, CA 92697-2625, USA,
Email:rui@uci.edu

1. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has several desirable attributes, such as high immunity to inter-symbol interference, robustness with respect to multi-path fading, and ability for high data rates, all of which are making OFDM to be incorporated in wireless standards like IEEE 802.11a/g/n WLAN and ETSI terrestrial broadcasting. However one of the major problems posed by OFDM is its high Peak-to-Average-Power Ratio (PAPR), which seriously limits the power efficiency of the transmitter's High Power Amplifier (HPA). This is because PAPR forces the HPA to operate beyond its linear range with a consequent nonlinear distortion in the transmitted signal.

One of good solutions to mitigate this nonlinear distortion is put a Pre-Distorter before the High Power Amplifier and increase linear dynamic range up to a saturation region (1) (2) (3). However, the main disadvantage of Pre-Distorter technique is that these PD techniques only work in a limited range, that is, up to the saturation region of the amplifier. In this situation, Peak-to-Average Power Ratio (PAPR) reduction techniques which pull down high PAPR of OFDM signal to an acceptable range can be a good complementary solution. Due to practical importance of this, there are various PAPR reduction techniques for OFDM signals (4) (5) (6) (7) (8) (9). Among them, the PTS (Partial Transmit Sequence) technique is very promising because it does not give rise to any signal distortion (9). However, its high complexity makes it difficult to use in a practical system. To solve the complexity problem of the PTS technique, Cimini and Sollenberger proposed an iterative flipping algorithm (10). Even though the iterative flipping algorithm greatly reduces the complexity of the PTS technique, there is still some performance gap between the ordinary PTS and the iterative flipping algorithm.

In this chapter, we propose an enhanced version of the iterative flipping algorithm to reduce the performance gap between the iterative flipping algorithm and the ordinary PTS technique. In the proposed algorithm, there is an adjustable parameter to allow a performance/complexity trade-off.

2. OFDM and Peak-to-Average Power Ratio (PAPR)

An OFDM signal of N subcarriers can be represented as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] e^{j2\pi f_k t}, \; 0 \le t \le T_s \tag{1}$$

where T_s is the duration of the OFDM signal and $f_k = \frac{k}{T_s}$.
The high PAPR of the OFDM signal arises from the summation in the above IDFT expression. The PAPR of the OFDM signal in the analog domain can be represented as

$$PAPR_c = \frac{\max_{0 \le t \le T_s} |x(t)|^2}{E(|x(t)|^2)} \tag{2}$$

Nonlinear distortion in HPA occurs in the analog domain, but most of the signal processing for PAPR reduction is performed in the digital domain. The PAPR of digital domain is not necessarily the same as the PAPR in the analog domain. However, in some literature (11) (12) (13) , it is shown that one can closely approximate the PAPR in the analog domain by oversampling the signal in the digital domain. Usually, an oversampling factor $L = 4$ is sufficient to satisfactorily approximate the PAPR in the analog domain. For these reasons, we express PAPR of the OFDM signal as follows.

$$PAPR = \frac{\max_{0 \le n \le LN} |x(n)|^2}{E(|x(n)|^2)} \tag{3}$$

3. Existing PTS Techniques

The PTS technique is a powerful PAPR reduction technique first proposed by Muller and Huber in (9). Thereafter various related papers have been published. In this section, we show two representative PTS techniques, the original PTS technique and Cimini and Sollenberger's iterative flipping technique (10).

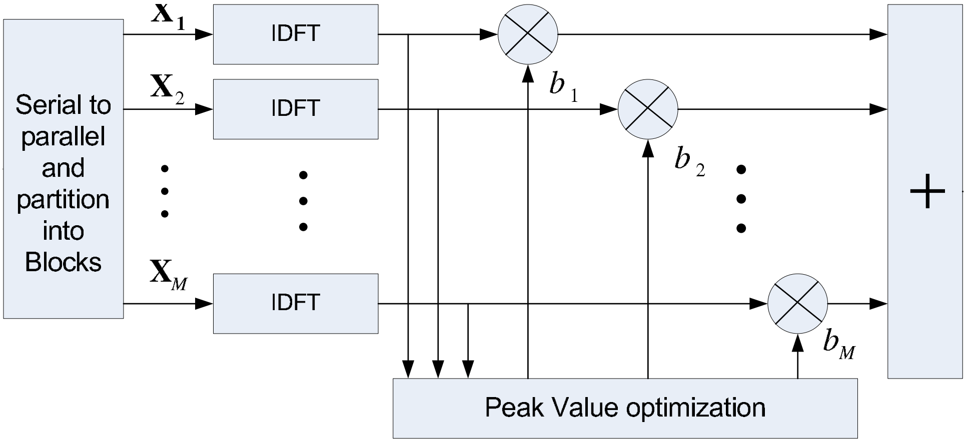


Fig. 1. Block diagram of the PTS scheme

3.1 Ordinary PTS Technique

A block diagram of the PTS technique is shown in Figure 1. The algorithm of the original PTS technique can be explained as follows.
First, the signal vector is partitioned into M disjoint subblocks which can be represented as

$$\mathbf{X_m} = [X_{m,0},\ X_{m,1},\ \cdots,\ X_{m,N-1}]^T,\ m = 1,\ 2,\ \cdots,\ M \tag{4}$$

All the subcarrier positions which are presented in other subblocks must be zero so that the sum of all the subblocks constitutes the original signal, i.e,

$$\sum_{m=1}^{M} \mathbf{X_m} = \mathbf{X} \tag{5}$$

Each subblock is converted through IDFT into an OFDM signal $\mathbf{x_m}$ with oversampling, which can be represented as

$$\mathbf{x_m} = [x_{m,0},\ x_{m,1},\ \cdots,\ x_{m,NL-1}]^T,\ m = 1,\ 2,\ \cdots,\ M \tag{6}$$

where L is the oversampling factor. After that, each subblock is multiplied by a different phase factor b_m to reduce PAPR of the OFDM signal. The phase set can be represented as

$$P = \{e^{j2\pi w/W} | w = 0,\ 1,\ \cdots,\ W-1\} \tag{7}$$

where W is the number of phases.
Because of the high computational complexity of the PTS technique, one generally uses only a few phase factors. The choice, $b_m \in \{\pm 1,\ \pm j\}$, is very interesting since actually no multiplication is performed to rotate the phase (14). The peak value optimization block in Figure 1 iteratively searches the optimal phase sequence which shows minimum PAPR. Finding optimal PAPR using PTS PAPR reduction technique can be represented as

$$PAPR_{optimal} = \frac{\min\limits_{b_1,\cdots b_M}\left(\max_{0\le n\le LN}\left|\sum_{m=1}^{M} b_m x_{m,n}\right|^2\right)}{E(|x(n)|^2)} \tag{8}$$

This process usually requires large computational power. After finding the optimal phase sequence which minimizes PAPR of the OFDM signal, all the subblocks are summed as in the last block of Figure 1 with multiplication of the optimal phase sequence. Then the transmit sequence can be represented as

$$\begin{aligned}\mathbf{x}'(\mathbf{b}) &= [\mathbf{x_1},\ \mathbf{x_2},\ \cdots,\ \mathbf{x_M}]\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} \\ &= \sum_{m=1}^{M} b_m \cdot \mathbf{x_m}\end{aligned} \tag{9}$$

Here we assume $\mathbf{b^T} = [b_1\ \ b_2\ \ \cdots\ \ b_M]$ is an optimal phase set which gives minimum PAPR among various phase sets.

3.2 Iterative Flipping PTS Technique

Cimini and Sollenberger's iterative flipping technique is developed as a sub-optimal technique for the PTS algorithm. In their original paper (10), they only use binary weighting factors. That is $b_m = 1$ or $b_m = -1$. These can be expanded to more phase factors. The algorithm is as follows. After dividing the data block into M disjoint subblocks, one assumes that $b_m = 1,\ (m = 1, 2, \cdots, M)$ for all of subblocks and calculates PAPR of the OFDM signal. Then one changes the sign of the first subblock phase factor from 1 to -1 $(b_1 = -1)$, and calculates the PAPR of the signal again. If the PAPR of the previously calculated signal is larger than that of the current signal, keep $b_1 = -1$. Otherwise, revert to the previous phase factor, $b_1 = 1$. Suppose one chooses $b_1 = -1$. Then the first phase factor is decided, and thus kept fixed for the remaining part of the algorithm. Next, we follow the same procedure for the second subblock. Since one assumed all of the phase factors were 1, in the second subblock, one also changes $b_2 = 1$ to $b_2 = -1$, and calculates the PAPR of the OFDM signal. If the PAPR of the previously calculated signal is larger than that of the current signal, keep $b_2 = -1$. Otherwise, revert to the previous phase factor, $b_2 = 1$. This means the procedure with the second subblock is the same as that with the first subblock. One continues performing this procedure iteratively until one reaches the end of subblocks (M^{th} subblock and phase factor b_M). A similar technique was also proposed by Jayalath and Tellambura (16). The difference between the Jayalath and Tellambura's technique and that of Cimini and Sollenberger is that, in the former, the flipping procedure does not necessarily go to the end of subblocks (M^{th} block). To reduce computational complexity, the flipping is stopped before the end of the entire procedure if the desired PAPR OFDM signal achieved at that point.

4. Enhanced Iterative Flipping PTS Technique

In this section, we present an Enhanced Iterative Flipping PTS (defined by EIF-PTS) technique which is a modified version of the Cimini and Sollenberger's Iterative Flipping PTS (IF-PTS) technique. We use, in this chapter, 4 phase factors to reduce the PAPR of the OFDM signal, that is, $W = 4\ \ (b_m \in \{\pm 1,\ \pm j\})$.

As explained earlier, in the iterative flipping algorithm, one keeps only one phase set in each subblock. Even though the phase set chosen in the first subblock shows minimum in the first subblock, that is not necessarily minimum if we allow it to change until we continue the procedure up to the end subblock. The basic idea of our proposed algorithm is that we keep more phase factors in the first subblock rather than keep only one phase factor, and delay the final decision to the end of subblock. We can choose the number of phase factors that we will keep by adjusting a parameter, S where S is the number of phase factors which we will keep in the first subblock. The larger S, the better performance we get but with higher complexity.

The basic structure of the Enhanced Iterative Flipping Partial Transmit Sequence (EIF-PTS) is illustrated in Figure 2, for the case in which $S = W = 4$. In this illustration, each of four phases $b_{11} = 1, b_{12} = -1, b_{13} = j, b_{14} = -j$ is multiplied successively by the first subblock of the signal thus generating four phase sequences, S_1, S_2, S_3 and S_4. Then for each S_i, from the second subblock, the IF (Iterative Flipping) algorithm of Cimini and Sollenberger is performed. At the end of application of this procedure up to the end subblock for respectively S_1, S_2, S_3 and S_4, there will be four sequences $\tilde{S}_1, \tilde{S}_2, \tilde{S}_3$ and $\tilde{S}_4$, each having respectively b_{1i} for the first sbublock of $\tilde{S}_i$, and different phases generated by the application of the IF procedure to each of the four sequences. At the conclusion of this procedure, the EIF-PTS algorithm chooses the $\tilde{S}_i, i = 1, 2, 3, 4$ which gives rise to the lowest PAPR. For the clarity, we provide an example in Table 1, Table 2 and Table 3.

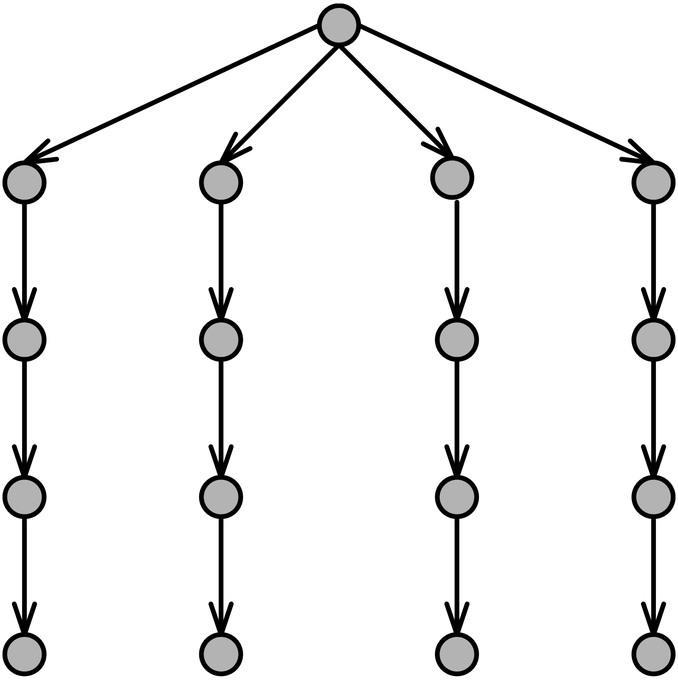


Fig. 2. Structure of an Enhanced iterative flipping algorithm ($S = 4$)

In summary, we perform following procedure to efficiently improve the iterative flipping algorithm.

1. Choose the parameter, S to decide how many phase factors we will keep in the first subblock depending on the performance/complexity, where $1 \leq S \leq W$.
2. Keep the S phase sequences which show minimum PAPRs in the first subblock.
3. From each node which was kept in the first subblock, do iterative flipping algorithm until you reach the end of subblock.
4. At the end of subblock, find the phase sequence and signal which show minimum PAPR and choose it as a final decision.

It is also worth noting that when $S = 1$, the proposed algorithm is equivalent to the iterative flipping algorithm.

5. Simulation Results and Discussion

In this section, we show simulation results of the proposed EIF (Enhanced Iterative Flipping) PTS algorithm. We use 16QAM OFDM with $N = 64$ subcarriers. We divide the one signal block as $M = 4$ adjacent/disjoint subblocks and use $W = 4$ $(b_m \in \{\pm 1, \pm j\})$ phase factors. We oversampled the data by $L = 4$ to estimate PAPR of the continuous time signal. The first simulation result is shown in Figure 3. In this figure, the x-axis denotes PAPR value in dB scale while the y-axis, the respective Complementary Cumulative Distribution Function (CCDF) or

Given:	• The number of subblocks, $M = 4$. • 4 phase factors, $b_{11} = 1, b_{12} = -1, b_{13} = j, b_{14} = -j$.
Step 0:	• Choose $S = 2$.
Step I-a:	• Complete PAPR for four sequences S_1, S_2, S_3, and S_4, each multiplied respectively by the respective phase factor to the first subblock. The phases for successive blocks are indicated below. $\begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \\ 1 & -1 & j & -j \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array}$ (10)
Step I-b:	• Choose 2 sequences corresponding to the lowest PAPR. Assume they are S_2 and S_3, so we have $\begin{array}{cc} S_2 & S_3 \\ -1 & j \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{array}$ (11)

Table 1. Example of EIF-PTS technique ($S = 2$) (1)

clipping probability. As we can see in Figure 3, the proposed algorithm reduces the PAPR of the OFDM signal by more than 2 dB at the 0.1% of CCDF. The performance degradation between the EIF-PTS and ordinary PTS is only less than 0.5dB. The complexity of ordinary PTS can be represented as

$$\text{The number of iterations of ordinary PTS} = W^{(M-1)} \tag{17}$$

In this chapter, we assume the complexity is only dependent on the number of iterations. The reason, for the number of iterations of ordinary PTS is W^{M-1}, and not W^M is that ordinary PTS can fix the phase factor of the first subblock without any performance penalty. The complexity

Step II-a:	• From now on we use the Cimini-Sollenberger procedure with the first element of S_2 and S_3 kept fixed. • Form sequences. $\begin{array}{cccccccc} S_{21} & S_{22} & S_{23} & S_{24} & S_{31} & S_{32} & S_{33} & S_{34} \\ -1 & -1 & -1 & -1 & j & j & j & j \\ 1 & -1 & j & -j & 1 & -1 & j & -j \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \qquad (12)$
Step II-b:	• Choose one sequence among S_{21}, S_{22}, S_{23} and S_{24} which has lowest PAPR. Assume that sequence S_{23}. Do the same S_{31}, S_{32}, S_{33} and S_{34}. Assume the with lowest PAPR is S_{31}. $\begin{array}{cc} S_{23} & S_{31} \\ -1 & j \\ j & 1 \\ 1 & 1 \\ 1 & 1 \end{array} \qquad (13)$
Step III-a:	• Form sequences $\begin{array}{cccccccc} S_{231} & S_{232} & S_{233} & S_{234} & S_{311} & S_{312} & S_{313} & S_{314} \\ -1 & -1 & -1 & -1 & j & j & j & j \\ j & j & j & j & 1 & 1 & 1 & 1 \\ 1 & -1 & j & -j & 1 & -1 & j & -j \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \qquad (14)$

Table 2. Example of EIF-PTS technique ($S = 2$) (2)

of the proposed EIF-PTS can be represented as

$$\begin{array}{l} \text{The Number of Iterations of Proposed Algorithm} = \\ W + (W-1)\cdot(M-1)\cdot S \end{array} \qquad (18)$$

We organize complexities of the proposed Enhanced Iterative Flipping (EIF) PTS and ordinary PTS in Table 4. The proposed EIF-PTS algorithm also can fix the first subblock (F-EIF-PTS).

Step III-b:	- Suppose that S_{232} and S_{314} have lowest PAPR.
Step IV-a:	- Form sequences $\begin{array}{cccccccc} S_{2321} & S_{2322} & S_{2323} & S_{2324} & S_{3141} & S_{3142} & S_{3143} & S_{3144} \\ -1 & -1 & -1 & -1 & j & j & j & j \\ j & j & j & j & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -j & -j & -j & -j \\ 1 & -1 & j & -j & 1 & -1 & j & -j \end{array} \qquad (15)$
Step IV-b:	- From the first 4, select the one with lowest PAPR, say S_{2323}. Do the same among the remaining 4. Say it is S_{3142}. Then you have $\begin{array}{cc} S_{2323} & S_{3142} \\ -1 & j \\ j & 1 \\ -1 & -j \\ j & -1 \end{array} \qquad (16)$
Step V:	- Select the one with the lowest PAPR in Step IV. Say it is S_{3142}. - Final solution, $S_{3142} = \{j, 1, -j, -1\}$

Table 3. Example of EIF-PTS technique ($S = 2$) (3)

	Number of iterations
EIF-PTS, $S = 1$	13
EIF-PTS, $S = 2$	22
EIF-PTS, $S = 3$	31
EIF-PTS, $S = 4$	40
Ordinary PTS	64

Table 4. Comparison of complexities between EIF-PTS and Ordinary PTS

However, in this case, we get some performance penalty. The simulation results and comparison of complexity of this case is in Figure 4 and Table 5. It is obvious that, in this case, we can

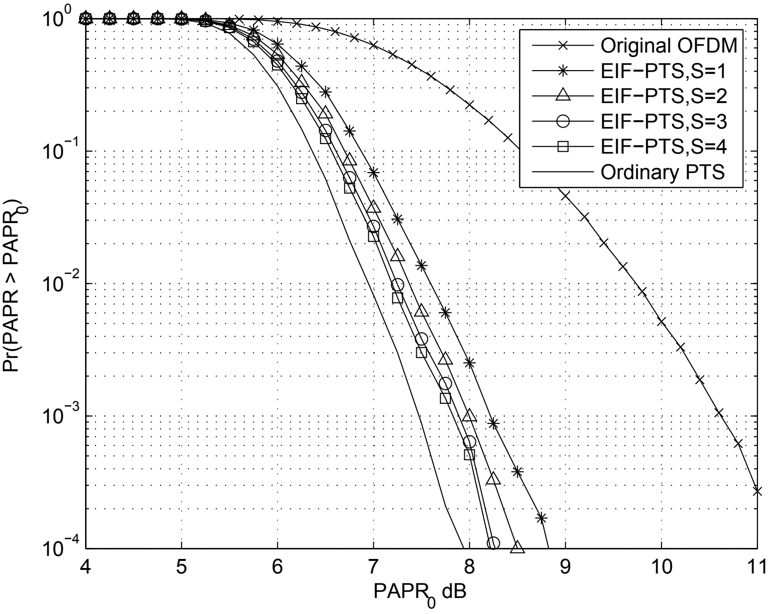


Fig. 3. Performance of EIF-PTS, $M = 4$

represent the number of iterations as follows.

$$\begin{aligned}&\text{The Number of Iterations of Proposed Algorithm} =\\&W + (W-1)\cdot(M-2)\cdot S\end{aligned} \tag{19}$$

This fixed technique, which we call F-EIF-PTS, is needed if we try to send SI (Side Information) to the receiver. To embed SI, at least one block of phase should not be changed.

	Number of iterations
EIF-PTS, $S = 1$	10
EIF-PTS, $S = 2$	16
EIF-PTS, $S = 3$	22
EIF-PTS, $S = 4$	28
Ordinary PTS	64

Table 5. Comparison of complexities between EIF-PTS and Ordinary PTS, when the first phase factor is fixed as 1

Now we increase the number of subblocks from $M = 4$ to $M = 8$. In Figure 5, as an ordinary PTS approximation, we refer (15).

The performance gap is larger than that of the previous case ($M = 4$). However, the complexity gap is much larger than the performance gap. The comparison of the complexity is provided in Table 6. The performance difference between EIF-PTS, $S = 4$ and ordinary PTS is less

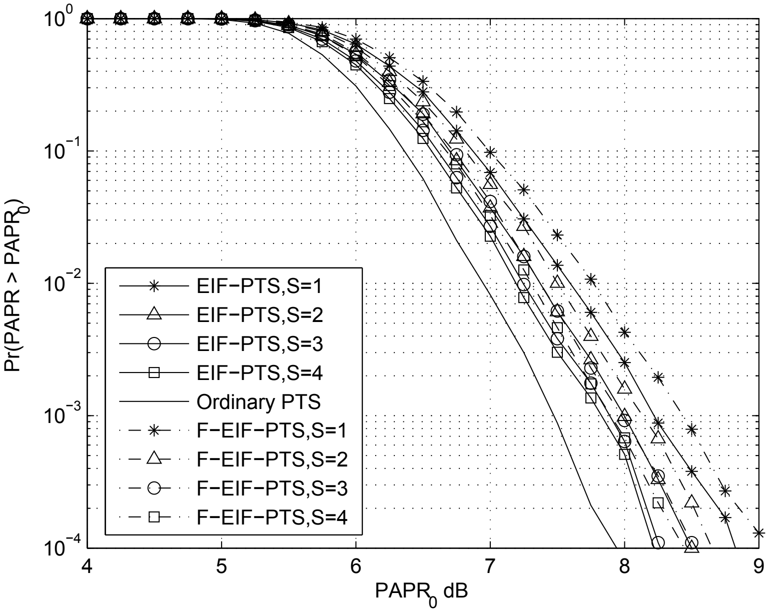


Fig. 4. Performance of EIF-PTS, when fixed the first phase factor (F-EIF-PTS) and normal EIF-PTS case

	Number of iterations
EIF-PTS, $S = 1$	25
EIF-PTS, $S = 2$	46
EIF-PTS, $S = 3$	67
EIF-PTS, $S = 4$	88
Ordinary PTS	16384

Table 6. Comparison of complexities between EIF-PTS and Ordinary PTS, $M = 8$

than 1 dB at 0.1% of CCDF. However, we get this performance with only $88/16384 = 0.54\%$ of computational complexity by using the proposed algorithm. The complexity will be further reduced, if we use the simple adaptive technique which was proposed in (16).

To visualize the increase of complexity, we provide two figures, Figure 6 and Figure 7. We use (17) and (18) to plot Figure 6 and Figure 7. As the number of subblocks or phase factors is increased, the complexity of ordinary PTS technique is increased dramatically. However, the complexity of the proposed EIF-PTS technique is not increased so dramatically when compared with the ordinary PTS technique.

5.1 Power Spectral Density Analysis

In this subsection, we present Power Spectral Density (PSD) analysis of the proposed algorithm. To show spectral leakage, we combine an ideal Pre-Distorter (PD) with High Power

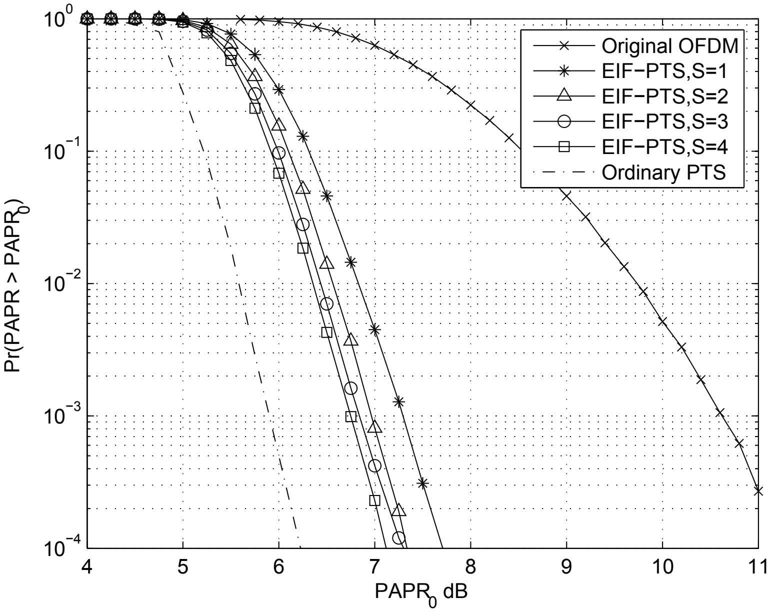


Fig. 5. Performance of EIF-PTS, $M = 8$

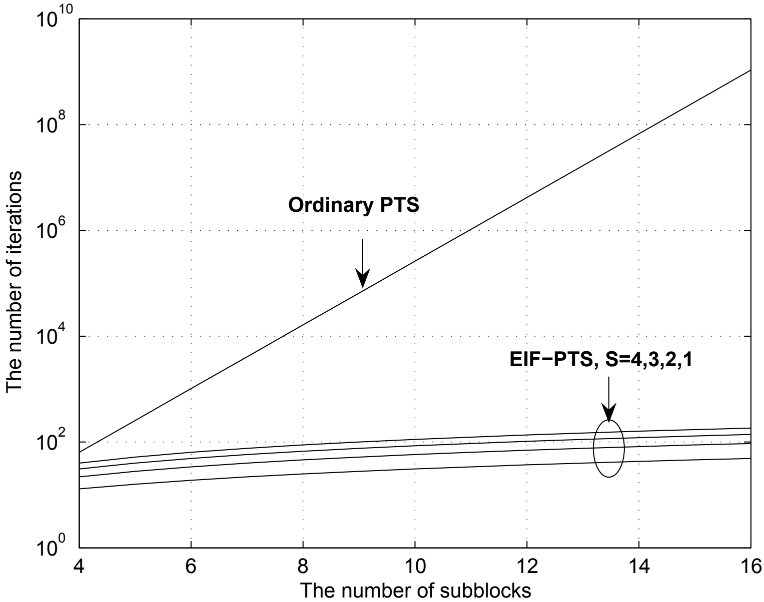


Fig. 6. Comparison of complexities between ordinary PTS and proposed EIT-PTS, $W = 4$

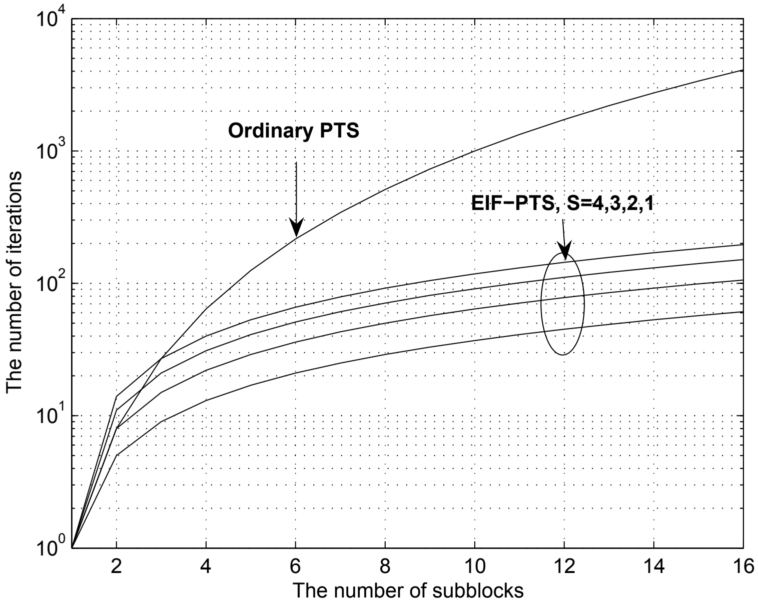


Fig. 7. Comparison of complexities between ordinary PTS and proposed EIT-PTS, $M = 4$

Amplifier (HPA), as we did in (1). Thus the amplitude transfer function becomes a Soft Envelope Limiter (Please refer (1)). That is, the transfer function is linear up to a certain range and beyond that range, the signal is clipped. The spectral leakage is due to this clipping process. In Figure 8, we use $M = 4$ subblocks and set Input Back-Off (IBO) = 8 dB. As we can see, without performing any PAPR reduction technique, even though we use an ideal PD and set high IBO which reduces power efficiency, we cannot avoid large spectral leakage. The proposed EIF-PTS algorithm can significantly reduce the spectral leakage and moreover it can also adjust the performance by adjusting the parameter, S. If we increase the number of subblocks from $M = 4$ to $M = 8$, we can get better performance even though we reduce the IBO from 8 to 7 (Figure 9).

6. Conclusion

One of the major problems associated with OFDM is its high PAPR. In this chapter, we proposed an enhanced version of the iterative flipping algorithm to efficiently reduce the PAPR of OFDM signal. There is an adjustable parameter so that one can choose based on performance/complexity trade-off. Simulation results show that this new technique gives good performance with significantly lower complexity compared with the ordinary PTS technique.

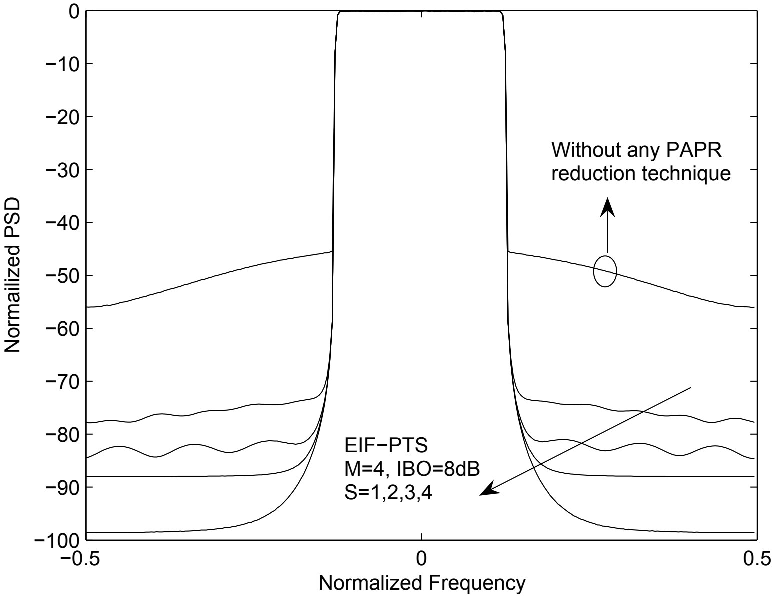


Fig. 8. Power Spectral Density of the EIT-PTS technique, when $IBO = 8dB$, $M = 4$

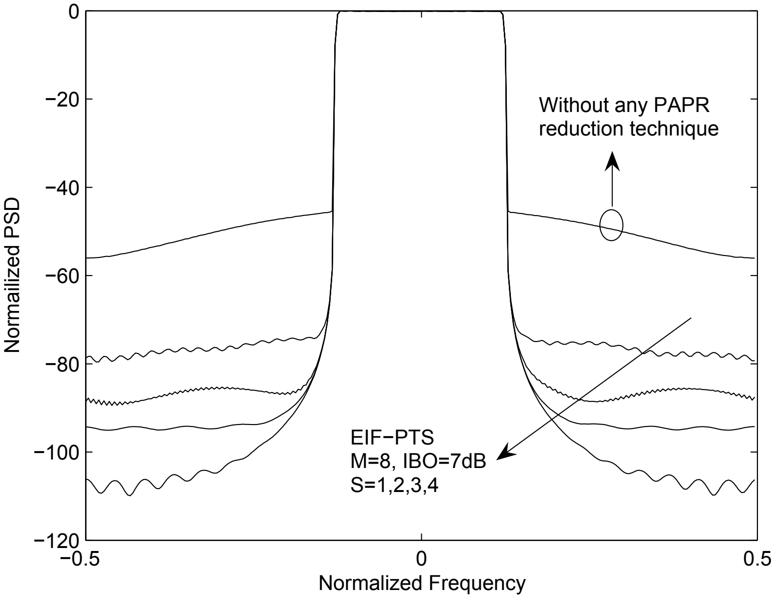


Fig. 9. Power Spectral Density of the EIT-PTS technique, when $IBO = 7dB$, $M = 8$

7. References

[1] Byung Moo Lee and Rui J.P. de Figueiredo, "Adaptive Pre-Distorters for Linearization of High Power Amplifiers in OFDM Wireless Communications," *Circuits, Systems & Signal Processing*, Birkhauser Boston, vol. 25, no. 1, 2006, pp. 59-80.

[2] Byung Moo Lee and Rui J.P. de Figueiredo, "A Tunable Pre-Distorter for Linearization of Solid State Power Amplifier in Mobile Wireless OFDM," *IEEE 7th Emerging Technologies Workshop*, pp. 84-87, St.Petersburg, Russia, June 23 - 24, 2005.

[3] Rui J. P. de Figueiredo and Byung Moo Lee, "A New Pre-Distortion Approach to TWTA Compensation for Wireless OFDM Systems," *2nd. IEEE International Conference on Circuits and Systems for Communications, ICCSC-2004*, Moscow, Russia, No. 130, June 30 - July 2, 2004. (Invited Plenary Lecture).

[4] Y. Kou, W. Lu and A. Antoniou, "New Peak-to-Average Power-Ratio Reduction Algorithms for Multicarrier Communications," *IEEE Transactions on Circuits and Systems I*, Vol. 51, No. 9, September 2004, pp. 1790-1800.

[5] X. Li and L. J. Cimini, "Effect of Clipping and Filtering on the performance of OFDM," *IEEE Communication Letters*, Vol. 2 No. 5, May 1998, pp.131-133.

[6] A.E.Jones, T.A.Wilkinson, and S.K.Barton, "Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission scheme," *Electronics Letters*, vol 30, pp. 2098-2099, December 1994.

[7] J. Tellado and J. M. Cioffi, "Efficient algorithms for reducing PAR in multicarrier systems," *IEEE International Symposium on Information Theory*, Cambridge, MA, pp. 191, Aug. 1998.

[8] V. Tarokh and H. Jafarkhani, "On the computation and reduction of the peak-to-average power ratio in multicarrier communications," *IEEE Trans. Commun.*, vol. 48, pp. 37-44, Jan. 2000.

[9] S. H. Muller and H. B. Huber, "OFDM with reduced peak-to-mean power ratio by optimum combination of partial transmit sequences," *Electronics Letters*, vol. 33, pp. 368-369, Feb. 1997.

[10] L. J. Cimini, Jr. and N.R. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences," *IEEE Communication Letters*, vol. 4, pp. 86-88, Mar. 2000.

[11] R. O'Neil and L. N. Lopes, "Envelop variations and spectral splatter in clipped multicarrier signals," *in Proc. of PIMRC'95*, Sept. 1995 pp. 71-75.

[12] L. Wang and C. Tellambura, "A Simplified Clipping and Filtering Technique for PAR Reduction in OFDM Systems," *IEEE Signal Processing Letters*, Vol. 12, No. 6, June 2005, pp. 453-456.

[13] J. Tellado, "Peak to average power reduction for multicarrier modulation," *Ph.D dissertation*, Stanford university, Sept. 1999.

[14] S. H. Muller and J. B. Huber "A Comparison of Peak Power Reduction Schemes for OFDM," *IEEE GLOBECOM'97*, Phoenix, Arizona, pp. 1-5, Nov. 1997.

[15] S. H. Han and J. H. Lee "PAPR reduction of OFDM signals using reduced complexity PTS technique," *IEEE Signal Processing Letters*, Vol. 11, No. 11, Nov. 2004.

[16] A. D. S. Jayalath and C. Tellambura "Adaptive PTS approach for reduction of peak-to-average power ratio of OFDM signal," *Electronics Letters*, pp. 1226-1228, vol. 36, no. 14 6th July 2000.

11

Downlink Resource Scheduling in an LTE System

Raymond Kwan†, Cyril Leung††, and Jie Zhang†
†*Centre for Wireless Network Designs (CWiND), University of Bedfordshire, Luton, UK*
††*Dept. of Electrical & Computer Engineering, The University of British Columbia,*
Vancouver, B.C., Canada
Emails: raymond.kwan@beds.ac.uk, cleung@ece.ubc.ca, Jie.zhang@beds.ac.uk

Abstract

The problem of allocating resources to multiple users on the downlink of a Long Term Evolution (LTE) cellular communication system is discussed. An optimal (maximum throughput) multiuser scheduler is proposed and its performance is evaluated. Numerical results show that the system performance improves with increasing correlation among OFDMA subcarriers. It is found that a limited amount of feedback information can provide a relatively good performance. A sub-optimal scheduler with a lower computational complexity is also proposed, and shown to provide good performance. The sub-optimal scheme is especially attractive when the number of users is large, as the complexity of the optimal scheme may then be unacceptably high in many practical situations. The performance of a scheduler which addresses fairness among users is also presented.

1. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is an attractive modulation technique that is used in a variety of communication systems such as Digital Subscriber Lines (DSLs), Wireless Local Area Networks (WLANs), Worldwide Interoperability for Microwave Access (WiMAX) Andrews et al. (2007), and Long Term Evolution (LTE) cellular networks. In order to exploit multiuser diversity and to provide greater flexibility in resource allocation (scheduling), Orthogonal Frequency Division Multiple Access (OFDMA), which allows multiple users to simultaneously share the OFDM sub-carriers, can be employed. The problem of power and sub-carrier allocation in OFDMA systems has been extensively studied, e.g. Liu & Li (2005); Wunder et al. (2008), and the references therein.

What distinguishes packet scheduling in LTE from that in earlier radio access technologies, such as High Speed Downlink Packet Access (HSDPA), is that LTE schedules resources for users in both the time domain (TD) and the frequency domain (FD) whereas HSDPA only involves the time domain. This additional flexibility has been shown to provide throughput and coverage gains of around 40% Pokhariyal et al. (2006). Because packet scheduling for LTE involves scheduling users in both TD and FD, various TD and FD schemes have been proposed in Pokhariyal et al. (2006)-Monghal et al. (2008). Assume that we have packets for N_{users} users waiting in the queue and that resources can only be allocated at the beginning of a pre-defined time period known as the Transmission Time Interval (TTI) or scheduling period. In TD scheduling, U users from the total of N_{users} users are selected based on some priority metric. After the U users have been selected, appropriate subcarrier frequencies and modulation and coding schemes (MCSs) are then assigned by the FD scheduler. Note that the metrics used for TD and FD scheduling can be different in order to provide a greater degree of design flexibility. Examples of TD/FD scheduling metrics have been proposed in Kela et al. (2008); Monghal et al. (2008).

In order to make good scheduling decisions, a scheduler should be aware of channel quality in the time domain as well as the frequency domain. Ideally, the scheduler should have knowledge of the channel quality for each sub-carrier and each user. In practice, due to limited signalling resources, sub-carriers in an OFDMA system are often allocated in groups. On the downlink in LTE systems, sub-carriers are grouped into resource blocks (RBs) of 12 adjacent sub-carriers with an inter sub-carrier spacing of 15 kHz Dahlman et al. (2008); *Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8)* (2007). Each RB has a time slot duration of 0.5 ms, which corresponds to 6 or 7 OFDM symbols depending on whether an extended or normal cyclic prefix is used. The smallest resource unit that a scheduler can assign to a user is a Scheduling Block (SB), which consists of two consecutive RBs, spanning a sub-frame time duration of 1 ms Dahlman et al. (2008); *Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8)* (2007) (see Fig. 1).

From the perspective of downlink scheduling, the channel quality is reported by the user via a Channel Quality Indicator (CQI) over the uplink. If a single CQI value is used to convey the channel quality over a large number of SBs, the scheduler may not be able to distinguish the quality variations within the reported range of subcarriers. This is a severe problem for highly frequency-selective channels. On the other hand, if a CQI value is used to represent each SB, many CQI values may need to be reported back, resulting in a high signalling overhead. A number of CQI reporting schemes and associated trade-offs are discussed in Kolehmainen (2008).

Given a set of CQI values from different users, the multiuser scheduling problem in LTE involves the allocation of SBs to a subset of users in such a way as to maximize some objective function, e.g. overall system throughput or other fairness-sensitive metrics. The identities of the assigned SBs and the MCSs are then conveyed to the users via a downlink control channel. Studies on LTE-related scheduling have been reported in Kwan et al. (2008); Liu et al. (2007); Ning et al. (2006); Pedersen et al. (2007); Pokhariyal et al. (2007) and the references therein.

As pointed out in Jiang et al. (2007), the type of traffic plays an important role in how scheduling should be done. For example, Voice-over IP (VoIP) users are active only half of the time. Also, the size of VoIP packets is small, and the corresponding inter-arrival time is fairly constant. While dynamic scheduling based on frequent downlink transmit format signalling and uplink CQI feedback can exploit user channel diversity in both frequency and time domains, it requires a large signalling overhead. This overhead consumes time-frequency resources, thereby reducing the system capacity. In order to lower signalling overhead for VoIP-type traffic, *persistent scheduling* has been proposed *Discussion on Control Signalling for Persistent Scheduling of VoIP* (2006); *Persistent Scheduling in E-UTRA* (2007). The idea behind persistent scheduling is to pre-allocate a sequence of frequency-time resources with a fixed MCS to a VoIP user at the beginning of a specified period. This allocation remains valid until the user receives another allocation due to a change in channel quality or an expiration of a timer. The main disadvantage of such a scheme is the lack of flexibility in the time domain. This shortcoming has led to semi-persistent scheduling which represents a compromise between rigid persistent scheduling on the one hand, and fully flexible dynamic scheduling on the other. In semi-persistent scheduling, initial transmissions are persistently scheduled so as to

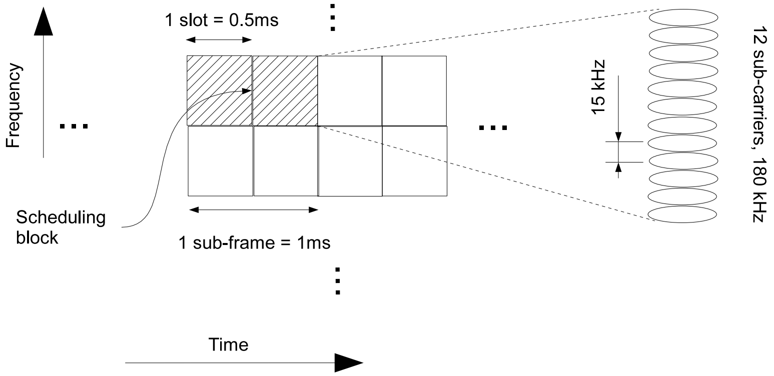


Fig. 1. LTE downlink time-frequency domain structure.

reduce signalling overhead and retransmissions are dynamically scheduled so as to provide adaptability. The benefits of semi-persistent scheduling are described in Jiang et al. (2007). An important constraint in LTE downlink scheduling is that all SBs for a given user need to use the same MCS in any given TTI[1] Dahlman et al. (2008). In the rest of this chapter, we focus on the FD aspect of dynamic scheduling. Specifically, the challenging problem of multiuser FD scheduling is formulated as an optimization problem, taking into account this MCS restriction. Simpler sub-optimal solutions are also discussed.

2. System Model

The system model we will use to study the resource allocation problem is now described[2]. An SB consists of a number, N_{sb}, of consecutive OFDM symbols. Let L be the total number of sub-carriers and $L_d(\nu) \leq L$ be the number of data-carrying sub-carriers for symbol ν, where $\nu = 1, 2, \ldots, N_{sb}$. Also, let $R_j^{(c)}$ be the code rate associated with the MCS $j \in \{1, 2, \ldots, J\}$, M_j be the constellation size of the MCS j and T_s be the OFDM symbol duration. Then, the bit rate, r_j, that corresponds to a single SB is given by

$$r_j = \frac{R_j^{(c)} \log_2\left(M_j\right)}{T_s N_{sb}} \sum_{\nu=1}^{N_{sb}} L_d(\nu). \tag{1}$$

Let U be the number of simultaneous users, and N_{tot} be the total number of SBs that are available during each TTI. In addition, let $\mathcal{N}_i$ be a subset of the N_{tot} SBs whose Channel Quality

---

[1] This applies in the non multiple-input-multiple-output (MIMO) configuration. For the MIMO configuration, a maximum of two different MCSs can be used for data belonging to two different transport blocks.

[2] The material in this section is based in part on Kwan et al. (2009b)

Indicator (CQI) values are to be reported by user i; the size of $\mathcal{N}_i$ is denoted by N_i. It is assumed that the N_i highest SB CQI values are fed back. Such a limited feedback scheme would necessitate a smaller bandwidth albeit at the cost of a degraded system performance. We also assume that the total available power is shared equally among the users. It is noted in Chung & Goldsmith (2001); Miki et al. (2007) that the throughput degradation resulting from such an assumption is small when adaptive modulation and coding (AMC) is used, as is the case in LTE.

Let $\mathbf{x}_{i,n}, n = 1, 2, \ldots, N_i$ be a real scalar or vector reported (via a feedback channel) by user i to indicate the collective channel qualities of all the sub-carriers within the n-th reported SB. The exact nature of $\mathbf{x}_{i,n}$ depends on the feedback method adopted. Furthermore, let $q_{i,max}(\mathbf{x}_{i,n}) \in \{1, 2, \ldots, J\}$ be the index of the highest-rate MCS that can be supported by user i for the n-th SB at CQI value $\mathbf{x}_{i,n}$, i.e. $q_{i,max}(\mathbf{x}_{i,n}) = \arg\max_j \left(R_j^{(c)} \log_2 \left(M_j \right) |\mathbf{x}_{i,n} \right)$. Due to frequency selectivity, the qualities of the sub-carriers within a SB may differ; the indicator $\mathbf{x}_{i,n}$ should provide a good collective representation of the qualities for all the sub-carriers within the n-th SB Blankenship et al. (2004); Ericsson (2003); Lampe & Rohling (2003). For convenience, we assume that the MCS rate $R_j^{(c)} \log_2 \left(M_j \right)$ increases monotonically with j and that the rate of MCS 1 is zero. SBs whose CQI values are not reported back are assigned to MCS 1.

Let $1 \leq q_{i,max}(x_{i,n}) \leq J$, where $q_{i,max}(x_{i,n}) = 1$ corresponds to the index of a zero bit rate, and J the index of the highest bit rate. The rate indices for the un-reported SBs are assigned zero bit rates, i.e. $q_{max,i}(x_{i,n}) = 1, \forall n \notin \mathcal{N}_i$. Fig. 2 shows the relationship between $r_{i,j}$ and $x_{i,n}$ in the case when $x_{i,n}$ is assumed to represent an effective Signal-to-Interference and Noise Ratio (SINR) over a single SB. It might be noted that $x_{i,n}$ may not always take the form of a SINR Ericsson (2003)-*Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems* (2005). In Fig. 2, $x_{i,n}$ maps to $j = q_{i,max}(x_{i,n}) = 3$, which corresponds to $r_{i,3}$. It is important to note that $q_{i,max}(x_{i,n})$ is assumed to be known at the scheduler once $x_{i,n}$ is known. The exact form of $x_{i,n}$ is not the focus of this paper.

As mentioned in Section 1, in a non-MIMO configuration, all SBs scheduled for a given user within the same TTI must use the same MCS. If MCS j is to be used for user i, then only certain SBs can be assigned to the user. For example, suppose $N_i = 4$, and

$$\begin{aligned} 1 \leq q_{i,max}(\mathbf{x}_{i,2}) \quad < \quad & q_{i,max}(\mathbf{x}_{i,1}) < q_{i,max}(\mathbf{x}_{i,3}) \\ & < q_{i,max}(\mathbf{x}_{i,5}) \leq J. \end{aligned} \tag{2}$$

Then, if MCS $j = q_{i,max}(\mathbf{x}_{i,3})$ is used, only SBs $n = 3$ and 5 can be allocated to user i since only these SBs have good enough channel qualities to support an MCS index of $q_{i,max}(\mathbf{x}_{i,3})$ or higher. Selecting SBs $n = 1$ or 2 with MCS $j = q_{i,max}(\mathbf{x}_{i,3})$ would result in unacceptably high error rates for these SBs. On the other hand, if $j = q_{i,max}(\mathbf{x}_{i,2})$, all 4 SBs can be selected, at the expense of a lower bit rate for SBs 1, 3 and 5. This suggests that there is an optimal value of j which maximizes the total bit rate for user i.

3. Single User Optimization

In single user optimization, the aim is to determine the MCS (rate) index, j^* and the set of SBs to be allocated to user i so as to maximize the assigned bit rate, R_i, given the set of channel qualities $\{q_{i,max}(\mathbf{x}_{i,n}), n \in \mathcal{N}_i\}$[3].

---

[3] The material in this section is based in part on Kwan et al. (2009a)

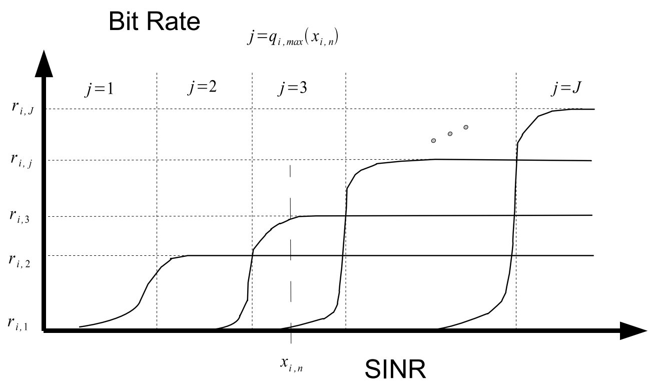


Fig. 2. Relationship between $r_{i,j}$ and $x_{i,n}$ in the case when $x_{i,n}$ takes on the form of an effective SINR Kwan et al. (2009b)©[2009] IEEE.

Let $Q_{max}(i) = \max_{n \in \mathcal{N}_i}\{q_{i,max}(\mathbf{x}_{i,n})\}$, and let $\mathbf{b}_i = [b_{i,1},\ b_{i,2},\ \ldots, b_{i,Q_{max}(i)}]$ be the MCS vector for user i, where

$$b_{i,j} = \begin{cases} 0, & \text{if user } i \text{ is not assigned MCS } j \\ 1, & \text{if user } i \text{ is assigned MCS } j. \end{cases}$$

The optimal $\mathbf{b}_i$, which maximizes the total bit rate for user i, is obtained by solving Problem $(P1)$.

$$(P0): \qquad \max_{\mathbf{b}_i} \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} b_{i,j} r_j \tag{3}$$

subject to

$$\sum_{j=1}^{Q_{max}(i)} b_{i,j} = 1, \forall\, i, \tag{4}$$

$$b_{i,j} \in \{0,1\}, \forall\, i,j. \tag{5}$$

The formulation in (3) allows the selected bit rate for SB n to be less than what $\mathbf{x}_{i,n}$ can potentially support, as may be the case if user i is assigned more than one SB during a TTI. Constraint (4) ensures that the MCS for user i can only take on a single value between 1 and $Q_{max}(i)$.

The above optimization problem can be easily solved as follows. Let $\mathbf{R}^{(i)}$ be an $N_i \times Q_{max}(i)$ matrix with (n,j)-th element $\{r_{n,j}^{(i)} = r_j,\ \ j = 1,2,\ldots,q_{i,max}(\mathbf{x}_{i,n})\}$. Denote the sum of the elements in the j-th column of $\mathbf{R}^{(i)}$ by

$$c_j^{(i)} = \sum_{n=1}^{N_i} r_{n,j}^{(i)}. \tag{6}$$

Then the optimal MCS for user i is

$$j^* = \underset{1\le j\le Q_{max}(i)}{\operatorname{argmax}} \; c_j^{(i)} \tag{7}$$

and the corresponding maximum bit rate is $c_{j^*}^{(i)}$. The set, $\mathcal{K}_i$, of SBs allocated to user i is given by

$$\mathcal{K}_i = \{k \mid q_{i,max}(\mathbf{x}_{i,k}) \ge j^*,\ k \in \mathcal{N}_i\}. \tag{8}$$

4. Throughput Optimal Scheduler

In this section, the issue of multiuser scheduling is addressed, in which radio resources are jointly allocated in maximizing the total system throughput[4].

4.1 Multi-User Optimal Scheduling

If there are many users, the optimization problem becomes harder to solve. In addition, each SB can be allocated to at most one user Dahlman et al. (2008). Let

$$v_{i,n}(\mathbf{x}_{i,n}) = \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} b_{i,j} r_j \tag{9}$$

be the bit rate of SB n selected for user i given the channel quality $\mathbf{x}_{i,n}$, where $b_{i,j} \in \{0,1\}$ is a binary decision variable. Let $Q_{max}(i) = \max_{n\in\mathcal{N}_i}\{q_{i,max}(\mathbf{x}_{i,n})\}$. The constraint

$$\sum_{j=1}^{Q_{max}(i)} b_{i,j} = 1 \tag{10}$$

is introduced to ensure that the MCS for user i can only take on a single value between 1 and $Q_{max}(i)$. The formulation in (9) allows the selected bit rate for SB n to be less than what $\mathbf{x}_{i,n}$ can potentially support, as may be the case if user i is assigned more than one SB during a TTI. From (9) and (10), it can be seen that SB n might be selected for user i only if the MCS j^* chosen for user i satisfies $j^* \le q_{i,max}(\mathbf{x}_{i,n})$, as illustrated in Fig. 3.

The problem of jointly maximizing the sum of the bit rates for all users can be formulated as

$$(P1): \qquad \max_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{U} \sum_{n\in\mathcal{N}_i} a_{i,n} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} b_{i,j} r_j \tag{11}$$

subject to (10) and

$$\sum_{i=1}^{U} a_{i,n} = 1,\ n \in \bigcup_{i=1}^{U} \mathcal{N}_i \tag{12}$$

$$a_{i,n}, b_{i,j} \in \{0,1\},\ \forall\, i,j,n. \tag{13}$$

In problem $(P1)$, $\mathbf{A} = \{a_{i,n}, i = 1,\ldots,U, n \in \mathcal{N}_i\}$, $\mathbf{B} = \{b_{i,j}, i = 1,\ldots,U,\ j = 1,\ldots,Q_{max}(i)\}$, and $a_{i,n}$ is a binary decision variable, with value 1 if SB n is assigned to user i and 0 otherwise. The objective in (11) is to select optimal values for $\mathbf{A}$ and $\mathbf{B}$ to maximize the aggregate bit rate $\sum_{i=1}^{U}\sum_{n\in\mathcal{N}_i} a_{i,n} v_{i,n}(\mathbf{x}_{i,n})$.

---

[4] The material in this section is based in part on Kwan et al. (2009b)

4.2 Linearized Model

Problem $(P1)$ is non-linear due to the product term $a_{i,n}b_{i,j}$ in (11). Although solutions can be obtained using optimization techniques such as Branch-and-Bound Rardin (1998), a globally optimal solution cannot be guaranteed. To avoid this difficulty, the problem can be transformed into an *equivalent* linear problem by introducing an auxiliary variable $t_{n,i,j} = a_{i,n}b_{i,j}$. Then, Problem $(P1)$ can be linearized as follows:

$$(P1'): \qquad \max_{\mathbf{A},\mathbf{B},\mathbf{T}} \sum_{i=1}^{U} \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} t_{n,i,j} r_j \tag{14}$$

subject to (10), (12), (13) and

$$t_{n,i,j} \leq b_{i,j}, \tag{15}$$

$$t_{n,i,j} \leq a_{i,n}M, \tag{16}$$

$$t_{n,i,j} \geq b_{i,j} - (1 - a_{i,n})M, \tag{17}$$

where M is a large positive real value. Problem $(P1')$ can then be solved using well-known integer linear programming techniques Rardin (1998).

4.3 Multi-user Suboptimal Scheduling

In the optimal scheduler formulations in $(P1)$ and $(P1')$, the MCSs, SBs, and users are jointly assigned. To reduce the computational complexity, the proposed sub-optimal scheduler performs the assignment in two stages. In the first stage, each SB is assigned to the user who can support the highest bit rate. In the second stage, the best MCS for each user is determined. The idea behind the sub-optimal scheduler is to assign a disjoint subset of SBs to each user, thereby reducing a joint multiuser optimization problem into U parallel single-user optimization problems. We will refer to this two-stage problem as Problem $(P2)$.

Let φ_n be the index of the user which can support the highest-rate MCS for SB n, i.e. $\varphi_n = \arg\max_{i \in \{1,2,\ldots,U\}} q_{i,max}(\mathbf{x}_{i,n})$. Furthermore, let $\tilde{\mathcal{N}_i}$ be the (disjoint) set of SBs assigned to user i, i.e. $\{n$ such that $\varphi_n = i\}$. In the first stage, the sub-optimal scheduler determines $\{\tilde{\mathcal{N}_i}\}_{i=1}^{U}$. In other words, the best user is selected for each SB, and information regarding the set of SBs associated with each user is collected and represented by $\tilde{\mathcal{N}_i}$, as summarized in the following pseudo-code listing.

Algorithm 1 Sub-optimal algorithm for solving Problem $(P2)$.

```
for $i = 1$ to $U$ do
    for $n \in \bigcup_{i=1}^{U} \mathcal{N}_i$ do
        $\varphi_n = \arg\max_i q_{i,max}(\mathbf{x}_{i,n})$
    end for
    $\tilde{\mathcal{N}_i}$ = the set of values of $n$ such that $\varphi_n = i$
end for
return $\{\tilde{\mathcal{N}_i}\}_{i=1}^{U}$.
```

Let $Q'_{max}(i) = \max_{n \in \tilde{\mathcal{N}_i}} \{q_{i,max}(\mathbf{x}_{i,n})\}$ be the index of the highest MCS among all SBs associated with user i, and let the MCS vector, $\mathbf{b}_i$, for user i be

$$\mathbf{b}_i = [b_{i,1},\, b_{i,2},\, \ldots, b_{i,Q'_{max}(i)}]. \tag{18}$$

Note that the length of $\mathbf{b}_i$ may not necessarily be the same for all users. In the second stage, the sub-optimal scheduler determines $\mathbf{b}_i$ which maximizes the total bit rate for user i. Similar to the approach in Section 4, the optimal $\mathbf{b}_i$ can be obtained by solving the problem (P0) as described in section 3. Since the association between the SBs and the users are done in the first stage, the second stage only involves the selection of the best MCS given the selected SBs for each user. Compared to $(P1)$ or $(P1')$, $(P2)$ is a much simpler problem, due to the decoupled selection between SBs and MCSs.

4.4 Numerical Results

For illustration purposes, we assume $N_{tot} = 12$ SBs per TTI, $L = 12$ sub-carriers per SB, $N_1 = N_2 = \ldots = N_U = N$ and that the normal cyclic prefix configuration is used Dahlman et al. (2008). The received amplitude for each subcarrier and user undergoes Nakagami-m fading Simon & Alouini (2005); unless otherwise indicated, a fading figure, m, value of 1 is assumed. The average signal-to-interference plus noise ratios (SINRs) for the $U = 3$ users are 10 dB, 11 dB, and 12 dB respectively. It is assumed that the SINRs for all sub-carriers of each user are correlated, but identically distributed (c.i.d.), and that the resource blocks follow the localized configuration *Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8)* (2007). The correlation coefficient between a pair of sub-carriers is given by $\rho^{|i-j|}$, where i and j are the sub-carrier indices. The SINR of each sub-carrier is assumed to be independent at the beginning of each scheduling period, and constant throughout the entire period. For simplicity, it is assumed that the set of MCSs consists of QPSK 1/2 and 3/4, 16-QAM 1/2 and 3/4, as well as 64-QAM 3/4 Andrews et al. (2007), and the L1/L2 control channels are mapped to the first OFDM symbol within each sub-frame. Furthermore, each sub-frame consists of 8 reference symbols Dahlman et al. (2008). The feedback method is based on the Exponential Effective SINR Mapping (EESM) Ericsson (2003), with parameter values obtained from Westman (2006). Let R^*_{tot} be the total bit rate defined in (11) or (14), and $E[R^*_{tot}]$ be the value of R^*_{tot} averaged over 2500 channel realizations.

Fig. 4 shows the average total bit rate, $E[R^*_{tot}]$, as a function of ρ for $N = 5$ and 12. It can be observed that the performance improves with the level of correlation among sub-carriers. Recall that the idea behind EESM is to map a set of sub-carrier SINRs, $\{\Gamma_i\}_{i=1}^{L}$, to a single effective SINR, Γ^*, in such a way that the block error probability (BLEP) due to $\{\Gamma_i\}_{i=1}^{L}$ can be well approximated by that at Γ^* in additive white Gaussian noise (AWGN) Andrews et al. (2007); Ericsson (2003). The value of Γ^* tends to be skewed towards the weaker sub-carriers in order to maintain an acceptable BLEP. For small values of ρ, sub-carriers with large SINRs are not effectively utilized, leading to a relatively poor performance. For $N = 5$, the difference in $E[R^*_{tot}]$ values achieved by the optimal scheduler at $\rho = 0.1$ and $\rho = 0.9$ is about 25%; the corresponding difference is similar for the sub-optimal scheduler.

Fig. 5 shows the average total bit rate, $E[R^*_{tot}]$, as a function of N for $\rho = 0.5$ and 0.9. It can be seen that the performance improves with N as to be expected, but the rate of improvement decreases. There is little performance improvement as N increases beyond 8. For $\rho = 0.5$, the difference in $E[R^*_{tot}]$ values achieved by the optimal scheduler at $N = 6$ and $N = 12$ is about 15%; the corresponding difference is about 10% for the sub-optimal scheduler.

Fig. 6 shows $E[R^*_{tot}]$ as a function of the fading figure, m, with $U = 3$, $\rho = 0.9$ and $N = 12$. It can be seen that the throughput performance improves with m. As m increases, there is less variation in channel quality which allows a more efficient resource allocation among users. The difference in $E[R^*_{tot}]$ values achieved by the optimal scheduler for $m = 1$ and $m = 10$ is about 15%; the corresponding difference is similar for the sub-optimal scheduler.

Fig. 7 shows $E[R_{tot}^*]$ as a function of the number, U, of users for $\rho = 0.9$ and $N = 12$. The average SINRs for all users are set to 10 dB. As U increases, $E[R_{tot}^*]$ increases due to the more pronounced benefits from multiuser diversity. Fig. 8 shows the percentage gain in $E[R_{tot}^*]$ for the optimal scheduler relative to the sub-optimal scheduler as a function of U. As U increases, it becomes increasingly likely that a given user will be assigned at most one SB in the first stage operation of the sub-optimal scheduler. In this event, the sub-optimal scheduler is actually optimal. It is therefore expected that the difference in performance between the optimal and sub-optimal schedulers will be small when U is large, as illustrated in Fig. 8. The result indicates that the sub-optimal scheduler is especially attractive for large values of U since it provides a significant reduction in complexity and its performance approaches that of the optimal scheduler.

5. Optimal Scheduler with Fairness

In Section 5[5], the objective is to maximize system throughput. In this Section, we consider a proportional-rate scheduler which is intended to improve fairness among users.

5.1 Multi-User Optimal Scheduling

The joint optimization problem can be formulated as

$$(P3): \qquad \max_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{U} \sum_{n \in \mathcal{N}_i} a_{i,n} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n}(t))} b_{i,j} \left(\frac{r_j}{\psi_i(t)} \right) \tag{19}$$

subject to (4), (12), (13). The term $\psi_i(t)$ is given by

$$\psi_i(t) = \begin{cases} \overline{R}_i(t) & , \text{ PF Scheduling} \\ 1 & , \text{ Max-Rate Scheduling.} \end{cases} \tag{20}$$

In (20), $\overline{R}_i(t) = (1-\alpha)\overline{R}_i(t-1) + \alpha R_i(t-1)$ is the average bit rate up to time $t-1$, $\alpha \in [0,1]$, and $R_i(t)$ is the bit rate assigned to user i at time t. It is known that the PF scheduler is asymptotically optimal Kelly (1997); Stolyar (2005). Other multi-user schedulers have been proposed in Bennett & Zhang (1996); Shakkottai & Stolyar (2002).

In problem $(P3)$, $\mathbf{A} = \{a_{i,n}, i = 1,\ldots,U, n \in \bigcup_{i=1}^{U} \mathcal{N}_i\}$, $\mathbf{B} = \{b_{i,j}, i = 1,\ldots,U,\ j = 1,\ldots,Q_{max}(i)\}$, and $a_{i,n}$ is a binary decision variable, with value 1 if SB n is assigned to user i and 0 otherwise. Problem $(P3)$ is non-linear due to the product $a_{i,n}b_{i,j}$ in (19). Although solutions can be obtained using optimization techniques such as Branch-and-Bound Rardin (1998), global optimality cannot be guaranteed. To overcome this difficulty, Problem $(P2)$ can be transformed into an *equivalent* linear problem $(P3')$ by introducing an auxiliary variable $t_{n,i,j} = a_{i,n}b_{i,j}$, i.e.

$$(P3'): \qquad \max_{\mathbf{A},\mathbf{B},\mathbf{T}} \sum_{i=1}^{U} \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n}(t))} t_{n,i,j} \left(\frac{r_j}{\psi_i(t)} \right) \tag{21}$$

subject to (4), (12), (13), together with (15)-(17). Problem $(P3')$ can then be solved using standard integer linear programming techniques Rardin (1998).

---

[5] The material in this section is based in part on Kwan et al. (2009a)

5.2 Multi-user Suboptimal Scheduling

The proposed sub-optimal multiuser scheduler consists of two stages. In the first stage, the scheduler determines the set,

$$\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i) = \{q_{i,max}(\mathbf{x}_{i,n}(t)) | n \in \mathcal{N}_i\}, \tag{22}$$

of maximum rate indices, one for each SB for user i at time t. The users are then ranked according to their priority index values, $\{\varphi_i, i = 1, 2, \ldots, U\}$,

$$\varphi_i = \begin{cases} g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i))/\overline{R}_i(t) & , \text{ PF Scheduling} \\ g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i)) & , \text{ Max-Rate Scheduling} \end{cases} \tag{23}$$

The first line in the RHS of (23) corresponds to proportional fair (PR) scheduling (Dahlman et al., 2008, Page 113), Kelly (1997); Wengerter et al. (2005), whereas the second line corresponds to maximum rate scheduling (Dahlman et al., 2008, Page 111). The term $g(.)$ is a function which returns the highest bit rate that user i can support based on $\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i)$, as discussed in Section 3, i.e. $c_{j^*}^{(i)}(t) = g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i))$.

For notational convenience, let $\varphi_{(1)} \geq \varphi_{(2)} \geq \ldots \geq \varphi_{(U)}$ be the ranked version of $\{\varphi_1, \varphi_2, \ldots, \varphi_U\}$, and $\theta(j) \mapsto i$ be a function which maps the ordered user index j back to the original user index i. In the second stage, the allocation of resources is done in a sequential fashion, one user at a time, according to the following user order: $\theta(1), \theta(2), \ldots$. Thus, starting with user $\theta(1)$, and the initial set of SBs, $\mathcal{N}_{\theta(1)} = \mathcal{N}$, where $\mathcal{N}$ corresponds to the complete set of available SBs, the MCS index and the set of SBs, $\mathcal{K}_{\theta(1)}$, are determined as described in Section 3, and assigned to user $\theta(1)$. The remaining SBs, $\mathcal{N}_{\theta(2)} = \mathcal{N}_{\theta(1)} - \mathcal{K}_{\theta(1)}$, are then made available to user $\theta(2)$. The resource allocation process continues until all SBs have been assigned.

5.3 Numerical Results

For illustration purposes, we assume $N_{tot} = 12$ SBs per TTI, $L = 12$ sub-carriers per SB, $N_1 = N_2 = \ldots = N_U = N_{tot}$ and that the normal cyclic prefix configuration is used Dahlman et al. (2008). The fading amplitude for each subcarrier of any user follows the Nakagami-m model Simon & Alouini (2005), with a fading figure m equal to 1. It is assumed that the signal-to-interference plus noise ratios (SINRs) for all sub-carriers of any user are correlated, but identically distributed (c.i.d.), and that the resource blocks follow the localized configuration Dahlman et al. (2008). The correlation coefficient between a pair of sub-carriers is given by $\rho^{|i-j|}$, where i and j are the sub-carrier indices.

For simplicity, the SINR of a given sub-carrier is assumed to be independent at every scheduling period, and constant within a scheduling period. This independent assumption is reasonable for the purpose of comparing the long-term fairness for the Max-Rate and PF schedulers. The set of MCSs consists of QPSK 1/2 and 3/4, 16-QAM 1/2 and 3/4, as well as 64-QAM 3/4 Andrews et al. (2007), and the L1/L2 control channels are mapped to the first OFDM symbol within each sub-frame. Furthermore, each sub-frame consists of 8 reference symbols Dahlman et al. (2008). The feedback method is based on the Exponential Effective SINR Mapping (EESM) Ericsson (2003), with parameter values obtained from Westman (2006).

Let $R_{tot}(t) = \sum_{i=1}^{U} R_i(t)$ be the total bit rate at time t, and $\overline{R}_{tot} = \sum_{t=1}^{T} R_{tot}(t)/T$ be the corresponding value averaged over $T = 2500$ channel realizations. Similarly, let $\overline{R}_i = \sum_{t=1}^{T} R_i(t)/T$

be the average bit rate for user i, and $\eta = \frac{\left(\sum_{i=1}^{U} \overline{R}_i\right)^2}{U \sum_{i=1}^{U} \overline{R}_i^2}$ be the Jain's fairness index Chiu & Jain (1989) for the average user bit rates. The value of η lies in the range [0,1]; an η value of 1 corresponds to all users having the same average (over T scheduling periods) bit rates.

Figs. 9 and 10 show the average total bit rate, $\overline{R}_{tot}$, and fairness index, η, as a function of ρ for three users, with average user SINRs of 14 dB, 15 dB, and 16 dB. It can be seen from Fig. 9 that the bit rates for all schedulers increase with ρ. This can be explained as follows. The motivation behind EESM is to map a set of sub-carrier SINRs, $\{\Gamma_i\}_{i=1}^{L}$, to a single effective SINR, Γ^*, in such a way that the block error probability (BLEP) due to $\{\Gamma_i\}_{i=1}^{L}$ can be well approximated by that at Γ^* in additive white Gaussian noise (AWGN) Andrews et al. (2007); Ericsson (2003). The value of Γ^* tends to be skewed towards the weaker sub-carriers in order to maintain an acceptable BLEP. At a low value of ρ, sub-carriers with large SINRs are not effectively utilized, leading to a relatively poor performance. It can also be seen that the bit rate for the jointly optimal PF scheduler is almost as good as that for the jointly optimal Max-Rate scheduler. In comparison, the bit rates for the sequential Max-Rate and PF schedulers are about 5% and 10% lower. Fig. 10 shows that the fairness index, η, is significantly higher for the two PF schedulers than for their Max-Rate counterparts, indicating that the PF schedulers are quite effective in promoting fairness among users.

Similar plots are shown in Figs. 11 and 12, for user average SINRs of 10 dB, 15 dB, and 20 dB respectively. Here, the variation among user average SINRs is larger than in Figs. 9 and 10. Fig. 11 shows that there is now a larger gap between the bit rates for the jointly optimal PF and Max-Rate schedulers. This is due to the increased effort needed to maintain fairness. It can be seen from Fig. 12 that the two PF schedulers provide significantly better user fairness than the Max-Rate schedulers.

The average total bit rate and fairness index are plotted as a function of the number of users in Figs. 13 and 14 respectively, with an average SINR of 7 dB for all users. In this case, the results show that the jointly optimized Max-rate and PF schedulers provide similar performances. The sequential PF scheduler has a slightly lower throughput than the sequential Max-rate scheduler but a higher fairness index.

6. Conclusion

The problem of allocating resources to multiple users on the downlink in an LTE cellular communication system in order to maximize system throughput was studied. Numerical results show that both the correlation among sub-carriers and the amount of information fed back play important roles in determining the system throughput. It was found that a limited amount of feedback may be sufficient to provide a good performance. A reduced complexity sub-optimal scheduler was proposed and found to perform quite well relative to the optimal scheduler. The sub-optimal scheduler becomes especially attractive as the number of users increases. A scheduler which takes fairness among users into account was also discussed.

Acknowledgment: This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant OGP0001731, by the UBC PMC-Sierra Professorship in Networking and Communications, and a EU FP7 Marie Curie International Incoming Fellowship.

7. References

Andrews, J. G., Ghosh, A. & Muhamed, R. (2007). *Fundamentals of WiMAX - Understanding Broadband Wireless Networking*, Prentice Hall.

Bennett, J. C. R. & Zhang, H. (1996). WF2Q: Worst-case Fair Weighted Fair Queuing, *INFOCOM*, pp. 120–128.

Blankenship, Y., Sartori, P. J., Classon, B. K., Desai, V. & Baum, K. L. (2004). Link Error Prediction Methods for Multicarrier Systems, *Proc. of IEEE Vehiclar Technology Conference, Fall*, Vol. 6, pp. 4175 – 4179.

Chiu, D. & Jain, R. (1989). Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks, *Computer Networks and ISDN Systems* .

Chung, S. T. & Goldsmith, A. J. (2001). Degrees of Freedom in Adaptive Modulation: A Unified View, *IEEE Transactions on Communications* **49**(9): 1561 – 1571.

Dahlman, E., Parkvall, S., Sköld, J. & Beming, P. (2008). *3G HSPA and LTE for Mobile Broadband*, Academic Press.

Discussion on Control Signalling for Persistent Scheduling of VoIP (2006). *Technical Document R1-063275*, 3rd Generation Partnership Project, Riga, Latvia.

Ericsson (2003). System-Level Evaluation of OFDM - Further Considerations, *TR R1-031303*, TSG-RAN WG1 #35, Lisbon, Portugal.

Evolved Universal Terrestrial Radio Access (E-UTRA);Physical Channels and Modulation (Release 8) (2007). *Technical Report 3G TS36.211*, 3rd Generation Partnership Project.

Jiang, D., Wang, H., Malkamaki, E. & Tuomaala, E. (2007). Principle and Performance of Semi-persistent Scheduling for VoIP in LTE System, *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 2861 – 2864.

Kela, P., Puttonen, J., Kolehmainen, N., Ristaniemi, T., Henttonen, T. & Moisio, M. (2008). Dynamic Packet Scheduling Performance in UTRA Long Term Evolution Downlink, *Proc. of Wireless Pervasive Computing*, pp. 308 – 313.

Kelly, F. (1997). Charging and rate control for elastic traffic, *European Transactions on Telecommunications* **8**: 33 – 37.

Kolehmainen, N. (2008). Channel Quality Indication Reporting Schemes for UTRAN Long Term Evolution Downlink, *Proc. of IEEE Vehicular Technology Conference, Spring*, pp. 2522 – 2526.

Kwan, R., Leung, C. & Zhang, J. (2008). Multiuser Scheduling on the Downlink of an LTE Cellular System, *Research Letters in Communications* .

Kwan, R., Leung, C. & Zhang, J. (2009a). Proportional Fair Multiuser Scheduling in LTE, *IEEE Signal Processing Letters* **16**(6): 461 – 464.

Kwan, R., Leung, C. & Zhang, J. (2009b). Resource Allocation in an LTE Cellular Communication System, *Proc. of IEEE International Conference on Communications*, Dresden, Germany.

Lampe, M. & Rohling, H. (2003). PER-Prediction for PHY Mode Selection in OFDM Communication Systems, *Proc. of IEEE Globecom*, Vol. 1, pp. 25 – 29.

Liu, G., Zhang, J., Zhou, B. & Wang, W. (2007). Scheduling Performance of Real Time Service in Multiuser OFDM System, *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, Shanghai, China, pp. 504 – 507.

Liu, H. & Li, G. (2005). *OFDM-Based Broadband Wireless Networks, Design and Optimization*, John Wiley & Sons.

Miki, N., Kishiyama, Y., K.Higuchi & Sawahashi, M. (2007). Optimum Adaptive Modulation and Channel Coding Scheme for Frequency Domain Channel-Dependent Scheduling in OFDMA Based Evolved UTRA Downlink, *Proc. of IEEE Wireless Communications and Networking Conference WCNC*, pp. 1785 – 1790.

Monghal, G., Pedersen, K. I., Kovacs, I. Z. & Mogensen, P. E. (2008). QoS Oriented Time and Frequency Domain Packet Schdulers for the UTRAN Long Term Evolution, *Proc. of IEEE Vehicular Technology Conference, Spring*, pp. 2532 – 2536.

Ning, X., Ting, Z., Ying, W. & Ping, Z. (2006). A MC-GMR Scheduler for Shared Data Channel in 3GPP LTE System, *IEEE Proc. of Vehicular Technology Conference (VTC), Fall*, pp. 1 – 5.

Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (2005). *TS IEEE Std 802.16e-2005*, TSG-RAN WG1 #35.

Pedersen, K. I., Monghal, G., Kovacs, I. Z., Troels, E., Pokhariyal, A., Frederiksen, F. & Mogensen, P. (2007). Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback, *Proc. of IEEE Vehicular Technology Conference (VTC), Fall*, Baltimore, USA, pp. 1792 – 1796.

Persistent Scheduling in E-UTRA (2007). *Technical Document R1-070098*, 3rd Generation Partnership Project, Sorrento, Italy.

Pokhariyal, A., Kolding, T. E. & Mogensen, P. E. (2006). Performance of Downlink Frequency Domain Packet Scheduling For the UTRAN Long Term Evolution, *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*.

Pokhariyal, A., Pedersen, K. I., Monghal, G., Kovacs, I. Z., Rosa, C., Kolding, T. E. & Mogensen, P. E. (2007). HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution, *Proc. of IEEE Vehicular Technology Conference (VTC), Spring*, Dublin, Ireland, pp. 2761 – 2765.

Rardin, R. (1998). *Optimization in Operations Research*, Prentice Hall, Upper Saddle River, NJ.

Shakkottai, S. & Stolyar, A. L. (2002). Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule, *American Mathematical Society Translations* **207**: 185–202.

Simon, M. K. & Alouini, M.-S. (2005). *Digital Communication over Fading Channels*, 2 edn, John Wiley & Sons.

Stolyar, A. L. (2005). On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation, *Operations Research* **53**(1): 12 – 25.

Wengerter, C., Ohlhorst, J. & A.G.E Von Elbwert (2005). Fairness and Throughput Analysis for Generalized Proportional Fair Frequency Scheduling in OFDMA, *Proc. of IEEE Vehicular Technology Conference*.

Westman, E. (2006). *Calibration and Evaluation of the Exponential Effective SINR Mapping (EESM) in 802.16*, Master's thesis, The Royal Institute of Technology (KTH).

Wunder, G., Zhou, C., Bakker, H.-E. & Kaminsk, S. (2008). Thoughput Maximization under Rate Requirements for the OFDMA Downlink Channel with Limited Feedback, *Article ID 437921*, EURASIP Journal on Wireless Communications and Networking.

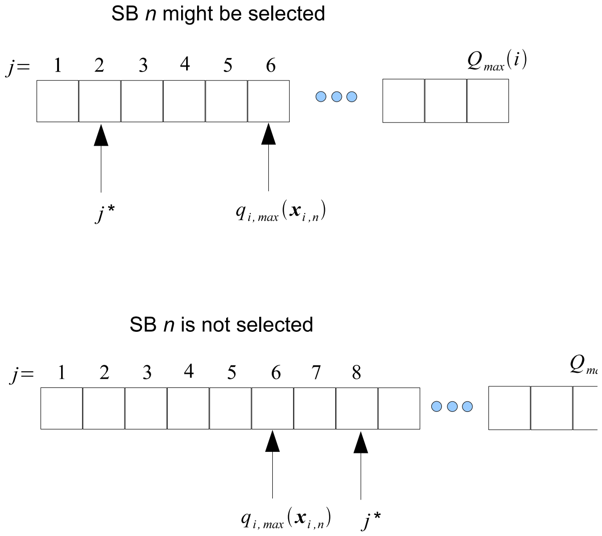


Fig. 3. MCS Selection. In the upper illustration, SB n could be selected, since $j^* \leq q_{i,max}(\mathbf{x}_{i,n})$. On the other hand, when $j^* > q_{i,max}(\mathbf{x}_{i,n})$, SB n cannot be selected, as shown in the bottom illustration Kwan et al. (2009b)©[2009] IEEE.

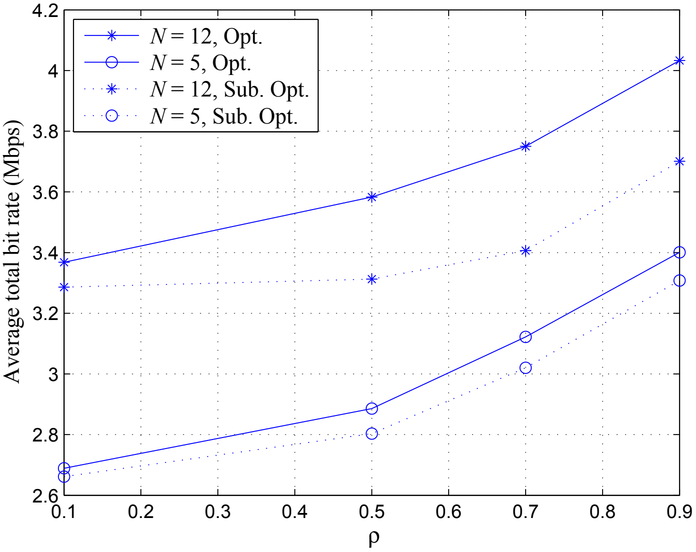


Fig. 4. Average total bit rate as a function of ρ with $N = 5$ and 12 and $U = 3$ Kwan et al. (2009b)©[2009] IEEE.

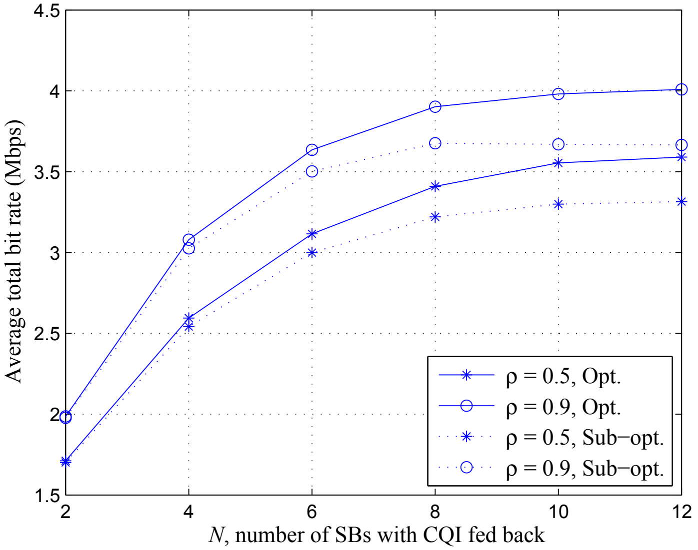


Fig. 5. Average total bit rate as a function of N for $\rho = 0.5$ and 0.9 with $U = 3$ Kwan et al. (2009b)©[2009] IEEE.

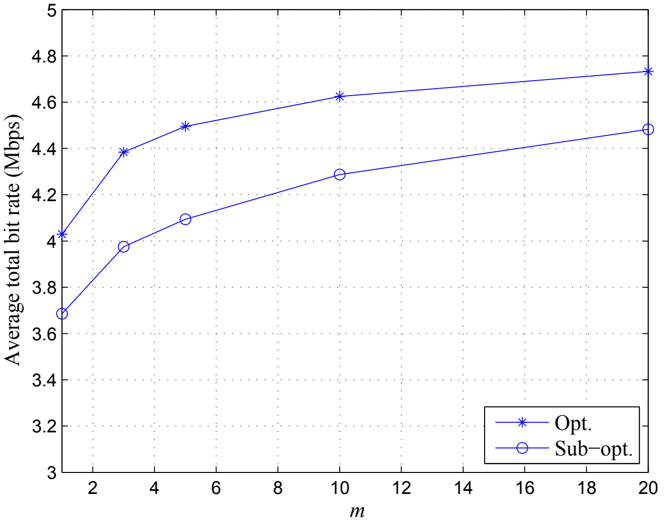


Fig. 6. Average total bit rate as a function of the fading figure, m, with $U = 3$, $\rho = 0.9$ and $N = 12$ Kwan et al. (2009b)©[2009] IEEE.

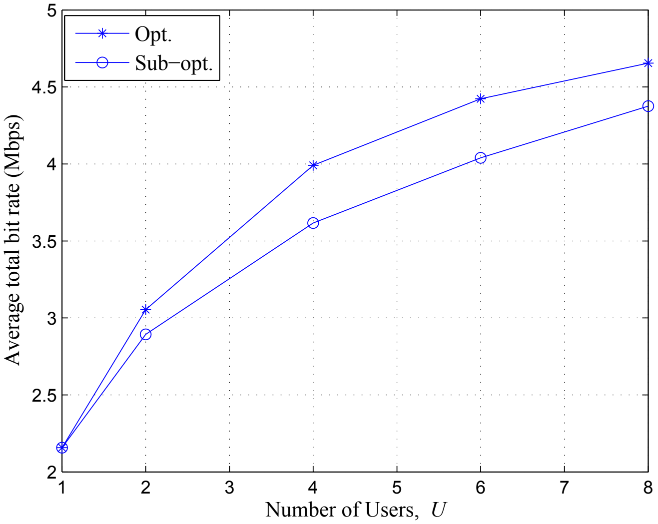


Fig. 7. Average total bit rate as a function of the number of users, U, with $\rho = 0.9$ and $N = 12$ Kwan et al. (2009b)©[2009] IEEE.

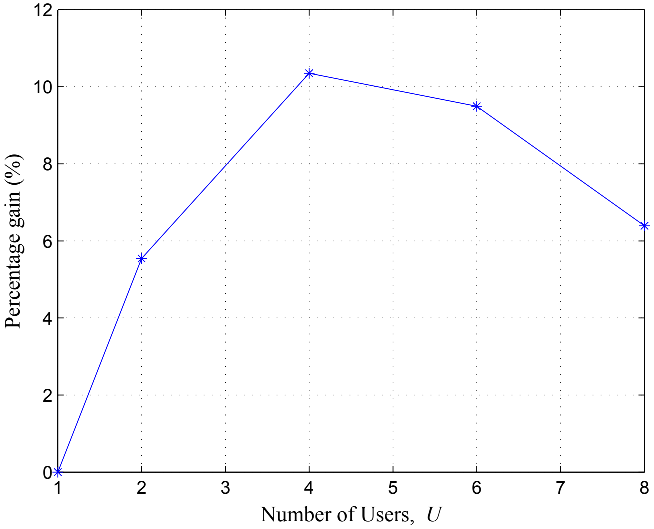


Fig. 8. Percentage gain of the optimal scheduler over the suboptimal scheduler as a function of the number, U, of users with $\rho = 0.9$ and $N = 12$ Kwan et al. (2009b)©[2009] IEEE.

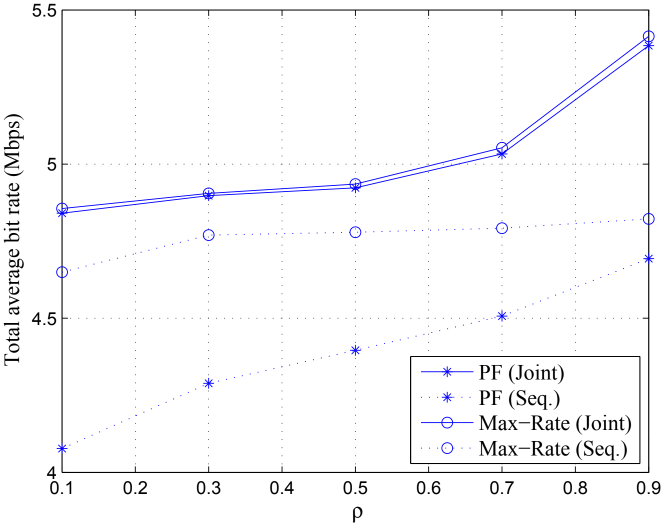


Fig. 9. Average total bit rate as a function of ρ for three users with average SINRs of 14 dB, 15 dB and 16 dB Kwan et al. (2009a)©[2009] IEEE.

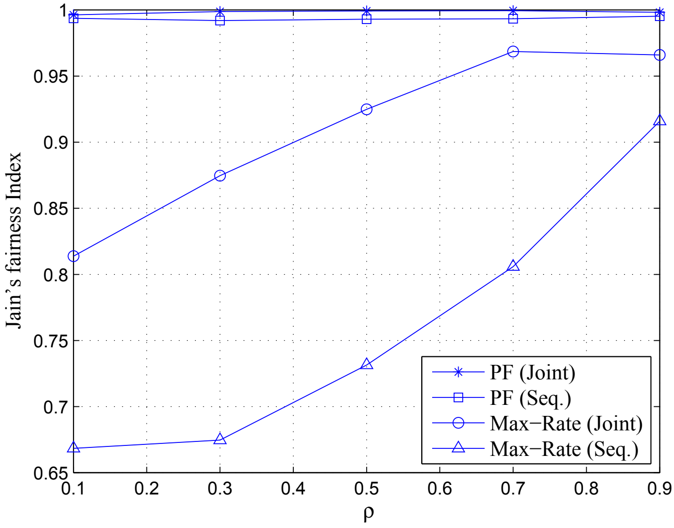


Fig. 10. Fairness index as a function of ρ for three users with average SINRs of 14 dB, 15 dB and 16 dB Kwan et al. (2009a)©[2009] IEEE.

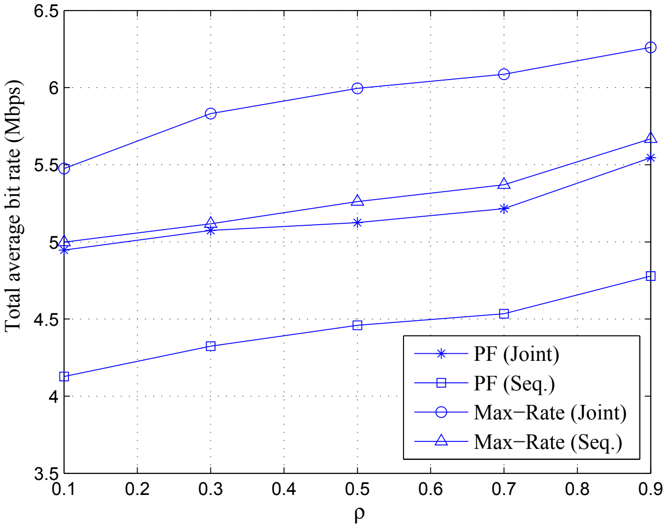


Fig. 11. Average total bit rate as a function of ρ for three users with average SINRs of 10 dB, 15 dB and 20 dB Kwan et al. (2009a)©[2009] IEEE.

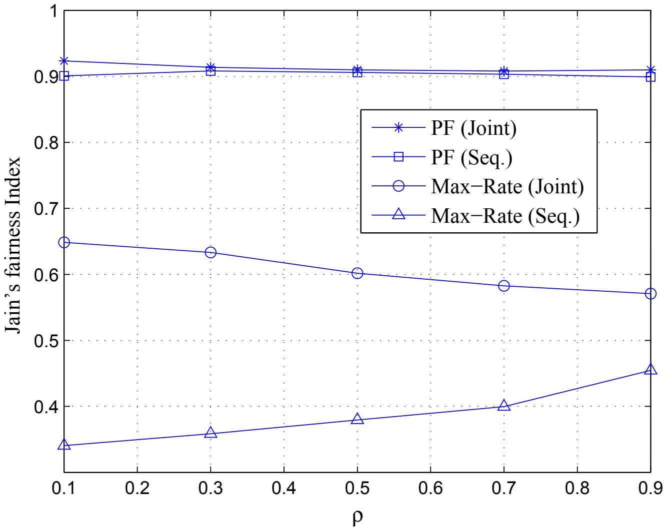


Fig. 12. Fairness index as a function of ρ for three users with average SINRs of 10 dB, 15 dB and 20 dB Kwan et al. (2009a)©[2009] IEEE.

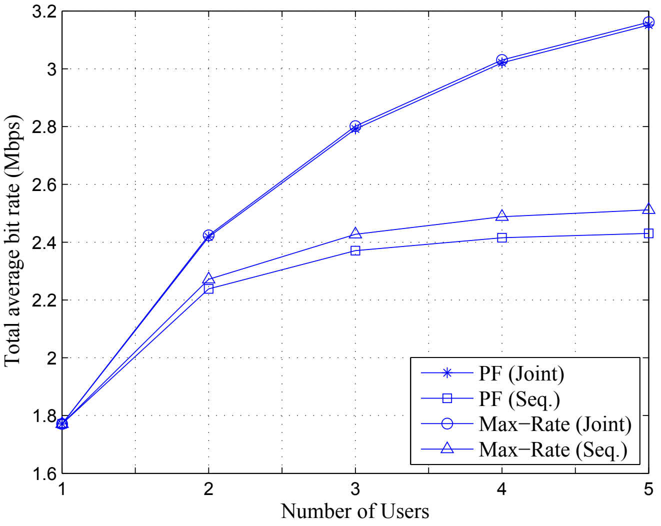


Fig. 13. Average total bit rate as a function of the number of users with $\rho = 0.9$ and average SINRs of 7 dB for all users Kwan et al. (2009a)©[2009] IEEE.

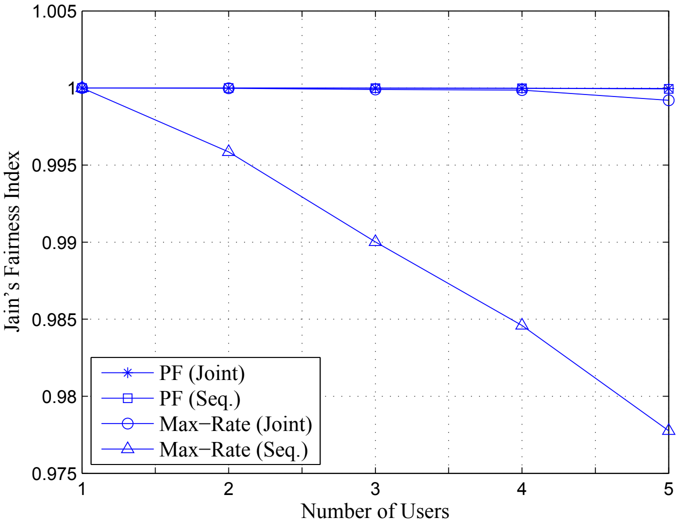


Fig. 14. Fairness index as a function of the number of users with $\rho = 0.9$ and average SINRs of 7 dB for all users Kwan et al. (2009a)©[2009] IEEE.

12

Joint Cooperative Diversity and Scheduling in OFDMA Relay System

Salma Ait Fares[1], Fumiyuki Adachi[1] and Eisuke Kudoh[2]
1-Tohoku University, Japan
2-Tohoku Institute of Technology, Japan

1. Introduction

The most important technical challenge for the realization of 4G mobile networks is to overcome the frequency selective fading channel and significantly reduce the transmitted power from mobile terminals. To meet the challenges of this next generation, wireless *multi-hop virtual cellular network* (VCN) technology is newly proposed (Fitzek & Katz, 2006; Adachi & Kudoh, 2007). Multi-hop VCN is an on-going research and its study and development are still in progress for performance improvements. Before the deployment of Multi-hop VCN technology, many attractive and motivating studies are left in front of researchers. For instance, to avoid large transmission delay, a novel quantitative criterion to determine the optimal number and locations of fixed relay nodes, under the assumption of using orthogonal channels for the hops, has been developed in (Florea & Yanikomeroglu, 2005). To increase the diversity gain, a cooperative diversity between relays and destination has been introduced in (Bletsas et al., 2006) by selecting the best relay to cooperate between the source and the destination. Moreover, to avoid the call blocking when a channel allocation failure occurs, a relay skipping method applied in DS-CDMA system has been presented in (El-Alami et al., 2006). An adaptive antenna array (AAA) has been as well studied in the virtual cellular network (VCN) to optimize the number of relays required in a network area (Adinoyi & Yanikomeroglu, 2006). Also, to enhance the transmit signals from relays to the destination under low transmit power condition, the authors in (AitFares et al., 2009 a) proposed a novel scheme of cooperative relaying network based on data exchange between relays before forwarding data to destination using AAA and transmit beamforming techniques.

In multiuser system, OFDMA based-relay networks have recently received much renewed research interest and recognized as enabling techniques to achieve greater coverage and capacity by exploiting multiuser diversity and allowing efficient sharing of limited resources such as spectrum and transmit power among multiple users (Wittneben et al., 2004; Hammerstrom et al., 2004; Tarasak & Lee, 2007; Tarasak & Lee, 2008). At a particular instant of time, since several users' channels fade differently, the scheduler offer different users the access to the channel based on their channel conditions. In fast-varying fading scenario, if the users have the same average signal-to-noise power ratio (SNR) at their destinations, they will have approximately the same chance to access the channel. However, in slow-varying fading scenario, some users with highest SNR at the destination will access the channel for a long time while unfortunately others have to wait until their channel condition improves.

For such slowly time-varying channel environment, joint cooperative diversity and scheduling (JCDS) technique has been proposed in (Wittneben et al., 2004; Hammerstrom et al., 2004; Tarasak & Lee, 2007; Tarasak & Lee, 2008) to improve the capacity performance. The authors in (Wittneben et al., 2004; Hammerstrom et al., 2004) introduced a time-varying phase rotation in time domain at relay nodes by multiplying each transmit relay signal by a specific phase rotation. This latter creates a time-variant relay fading channel which can be exploited to provide opportunity for every user to be scheduled. For frequency selective fading channel, the works in (Tarasak & Lee, 2007; Tarasak & Lee, 2008) have extended the JCDS technique by introducing cyclic delay diversity (CDD) at the relay nodes in OFDMA system. Using CDD technique, additional fluctuation among the sub-carriers is produced and as a result the scheduler can successfully provide more chance to users to have access to the channel by allocating subcarriers to users whose SNR are highest.

However, the performance of the JCDS depends as well on the cooperative diversity technique used at relay nodes. It has been shown in (Laneman et al., 2003) that using single Amplify and Forward (AF) relay, second order diversity can be achieved. But, it is not necessarily evident to achieve higher order diversity by using several AF-relays. For instance, if some relays receive noisy signals then the noises contained in these received signals are also amplified during a retransmission process. Without any further signal processing, except amplification relay gain, these noisy signals may disturb the received signal at the destination and hence diversity order is reduced. With proper processing of the received signals at the relay nodes, the performance of the JCDS system may perform better by improving the quality of communication links between relays and destinations. For this aim, several algorithms have been proposed in literature known as cooperative distributed transmit beamforming (DTB) for single carrier transmission (AitFares et al., 2009 a; Wang et al., 2007; Yi & Kim, 2007).

In this Chapter, we will introduce the DTB approach to JCDS OFDMA-based relay network in multi source-destination pair's environment and we will highlight its potential to increase the diversity order and the system throughput performance. By jointly employing the JCDS with DTB, the *aggregate throughput*, defined as the total throughput in given physical resources, is enhanced. On the other hand, the *per-link throughput*, defined as the user throughput in a given transmission cycle, is not significantly improved, since the performance of this per-link throughput depends on how many subcarriers are allocated to the user during a given transmission cycle. In addition, to trade-off a small quantity of the aggregate throughput in return for significant improvement in the per-link throughput, we introduce also the fixed CDD approach at relay stations to the proposed JCDS-DTB. Also we prove that in multi source-destination pairs system, combining DTB with CDD at relay nodes creates more fluctuation among subcarriers resulting in time-variant SNR at each destination and consequently gives more opportunity to users to access to the channel.

2. Evolution of wireless mobile communication technology

In the 1980s, first generation (1G) cellular mobile phone, consisted of voice-only analog devices with limited range and features, was introduced. In the 1990s, a second generation (2G) of mobile phones was presented with digital voice/data and with higher data transfer rates, expanded range, and more features. 2G networks saw their first commercial light of day on the *global system* for *mobile* (GSM) standard. In addition to GSM protocol, 2G also utilizes various other digital protocols including CDMA, TDMA, iDEN and PDC. Afterwards, 2.5G wireless technology was established as a stepping stone that bridged 2G to 3G wireless technology. 3G technology was introduced to enable faster data-transmission speeds, greater network capacity and more advanced network services. The first pre-commercial 3G was launched by NTT DoCoMo in Japan in May 2001.

Actually, wireless mobile communications have become very persistent. The number of mobile phones and wireless internet users has increased significantly. The growth of the number of mobile subscribers over the last years led to a saturation of voice-oriented wireless telephony. From a number of 214 million subscribers in 1997 to 4 billion cellular mobile subscribers in 2008 (Acharya, 2008).

However, modern cellular networks need to provide not only high quality voice service for users, but a large amount of data transfer services as well. Users want to be connected with the networks not only for making voice conversations anytime and anywhere with people but also for data downloading/uploading. It is now time to explore new demands and to find new ways to extend the mobile wireless concept.

The evolution of 3G mobile networks will be followed by the development of next generation mobile networks, called 4th generation (4G) or "beyond 3G" mobile phone technology. 4G refers to the entirely new evolution in wireless communications and will support extremely high-speed packet data service 100M–1Gbps (Adachi & Kudoh, 2007) as shown in Fig. 1.

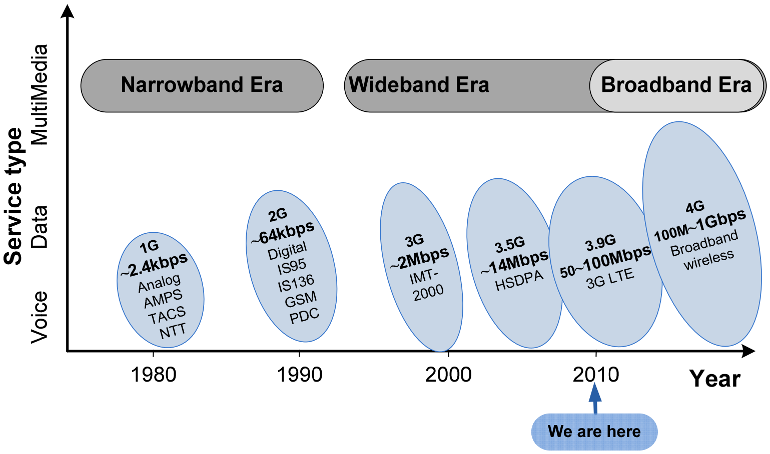


Fig. 1. Wireless mobile communication network evolution.

Although 4G wireless communication systems are expected to offer considerably higher data-rate services and larger coverage areas compared to these older generations, these expectations about wireless communication systems performance appear to be unfeasible in the conventional cellular architecture due to limited transmission capabilities and spectrum efficiency (Adachi & Kudoh, 2007; Adachi, 2008). Indeed, for a peak data rate of

~1Gbps/Base Station (BS), there are two important technical issues to address: (1) to overcome the highly frequency-selective fading channel, and (2) to significantly reduce the transmit power from mobile terminals.

2.1 Spectrum Efficiency Problem

In terrestrial wireless communications, the transmitted signal is reflected or diffracted by large buildings between transmitter and receiver, creating propagation paths having different time delays. For instant, for 1Gbps transmission, 1bit time length is equivalent to the distance of 0.3 m (Adachi, 2008). Then, many distinct multipaths are created, where strong inter-symbol interference (ISI) may be produced. Consequently, the challenge of 4G realization is to transmit broadband data close to 1 Gbps with high quality over such a severe frequency-selective fading channel. In this case, some advanced equalization techniques are necessary to overcome the highly frequency-selective fading channel (Adachi, 2008).

2.2 Transmit Power Problem

In fact, the peak transmit power is in proportion to "transmission rate". Hence, for a very high rate transmission, a prohibitively high transmit power is required if the same communication range in distance is kept as in the present cellular systems. Ignoring the shadowing loss and multipath fading, the energy per bit-to-AWGN (additive white Gaussian noise) power spectrum density ratio E_b/N_0 is given by (Adachi & Kudoh, 2007)

$$\frac{E_b}{N_0} = \left(\frac{P_T/B}{N_0}\right) r_0^{-\alpha}, \tag{1}$$

where P_T is the transmit power, B is the bit rate, r_0 is the cell radius, α is the path loss exponent. We can notice from (1), for a given cell radius r_0, as the bit rate B increases, the transmit power should be increased in order to satisfy the required E_b/N_0. Therefore, keeping the transmit power the same as in the present conventional cellular network, will result in decreasing of coverage of the BS to r_0' as shown in Fig. 2.

For instant, assume that the required transmit power for 8*kbps* at 2*GHz* is 1Watt for a communication range of r_0 =1,000*m*. Since the peak power is in proportion to (transmission rate) x (f_c $^{2.6}$)[Hata-formula] (Kitao & Ichitsubo, 2004) where f_c is the carrier frequency, then, the required peak transmission power for 1Gbps at 3.5GHz needs to be increased by 1Gbps/8kbps x (3.5GHz/2GHz)$^{2.6}$ = 535,561 times, that is, P_T=536kWatt. Obviously, this cannot be allowed. Hence, to keep the 1W power, the communication range should be reduced by 43 times if the propagation path loss exponent is α=3.5. Hence, the cell size should be significantly reduced to r_0'=23m and that leads to increase in the number of BS and consequently gives rise to high infrastructure cost (Adachi, 2008). However, to extend the coverage of BS even at high transmission rate while keeping the transmit power the same as in the present cellular systems, fundamental change in wireless access network is required.

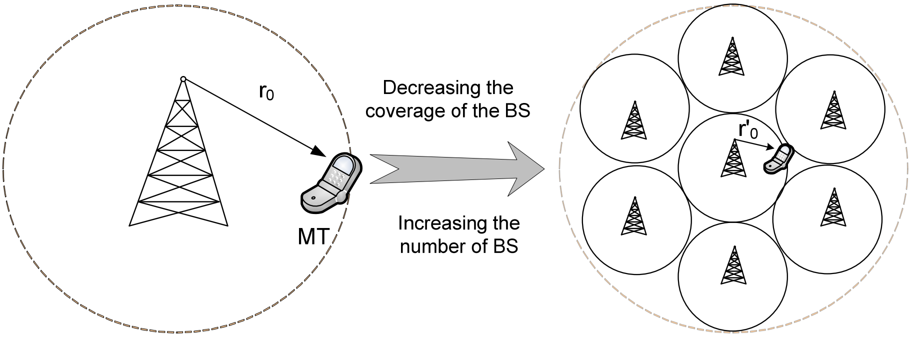


Fig. 2. Decreasing the coverage of BS in the case of keeping the transmit power the same as in the present conventional cellular network for high data rate transmission.

Without reducing the cell size, the direct transmission between widely separated BS and mobile terminal (MT) can be extremely expensive in terms of transmitted power required for reliable communication. Actually, the need of high-power transmissions may increase the co-channel interferences as well as lead to faster battery drain (shorter network life). An alternative approach to direct transmission is to employ *relay* stations as '*intermediate*' nodes to establish *multi-hop* communication links between BS and MT. Such strategies are named as wireless *multi-hop Virtual Cellular Network* (VCN). This architecture consists of a central port (CP), which is the gateway to the network, and many distributed wireless ports (WP) which directly communicate with the mobile terminals. These WPs, often referred to as relay nodes, are used to forward the information of the users having poor coverage to the CP as shown in Fig. 3. The wireless multi-hop VCN will play key roles in future infrastructure-based wireless networks owing to its considerable economical and technical advantages, including: increase system capacity and spectral efficiency, and reduce transmission energy, compared to other network architectures (Dau et al., 2008; Fitzek & Katz, 2006; Adachi & Kudoh, 2007).

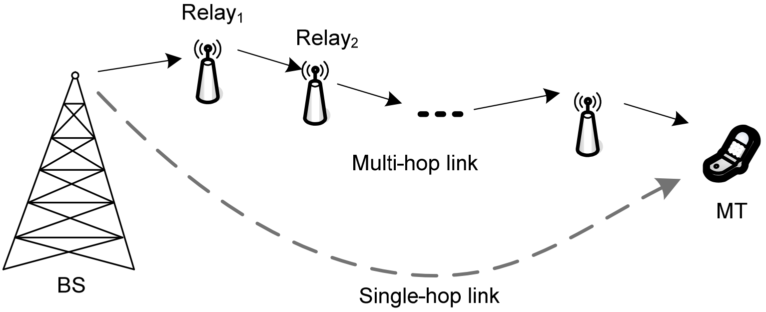


Fig. 3. Multi-hop VCN technology and coverage extension of a multi-hop VCN.

Cooperative relay network is an upgrade technology of multi-hop VCN systems, where relays have to cooperate in relaying information as shown in Fig.4 for 2-hop VCN technology. One advantage of these structures is that it is possible to unite multiple relays in the cellular network as a "*virtual antenna array*" to forward the information cooperatively while an appropriate combining at the destination realizes diversity gain. Therefore cooperative relaying is regarded as a promising method to the challenging of throughput and high data rate coverage requirements of future wireless networks as it provides flexible extension, capacity increase to the conventional wireless systems (Adachi & Kudoh, 2007).

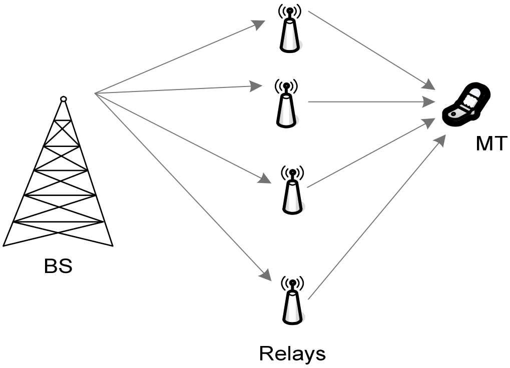


Fig. 4. Cooperative relay network using 2-hop VCN technology.

2.3 OFDMA - based relay in 2-hop VCN technology

OFDM modulation is a bandwidth-efficient technique to obviate inter-symbol interference arising from multipath fading by transmitting multiple narrowband subcarriers. However, in a multipath fading environment, these subcarriers can experience different fading levels; thus, some of them may be completely lost due to deep fading. Cooperative relay network technique may enhance the reliability of subcarriers through redundancy by exploiting the spatial diversity. In fact, since cooperative relay technique provides spatial diversity gain for each subcarrier, the total number of lost subcarriers due to deep fading may be reduced.

On the other hand, in multi-user system, Orthogonal Frequency Division Multiple Access (OFDMA) based relay networks have recently received much renewed research interest and recognized as enabling techniques to achieve greater coverage and capacity by exploiting multi-user diversity and allowing efficient sharing of limited resources such as spectrum and transmit power among multiple users (Tarasak & Lee, 2007; Tarasak & Lee, 2008; AitFares et al., 2009 b). For instance, OFDMA is very flexible since different subcarriers to different users depending on their channel conditions and as several users' channels fade differently, the scheduler offer the access to the channel to different users based on their channel conditions to increase the system capacity.

In multi-user scheduling, the subcarriers can be allocated using private subcarrier assignment (i.e., one user uses private multiple subcarriers at any given time) or shared subcarrier assignment (i.e., several users use a given subcarrier). The subcarriers can be assigned based on each user-destination's SNR or rate maximization technique (Wong et al., 2004). Allocating carriers based on each user's SNR maximizes the total capacity but without being fair to each user. An example is shown in Fig.5 using three user-destination pairs with total number of subcarriers N_c=12.

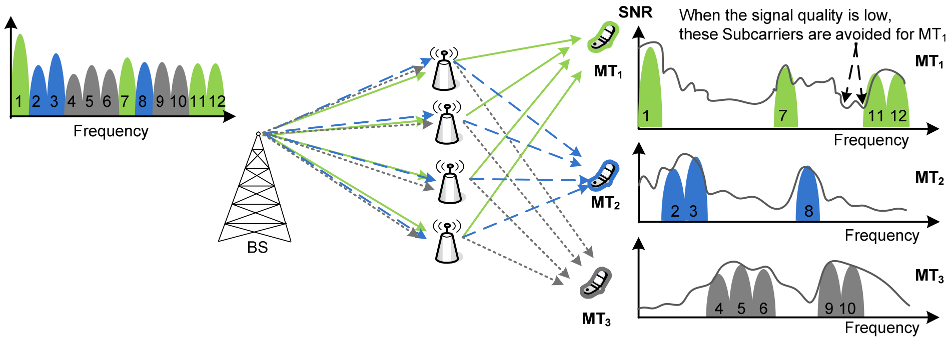


Fig. 5. OFDMA network architecture and Scheduling technique based on SNR assignement approach.

Fig. 6 illustrates an example of the OFDMA transmitter structure for the system at the BS studied in Fig.5 where the subcarrier and power allocations are carried out relying on the feedback information from the scheduler. As shown in this example over one OFDMA symbol, the scheduler chooses the best link (highest SNR) in each subcarrier taking into consideration the channel information at each destination.
OFDMA technology faces several challenges to present efficiency realizations. For instance, if many users in the same geographic area are requiring high on-demand data rates in a finite bandwidth with low latency, a fair and efficient scheduler is required. In addition, to carry out this scheduling, the transmitter needs the channel state information for the different users, and the receiver need information about its assigned subcarriers and all information exchange should be carried out with low overhead.

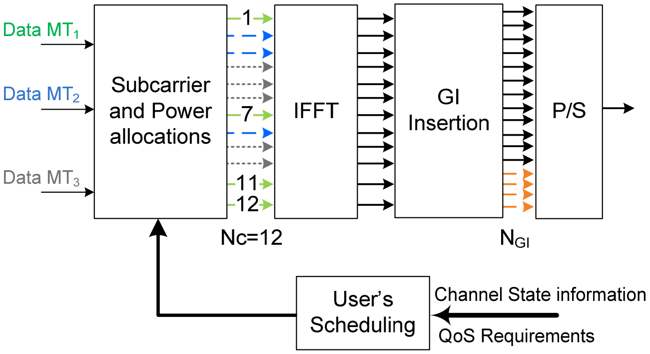


Fig. 6. OFDMA transmitter structure for subcarrier and power allocations at the BS.

2.4 Multi source-destination pairs in OFDMA – based relay in 2-hop VCN technology

OFDMA wireless network architecture in 2-hop VCN technology, illustrated in Fig. 5, can be extended and applied for multi-source destination pairs, where multiple sources communicating with their corresponding destinations utilizing same half-duplex relays as shown in Fig. 7. This kind of network architecture, typically applied in ad-hoc network, presented promising techniques to achieve greater capacity. Analyzing and evaluating the capacity of wireless OFDMA-based relay in multi-source destination pair's networks is one of the most important issues. However, if the wireless nodes are using the same physical resources (i.e., same subcarriers), the problem of evaluating the throughput becomes much more challenging since the transmission of other sources acts as co-channel interference for the others destinations.

In this Chapter, we are interested to study the OFDMA-based relay network in multi-source destination pair's system. In addition, to avoid interferences, instead of using all the orthogonal subcarriers, according to the rate of transmission required by an MT, only the subcarriers with highest received SNR can be allocated independently to the source-destination links.

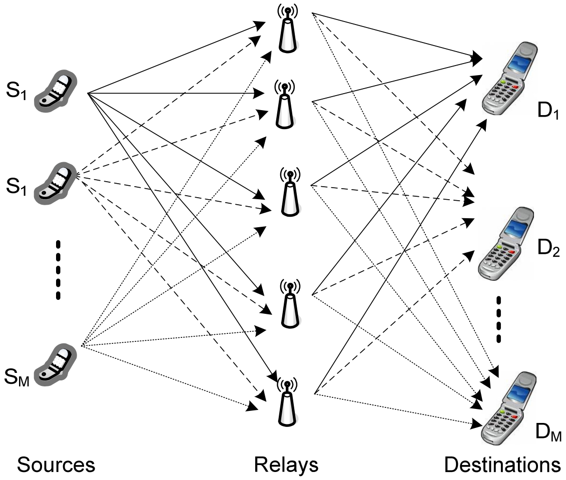


Fig. 7. Multi source-destination pairs via relay routes.

3. JCDS with Distributed Transmit Beamforming and fixed Cyclic Delay Diversity

3.1 System Model

Consider a wireless system composed of M user-destination pairs. R relays are assisting the communication link. Each source needs to communicate with its own destination with the help of these relays. We assume the destinations are far away from sources and there are no direct paths between source-destination pairs. Fig. 8 illustrates an example of the system model with two source-destination pairs (M=2) using four relays (R=4). We assume that the relays operate in duplex mode where in the first time slot, they receive the OFDMA signals from sources that are transmitting simultaneously but with different non-overlapping sub-channels (i.e., a set of OFDM subcarriers), while in the second slot they forward concurrently their received signals to destinations. The channels are assumed time-invariant over one OFDMA block and i.i.d. frequency selective Rayleigh fading with the channel order L. The l-*th* path complex-valued gains of the channels between the i-*th* user and the r-*th* relay and between the r-*th* relay and the i-*th* destination are denoted by $h_{i,r}(l)$ and $g_{r,i}(l)$, respectively. Both $h_{i,r}(l)$ and $g_{r,i}(l)$ are zero mean complex Gaussian random and their variances follow an exponential delay profile such as $E\left[\left|h_{i,r}\right|^2\right] = E\left[\left|g_{r,i}\right|^2\right] = e^{-l/\tau_{rms}} / \sum_{l=0}^{L} e^{-l/\tau_{rms}}$.

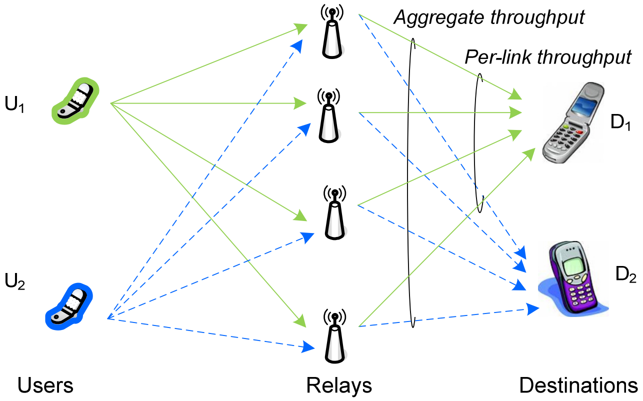


Fig. 8. Multi source-destination pairs in OFDMA 2-hop VCN technology.

The structure of the OFDMA signal transmitted from user U_i is depicted in Fig.9 where N_c represents the N_c-point (I) FFT in the OFDMA transmitters and receivers, N_{ci} is the number of subcarriers allocated to the user U_i, where the remaining subcarriers (N_c-N_{ci}) are padded (e.g., zero padding) and N_{GI} is the guard interval (GI) length and assumed to be longer than the maximum channel delay spread.

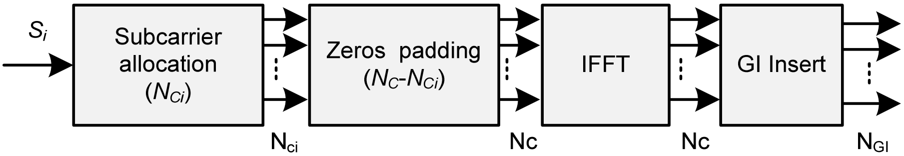


Fig. 9. Transmit OFDMA signal structure and subcarrier allocation scheme.

After removing GI and applying FFT transform the received signal of the p-*th* subcarrier at the r-*th* relay is given by

$$Y_r(p) = \sqrt{P_s} \cdot H_{i,r}(p) \cdot S_i(p) + \eta_r(p), \; r = 1, \cdots, R. \tag{2}$$

where $S_i(p)$ is a unit-energy data symbol transmitted from user U_i ($1 \le i \le M$) whose subcarrier p has been assigned by the scheduler, P_s is the transmit power used by the user U_i, $H_{i,r}(p)$ is the channel gain of the subcarrier p from the i-*th* user to the r-*th* relay and $\eta_r(p)$ is the AWGN's in the corresponding channels with variance σ_R^2. Before forwarding the received signals to the destination, the relays may perform some signal processing as shown in Fig.10

(a, b, and c), such as jointly AF and CDD proposed in (Tarasak & Lee, 2008), jointly AF and DTB or jointly AF, DTB and fixed CDD as will be studied in the following.

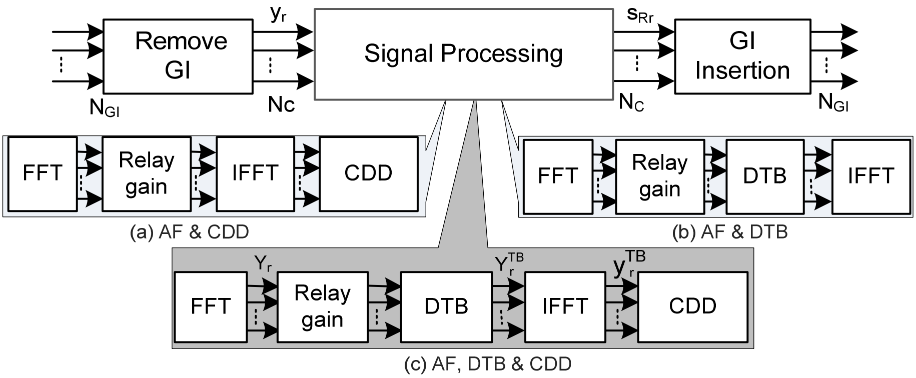


Fig. 10. Relay node structure using different cooperative techniques.

In AF scheme, the relay normalizes its received signal by multiplying it with a relay gain given by

$$\beta_{i,r}(p) = 1/\sqrt{P_s \cdot \left|H_{i,r}(p)\right|^2 + \sigma_R^2}\,,\ r = 1, \cdots, R. \tag{3}$$

With channel order equal to L, the channel gain $H_{i,r}(p)$ at the *p-th* subcarrier can be written as

$$H_{i,r}(p) = \sum_{l=0}^{L} h_{i,r}(l) \cdot e^{-j2\pi lp/N_c}. \tag{4}$$

The output of the transmit beamforming can be expressed by

$$Y_r^{TB}(p) = \beta_{i,r}(p) \cdot W_{TB,r}(p) \cdot Y_r(p),\ r = 1, \cdots, R, \tag{5}$$

where $W_{TB,r}(p)$ represents the weight element of the *p-th* subcarrier at the *r-th* relay.
The received signal at the *i-th* destination after performing FFT is written as

$$Y_i(p) = \sum_{r=1}^{R} G_{r,i}(p) \cdot S_{Rr}(p) + \gamma_i(p), \quad i = 1, \cdots, M. \tag{6}$$

where $S_{Rr}(p)$ is the *p-th* subcarrier component of the OFDMA signal transmitted from the *r-th* relay, $G_{r,i}(p)$ denotes the channel gain at the *p*-th subcarrier from the *r-th* relay to the *i-th* destination, calculated using (4) by replacing $h_{i,r}(l)$ by $g_{r,i}(l)$, and $\gamma_i(p)$ is the AWGN's with variance σ_D^2.
By substituting (2) and (5) into (6), we obtain

$$Y_i(p) = \sqrt{P_s} \cdot S_i(p) \cdot \boldsymbol{a}(p) \cdot \boldsymbol{W}_{TB}^H(p) + v_i(p), \tag{7}$$

where

$$\boldsymbol{a} = \left[\beta_{i,1} \cdot H_{i,1} \cdot G_{1,i}, \dots, \beta_{i,R} \cdot H_{i,R} \cdot G_{R,i}\right], \tag{8}$$

$$v_i = \boldsymbol{\eta} \cdot \boldsymbol{A_m} \cdot \boldsymbol{W}_{TB}^H + \gamma_i, \tag{9}$$

$$\boldsymbol{A_m} = diag\left[\beta_{i,1} \cdot G_{1,i}, \dots, \beta_{i,R} \cdot G_{R,i}\right], \tag{10}$$

$\boldsymbol{W}_{TB} = [W_{TB,1}^{*}, \dots, W_{TB,R}^{*}]$, (.)* is the conjugate and $\boldsymbol{\eta} = [\eta_1, \dots, \eta_R]$.

To ensure that all relays transmit data with total energy *P_r*, the transmit beamforming weight vector should satisfy

$$\boldsymbol{W}_{TB} \cdot \boldsymbol{W}_{TB}^{H} = P_r. \tag{11}$$

From (7), the instantaneous SNR of the *p-th* subcarrier at the *i-th* destination can be expressed as

$$SNR_i(p) = P_s \cdot \frac{\boldsymbol{W}_{TB}(p) \cdot \boldsymbol{a}^H(p) \cdot \boldsymbol{a}(p) \cdot \boldsymbol{W}_{TB}^H(p)}{\boldsymbol{W}_{TB}(p) \cdot \boldsymbol{\Omega}_{\mathrm{m}}(p) \cdot \boldsymbol{W}_{TB}^H(p) + \sigma_D^2(p)}, \tag{12}$$

where

$$\boldsymbol{\Omega}_{\mathrm{m}} = \sigma_R^2 \cdot \boldsymbol{A}_m^H \cdot \boldsymbol{A}_m. \tag{13}$$

Let define $\boldsymbol{\Omega}_{1\mathrm{m}} = \boldsymbol{\Omega}_{\mathrm{m}} + \sigma_D^2/P_r \cdot I_{R\times R}$ and since $\boldsymbol{W}_{TB} \cdot \boldsymbol{W}_{TB}^H = P_r$ is assumed in (11), (12) can be written as

$$SNR_i(p) = P_s \cdot \frac{\boldsymbol{W}_{TB}(p) \cdot \boldsymbol{a}^H(p) \cdot \boldsymbol{a}(p) \cdot \boldsymbol{W}_{TB}^H(p)}{\boldsymbol{W}_{TB}(p) \cdot \boldsymbol{\Omega}_{1\mathrm{m}}(p) \cdot \boldsymbol{W}_{TB}^H(p)}. \tag{14}$$

From (14), the source destination channel capacity of the *p-th* subcarrier for the *i-th* user is given by

$$C_{i,p}(p) = \frac{B}{2N_c} \cdot log_2\big(1 + SNR_i(p)\big), \tag{15}$$

where B is the total bandwidth. It can be seen from (15) that in order to maximize the aggregate channel capacity, each destination's SNR should be maximized at each subcarrier. Therefore, we develop in the following section a transmit beamforming technique that maximizes the SNR at each destination and for each subcarrier.

3.2 Derivation of the distributed transmit beamforming weight

To combat fading effects and then improve the link level performance, the distributed spatial diversity created by the relay nodes can be effectively exploited using a transmit diversity weight technique. To determine the transmit beamforming vector we develop the optimal weight vector that maximizes the SNR at the destination given by (14), as

$$\boldsymbol{W}_{TB} = arg \max_{\boldsymbol{W}} \left(\frac{\boldsymbol{W} \cdot \boldsymbol{a}^H \cdot \boldsymbol{a} \cdot \boldsymbol{W}^H}{\boldsymbol{W} \cdot \boldsymbol{\Omega}_{1\mathrm{m}} \cdot \boldsymbol{W}^H} \right). \tag{16}$$

The weight optimization criterion expressed by (16) is in the form of Rayleigh quotient, and can be derived by solving the generalized Eigen-value problem (Yi & Kim, 2007; AitFares et al., 2009 a). Hence, for any weight vector $\boldsymbol{W}_{TB}$, we have

$$SNR_i \leq \lambda_{max}, \tag{17}$$

where λ_{max} is the largest Eigen-value of $\left(\boldsymbol{\Omega}_{1m}^{H/2}\right)^{-1} \boldsymbol{a}^H \boldsymbol{a} \left(\boldsymbol{\Omega}_{1m}^{1/2}\right)^{-1}$.

The equality holds if

$$\boldsymbol{W}_{TB}(p) = c \cdot \boldsymbol{a}(p) \cdot \boldsymbol{\Omega}_{1m}^{-1}(p), \tag{18}$$

where

$$c = 1/\sqrt{|\boldsymbol{a} \cdot \boldsymbol{\Omega}_{1m}^{-1}|^2}. \tag{19}$$

By using this derived optimal transmit beamforming that maximizes the SNR at the destination, the aggregate channel capacity is significantly enhanced while in parallel the per-link capacity is not much improved and in particularly in slow-varying fading scenario. To overcome this problem we applied the fixed cyclic delay diversity (CDD) approach (Tarasak & Lee, 2007; Tarasak & Lee, 2008) in the time domain (after IFFT) at relay nodes as shown in Fig. 10 (c) in order to create a phase rotation in frequency domain and hence the scheduler will offer opportunity to more users to get channel access. Hence, after performing IFFT at the *r-th* relay, the output of the fixed CCD block is given by

$$y_r^{CDD}(l) = y_r^{TB}(l - \Delta_r),\ r = 1, \cdots, R \tag{20}$$

where $y_r^{TB}(l)$ represents the *l-th* element of the IFFT of the $Y_r^{TB}(p)$ signal and Δ_r represents the cyclic delay value used at the *r-th* relay. Δ_r is selected as a fixed cyclic delay given by

$$\Delta_r = Int\left(\frac{N_c}{R} \cdot (r-1)\right), r = 1, \cdots, R \tag{21}$$

where $Int(x)$ represents the nearest integer function of x.

Subsequently by using the fixed CDD approach, the instantaneous SNR given in (14) is expressed as

$$SNR_i(p) = P_s \cdot \frac{\boldsymbol{W}(p) \cdot \boldsymbol{a}^H(p) \cdot \boldsymbol{a}(p) \cdot \boldsymbol{W}^H(p)}{\boldsymbol{W}(p) \cdot \boldsymbol{\Omega}_{1m}(p) \cdot \boldsymbol{W}^H(p)}. \tag{22}$$

where

$$\boldsymbol{W}(p) = \boldsymbol{W}_{TB}(p) \cdot diag(\boldsymbol{a}_1(p)), \tag{23}$$

and

$$\boldsymbol{a}_1(p) = \left[e^{-j\frac{2\pi p \Delta_1}{N_c}}, \cdots, e^{-j\frac{2\pi p \Delta_R}{N_c}}\right]. \tag{24}$$

An adaptive scheduling in OFDMA-based relay network is adopted to allocate the subcarriers to each source based on SNR channel assignment approach. This adaptive scheduler allocates the *p-th* subcarrier to the *i-th* user destination pair with the highest SNR such that

$$U_i(p) = arg \max_{i \epsilon \{1, \cdots, M\}} (SNR_i(p)), \qquad p\, \epsilon \{1, \cdots, N_c\} \tag{25}$$

Two significant measured performances, highlighted in Fig.8, are studied, the *aggregate throughput* and the *per-link throughput*. By ignoring the loss from GI, the aggregate throughput (in bit per complex dimension) is expressed by

$$C_{agr} = \frac{1}{2N_c} \cdot \sum_{p=0}^{N_c - 1} log_2(1 + SNR_i(p)). \tag{26}$$

While the per-link throughput or average user throughput is defined by

$$C_{per-link} = M \cdot \frac{1}{2N_c} \cdot \sum_{p \epsilon \Gamma} log_2(1 + SNR_i(p)), \tag{27}$$

where Γ is the set of subcarriers allocated to the *i-th* user and M represents the number of user-destination pairs.

It should be noticed from (26-27) that increasing the user-destination pairs increases the aggregate throughput while the per-link throughput is reduced since the number of allocated subcarriers for each user is largely reduced. Hence using our proposed JCDS with adaptive scheduling based on SNR channel assignment; a trade-off between aggregate throughput and per-link throughput is achieved and that guarantees the per-link throughout to have at least the same QoS as in the static scheduling (SS) where all users get an equal share of the allocated resources.

4. Computer simulation results

In this section, we compare the performance of the proposed JCDS using both DTB and fixed CDD with different cooperative diversity techniques such as JCDS with DTB, JCDS with AF and JCDS with fixed CDD where adaptive scheduling based on SNR channel assignment is employed. This adaptive scheduler allocates the subcarriers to the source whose SNR is highest as illustrated in the example shown in Fig. 5. Both techniques, JCDS-AF and JCDS-CDD, are using equal divided transmit power at relay stations, i.e., $P=P_r/R$. While, in JCDS-DTB the relays are using DTB under constraint of (11). We evaluate the system performance by taking the same simulation scenario presented in (Tarasak & Lee, 2007) for comparison purpose. In this scenario, two types of fading are studied, the flat fading where the normalized *rms* delay spread (τ_{rms}) is relatively short and equals to 0.3; corresponding to L=3, and the frequency selective fading where the normalized *rms* delay spread is relatively large and equal to 1.5; corresponding to L=15. The number N_c of subcarriers is equal to 256, R=20 and the average SNR at the relay and at the destination are defined to be the same 20dB which is equivalent to $\sigma_R^2 = \sigma_D^2 = 0.01$.

Fig. 11 illustrates the cumulative distribution functions (CDF) of the aggregate throughput, $P(C_{agr}<\text{throughput})$, and the per-link throughput, $P(C_{per\text{-}link}<\text{throughput})$ in short delay spread scenario ($\tau_{rms} = 0.3$) using our proposed method; i.e., JCDS with DTB and CDD for different user-destination pairs. Aggregate and per-link throughput's results are shown by solid and dashed lines, respectively. A comparison of the static scheduling with R=1 (single relay node), in which the aggregate throughput and per-link throughput are equal, is also studied. It should be noticed that when M=1 (single source-destination pair), the aggregate throughput is equal to the per-link throughput and the employed adaptive scheduler is equivalent to the static scheduling. Hence, from Fig. 11, by comparing the throughput using static scheduling and R=1 with that of our proposed method using M=1 and R=20, we can see clearly the cooperative relay diversity gain.

Furthermore, we can observe as well the user diversity effect in both aggregate and per-link throughputs. It is intuitively clear that when the number of users increases the aggregate throughput is improving since the scheduler switches to the user whose link is better. In contrast, the per-link throughput is decreasing when the number of source-destination pairs is getting higher. Thus the QoS of each source-destination pair is severely affected due to the reduced number of assigned subcarriers. In addition, at 1% outage per-link throughput, if we want to maintain the per-link throughput at least equal to that of static scheduling, it is seen that 5 users can be handled by this system.

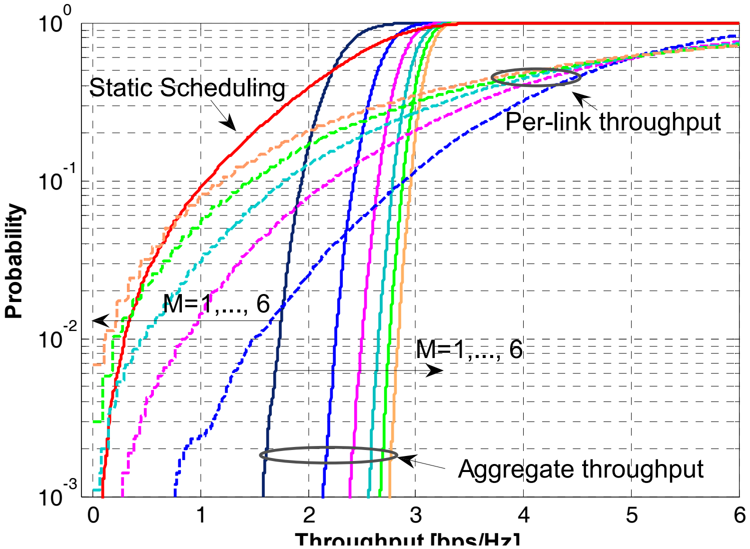


Fig. 11. CDF of the aggregate and per-link throughput for short delay spread ($\tau_{rms} = 0.3$) using JCDS with DTB and fixed CDD.

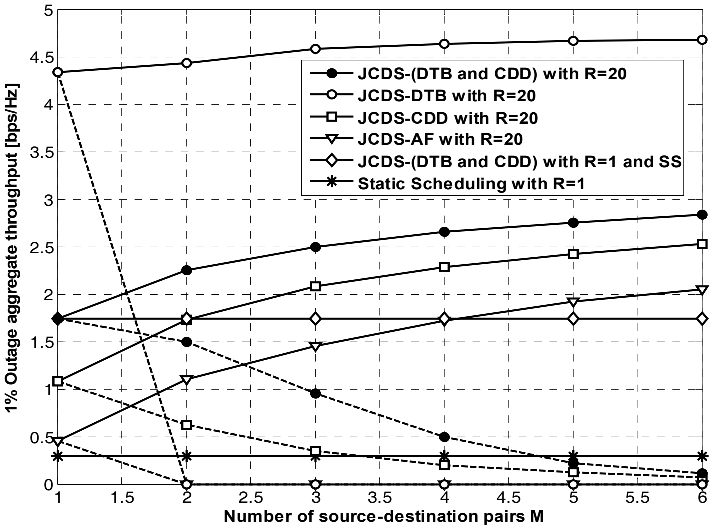


Fig. 12. 1% outage throughput comparaison for short delay spread ($\tau_{rms} = 0.3$) using different cooperation diversity methods.

Fig. 12 compares the 1% outage aggregate throughput and 1% outage per-link throughput, using different cooperative diversity and scheduling approach in short delay spread scenario in terms of the number *M* of source-destination pairs. Aggregate and per-link throughput's results are shown by solid and dashed lines, respectively. From this figure, and for aggregate throughput curves comparison, we notice that the aggregate throughput of all cooperative diversity techniques exceeds that of the static scheduling with R=1. In addition, the throughput using JCDS - DTB is the largest followed by JCDS with DTB & CDD and both of them significantly outperform the JCDS with DTB & CDD using static scheduling with R=20. Moreover, the aggregate throughputs obtained by using JCDS-CDD and JCDS-AF exceed that of using JCDS with DTB & CDD using static scheduling with R=20 when M>1 and M>3, respectively. However, for per-link throughput curves comparison in the same figure, we notice that the per-link throughputs of all cooperative diversity techniques are lower than those of the JCDS–(DTB & CDD) with R=20 and using static scheduling. This indicates the impact of the unfair SNR assignment relative to user throughput. In addition, the per-link throughput using JCDS-(DTB & CDD) is the largest followed by JCDS–CDD. The JCDS–(DTB & CDD) achieves the highest throughput while JCDS-DTB and JCDS-AF have the worst performance. By comparing aggregate and per-link throughputs of JCDS-(DTB & CDD) with JCDS–DTB, the proposed method JCDS–(DTB & CDD) sacrifices a small quantity of the aggregate throughput in return for significant improvement in the per-link throughput.

Using the same propagation environment but having long normalized delay spread (τ_{rms} = 1.5), Fig. 13 illustrates CDFs of the aggregate throughput and per-link throughput using our proposed JCDS with DTB and CDD method for different number *M* of source-destination pairs. Similar observation given in Fig.11 can be provided herein regarding the cooperative diversity gain and user diversity gain which are positive on the aggregate throughput and negative on the per-link throughput. However, we notice that all per-link throughput curves exceed largely that of static scheduling with $R=1$. This can be explained by the higher multipath diversity in the large delay spread scenario where more fluctuation in frequency domain is provided and hence more users are scheduled.

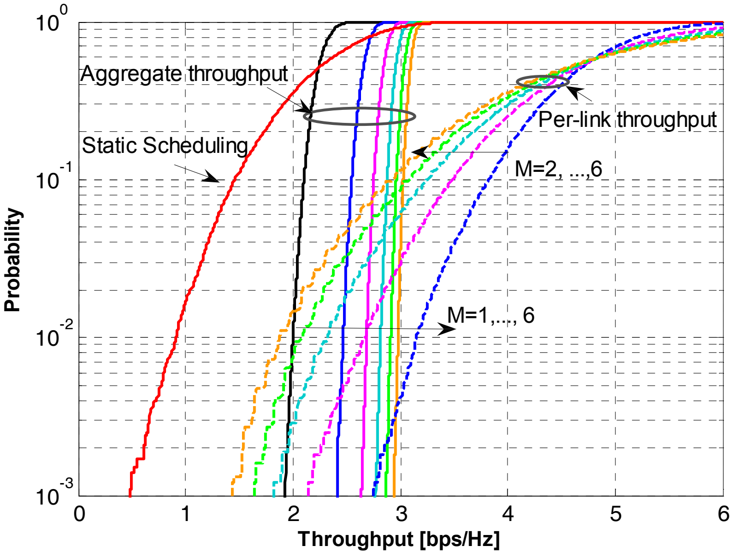


Fig. 13. CDF of the aggregate and per-link throughput for long delay spread (τ_{rms} = 1.5) using JCDS with DTB and CDD.

Fig. 14 compares the 1% outage aggregate and per-link throughputs, using different cooperative diversity and scheduling approaches in long delay spread scenario (τ_{rms} = 1.5) in terms of the number M of source-destination pairs. Aggregate and per-link throughput's results are shown by solid and dashed lines, respectively. From this figure, and for aggregate throughput curves comparison, the gain of the aggregate throughput using all JCDS techniques with static scheduling is smaller compared to short delay case (τ_{rms} = 0.3). Also, the aggregate throughput of all cooperative diversity techniques exceeds that of the static scheduling. However, for per-link throughput curves comparison in the same figure, we notice that the per-link throughput of JCDS-(DTB& CDD) is the best followed by JCDS-CDD and both of them outperform that of static scheduling with R=1 irrespective to the number M of source-destination pairs. This result is due to the increase in frequency selectivity in such long delay spread channel. Moreover, the JCDS-(DTB & CDD) achieves highest throughput while JCDS with DTB provides the lowest throughput when M>1. When M=2, the JCDS- (DTB & CDD) achieves the highest throughput due to the increase in path and user diversities.

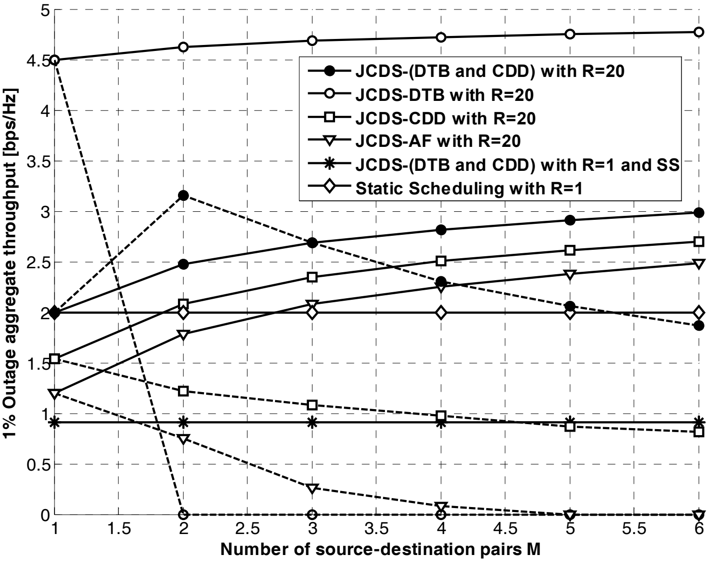


Fig. 14. 1% outage throughput comparaison for long delay spread ($\tau_{rms} = 1.5$) using different cooperation methods.

5. Conclusion

In this Chapter, we studied the JCDS for active source/destination pairs in OFDMA-based relay system. We investigated the performance of the JCDS technique by introducing and developing a distributed transmit beamforming approach jointly used with cyclic delay diversity (CDD) at the relay nodes. Combining transmit beamforming with fixed CDD approach creates more fluctuation among subcarriers and gives more opportunity to users to access to the channel and thus to increase considerably the diversity order and the throughput performance. Therefore, time-varying SNR at the destination is created and more users can be efficiently assigned by the scheduler. Simulation results show that the system performance using our proposed JCDS improves considerably compared with the other previously proposed cooperative diversity techniques.

6. References

Acharya, S. (2008). Worldwide mobile cellular subscribers to reach 4 billion mark late 2008, *International Telecommunication Union Statistics*, 2008, http://www.itu.int/ITU-D/ict/statistics.

Adachi, F. & Kudoh, E. (2007). New direction of broadband wireless technology, (invited) *Wireless Communications and Mobile Computing*, vol. 7, no. 8 (Special Issue on Asia-Pacific B3G R&D Activities and Technology Innovations), pp.969-983, Oct. 2007.

Adachi, F. (2008). Challenge for 4G wireless, *GCOE Workshop on Advanced Wireless Signal Processing and Networking Technology Tohoku University*, 20-21 Aug. 2008. http://www.mobile.ecei.tohoku.ac.jp/COE/workshop_2008_08/pdf/Prof.%20Adachi.pdf.

Adinoyi, A. & Yanikomeroglu, H. (2006). Multi antenna aspects of wireless fixed relays, in *Proc. of IEEE Wireless Communications* & *Networking*, vol.2, pp. 1021-1026, 2006.

AitFares, S.; Adachi, F & Kudoh,E. (2009). Novel Cooperative Relaying Network Scheme with Inter-Relay Data Exchange, *IEICE Transaction on Communications*, vol. E92-B, no.05, May 2009.

AitFares, S.; Adachi, F & Kudoh, E. (2009). Joint cooperative diversity and scheduling with distributed transmit beamforming in OFDMA relay system, Submitted to Personal, *Indoor and Mobile Radio Communications Symposium* 2009 (PIMRC'09), April 2009.

Bletsas et al., (2006). A simple cooperative diversity method based on network path selection, *IEEE Journal on Selected Areas in Comm.*, vol. 24, no. 3, pp. 659-672, 2006.

Dau, I; Kudoh, E. & Adachi, F. (2008). Multi-hop link capacity of multi-route multi-hop MRC diversity for a virtual cellular network, *IEICE Trans. Commun.*, vol.E91-B, no.5, pp. 1568-1574, May, 2008.

El-Alami, L.S.; Kudoh, E. & Adachi, F. (2006). Blocking probability of a DS-CDMA multi-hop virtual cellular network, *IEICE Trans. Fundamentals*, vol. E89-A, no.7, pp.1875-1883, July 2006.

Fitzek, F. H. P. & Katz, M. D. (2006). Cooperation in wireless networks: Principles and Applications, Springer, 2006.

Florea, A. & Yanikomeroglu, H. (2005). On the optimal number of hops in infrastructure-based fixed relay networks, in *Proc. of IEEE Globecom*, vol. 6, pp. 3242- 3247, 2005.

Hammerstrom, I.; Kuhm, M. & Wittneben, A. (2004). Cooperative diversity by relay phase rotations in block fading environments, in *Proc. Fifth IEEE Workshop on SPAWC*, pp. 293-297, July 2004.

Kitao, K.; Ichitsubo, S. (2004). Path loss prediction formula for microcell in 400 MHz to 8 GHz band, *Electronics Letters*, vol. 40, no. 11, pp. 685 – 687, May 2004.

Laneman, J. N; Tse, D. N. & Wornell, G. W. (2003). Cooperative diversity in wireless networks: Efficient protocols and outage behavior, *IEEE Trans. Inform*. Theory, 2003.

Tarasak, P. & Lee, Y. H. (2007). Joint cooperative diversity and scheduling in OFDMA relay systems, in *Proc. IEEE WCNC'2007*, pp. 980-984, March 2007.

Tarasak, P. & Lee, Y. H. (2008). Joint cooperative diversity and proportional fair scheduling in OFDMA relay systems, in *Proc. IEEE, VTC'2008*, pp.1-5, Sept. 2008.

Wang, C.; Yuan, T. & Yang, D. (2007). Cooperative relay network configuration with spatial multiplexing and beamforming, *International Conf. on Wireless Communications, Networking and Mobile Computing*, pp.137 – 140, Sept. 2007.

Wittneben, A. et al, (2004). Joint cooperative diversity and scheduling in low mobility wireless netwotks, in *Proc. IEEE Globecom*, vol. 2, pp. 780-784, Dec. 2004.

Wong, I.C.; Shen, Z.; Evans, B.L.& Andrews, J.G. (2004). Low complexity algorithm for proportional resource allocation in OFDMA systems, in *Proc. IEEE Workshop on Signal Processing Systems*, pp. 1-6, Dec. 2004.

Yi, Z. & Kim, I. (2007). Joint optimization of relay-precoders and decoders with partial channel side information in cooperative networks, *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 447-458, Feb. 2007.

13

Performance Modelling and Analysis of Mobile Wireless Networks

Carmen B. Rodríguez-Estrello[1], Genaro Hernández Valdez[2]
and Felipe A. Cruz Pérez[1]
Electric Engineering Department, CINVESTAV-IPN[1]
Electronics Department, UAM-A[2]
Mexico

1. Introduction

Nowadays, mobile wireless communications are continuously experiencing a growing demand. Therefore, communication systems must be designed and/or enhanced to increase their capacity. In that sense, mathematical analysis must be the first step of designing or improving a network. Thus, powerful mathematical tools which take into account most of the involved parameters in network performance are required to analyze mobile wireless networks.

Wireless networks could be analyzed from link level approach or from system level approach in terms of Quality of Service (QoS). Link level analysis is related to the physical channel characterization and involves statistics such as the probability distributions of the channel states duration, which, in general, are not easily obtained at real cellular networks. In contrast, system level analysis is associated with the characterization of the network's dynamic and involves variables such as channel holding times for successfully and forced terminated calls which are easily obtained at real networks.

QoS in mobile wireless networks means the level of usability and reliability of a network and its services. Consequently, QoS for mobile wireless networks are the basis of dimension and planning. The main concern for an operator is the accessibility and continuity of the connection. As a result, it has been widely accepted that call forced termination probability is one of the most important QoS performance metrics in cellular networks. Forced termination is due to two fundamental features: resource insufficiency and link unreliability. In order to adequately model mobile cellular networks at system level, its mathematical analysis should consider both causes of call forced termination: resource insufficiency and wireless link unreliability. In the literature, resource insufficiency has been widely studied at system level while link unreliability has not been included at system level analysis due to its inclusion entails.

On the other hand, Code Division Multiple Access (CDMA) has been selected as the main multiple access technology of several third generation cellular network standards (Dahlman et al., 2007). CDMA-based cellular systems employ universal frequency reuse factor which makes them interference limited. Consequently, capacity is a direct function of interference generated by the users. That is, the number of available radio resources depends on the interference. This special feature is known as “soft capacity”. Thus, to correctly asses the performance of a CDMA system it is imperative to consider the effect of interference. The aim of this chapter is to present the mathematical analysis of a CDMA-based system considering link unreliability in a system level analysis, which has been traditionally considered only at link level analysis.

2. Overview of call forced termination analysis in mobile wireless networks

One of the most important QoS metrics for the performance evaluation of present and future mobile wireless networks is call forced termination probability. The call forced termination probability is the probability that a call which is not initially blocked be interrupted. In mobile wireless networks, a call is forced to terminate because of two fundamental features: resource insufficiency and link unreliability. Moreover, in the context of packet switched mobile communication networks, call forced termination probability is especially important in the performance evaluation of conversational and real-time services (i.e., voice, audio, music, videophone, videoconference etc.) (Li et al., 2004; Wong et al., 2004).

2.1 Resource insufficiency

When a mobile user moves into a different cell during the course of a call, a handoff must be performed. If no radio resources are available in the target cell, the call is said to be forced to terminate due to resource insufficiency. However, in a well-established cellular network and from the call forced termination point of view, handoff failure can be usually a negligible event (Boggia et al., 2005).

2.2 Link unreliability

Physical link is said to be unreliable if the experienced signal-to-interference ratio (SIR) is below than a minimum required value (SIR threshold) for more than a specified period of time (time threshold). During the course of a call, the physical link between base station and mobile station may suffer link unreliability due to propagation impairments such as multi-path fading, shadowing or path loss, and interference. In CDMA-based systems, link unreliability is experimented due to the initial power allocation. The initial handoff decisions are made for individual connections independent of the other connections or the BS power availability. Therefore, current connections may suffer from link unreliability (Zhao et al., 2006). Hence, the call may be abnormally terminated. In this case, the call is said to be forced to terminate due to link unreliability. In particular, the analysis of measured data traffic supplied by Vodafone Italy is a good example for these phenomena (Boggia et al., 2005). However, only relatively few recent studies have addressed the effect of link unreliability on the performance of mobile wireless communication networks (Elsharabwy & Le-Ngoc, 2005; Zhang & Song, 2006; Zhang & Song, 2005; Zhao et al., 2006; Liu & Sule, 2004; Naraghi Pour & Chai, 2006; Kong, 2002).

Most of the work devoted to study the impact of link unreliability on system performance has considered a link level channel model in order to study how the channel impairments affect system performance. Gilbert Elliot and Fritchman channel models have been widely used for this purpose. For instance, authors of (Kong, 2002) proposed a queuing system with impaired wireless channel based on the Markov chain approach assuming that the unreliable wireless channel can be modeled by the Gilbert–Elliott channel model. Analysis is performed assuming that each state of the wireless channel is a node. Then, service rate becomes time-varying due to propagation impairments. A model to quantify the performance of a queue with respect to such impaired wireless channel is then developed. In a similar work (Elsharabwy & Le-Ngoc, 2005), the Gilbert-Elliot channel model is proposed for the downlink performance evaluation of WCDMA cellular systems. In addition, the Gilbert-Elliot channel parameters in terms of the mean fade and non-fade durations are obtained. The proposed QoS performance metrics are based on the satisfied-user criteria recommended by UMTS (i.e., satisfied user probability for speech services and satisfied user probability for data services). Specifically, (Elsharabwy & Le-Ngoc, 2005), deals with the UMTS QoS recommendation for packet based networks. Authors in (Elsharabwy & Le-Ngoc, 2005), propose a composite performance index (called satisfied-user probability) based on dropped-call probability and session outage percentage due to link unreliability in wireless communication networks. This performance index is calculated by considering a Gilbert-Elliot channel model with negative exponentially distributed state durations. However, none of the above papers considers users' mobility. Consequently, call forced termination due to resource insufficiency is not addressed.

Other related papers devoted to study the impact of link unreliability on system performance are (Zhang & Song, 2006), (Zhang & Song, 2005), which use either the Gilbert-Elliot or Fritchman model to characterize the time-variant wireless channel. Zhang et al. derived mathematical expressions for the probability that a call be successfully completed considering the concurrent impacts of bad quality in the channel and the lack of radio resources. Zhang studied the impact of Rayleigh fast-fading on various teletraffic QoS metrics in wireless networks (i.e., channel holding time, handoff probability, handoff call arrival rate, call blocking probability, call completion probability, and call forced termination probability) taking into account carrier frequency, maximum Doppler frequency and fade margin. From teletraffic point of view, system level-based modeling of link unreliability is preferred over link level-based modeling because fewer state variables are needed. However, mathematical models considered in (Elsharabwy & Le-Ngoc, 2005), (Zhang & Song, 2006), (Zhang & Song, 2005), (Zhao et al., 2006), (Naraghi Pour & Chai, 2006), (Kong, 2002) are based on link level statistics which are not easily obtained by direct measures. Contrary to the Zhang's works, here, the effect of link unreliability is captured through easily obtained system level quantities which allows including the effect of link unreliability on the channel occupancy directly in the teletraffic analysis.

Only few recently published studies have addressed system level analysis considering link unreliability for wireless networks (Liu & Sule, 2004; Rodríguez-Estrello et al., 2009). In (Liu & Sule, 2004), a queuing model to evaluate the performance of CDMA reverse link in a multiple cell scenario was developed. In that work, a quasi-birth-and-death process was used to capture the variation of traffic loads in cells. Then, authors of (Liu & Sule, 2004), obtained the stationary distribution of the system and some performance indicators, such as the outage probability of existing calls, blocking probability of new calls, average carried traffic in a cell, and dropping frequency of ongoing calls. Nonetheless, neither mobility of users nor soft handoff are modeled in (Liu & Sule, 2004). Recently, a teletraffic model to evaluate the performance of TDMA/FDMA-based cellular networks considering both resource insufficiency and link unreliability (Rodríguez-Estrello et al., 2009) was proposed.

The effect of link unreliability is captured by an interruption Poisson process, which is characterized by the mean time of the "*unencumbered call interruption time*". This interruption process is characterized via system level statistics, based on channel holding time, which is easily measured at base stations (BSs).

3. System model

To capture the main features of CDMA-based cellular systems in their performance evaluation, geometrical, users' mobility, interference, and call interruption characteristics are considered.

3.1 Soft handoff geometrical model

A homogeneous mobile multi-cellular system with omni-directional antennas located at the center of cells is assumed[1]. Base Stations (BSs) are assumed to use Frequency Division Duplexing (FDD). As in previously published related studies (Zhang & Lea, 2006; Ma et al., 2006; Hegde & Sohraby, 2002; Piao et al., 2006; Su et al., 1996 and Kim & Sung., 1999), we focus on the reverse link as it was found to be the link that limits system performance.

Soft handoff process is performed when a MS receives comparable pilot signal strengths from two or more BSs. At this moment, a communication path is established between the MS and all BSs with comparable pilot signal strengths. Consequently, two or more BSs receive independent streams from the MS. Independent streams can be combined (macro diversity) so that the bit stream is decoded much more reliable than if only one BS were receiving from the MS. When a pilot signal from one BS is considerably stronger than that from the others BSs, the MS is then served by only one BS. Then, soft handoff process can guarantee that the MS is always linked to the BS from which it receives the strongest pilot (Garg, 2000). Here, it is assumed that the MS can communicate with the two nearest BSs only, and, if only propagation path losses are considered[2], the region where the soft handoff process is performed is the ring area near the borders of the cell. Figure 1 depicts the geometry of the analyzed cells. Ring area is used to represent the area where the soft handoff process is performed.

Thus, an active user in the inner area, referred to as the hard region, is assumed to be connected only to the nearest BS; while mobiles in the outer area, referred to as the soft region, are assumed to be in soft handoff to its two nearest BS's only. As seen in Figure 1, it is clear that the "nominal" coverage of the analyzed cell is increased by the soft regions of the adjacent cells.

The ratio between the area of the hard region to the total cell area (including the overlapped area of the soft regions) is denoted by p while p' is the ratio between the area of the hard region to the nominal area of the cell, which is the area of the cell without considering overlapped areas of soft handoff regions.

---

[1] To simplify mathematical analysis, a first approach is to consider the use of omni-directional antennas. Nonetheless some minor modifications need to be done when directional case is considered.

[2] This consideration is acceptable under not severe conditions of shadowing.

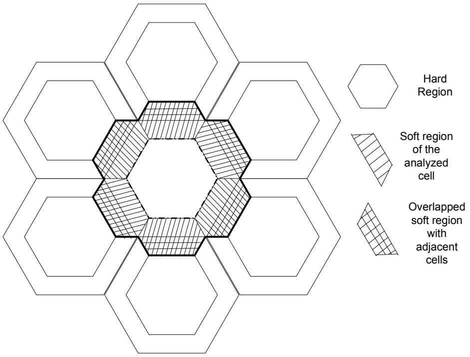


Fig. 1. Network Topology

3.2 Mobility model

Macroscopic modeling of mobility in cellular systems is used here. Since the geometrical model previously described considers differentiated coverage regions, users' mobility should be characterized in terms of mean residence times in different coverage zones and the probability q that a user carrying a call originated in the soft region moves to the hard region.

To determine the specific macroscopic statistics, the generalized smooth random mobility model proposed in (Zoonozi & Dassanayake, 1997) is used here due to its simplicity and versatility to represent several scenarios. The model proposed in (Zoonozi & Dassanayake, 1997) is characterized by the parameter α that limits the range of maximum variation of the current direction of a user, which allows for the representation of different mobility scenarios.

In order, to take into account correlated directions of movement characteristic of smooth random mobility models, probability q takes into consideration the current coverage region as well as the previously visited region (if any) before entering the region under analysis. Hence, three different cases are distinguished in the calculation of q:

1) calls that are originated in the soft region (q_s)
2) users that arrived to the soft region of a given cell coming from the hard region of the same cell (q_{sh})[3]
3) users that arrived to the soft region of a given cell coming from the soft region of another cell (q_{ss}).

Residence times in different coverage regions also depend on the correlated moving direction of users. Hence, five different cases are considered:

a) residence time in the soft region when users originate a call in this region (T_s),

---

[3] This probability is zero when a linear movement inside cells is considered, because there is no possibility to return to the previous region. As such, when linear movement is considered, it is not necessary to take into account this probability.

b) residence time in the soft region when users come from a hard region (T_{sh}),
c) residence time in the soft region when users come from the soft region of other cell (T_{ss}),
d) residence time in the hard region when users originate a call in this region (T_h),
e) residence time in the hard region when users come from the soft region (T_{hs}).

3.3 Interference model

Interference power is proportional to the system load and has random characteristics depending on several system variables such as shadowing losses and propagation characteristics. Interference in CDMA systems could be intra-cellular (interference caused by the users which are power controlled by the own-cell) and inter-cellular (caused by the users which are power controlled by another cell).

In this work, as in (Viterbi et al., 1994) interference power is modelled by means of a random variable, **Z**, which represents the power of the total interference caused by all users in the system.

Since CDMA systems are interference-limited, transmitted power from each mobile user must be controlled to limit interference. However, the received power level should be sufficient to guarantee an adequate energy per bit (Garg, 2000). Then, received power of the j-th user should be:

$$P_j = E_{b,j} R \tag{1}$$

where $E_{b,j}$ is the required energy per bit and R is the transmission rate (R). Normalizing the received power to the product of transmission rate and the maximum acceptable power spectral density of interference (I_0):

$$\frac{P_j}{I_0 R} = \frac{E_{b,j} R}{I_0 R} \tag{2}$$

Let us define the average required energy per bit normalized to the maximum acceptable power spectral density of interference (I_0) as follows:

$$\varepsilon_j = \frac{E_{b,j}}{I_0} \tag{3}$$

It was shown in (Viterbi et al., 1994) that ε_j could be approximated by a lognormal distribution, that is:

$$f_\varepsilon(\varepsilon) = 10^{\chi/10} \tag{4}$$

where χ is a Gaussian random variable with mean m_s and variance σ_s for users in soft handoff region and m_h and σ_h for users in hard region. Due to macro diversity in soft handoff region, users in hard region have different behavior in terms of required average energy per bit. Then, m_s and m_h and σ_s and σ_h are, in general, different.

The transmitted power of each MS is calculated assuming that the effects of fast fading can be ignored[4]. Hence, in order to determine which BS is controlling a MS, only path losses and shadowing are considered. Total attenuation is modeled as the product of the path loss r^m by $10^{\zeta/10}$, which is the log-normal random component representing shadowing losses. Then, for a user at a distance r from a base station, attenuation is proportional to:

$$a(r,\zeta) = r^m 10^{\zeta/10} \tag{5}$$

where ζ is a zero-mean Gaussian random variable.

In the reverse link, there are a significant number of mobiles that contribute to the interference. In accordance with the central limit-theorem, **Z** can be approximated by a Gaussian random variable which is completely characterized by its mean and variance.

Due to the soft handoff process, users in hard region have different behavior in terms of interference than users in the soft region. Hence, **Z** can be expressed as a sum of four components: intra-cellular and inter-cellular interference caused by the users in both the hard and soft regions.

Then, following the same mathematical procedure described in (Viterbi et al., 1994), the mean and the variance of inter-cell interference can be expressed as a factor of the mean and variance of the intra-cell interference. Let us denote by f_h and f_s the interference factor for the hard and soft regions, respectively. Thus, the mean and variance of **Z** as function of a given number of active users in the hard (h) and soft (s) handoff regions of the cell of interest can be expressed as follows

$$E\left[Z(h,s)\right] = E\left[I_h\right] + E\left[I_s^+\right] + f_h E\left[I_h\right] + f_s E\left[I_s^+\right] \tag{6}$$

$$\mathrm{Var}\left[Z(h,s)\right] = \mathrm{Var}\left[I_h\right] + \mathrm{Var}\left[I_s^+\right] + f_h \mathrm{Var}\left[I_h\right] + f_s \mathrm{Var}\left[I_s^+\right] \tag{7}$$

where $E[I_h]$, $E[I_s^+]$, $\mathrm{Var}[I_h]$, and $\mathrm{Var}\,[I_s^+]$ represent, respectively, mean value and variance of the normalized interference due to the active users in the hard and soft region of the analyzed cell which are power controlled by the analyzed cell. Values of interference factors depend on propagation conditions and are calculated as suggested in (Viterbi et al., 1994). These values are calculated in terms of the number of users in other cells. However, teletraffic analysis of a single cell cannot provide the exact number of users in the whole system and, as a consequence; inter-cell interference should be calculated in terms of the average number of users in neighbouring cells. Let us denote by **H** and **S** the associated random variables to h and s, respectively. Since a homogeneous system is considered and

---

[4] The impact of fast fading can be effectively reduced by using channel coding, interleaving, and signal processing techniques such as channel equalization. Additionally, if a spread spectrum transmission technique is employed, the frequency diversity property increases the mitigation effect over fast fading. Consequently, its effects can be ignored (Garg, 2000).

the cells are assumed to be statistically identical, some statistics of the inter-cellular interference power can be calculated considering the average number of active users in the hard region, $E[\mathbf{H}]$, and in the soft region, $E[\mathbf{S}]$, of the cell of interest. Section V shows how $E[\mathbf{H}]$ and $E[\mathbf{S}]$ can be calculated via teletraffic analysis.

3.4 Proposed interruption model

Mathematical model used to consider link unreliability is based on the proposed call interruption process in (Rodríguez-Estrello et al., 2009). In (Rodríguez-Estrello et al., 2009), an interruption model and a potential associated time to this process, which is called "*unencumbered call interruption time*", is proposed. Unencumbered call interruption time is defined as the period of time from the epoch the MS establish a link with a BS until the instant the call would be interrupted due to the wireless link unreliability assuming that the MS has neither successfully completed the call nor has been handed off to another cell. Thus, unencumbered call interruption time depends only on link reliability. Physically, this time represents the period of time in which a call would be terminated under the assumption that both the cell dwell time and the unencumbered service time are infinite. Call interruption time is said to be "*unencumbered*" because the interruption of a call in progress by link unreliability can or cannot occur, depending on the values of cell dwell time and unencumbered service time.

Unencumbered call interruption time could not be directly measured in real cellular networks, so, it is necessary to relate it with some parameters that could be measured at base stations (i.e., statistics of the unencumbered call service time and the channel holding time for calls forced terminated due to the link unreliability). As a first approach, unencumbered call interruption time is modeled as a negative exponentially distributed random variable (Rodríguez-Estrello et al., 2009).

It is worth mentioning that in CDMA systems, interruption process due to link unreliability of wireless channel in soft region is different from the interruption process due to link unreliability of wireless channel in hard region because of the diversity in the soft handoff region. So that, to differentiate mean unencumbered call interruption time is denoted by $1/\gamma_x$, where the sub index x represents the residence region. Then x could be $\{S, H\}$ for soft and hard regions, respectively.

In order to obtain some statistics, in (Rodríguez-Estrello et al., 2009) is shown that, if unencumbered service time and cell dwell time are considered negative exponentially distributed, γ_x can be calculated in terms of both channel holding time for calls forced to terminate ($1/\mu_{sdx}$) due to the link unreliability and unencumbered service time ($1/\mu$). Thus, γ_x is given by:

$$\gamma_x = \mu_{sd\_x} - \mu \tag{8}$$

Section 4 shows how the model for taking into account wireless channel unreliability is incorporated in the teletraffic analysis of CDMA-based cellular systems. Then, mathematical expressions for several system level performance metrics are derived.

4. Teletraffic analysis

4.1 General guidelines

The general guidelines of the model presented in (Lin et al., 1994) are adopted here to analyze the system under evaluation. The following assumptions have been widely accepted in the literature and allow the analysis presented to be cast in the framework of a multidimensional birth and death processes.

1) A homogeneous multi-cellular system is considered where each cell has associated capacity limits (in terms of the maximum number of simultaneous users in the soft and hard regions of the cell) that determine the valid state space[5].
2) Only voice service type is considered in the system.
3) New call arrival process follows a Poisson process with mean arrival rate λ_n per cell considering only the non-overlapped areas. Assuming uniform traffic over the system, the mean values of the new call arrival rate in the hard region is $p'\lambda_n$ (p' is considered since it is the ratio between the area of the hard region and the nominal area of the cell) and the new call arrival rate in the soft region $2^*(1-p')\lambda_n$, or $(1-p)\lambda_{n.}$ (p is considered since it considers the overlapped area), respectively.
4) Handoff call arrival process to every cell is also considered to be a Poisson process with mean arrival rate λ_{hI}[6].
5) Inter-cell handoff arrival rate and the average number of users in each region ($E[\mathbf{H}]$ and $E[\mathbf{S}]$) are iteratively calculated. The inter-cell handoff arrival rate is calculated using the methodology described in (Lin et al., 1994). On the other hand, the average number of users in each region is calculated as the carried traffic in each region. This value is used to calculate interference factor in the next stage. The process stops until the carried traffic in each region converges[7]. This algorithm is presented in Figure 2.
6) Unencumbered service time is considered to be a negative exponentially distributed random variable with mean $1/\mu$.
7) Residence time in each region is considered to be a negative exponentially distributed random variable with mean $1/\eta_{xy}$, where the sub indexes x and y represent the current residence region and the previous residence region, respectively. When only one sub-index is used, it means that the call was originated inside the region represented by that sub-index (i.e., soft region or hard region).
8) Unencumbered call interruption time is considered to be a negative exponentially distributed random variable with mean $1/\gamma_x$,

---

[5] Since the capacity limit (i.e., maximum number of simultaneous active users in the hard and soft regions) in CDMA-based cellular systems depends on the interference level in the system, each valid state should be determined depending on the level of interference assuming that the maximum number of codes has not been allocated.

[6] Handoff call arrival process generated by a single cell is clearly not Poisson. However, the combined process from the six different neighboring cells can be adequately approximated by a Poisson process (Cheblus & Ludwin, 1995).

[7] Valid state space depends on the values of $E[\mathbf{S}]$ y $E[\mathbf{H}]$ and they should be calculated iteratively.

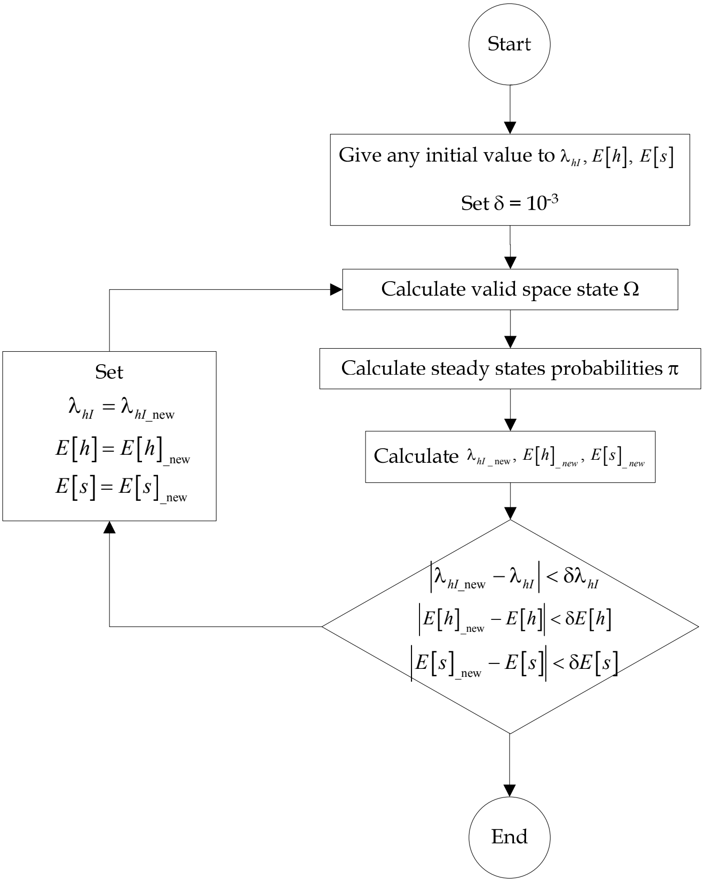


Fig. 2. Algorithm used to calculate valid space state

4.2 Valid space state

In order to analyze CDMA-based cellular systems considering link unreliability and soft handoff process by means of teletraffic analysis, a multi dimensional queuing analysis is necessary. Valid space state (Ω) is determined by the total number of users in the soft region s, and the total number of users in the hard region, h, given by the pairs $(\mathbf{s}, \mathbf{h})$ such that a given interference constraint is met. Pairs $(\mathbf{s}, \mathbf{h})$ which accomplish interference constraint depend on the system's characteristics such as processing gain (G) and interference margin (ϕ). Consequently, if $\mathbf{Z}$ is modelled as a Gaussian random variable and considering s and h users in the soft and hard regions, respectively, of the analyzed cell, the probability $P_{out}(s,h)$ that $\mathbf{Z}$ exceeds a given maximum acceptable threshold is given by:

$$P_{out}\left(s,h\right)=\frac{1}{2}-\frac{1}{2}erf\left(\frac{G\left(1-\phi\right)-E\left[Z\left(s,h\right)\right]}{\sqrt{2\operatorname{var}\left[Z\left(s,h\right)\right]}}\right) \tag{9}$$

As a result, a state is valid if it meets the condition $P_{out}(s,h) \leq P_{out\_max}$ (where P_{out_max} is the maximum allowed outage probability). Hence, for a given number h of users in the hard region a maximum number $M_s(h)$ of users in the soft region can exist, as shown in Figure 3. The values of $M_s(h)$ represent a *hard capacity limit* to guarantee an acceptable performance operation in terms of the outage probability.

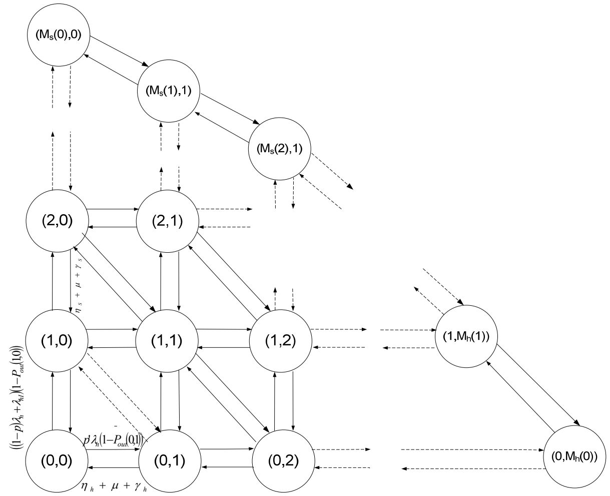


Fig. 3. Valid state space.

4.3 Call admission control strategy

The call admission strategy used works as follows:

When a **new call** or **handoff attempt** arrives at a given cell, the cell will try to serve the incoming request. Whether the call is accepted or not depends on the state of the system.

- If the **new call** is generated in the **soft** region or if it is an inter-cell **handoff request**, upon the call arrival, the condition $s+1 \leq M_s(h)$ is evaluated. If this condition is met, the request is accepted; otherwise, the call is blocked or dropped.
- If the **new call** request is generated in the **hard** region**;** upon the call arrival, the condition $s \leq M_s(h+1)$ is evaluated. If this condition is met, the request is accepted; otherwise, the call is blocked.
- If the request is performed by a user moving from the **soft region to the hard region** (intra-cell handoff)**,** upon the request, the condition $s-1 \leq M_s(h+1)$ is evaluated. If this condition is met, the request is accepted; otherwise, the call is dropped.

If the request is performed by a user coming from the **hard region to the soft region** (intra-cell handoff), upon the call arrival, the condition $s+1 \leq M_s(h\text{-}1)$ is evaluated. If this condition is met, the request is accepted; otherwise, the call is dropped.

4.4 Queuing formulation

Since a homogenous case is assumed where all cells are statistically identical, the overall system performance can be analyzed by focusing on only one given cell. Let us denote the state of a given cell as $\mathbf{K} = [k_H,\ k_S\ k_{SH},\ k_{SS}]$, where k_H and k_S represent, respectively, the number of active users in hard and soft regions that were originated in the hard and soft regions, respectively; k_{SH} represents the number of active users in the soft region of the cell coming from the hard region of the same cell, and k_{SS} represents the number of active users in the soft region of the cell coming from the soft region of other cell. Let us define $\mathbf{e}_i$ as a vector of the same dimension of $\mathbf{K}$ whose entries are all 0 except the i-th one which is 1. Then, equating rate out to rate in for each state, the statistical-equilibrium state equations are (Cooper, 1990):

$$P_{\mathbf{K}} = \frac{\left[\sum_{i=1}^{3} a_{i\mathbf{K}-\mathbf{e}_i} P_{\mathbf{K}-\mathbf{e}_i} + \sum_{i=1}^{4} b_{i\mathbf{K}+\mathbf{e}_i} P_{\mathbf{K}+\mathbf{e}_i} + c_{1\mathbf{K}+\mathbf{e}_1-\mathbf{e}_3} P_{\mathbf{K}+\mathbf{e}_1-\mathbf{e}_3} + \sum_{i=2}^{4} c_{i\mathbf{K}+\mathbf{e}_i-\mathbf{e}_1} P_{\mathbf{K}+\mathbf{e}_i-\mathbf{e}_1}\right]}{\left[\sum_{i=1}^{3} a_{i\mathbf{K}} + \sum_{i=1}^{4}\left(b_{i\mathbf{K}} + c_{i\mathbf{K}}\right)\right]} \tag{10}$$

For $[k_H,\ k_S,\ k_{SH},\ k_{SS}]$ such that $K \in \Omega$.

4.5 Transition rates

In order to simplify the notation, let us define $s = k_S+k_{SS}+k_{SH}$.

The call arrival rate for users in hard region that will generate a transition from $\mathbf{K}$ to $\mathbf{K}+\mathbf{e}_1$ is given by the product of the mean arrival rate of the new call arrival process weighted by the probability that a user is originated in the inner region and the probability that this arrival will not exceed outage probability:

$$a_{1\mathbf{K}} = \begin{cases} p'\lambda_n\left(1-P_{out}\left(s,k_H+1\right)\right) & ;\ s \leq M_s\left(k_H+1\right) \\ 0 & ;\ \text{otherwise} \end{cases} \tag{11}$$

The call arrival rate for users in soft region that will generate a transition from $\mathbf{K}$ to $\mathbf{K}+\mathbf{e}_2$ is given by the product of the mean arrival rate of the new call arrival process weighted by the probability that a user is originated in the outer region and the probability that this arrival will not exceed outage probability:

$$a_{2\mathbf{K}} = \begin{cases} (1-p)\lambda_n\left(1-P_{out}\left(s+1,k_H\right)\right) & ;\ (s+1) \le M_s\left(k_H\right) \\ 0 & ;\ \text{otherwise} \end{cases} \tag{12}$$

The call arrival rate for users in soft region that will generate a transition from **K** to **K+e**$_4$ is given by the product of the mean arrival rate of the handoff call arrival process and the probability that this arrival will not exceed outage probability:

$$a_{3\mathbf{K}} = \begin{cases} \lambda_{hI}\left(1-P_{out}\left(s+1,k_H\right)\right) & ;\ (s+1) \le M_s\left(k_H\right) \\ 0 & ;\ \text{otherwise} \end{cases} \tag{13}$$

The call departure rate for users in hard region that will generate a transition from **K** to **K-e**$_1$ is given by the sum of the mean departure rate of successfully terminated calls (μ), the mean departure rateof calls forced to terminate due to the link unreliability (γ_H), and the mean departure rate of calls that are handed off to the outer region (η_H), weighted by the probability that this handoff will not exceed outage probability

$$b_{1\mathbf{K}} = \begin{cases} k_H\left[\eta_H\left(1-P_{out}\left(s+1,k_H-1\right)\right)+\gamma_H+\mu\right] & ;\ (s+1) \le M_s\left(k_H-1\right) \\ k_H\left[\mu+\eta_H+\gamma_H\right] & ;\ \left(s+1,k_H-1\right) \notin \Omega \\ 0 & ;\ \text{otherwise} \end{cases} \tag{14}$$

The call departure rate of users in soft region that were originated in that region and will generate a transition from **K** to **K-e**$_2$**,** **K-e**$_3$ or **K-e**$_4$ is given by the sum of the mean departure rate of successfully terminated calls (μ), the mean departure rate of calls forced to terminate due to the link unreliability in the outer region (γ_S), and the mean departure rate of calls that are handed off to the inner region ($q_{Sx}\eta_{Sx}$), weighted by the probability that this handoff will not exceed outage probability

$$b_{i\mathbf{K}} = \begin{cases} k_{Sx}\left[\mu+\eta_{Sx}\left(\left(1-q_{Sx}\right)\left(1-P_{out}\left(s-1,k_H+1\right)\right)\right)+\gamma_S\right] & ;\ (s-1) \le M_s\left(k_H+1\right) \\ k_{Sx}\left[\mu+\eta_{Sx}+\gamma_S\right] & ;\ \left(s-1,k_H+1\right) \notin \Omega \\ 0 & ;\ \text{otherwise} \end{cases} \tag{15}$$

with i = 2, 3,4 and x = {S, SH, SS} for each case.

The call departure rate for users in hard region to soft region that will generate a transition from **K** to **K-e**$_1$**+e**$_2$ is given by the mean departure rate of calls that are handed off to the outer region (η_H) weighted by the probability that this handoff will not exceed outage probability:

$$c_{1\mathbf{K}} = \begin{cases} k_H\left[\eta_H\left(1-P_{out}\left(s+1,k_H-1\right)\right)\right] & ;\ \left(s+1\right)\le M_s\left(k_H-1\right) \\ 0 & ;\ \text{otherwise} \end{cases} \tag{16}$$

The call departure rate for users in soft region to hard region that will generate a transition from **K** to **K-e**$_2$**+e**$_1$, **K-e**$_3$**+e**$_1$, or **K-e**$_4$**+e**$_1$ is given by the mean departure rate of calls that are handed off to the inner region ($q_{Sx}\eta_{Sx}$) weighted by the probability that this handoff will not exceed outage probability:

$$c_{i\mathbf{K}} = \begin{cases} k_{Sx}\left[\eta_{Sx}q_{Sx}\left(1-P_{out}\left(s-1,k_H+1\right)\right)\right] & ;\ \left(s-1\right)\le M_s\left(k_H+1\right) \\ 0 & ;\ \text{otherwise} \end{cases} \tag{17}$$

with i = 2, 3, 4, and x = {S, SH, SS} for each case.
In addition, of course, the probabilities must satisfy the normalization equation given by (18).

$$\underset{\{\mathbf{K}\in\Omega\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} P_{\mathbf{K}} = 1 \tag{18}$$

The corresponding steady state probabilities are calculated by means of the Gauss-Seidel method (Cooper, 1990) by using the fixed point iteration described in Figure 2.

4.6 QoS performance metrics (KPIs)

The merit of this section is to derive general expressions for the QoS metrics used in this study to evaluate the performance of a CDMA-based cellular system considering both resource insufficiency and wireless link unreliability.

4.6.1 Blocking probability

Overall blocking probability is the weighted sum of the blocking probability of each region. New calls in soft region are blocked only if they find both cells in a blocking condition. Then, the overall blocking probability is given by (Zhang & Lea, 2006):

$$P_b = pP_{bH} + \left(1-p\right)P_{bS}^2 \tag{19}$$

Note that the probability that a call is blocked to avoid unacceptable interference (*interference-limited*) in every state (k_H, k_S, k_{SH}, k_{SS}), should be calculated by evaluating the outage probability considering that the call was accepted. For instance, if the new call blocking probability in soft handoff region is evaluated, then outage probability should be evaluated considering that s=k_S + k_{SH} + k_{SS} +1, and k_H. In general, the following cases can be distinguished.

- **New calls** in hard region could be blocked to avoid unacceptable interference (*interference-limited capacity*) in every state (k_H, k_S, k_{SH}, k_{SS}), with probability $P_{out}(s,k_H+1)$ even though the condition $s<M_S(k_H+1)$ is met. Additionally, because of resource insufficiency (*hard-limited capacity*), **new calls** in hard region are also blocked in the states where $s=M_S(k_H+1)$. Then, new call blocking probability for users originated in hard region is given by:

$$P_{bH} = \underset{\{s<Ms(k_H+1)\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} P_{\mathbf{K}} P_{out}\left(s, k_H+1\right) + \underset{\{s=Ms(k_H+1)\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} P_{\mathbf{K}} \qquad (20)$$

- **New calls** in soft region as well as **inter-cell handoff attempts** could be blocked to avoid unacceptable interference (*interference-limited*) in every state (k_H, k_S, k_{SH}, k_{SS}), with probability $P_{out}(s+1,k_H)$, even though the condition $(s+1)<M_S(k_H)$ is met. Additionally, due to resource insufficiency (*hard-limited capacity*), these calls are blocked in the states that meet the condition $(s+1)=M_S(k_H)$. Then, new call blocking probability for users originated in soft region is given by:

$$P_{bS} = \underset{\{(s+1)<Ms(k_H)\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} P_{\mathbf{K}} P_{out}\left(s+1, k_H\right) + \underset{\{(s+1)=Ms(k_H)\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} P_{\mathbf{K}} \qquad (21)$$

4.6.2 Intracell handoff rate

Intra-cell handoff attempt rates from hard region to soft one and from soft region to hard one (i.e., λ_{hiS} and λ_{hiH}) are, respectively, calculated as the weighted sum of the handoff rate in each state multiplied by the probability to be in each state.

$$\lambda_{hiS} = \underset{\{\mathbf{K}\in\Omega\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} k_H \eta_H P_{\mathbf{K}} \qquad (22)$$

$$\lambda_{hI} = \underset{\{\mathbf{K}\in\Omega\}}{\sum_{k_H}\sum_{k_S}\sum_{k_{SH}}\sum_{k_{SS}}} \left(k_S\eta_S\left(1-q_S\right) + k_{SH}\eta_{SH}\left(1-q_{SH}\right) + k_{SS}\eta_{SS}\left(1-q_{SS}\right)\right) P_{\mathbf{K}} \qquad (23)$$

4.6.3 Intracell handoff failure probability

Intra-cell handoff failure probability (in both directions: from hard region to soft region and vice versa) is given by the ratio of the intra-cell handoff failure rate to the intra-cell handoff attempt rate. Intra-cell handoff attempt rate is calculated as the weighted sum of the handoff rate in each state multiplied by the corresponding state probability considering only the states where handoff could be blocked. The Intra-cell handoff failure probability for subscribers moving from hard to soft regions (denoted by F_{hiS}) and for subscribers moving from soft to hard regions denoted by F_{hiH}) are given by (22) and (23), respectively.

$$F_{hiH} = \sum_{k_H} \sum_{\substack{k_S \\ \{(s+1)\le M_s(k_H-1)\}}} \sum_{k_{SH}} \sum_{k_{SS}} k_H \eta_H P_{\mathbf{K}} P_{out}\left(s+1, k_H - 1\right) \tag{24}$$

$$F_{hiS} = \sum_{k_H} \sum_{\substack{k_S \\ \{(s-1)\le M_s(k_H+1)\}}} \sum_{k_{SH}} \sum_{k_{SS}} \left(k_S \eta_S q_S + k_{SH} \eta_{SH} q_{SH} + k_{SS} \eta_{SS} q_{SS}\right) P_{\mathbf{K}} P_{out}\left(s-1, k_H + 1\right) \tag{25}$$

Then, intra-cell handoff failure probability P_{hiH} (P_{hiS}) from hard (soft) region to soft (hard) region is given by (26) and (27), respectively:

$$P_{hiH} = \frac{F_{hiH}}{\lambda_{hiH}} \tag{26}$$

$$P_{hiS} = \frac{F_{hiS}}{\lambda_{hiS}} \tag{27}$$

4.6.4 Interruption probability due to link unreliability

Link unreliability is the other fundamental cause of forced termination. Considering the proposed model, a call will be forced terminated due to the link unreliability in a given region x of a cell j when unencumbered call interruption time is smaller than both unencumbered service time and cell dwell time. Then, mathematically, it can be expressed as:

$$\beta_{xy}^{(j)} = P\left\{\mathbf{X}_{i\_xy}^{(j)} < \min(\mathbf{X}_s^{(j)}, \mathbf{X}_{r\_xy}^{(j)})\right\} \quad ; \; j = 0,\ 1,\ 2,\ \ldots \tag{28}$$

Due to the lack of memory of the negative exponential probability distribution of the unencumbered interruption time, unencumbered service time and cell dwell time, probability of interruption in region x is given by:

$$\beta_{xy} = \frac{\gamma_x}{\mu + \eta_{xy} + \gamma_x} \tag{29}$$

As for the residence time, it is necessary to distinguish new users in the region x and users handed off from region y to region x. Sub indexes x and y represent the actual residence region and the previous residence region, respectively.

4.6.5 Probability of successfully call termination

Considering the proposed model, a call will be successfully terminated in a given region x of a cell j when service time is smaller than both unencumbered call interruption time and cell dwell time. Then, mathematically, it can be expressed as:

$$\zeta_{xy}^{(j)} = P\left\{\mathbf{X}_s^{(j)} < \min(\mathbf{X}_{i\_xy}^{(j)}, \mathbf{X}_{r\_xy}^{(j)})\right\} \quad ; \; j = 0,\, 1,\, 2,\, ... \tag{30}$$

Due to the lack of memory of the negative exponential probability distribution of the unencumbered interruption time, unencumbered service time and cell dwell time, probability of interruption in region x is given by:

$$\zeta_{xy} = \frac{\mu}{\mu + \eta_{xy} + \gamma_x} \tag{31}$$

As for the residence time, it is necessary to distinguish new users in the region x and users handed off from region y to region x. Sub indexes x and y represent the actual residence region and the previous residence region, respectively.

4.6.6 Call forced termination probability

Call forced termination may result from intra-cell or inter-cell handoff failure or wireless channel unreliability. A signal flow diagram is used to describe the process of a call and the call forced termination probability is obtained by using Mason's Formula (Robichaud et al., 1962). Call forced termination probability is derived by means of graph theory and Mason's rule as a mathematical tool. The steps necessary to find the forced termination probability are listed below:

- Construct a graph, where each node of this graph represents one possible state of the call. Possible states for a call are:

  I: Call initiation.
  H: New call originated in hard region.
  S: New call originated in soft region.
  SH: Intracellular handed off call in soft region.
  SS: Intercellular handed off call in soft region
  HS: Intracellular handed off call in hard region
  ST: Successfully terminated call.
  FT: Forced terminated call.

- Connect each pair of nodes with a directed branch. Each connection is directed from the source point to the sink one. The weight of the connection is the transition probability between the states. The complete flux graph for this case is shown in Figure 4

As an illustrative example, let us assume that the weight between the '*S*' node and the '*FT*' node represents the probability that a call which was originated in soft region is forced to terminate. This probability is the probability that a call requires a handoff (i.e., the probability that residence time be less than the minimum of unencumbered service time and unencumbered interruption time which is represented by θ_S). The call performs an intracellular handoff with probability q_S and an intercellular handoff with probability (1-q_S). An intracellular handoff will fail from soft region to hard region occurs with probability P_{hiS} and an intercellular handoff will fail with probability and P_{bS}. Additionally, the other cause of force termination is link unreliability, (i.e., the probability that unencumbered interruption time is less than the minimum of unencumbered service time and residence time which is represented by β_S) and it occurs with probability β_S. Then, taking into account all of these different cases, the whole probability can be computed as $\theta_S[(1-q_S)\ P_{hI} + (1-q_S)\ P_{bS}]+ \beta_S$.

In order to obtain forced termination probability it is necessary to solve the graph by using the Mason's rule considering *I* as the source node and *FT* as the sink node (Robichaud et al., 1962). Final expression is not presented because it is very large.

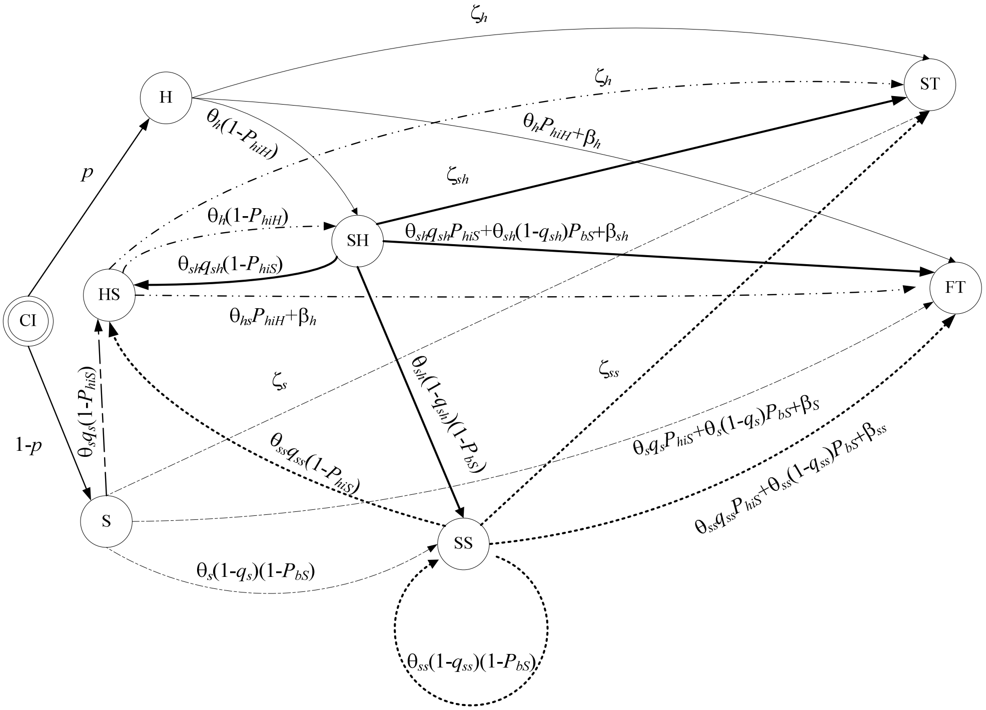


Fig. 4. Flux diagram for a call in a CDMA cellular system considering soft handoff characteristic.

5. Performance evaluation

The goal of the numerical evaluations presented in this section is to clarify, understand, and analyze the influence of link unreliability, user mobility, interference, soft capacity and soft handoff process on the performance metrics of CDMA-based cellular networks.

In numerical evaluations, it is assumed that p = 0.7, consequently, p' = 0.85. Mean service time is $1/\mu$ = 180 s and mean dwell time in the different regions is calculated by simulation using the mobility model of (Zoonozi & Dassanayake, 1997) using minimum and maximum values for parameter α established in [38], that is, $\alpha = 0°$ (linear movement) and $\alpha = 40°$. Circular shaped cells with radius 2 km are considered. Mean values of these times are presented in Table 1.

Mobility scenario/ Parameter	$\alpha = 0°$	$\alpha = 40°$
$1/\eta_H$	170 s	554 s
$1/\eta_S$	46 s	60 s
$1/\eta_{SS}$	55 s	130 s
$1/\eta_{SH}$	50 s	120 s
q_S	0.37052	0.46574
q_{SS}	0.62902	0.25394
q_{SH}	0	0.64470

Table 1. Users mobility parameters

The specific value of the mean unencumbered call interruption time is not easy to calculate because it closely depends on the channel and propagation conditions and the specific fading mitigation techniques. Thus, the impact of different values of mean unencumbered call interruption time are evaluated.

As in (Zhang & Lea, 2006), considered values of interference parameters for hard and soft regions are: m_h = 7 dB, $\sigma_h = \sigma_s$ = 2.5 dB. Additionally, it is assumed that the standard deviation of shadowing σ = 8 dB, a processing gain G = 32, and an interference threshold ϕ = 10 dB, an offered load is a = 2 Erlangs. Notice that the required energy per bit in soft region m_s is a random quantity because it closely depends on the macro-diversity gain of the soft handoff process.

In addition, there are different algorithms used to implement soft handoff, thus, diversity gain depends on the specific algorithm used for soft handoff (Da Silva & Farmine, 2005). Consequently in order to analyze the impact of this parameter on the system performance its value is varied in the range from 3.5 dB to 7 dB.

Figs. 5 and 6 plot, respectively, new call blocking probability and forced termination probability versus the inverse of the mean value of the unencumbered call interruption time in hard and soft regions for two different mobility scenarios ($\alpha = 0°$ and $\alpha = 40°$). Figures 7 and 8 show the global new call blocking probability and call forced termination probability versus the required bit energy to spectral interference density ratio for users in the soft region for two different users mobility scenarios ($\alpha = 0°$ and $\alpha = 40°$).

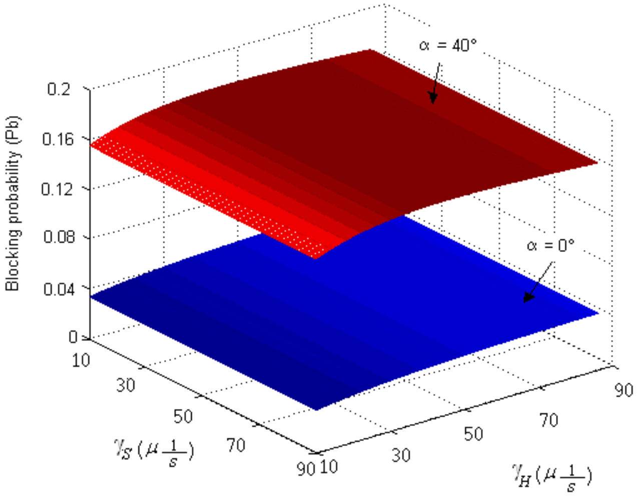


Fig. 5. New call blocking probability versus the inverse of the mean value of the unencumbered call interruption time in both regions for $\alpha = 0°$ and $\alpha = 40°$.

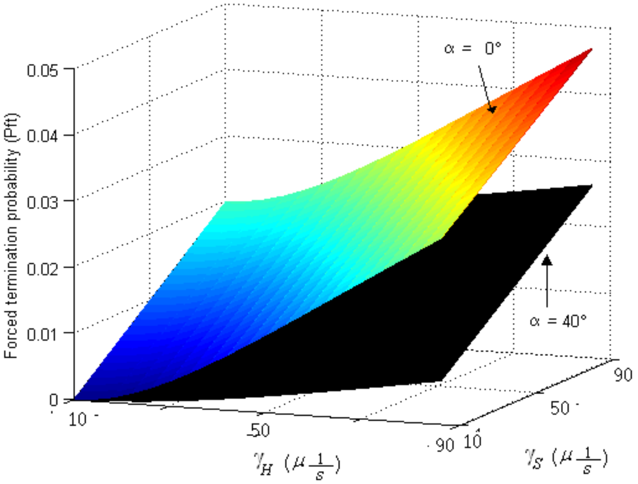


Fig. 6. Call forced termination probability versus the inverse of the mean value of the unencumbered call interruption time in both regions for $\alpha = 0°$ and $\alpha = 40°$.

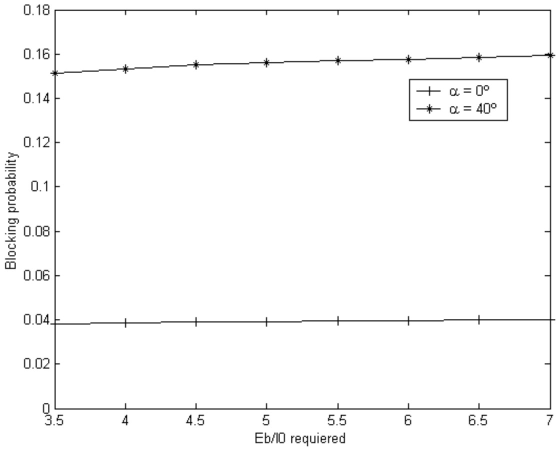


Fig. 7. New call blocking probability vs. mean value of the average requiered E_b/I_0 in soft handoff region m_s for $\alpha = 0°$ and $\alpha = 40°$.

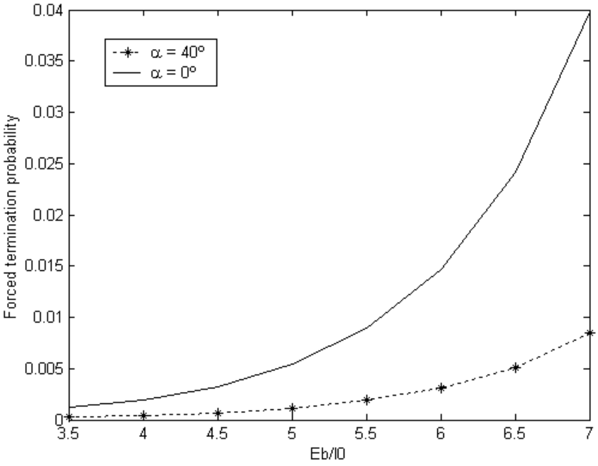


Fig. 8. Call forced termination probability vs. average requiered E_b/I_0 in soft handoff region m_s for $\alpha = 0°$ and $\alpha = 40°$.

5.1 Influence of link unreliability

As shown in Figure 5, new call blocking probability does not significantly vary with changes in the mean value of the unencumbered call interruption time in both regions. Nevertheless, it is observed that blocking probability increases as link unreliability in hard zones increases (i.e. when mean unencumbered call interruption time decreases). The call forced termination probability showed in Figure 6 has similar behavior. That is, the call forced termination probability increases as the link unreliability increases. However, the forced termination probability is more sensitive than blocking probability to changes in the mean value of the unencumbered call interruption time in both regions. This is because the interruption probability increases as link unreliability increases, directly affecting the call forced termination probability while the blocking probability remains unaffected because link unreliability is not considered to accept the call.

Call forced termination probability is the most affected QoS metric due link unreliability and also due to the soft handoff diversity. Thus, it is fundamental to consider link unreliability in the performance evaluation of CDMA-based cellular systems. Additionally, different resource allocation mechanisms designed in order to have reliable connections in CDMA-based systems have a great impact on the performance of these systems. In this work no specific strategies are considered because under the analytical framework developed in this study, the use of those reallocation mechanisms would be reflected in the reduction of the parameter γ_x of the proposed call interruption process and the energy per bit requirement in soft handoff region. Then, the mathematical tool proposed here is a general instrument that can be used to evaluate the impact of new resource allocation mechanisms by considering the main features in CDMA-based cellular systems.

5.2 Influence of mobility

Relative to the users' mobility: the blocking probability increases as the relative users' mobility decreases (or, equivalently, as mean cell dwell time increases). This is because the larger the mean cell dwell time the slower users with ongoing calls move and the rate at which resources are released, hence, decreases. Relative to the call forced termination probability, it increases as mobility increase because more handoff arrivals are generated and is reflected directly in call forced termination probability.

Observe that blocking probability is more sensitive to changes in link unreliability in case of $\alpha = 40°$. This is because in this particular scenario (i.e.: $\alpha = 40°$), the mean cell dwell time is greater than $\alpha = 0°$, thus a small number of handoffs are generated, and the blocking probability is affected mainly by interruption process. The opposite effect occurs with call forced termination, because the handoff failure probability closely depends on the handoff rate and handoff failure probability.

5.3 Influence of interference characteristics

From the presented figures it is observed that both QoS metrics are monotonically increasing functions of the required bit energy to spectral interference density ratio. This is because the higher the offered traffic the more interference is brought into the system. Since CDMA-based systems are interference limited, more interference translates into a higher outage probability in each state, therefore, fewer users can be accepted in the system and blocking probility increases.

6. Conclusion

A new system level mathematical performance analysis of mobile cellular CDMA networks was presented. Mathematical analysis takes into account main features of CDMA-based systems. Wireless channel unreliability was modeled by means of a Poisson call interruption process which allows an elegant teletraffic analysis considering both wireless link unreliability and resource insufficiency. Interference-limited capacity nature of CDMA systems was included in the teletraffic model as well as soft handoff capability considering their statistical parameters. Additionally, realistic users' movement was considered by characterizing users' mobility through the statistics of a smooth random mobility model and assuming that the user's movement is correlated.

More importantly, novel mathematical expressions for some significant performance metrics (i.e., new call blocking and call forced termination probabilities, and intra-cell and inter-cell handoff attempts rates) were derived. Graph theory and the Mason's formula were used as mathematical tools to derive a general expression for the call forced termination probability which includes both causes of forced termination: lack of resources due to the interference and wireless channel unreliability. Moreover, numerical results show that QoS metrics are evidently degraded when considering link unreliability. This results call again for the necessity of reexamining classical analytical results in traffic theory, which are used for the analysis and design of CDMA mobile networks that have not yet considered link unreliability.

On the other hand, in this work all involved times in the model were considered to be negative exponentially distributed random variables as a first step. However, future work includes the generalization of the mathematical model in order to consider general distribution for the involved time variables modelled as random variables with phase-type distributions. All these issues are subject of further future research.

7. References

Boggia, G.; Camarda, P.; D'Alconzo, A.; De Biasi, A.; & Siviero, M. (2005). Drop call probability in established cellular networks: from data analysis to modelling, *Proceedings of IEEE VTC'05-Spring*, pp. 2775-2779, Stockholm, Sweden, May-Jun. 2005.

Chlebus, E. & Ludwin, W. (1995). Is handoff traffic really Poissonian? *Proceedings of IEEE ICUPC'95*, pp. 348-353, Tokyo, Japan, November 1995.

Cooper, R.B. (1990). *Introduction to Queuing Theory*. CEE Press Books, Washington, D.C.

Da Silva, M. & Farmine, Y. W-CDMA uplink soft handover gain measurements. *Proceedings of IEEE VTC'05-Spring*, pp. 416-418, Stockolm, Sweden, June 2005.

Dahlman, E.; Parkvall, H.; Skold J. & Beming P. (2007). *3G Evolution.* Elsevier LTD. 2007.

Elshabrawy, T. & Le-Ngoc, T. (2005). Gilbert channel approximation for downlink performance evaluation of WCDMA systems. *Proceedings of IEEE VTC'05-Fall*, pp. 382-386, Dallas, TX, September 2005.

Garg V. (2000). *IS-95CDMA and cdma2000: cellular/PCS implementation.* Prentice Hall, New York.

Hegde, N. & Sohraby, K. (2002). On the impact of soft handoff in cellular systems, *Computer Networks*, Vol. 38, No.2, February 2002, pp. 257-271.

Kim D.K. & Sung, D.K. (1999). Characterization of soft handoff in CDMA systems. *IEEE Transactions on Vehicular Technology,* Vol. 48, No. 4, July 1999, pp.1195-1202.

Kong, P. Y. (2002). Performance of queue in impaired wireless channel. *IEE Electronics Letters*, Vol. 38, No. 22, October 2002, pp. 1342–1343.

Li, B.; Li, L.; Li, B.; Sivalingam, K.M. & Cao, X.R. (2004). Call admission control for voice/data integrated cellular networks: performance analysis and comparative study. *IEEE Journal on Selected Areas of Communications*, Vol. 22, No. 4, May 2004, pp. 706-718.

Lin, Y.B.; Mohan, S. & Noerpel, A. (1994). Queuing priority channel assignment strategies for PCS and handoff initial access. *IEEE Transactions on Vehicular Technology*, Vol. 43. No. 3, August 1994, pp. 704-712.

Liu, B. & Sule Alfa, A., (2004). A queueing model with time-varying QoS and call dropping for evaluating the performance of CDMA cellular systems, *Wireless Communication and Mobile Computing*, No. 4, March 2004, pp. 439-447.

Ma, X.; Liu, Y. & Trivedi, K. S. (2006). Modeling and performance analysis for soft handoff schemes in CDMA cellular systems. *IEEE Transaction on Vehicular Technology*, Vol. 55, No. 2, March 2006, pp. 670-680.

Naraghi-Pour, M. & Chai, Y. (2006). Call admission control for CDMA cellular networks supporting multimedia services. *Proceedings of IEEE WCNC'*06, pp. 997-1002, Las Vegas, NV, May 2006.

Piao, S.; Park, J. & Park, Y. (2006). Outage probability analysis for DS-CDMA systems with call admission control schemes. *IEICE Transactions on Communications,* Vol. E89B, No. 8, August 2006, pp. 2226-2229.

Robichaud, L.P.A.; Boisvert, M. & Robert, J. (1962). *Signal Flow Graphs and Applications.* Prentice Hall, New Jersey.

Rodriguez-Estrello, C.B.; Cruz-Pérez, F.A. & Hernández-Valdez, G. (2009). System level analysis of mobile cellular networks considering link unreliability. *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 2, , February 2009, pp. 926-940.

Su, L.; Chen, J.Y. & Huang, J.H. (1996) Performance analysis of soft handoff in CDMA Cellular Networks. *IEEE Journal on Selected Areas of Communication*, Vol. 14, No. 9, December 1996, pp. 1762-1769.

Viterbi, A.J.; Viterbi, A.M. & Zehavi, E. (1994). Other cell interference in cellular power controlled CDMA. *IEEE Transactions on Communications.* Vol. 42, No. 2/3/4, February/March/April 1994, pp. 1501-1504.

Wong, K.Y.M.; Keung, Y.; Li, C. & Li, B. (2004). Dynamic call admission control with connection level service guarantee in multi-service wireless cellular networks. *Proceedings of IEEE ICC'*04, pp. 4248-4252, Paris, France, June 2004.

Zhang, M. & Lea, C.T. (2006). The impact of mobility on CDMA call admission control. *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 6, November 2006, pp. 1908-1920.

Zhang, Y. & Soong, B. (2005). The effect of unreliable wireless channel on the call performance in mobile network. *IEEE Transanctions on Wireless Communications*, Vol. 4, No.2, March 2005, pp. 653-661.

Zhang, Y. & Soong, B. (2006). Performance of mobile network with wireless channel unreliability and resource insufficiency. *IEEE Transanctions on Wireless Communications*, Vol. 5, No. 5, May 2006, pp. 990-995.

Zhao, D.; Shen, X. (Sherman); & Mark, J.W. (2006). Soft handoff and connection reliability in cellular CDMA downlinks. *IEEE Transactions on Wireless Communications*, Vol. 5, No. 2, February 2006, pp. 354-365

Zonoozi, M.M. & Dassanayake, P. (1997). User mobility modelling and characterization of mobility patterns. *IEEE Selected Areas on Communication*, Vol. 15, No. 7, September 1997, pp. 1239-1252.

14

A Unified Data and Energy Model for Wireless Communication with Moving Senders and Fixed Receivers

Armin Veichtlbauer and Peter Dorfinger
Salzburg Research Forschungsgesellschaft mbH
Austria

1. Introduction

In recent years, the question of energy efficiency in ICT solutions has grown to a hot topic, both in research and in product development. Especially for applications in the field the efficient use of the available (stored or newly generated) energy is a precondition for the desired functionality. Energy wasting is not only a question of expenses or of impacts to the environment, but in many cases simply precludes the proper working of a sensor/actuator control system.

Our research group has conducted several research projects during the last years in the area of protocol optimisation in order to increase energy efficiency of wireless communication. First we developed an energy model to conduct simulations which describe the energy consumption of sending a well defined amount of data over a wireless link with fixed properties. As variable parameters of this model we used the transmission power of the sending antenna and the packet length of the transmitted data. This model already included a stochastic part: The loss of the transmitted packets. The packet loss probability was evidently dependent on the sending power. So far we followed the model of the group around J.P. Ebert and A. Wolisz (Ebert et al., 2000; Ebert et al., 2002).

We then integrated a data model to simulate the amount of newly produced data respectively data that has remained in the sending buffer, thus we generated a unified data and energy model. Finally we integrated a distance model to simulate the changing distances between the sender and the receiver. As a matter of simplicity (but without spoiling the capabilities of the model) we assumed that the receiver is fixed, and the sender is moving (Veichtlbauer & Dorfinger, 2007).

We conducted our research work within funded research projects: Autarchic Ski (ASki), GI Platform Salzburg and the GI Tech Lab, all of them funded by the Austrian Federal Ministry for Transport, Innovation, and Technology, in different funding schemes. Along with the different projects came different application scenarios, e.g. the communication of intelligent skis (which have sensors on board to measure for instance temperature or pressure during runs) with base stations which analyse the collected sensor data (Veichtlbauer & Dorfinger, 2008; Veichtlbauer & Dorfinger, 2009) or the collaboration of a swarm of flying sensors (Dorfinger & Veichtlbauer, 2008) for weather or gas density measurements.

2. Description of the Model

Our MATLAB/Simulink based „Unified Data and Energy Model" for wireless communication takes into account both its energy and its data balance, i.e. it calculates the amount of successfully transmitted and lost data per time unit and contrasts these values with the consumed energy.

2.1 Modelling Approach

The goal in our setting was to maximize the amount of successfully transmitted data in surroundings where energy is a scarce resource. For static scenarios (constant distance between sender and receiver) a well proven model can be found in literature: The model of J. P. Ebert and his team. Their mathematical analysis of wireless communication is based on the Link Budget Analysis of Zyren and Petrik (Zyren & Petrik, 1998) and the Gilbert-Elliot Bit Error Model (Gilbert, 1960).

The basic idea of Ebert's model is to calculate an "energy per bit" value to quantify the needed energy for the successful transmission of one bit, and to minimize this energy by changing the sending power. He proves that with variation of sending power and keeping all other parameters (like packet length, distance between sender and receiver, receiver gain, etc.) constant, such a minimum can be found: Obviously, increasing sending power leads to higher energy consumption of the sending attempts. On the other hand decreasing sending power leads to increasing loss probability of a transmitted packet, thus causing retransmissions of the lost packets (Ebert & Wolisz, 1999; Ebert & Wolisz, 2000; Burns & Ebert, 2001). Using appropriate simulations, an optimum can be found easily.

This approach can be applied for multi-hop ad-hoc networks (Matzen et al., 2003; Ebert, 2004), considering different routes and using the shortest links to save energy (the energy per bit value is lower for shorter distances), yet the dynamics (changing distances between nodes) are still not considered. It is possible to send packets with well calculated sending power at any time, but all data are sent immediately after their "production" (e.g. by sensors which measure periodically some environmental parameters).

In our scenarios we considered a moving sender and (one or more) fixed receiver(s). For a moving sender, it is profitable to consider also the sending times: Sending at the moment of minimal distance will optimise the energy per bit value. Thus, we integrated a distance model into our approach. The idea is to predict the further movement and to send during the time(s), when the sender is closest to the receiver(s).

We used a time discrete approach for our model, as the data generation is done that way by the sensors (depending on their sampling rate). Although we use the packet length as an input factor, we do not use packet simulations. Bit errors influence the data flows in a statistical manner, thus our model complies with the approach of Haber et al. (Haber et al., 2003) for fluid simulations of data streams.

2.2 Model Assumptions

The basic assumptions for our model are:

- Energy is stored in capacitors of a defined size; the efficiency of storing energy is dependent on the filling level of the capacitors.
- A data buffer storage of a defined size is used on the sender side to store some sensor data.
- The data storage is organised as a ring buffer, thus a full storage will lead to data loss (new data is written over old data which has not been successfully transmitted on time).
- The optimization criterion is given by amount of successfully transmitted data (with given energy).
- The adjustable parameters are: The sending power, the packet length and the sending time(s).

Sending power and packet length are optimized according to the Ebert model. To take into account the dynamics of the movement, we do not send immediately, but store the produced data in the local buffer and calculate the optimal sending times according to the distance model. Our approach is simple, but effective: We calculate whether the sender is approaching or departing a base station. In the first case we are waiting, in the latter case we are sending data (with some constraints, see below: sending strategy).

Additionally we integrated a sub-model for the energy production side, although being logically independent from the optimisation strategy. The reasons for this are first the fact that the time of energy generation has direct influence on the optimisation result and second the complex constraints in storing energy, especially when using capacitors.

2.3 Sending Strategy

This strategy makes implicit predictions about the further movement: If the sender has been approaching a base station during the last period, the predicted value for the further movement in the next period is a further approach (thus, sending later will be more efficient due to lower distances). If the sender has been departing during the last period, the predicted value for the further movement in the next period is a further departure (thus, sending later will be less efficient due to higher distances).

The downside of this strategy is the transmission delay of the sensor data. As we are waiting for energy optimal conditions, we can not guarantee maximum delay values, thus this approach is clearly not real-time capable. However in field surroundings which are naturally unsafe (the successful transmission can not be guaranteed anyway due to the sparse available energy) this drawback seems acceptable for us.

There are some other constraints in our sending strategy which shall ensure an efficient use of the available energy:

- Loss Threshold: If the probability of a packet loss is above a predefined threshold (which is the case for instance if the distance between sender and receiver is too long), we do not attempt to send.
- Data Threshold: If the amount of stored data increases a threshold (which is set to data buffer capacity minus the amount of newly produced data per time unit here, meaning that after the next cycle data loss can be expected, if no data can be successfully transmitted), we are sending data regardless the movement to or from a receiving base stations.
- Upper and Lower Energy Threshold: If the filling level of the energy storage exceeds an upper energy threshold, we make a sending attempt regardless the

movement of the sender, provided that energy level after sending is not expected to fall below a lower energy threshold. The reason for the upper threshold is that we might not be able to store the newly produced energy in the energy storage (e.g. capacitors), when the storage is already charged too high (see below: energy management). The reason for the lower threshold is that sending attempts at great distances would lead to almost emptying the storage at just one cycle tick. Especially in scenarios with few newly produced energy (see below: simulation scenarios) this could cause a sending inability even at energetically auspicious situations.

Figure 1 shows the flow chart of the sending strategy:

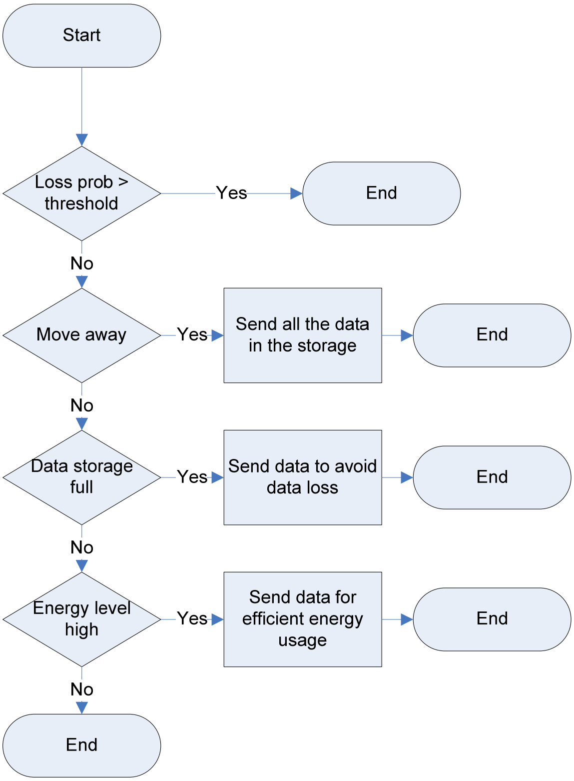


Fig. 1. Sending strategy flow chart

2.4 Simulation Scenarios

We applied our model to several practical application scenarios:

- The skiing scenario (Veichtlbauer & Dorfinger, 2007): A skier is equipped with intelligent skis with integrated sensors and energy harvesters. The sensors collect data in regularly intervals and store them in the local buffer. The energy harvesters produce energy during the run, e.g. by electromagnetic induction (EnOcean, 2007).

The energy generation is dependent on the movement (see fig.2). The energy is used to transmit the sensor data to a single fixed receiver.

- The cloud scenario (Dorfinger & Veichtlbauer, 2008): 20 Sensors are placed by an aeroplane to perform several measurement tasks in the air. They communicate with a grid of 16 fixed receivers on the ground, forming a 4.5 x 4.5 km square in total. Energy is stored in capacitors with total capacity of 600 µF. They are fully loaded at the start of their operation, i.e. they have an initial voltage of 12 V. No new energy is generated during the operation.

In order to examine the results of our model approach in different environments, we conducted several simulations with these scenarios. For the skiing scenario we made some additional assumptions (see above: model assumptions):

- The sender moves in different moving patterns along the fixed receiver (WLAN base station): We used straight moves, 2 different sine curves and a combination of sine and straight movement (see fig. 2).
- Energy is generated only at the sine parts (with 4 “passes” per second). The amount of produced energy per pass (see below: energy management) on the sender side is constant.
- For storing the energy (see below: energy management) we used 5 capacitors with 47 µF capacity each.
- The amount of produced (sensor) data per pass (and thus per time unit) on the sender side is constant.

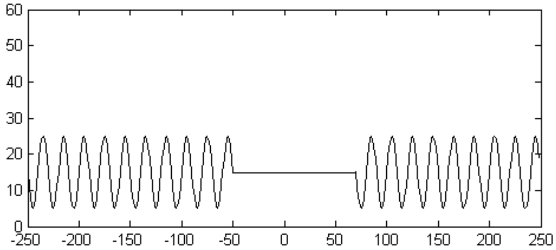


Fig. 2. Movement pattern of skiing scenario

2.5 Energy Management

For those scenarios where new energy is produced during operation (e.g. the skiing scenario) we assumed that the energy is provided by an energy harvester, e.g. the ECO 100 from EnOcean (EnOcean, 2007). This was motivated by our work in the project ASki where we built a prototype for the skiing scenario with an energy harvester placed on a ski. For those scenarios where all energy is pre-loaded (e.g. the cloud scenario) we used the same model, just setting the amount of energy generated during operation to zero.

The energy harvester is able to provide a voltage (see fig. 3) showing periodical peaks (“passes”). The original voltage pulse (green) is approximated by a triangle voltage (yellow), which is assumed to be our input voltage curve. The triangle voltage is described by the maximum input voltage and the duration of the pass. This model can be easily adapted to work with any kind of periodical energy source.

When using capacitors, energy can only be stored provided that the voltage of the produced energy is higher than the current voltage level in the capacitor (red). Thus, for all scenarios where we are able to produce new energy in the field, it is beneficial to keep the energy filling status on a lower level, as it is easier to charge the capacitors then. This can be done by setting the upper energy threshold to a comparatively lower level. The amount of energy which can be stored in capacitors is modelled in an extra sub-model (see below: energy storage model).

If we do not produce new energy, but use only stored energy from external sources, this constraint will be kept inactive by setting the upper energy threshold to the energy storing capacity (see above: sending strategy). Hence it is possible to use the same model without changes.

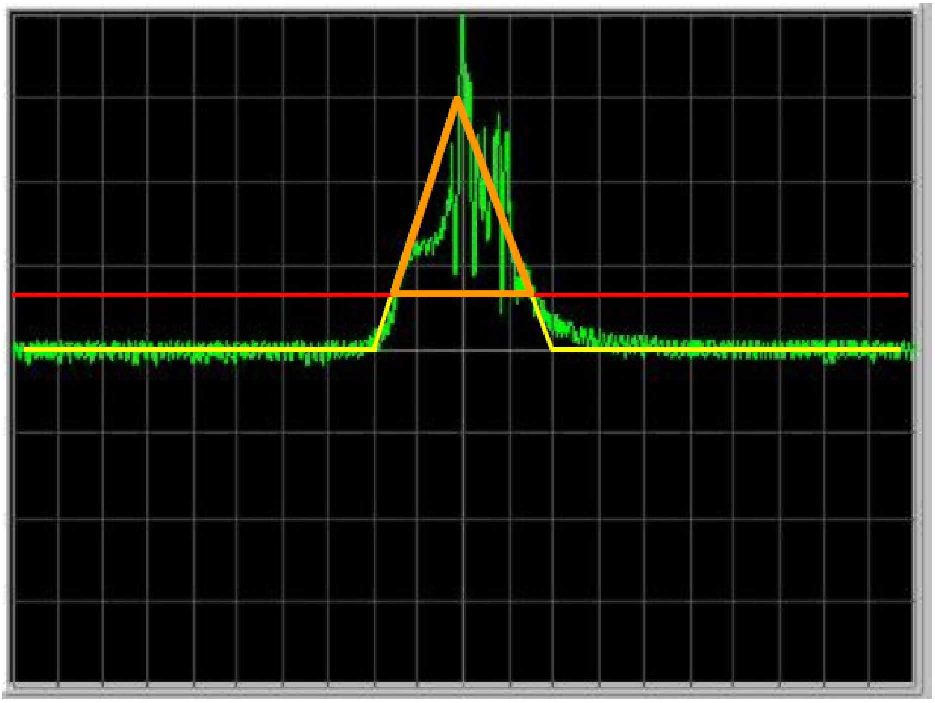


Fig. 3. Useable energy of triangle voltage

The amount of consumed energy per transferred bit is first dependent on the sending power. Second the packet loss probability has influence, because lost packets have to be retransmitted. The occurrence of a packet loss is dependent on the distance between sender and receiver, the packet length (`Pl`) as well as on the sending power. Yet it is a stochastic event, which has to be modelled properly (see below: loss model).

The probability of a packet loss is called packet error rate (`PER`). It is calculated based on the bit error rate (`BER`): $\mathtt{PER} = (1-(1-\mathtt{BER})^{\mathtt{Pl}})$. In the simulations we used a random number based on `PER` to determine whether the packet has been transmitted correctly or not. If the data is received correctly, it can be deleted from the sender's data storage.

3. Implementation of the Model

In the following our basic model and all of its sub-components (blocks) are described in detail. As model description language MATLAB/Simulink was used.

3.1 Basic Model

Our basic model consists of two main blocks (see fig. 4): The `Energy Storage` block, where the energy generation and energy storage behaviour is modelled (see below: Energy storage model), and the `Energy Cons` block (see below: Energy consumption model) modelling the energy consumption of the WLAN sender. The model has three input parameters:

- The energy produced during the last time interval
- The data produced by the sensors during the last time interval
- The current distance between the WLAN sender and the base station

The main interest is to successfully transmit as many data as possible. Furthermore we want to keep the amount of data that is overwritten in the data storage before being successfully transmitted (which is lost then) minimal. Consequently the output parameters of our basic model are:

- The aggregate of received data over simulation time
- The aggregate of overwritten (lost) data over simulation time

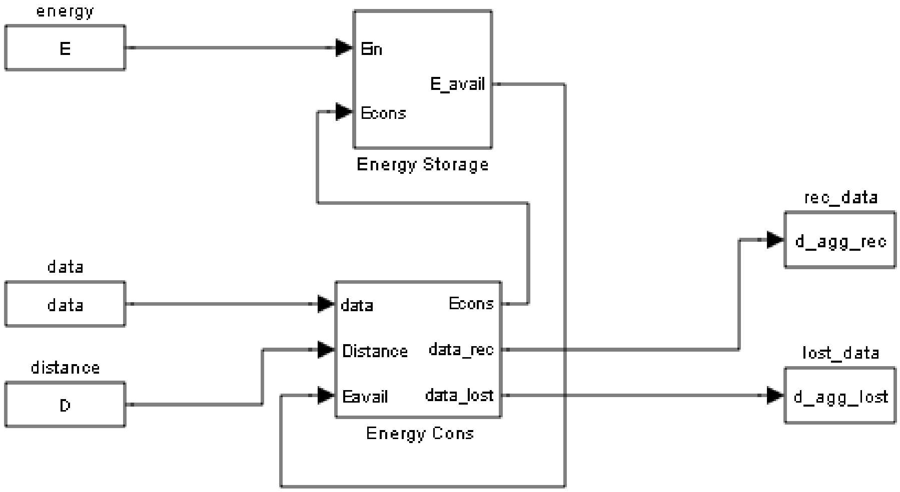


Fig. 4. Basic Model

3.2 Energy Storage Model

The main building block of the energy storage model (see fig.5) is a MATLAB function that calculates the current energy in the storage. As input parameter the model gets the energy produced during the last time interval (`Ein`) and the energy consumed during the last interval (`Econs`). The output is the available energy for transmission (`E_avail`).

For energy production we use an energy harvester (EnOcean, 2007); for energy storage we use common capacitors. The model uses the following parameters:

- Total capacity of the capacitors (`C`)
- Resistance of capacitor (`Rc`)
- Maximum voltage of energy triangle (`Ugmax`)
- Duration of the energy pass (`dur_pass`)
- Minimum voltage difference between energy source and capacitor that is needed to load the capacitors (`Uckorr`)
- Energy per pass (`Ep`)
- Maximum energy that can be stored in the capacitors (`Estoremax`)
- Minimum energy in capacitors, i.e. energy that remains in capacitors and can not be used by energy consumers (`Estoremin`)

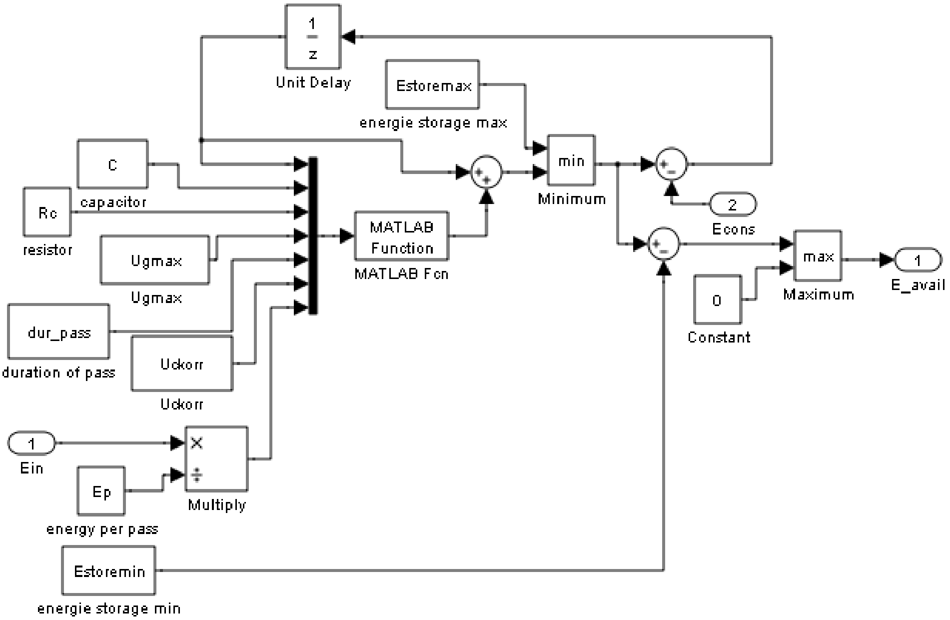


Fig. 5. Energy Storage Model

3.3 Energy Consumption Model

The energy consumption model (see fig. 6) consists of 6 main blocks:

- Distance model (`Dist_model`): Prediction of the further movement of the sender and calculation of the sending position

- Parameter model (`ideal send param`): Calculation of ideal parameters for data transmission
- Data storage (`data storage`): Calculation of the current filling level of the data buffer storage
- Sending decision (`send data?`): Decision whether to send data in the next time slot or not
- Link loss model (`link loss`): Determination of successfully transmitted and corrupted data packets (which have to be retransmitted and can not be deleted from the data storage)
- Data aggregation (`Aggregate`): Aggregation of successfully transmitted and lost data bits

Input signals for the energy consumption model are: The current distance (`Distance`), the data produced during the last interval (`data`) and the available energy from the energy storage (`Eavail`).

Output signals are: The consumed energy (`Econs`), the data successfully transmitted to the base station (`data_rec`) and the data lost by overwriting them in the data storage (`data_lost`).

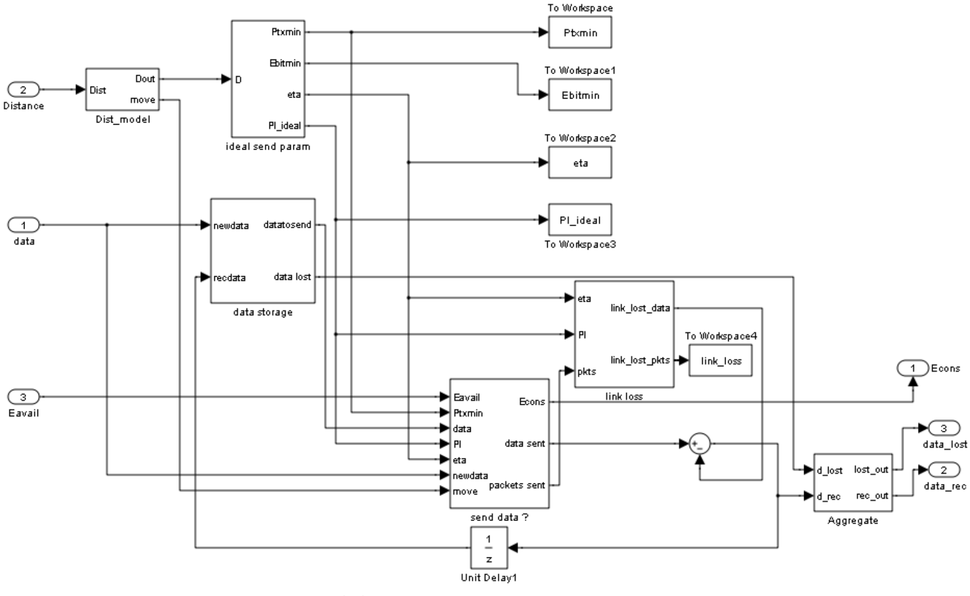


Fig. 6. Energy Consumption Model

3.4 Distance Model

The distance model (see fig. 7) calculates whether the sender is moving towards the base station or departing from the receiver by comparing the current distance with the distance of the previous clock cycle and assuming that the movement continues that way also for the upcoming cycle time. From that movement prediction the sending distance (which is then used for the calculation of the other sending parameters) is derived.

As argued by Ebert (Ebert, 2004), it is better to overestimate the distance than to underestimate it, because the sending power adaptation is not symmetric: If the sending power is too low, the loss probability (and thus the energy per *correct* transmitted bit) increases much faster than the energy per sent packet increases in the case when the sending power is too high.

Consequently for a movement towards the base station the output value for the distance is the current position, whereas for a movement departing from the base station the output value is an estimation of the position at the end of the time interval. As it is assumed that the movement continues the same way as in the last time interval, the estimated position is the current position plus the movement during the last time interval.

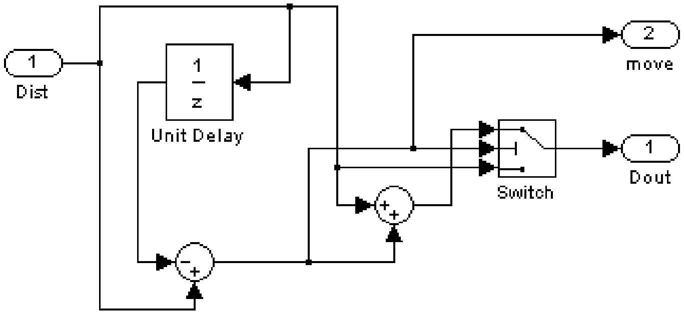


Fig. 7. Distance Model

3.5 Parameter Model

The parameter model consists basically of a MATLAB function which calculates the ideal sending parameters based on the Ebert model (Ebert, 2004).

As input parameters the MATLAB function receives technical parameters describing the WLAN connection: Sender gain, receiver gain, fade margin, receiver noise, bandwidth, sending rate, loss threshold, sending duration for 1 bit, wave length, noise, maximum packet size without header, overhead, and a correction constant. We kept these parameters constant in our simulations, yet they could easily be varied over time by setting appropriate values in the MATLAB configuration file. Furthermore the distance between sender and receiver is used as variable input parameter to the parameter model.

As output parameter we retrieve the ideal sending power (`Ptxmin`), the energy needed for transmission of one bit (`Ebitmin`), the probability that a packet is successfully transmitted (`eta`) and the ideal packet length for the transmission (`Pl_ideal`).

3.6 Data Storage Model

The data storage model calculates the current filling status of the data buffer storage by subtracting the data which has been successfully transmitted in the last time interval (`rec_data`) from last cycle's filling level and adding the data which has been newly produced during the last time interval (`newdata`). These two values are the input parameters of the data storage model.

The storage has a maximum size (`datamax`), and is organised as a ring buffer, i.e. exceeding the maximum value leads to data loss by overwriting the oldest stored data with the newly produced data. Hence the output parameters are the filling level, i.e. the amount of data which can be transmitted in this time interval (`datatosend`), and the amount of overwritten data (`data lost`).

3.7 Sending Decision Model

The sending decision model (see fig. 8) calculates the amount of data that are sent in the upcoming time interval.

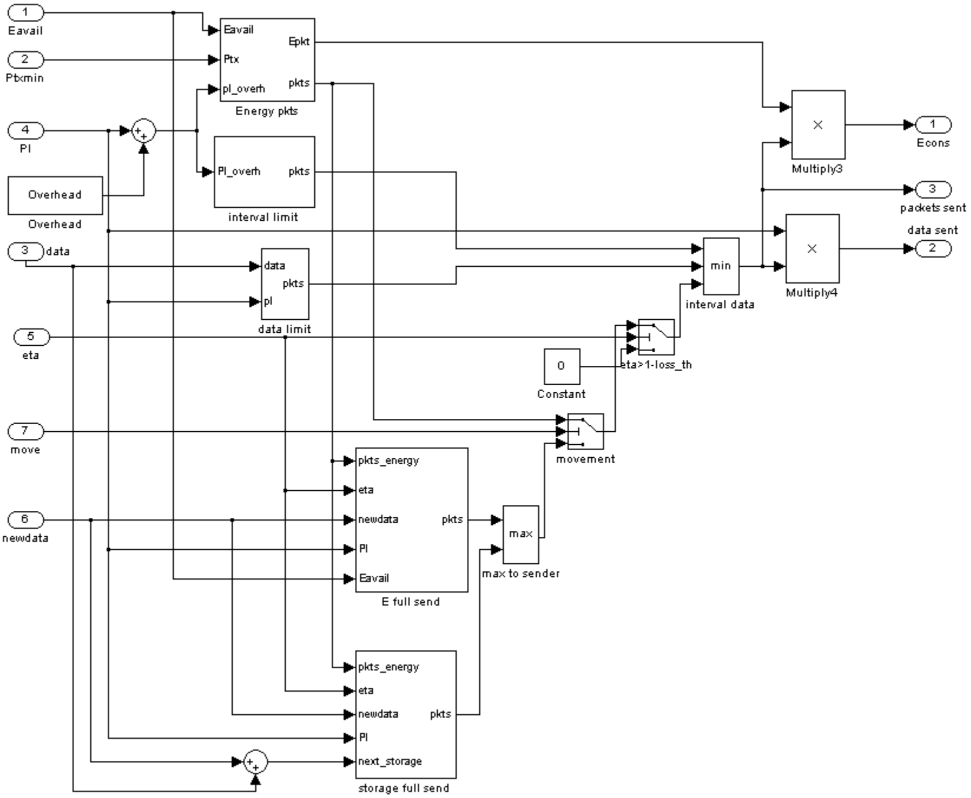


Fig. 8. Sending Decision Model

It consists of five sub-models; each of them determines the number of packets that could be sent taking into account different premises:

- How many packets can be transmitted if all the available energy is spent for transmission?
- How many packets can be transmitted within one time interval?
- How many packets can be filled with data from the storage?

- How many packets should be transmitted to allow efficient usage of the energy storage?
- How many packets should be transmitted to prevent overwriting data in the data storage?

The model provides us with three outputs:

- The consumed energy
- The sent data
- The number of sent packets

The number of sent packets is zero if the probability of a packet loss on the link is greater than a given threshold (see above: sending strategy).

If we are moving away from the base station, all data in the data storage are sent, except for the rest that does not fill a full packet with ideal packet length. Thereby we are taking into account the maximum amount of data that can be sent with the available energy and within one time interval. If we are moving towards the base station the same energy and time constraints are taken into account; furthermore we pay attention to the objectives to prevent data loss in the data storage and to allow efficient energy storage (see above: energy management).

In the next sub-sections some details about the main building blocks, including their input and output parameters, are given.

3.8 Packet Energy Model

The packet energy building block (see fig. 9) receives the following input signals: The available energy (`Eavail`), the ideal transmission power (`Ptx`), and the packet length including overhead (`pl_overh`).

Output variables are: The energy per packet (`Epkt`), and the number of packets that can be transmitted when consuming all available energy in the energy storage (pkts).

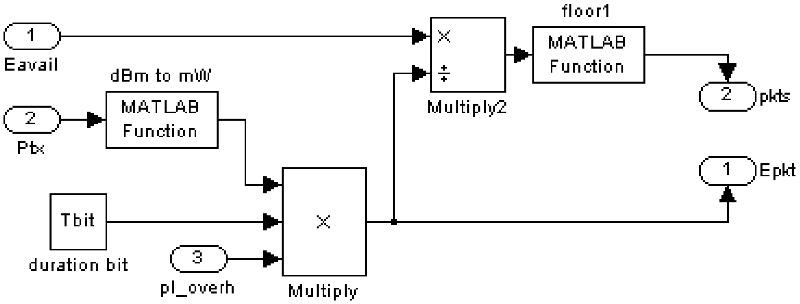


Fig. 9. Packet Energy Model

3.9 Interval Limit Model

The interval limit building block receives the packet length including overhead as input parameter. It calculates the number of whole packets that can be sent within one time interval, which is also the only output parameter.

3.10 Data Limit Model

The data limit building block has the following input parameters: The current level of data in the data buffer storage, and the packet length without header. It calculates the number of packets that can be filled with data from the storage. This is the only output parameter of the data limit model.

3.11 Energy Efficiency Model

To make energy usage more efficient (see above: sending strategy), we use the energy efficiency model (see fig. 10).

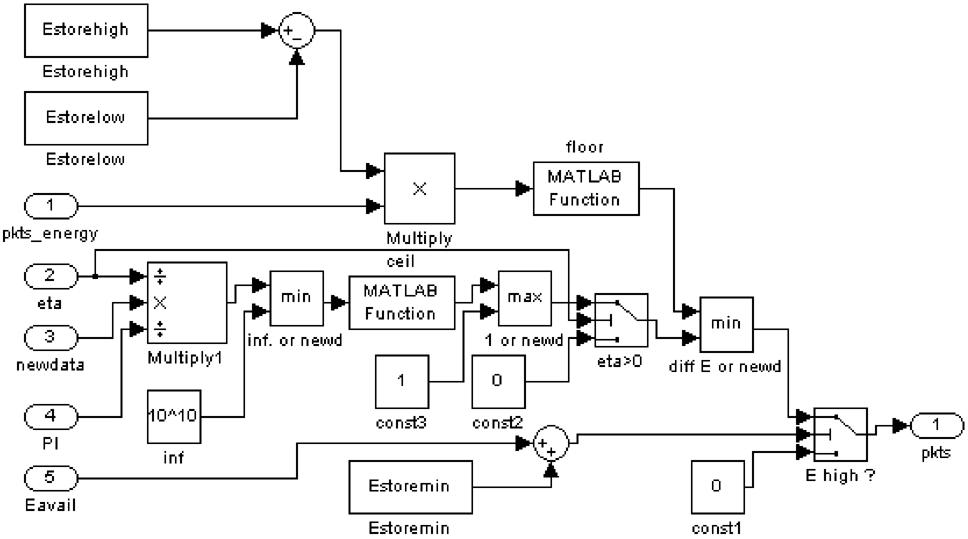


Fig. 10. Energy Efficiency Model

If the energy in the storage is above an upper threshold (`Estorehigh`), we transmit `ceil(1/eta*newdata/Pl)` packets, where `eta` is the probability that a transmission is successful, `newdata` is the amount of data stored in the last interval and `Pl` is the ideal packet length without header.

Thereby we have to guarantee, that the energy stored in the capacitors does not fall below a lower threshold (`Estorelow`) after data transmission, i.e. we transmit the maximum possible number of packets such that the energy consumption by the data transmission is low enough to keep this constraint.

3.12 Data Efficiency Model

The data efficiency model (see fig. 11) is used to prevent data loss in the storage during the time when the sender is moving towards the base station.

If the amount of data in the storage plus the amount of data received in the upcoming time interval is expected to exceed the capacity of the storage, we transmit a number of `ceil(1/eta*newdata/Pl)` packets.

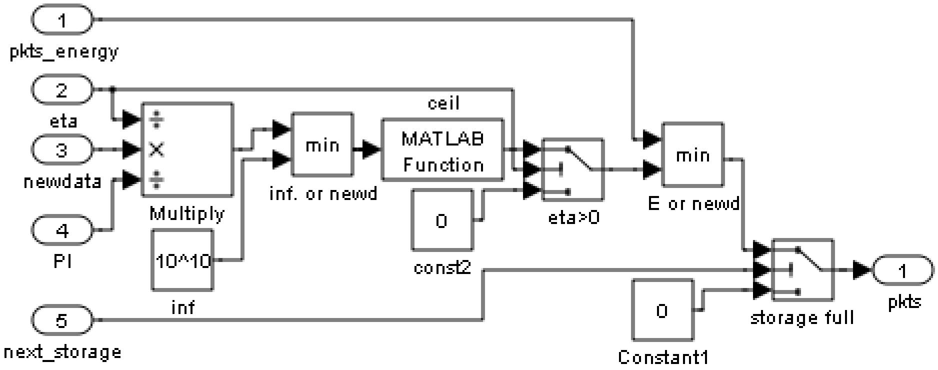


Fig. 11. Data Efficiency Model

The data efficiency model receives as input parameters: The number of packets that can be sent with all available energy (`pkts_energy`), the probability that a packet is successfully transmitted (`eta`), the data received in the last interval (`newdata`), the packet length (`Pl`) and a prediction of the next filling level of the data storage (`next_storage`). The output parameter is the number of packets that should be transmitted (`pkts`).

4. Parameter Tuning

In a number of simulations we have investigated the advantages of this model approach compared to the Ebert model and to a non-optimised episodic protocol (Veichtlbauer & Dorfinger, 2007; Dorfinger & Veichtlbauer, 2008; Veichtlbauer & Dorfinger, 2008). With optimal parameter settings however, some percent additional efficiency gain could be achieved.

To investigate the influence of different settings, several studies in the skiing environment have been performed. Thereby the setup of the main factors that influence throughput and data loss has been studied:

- Capacity of energy storage
- Size of data storage
- Energy threshold
- Loss threshold

For the simulation with different sizes of energy and data storages we got the expected results: The bigger the storage, the greater the number of successfully transmitted packets, and the lower the packet loss. For the setting of the energy thresholds we got similar results for different parameter sets. In the performed scenarios there is no strong argument for a certain parameter set of the energy thresholds.

The most interesting parameter in our simulations of the skiing scenario was the loss threshold. We conducted simulation runs with several different movements, e.g. a straight movement (see fig. 12, table 1) and a sine movement (see fig. 13, table 2).

Table 1 shows statistical results for different values of loss threshold in the skiing scenario with straight movement. For each value 100 simulation runs have been performed.

Loss_th	mean	standard dev.	95% confidence interval	min	max
1.0	4.373.680	22.115	[4.369.346, 4.378.014]	4.330.832	4.443.256
0.9	4.365.656	22.223	[4.361.300, 4.370.011]	4.313.576	4.444.368
0.5	4.342.520	25.479	[4.337.526, 4.347.514]	4.273.690	4.401.360
0.3	4.292.992	22.868	[4.288.509, 4.297.474]	4.241.280	4.347.520
0.1	4.136.336	19.153	[4.132.582, 4.140.897]	4.090.376	4.192.672

Table 1. Throughput for different values of loss threshold with straight movement

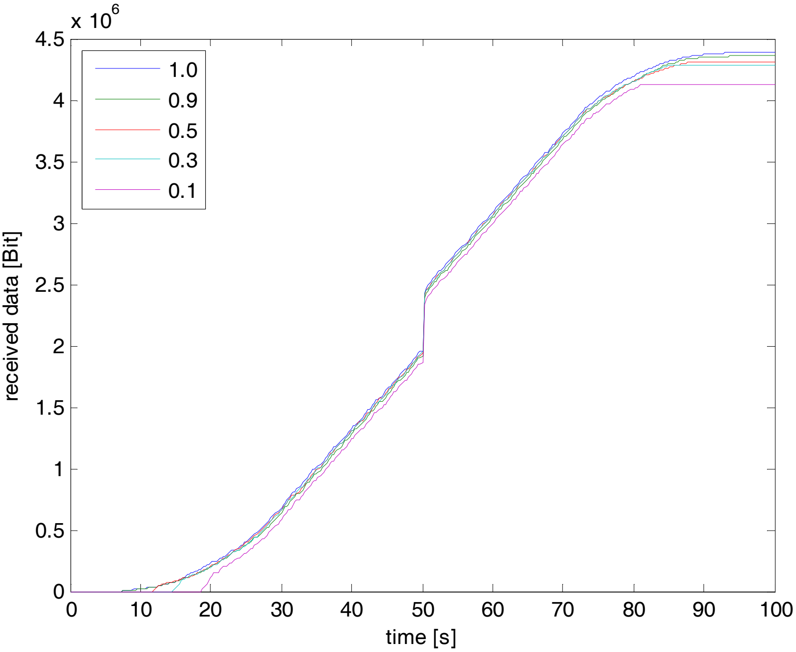


Fig. 12. Received data for different values of loss threshold with straight movement

As it can be seen, the lower the loss threshold is set, the less data is received. A detailed analysis has shown that it would be beneficial to use different settings for loss threshold in the approaching phase and in the departing phase of a simulation of the skiing scenario: During the approaching phase a loss threshold of about 0.9 would perform best. During the departure phase transmission attempts should be performed as long as there is a possibility to successfully transmit data packets, thus the loss threshold should be set to 1.

Table 2 shows statistical results for different values of loss threshold in the skiing scenario with sine movement. Again, for each value 100 simulation runs have been performed.

Loss_th	mean	Standard dev.	95% confidence interval	min	max
1.0	4.357.283	22.628	[4.352.848, 4.361.718]	4.304.752	4.401.944
0.9	4.356.078	21.182	[4.351.926, 4.360.230]	4.311.872	4.400.640
0.5	4.336.494	23.374	[4.331.913, 4.341.076]	4.269.368	4.386.384
0.3	4.277.693	22.323	[4.273.317, 4.282.068]	4.215.872	4.343.960
0.1	4.118.336	21.435	[4.114.134, 4.122.537]	4.038.416	4.154.080

Table 2. Throughput for different values of loss threshold with sine movement

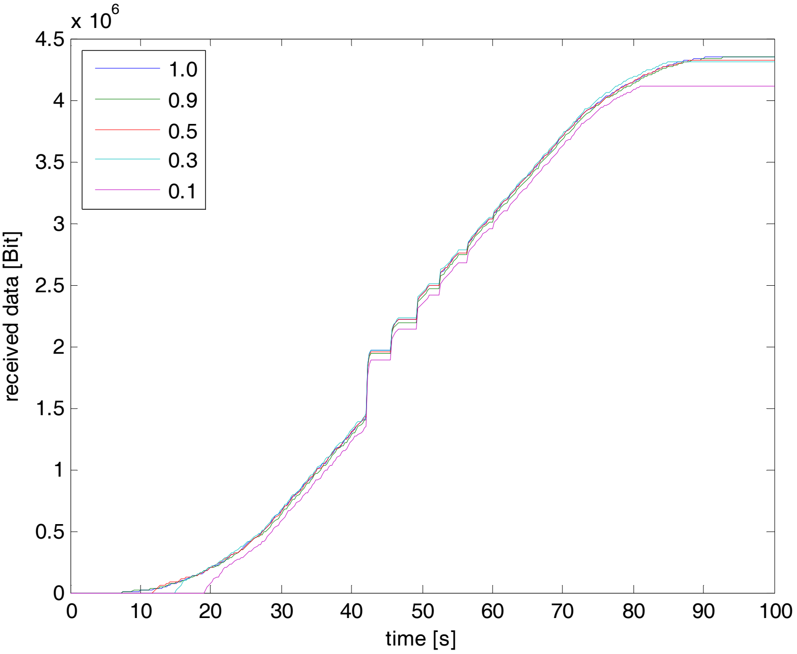


Fig. 13. Received data for different values of loss threshold with sine movement

Also for the sine movement pattern in the skiing scenario a loss threshold of 1 performs best. This is in contrast to our findings for the cloud scenario (Veichtlbauer & Dorfinger, 2008), where we concluded that for pre-loaded energy sources smaller values for the loss threshold lead to better performance.

An overall conclusion of our investigations in setting the loss threshold parameter is that it depends very strongly on the kind of energy source how to optimise the parameter setting. For pre-loaded energy sources with no further energy generation during the simulation a small value for the loss threshold is advisable, whereas for energy sources that supply energy during the run (energy harvesters) a loss threshold near to 1 should be used.

5. Conclusion

As a result of our simulations we can see a remarkable improvement (Veichtlbauer & Dorfinger, 2007; Dorfinger & Veichtlbauer, 2008) of the use of energy compared with the underlying Ebert model (Ebert, 2004). Yet the efficiency gain is very much dependent on the applied scenario. Especially in scenarios where energy is produced regularly during the operation of the communication system, the gain is only a few percent.

However, energy efficiency is a much more critical issue in scenarios where no or only sporadic energy production is possible. Our model has been developed for mobile scenarios with sparse energy. Here the strengths of our approach come into effect, as we have proved in the mentioned examples.

6. Future Work

Obviously, energy is consumed not only by (sending and receiving) antennas, but also in other parts of embedded systems (especially microcontrollers/microprocessors) – yet our focus was set on the communication aspects, and we disregarded all other energy consumers. Furthermore, we just touched on the topic of energy generation. Basically, we assumed that energy is either stored (in capacitors or batteries) or produced live according to the movement pattern of the sender. In both areas very interesting future research topics can be defined.

We are especially interested in the question of the "distribution of intelligence" in the network (i.e. should calculations be performed locally and their results be transmitted to a data base, or should just the raw data be transmitted and the calculation be performed centrally?). We consider that this topic has the potential for several years of research in future research projects. We have already made some effort in the application domain of ICT support for dynamic evacuation.

The challenge is to decentralise the intelligence of an evacuation support system for emergency cases (fire or gas in a building) in order to provide situational and personalised information for evacuees without overloading the network nodes. Besides energy aspects (in case of a breakdown of the regular power supply) one has to face real-time, safety and security aspects. Thus policies determining which data have to be transmitted when and where have to be defined (Hofmann et al., 2009).

Another research challenge is to define standards for open sensor/actuator systems for building automation. Our goal is the prototypical realisation of a generic in-house communication infrastructure providing a multi-user/multi-application approach, i.e. every registered user has access to sensor data (if allowed; also a billing system is possible here) and to applications that perform control tasks (e.g. remote heating/cooling). Similar solutions can be thought of also for traffic control. For instance a driver could access traffic data and plan/modify the route of the journey according to the collected sensor data.

7. References

Burns, B.; and Ebert, J. P. (2001) Power Consumption, Throughput and Packet Error Measurement of IEEE 802.11 WLAN Interface2. *TKN Technical Report TKN-01-007*, Berlin, August 2001.

Dorfinger, P.; and Veichtlbauer, A. (2008) Simulation of Energy Efficient Communication from Flying Sensors to a Grid of Base Stations on the Ground. *Proceedings of the 15th International Conference on Telecommunications (ICT 2008)*, St. Petersburg, June 2008

Ebert, J. P.; and Wolisz, A. (1999) Power Saving in Wireless LANs: Analysing the RF Transmission Power and MAC Retransmission Trade-Off. *ITG Fachbericht 157*, pp 187- 192, Munich, October 1999.

Ebert, J. P.; Trammel, B.; Wiederhold, E.; and Wolisz, A. (2000) Energy-efficient Power Control Approach for WLANs. *Journal of Communications and Networks (JCN)*, September 2000.

Ebert, J. P.; and Wolisz, A. (2000) Combined Tuning of RF Power and Medium Access Control for WLANs. *Journal of Mobile Networks & Applications*, vol. 6, no 5, pp 417-426, Berlin, September 2000.

Ebert, J. P.; Aier, S.; Kofahl, G.; Becker, A.; Burns B.; and Wolisz, A. (2002) Measurement and Simulation of the Energy Consumption of WLAN Interface. *TKN Technical Report TKN-02-010*, Berlin, June 2002.

Ebert, J. P. (2004) Energies-efficient Communication in Ad Hoc Wireless Local Area Networks. *Dissertation*, TU Berlin

EnOcean GmbH (2007): Energy harvester ECO 100. *Internet Document*, May 2007 http://www.enocean.com/php/upload/pdf/DB_DE20.pdf

Gilbert, E. N. (1960) Capacity of a burst-noise channel. *Bell Systems Technical Journal*, 39:1253–1265, September 1960

Haber, P.; Bergholz, G.; Hofmann, U.; Miloucheva, I. (2003) Time and Rate Continuous Multiclass Fluid Simulation Model for Inter-domain traffic flow simulation. *Proceedings of the first international workshop on Inter-domain performance and simulation (IPS 2003)*, Salzburg, Februar 2003

Hofmann, U.; Miloucheva, I.; and Veichtlbauer, A. (2009) Dynamic Evacuation Architecture using Context-Aware Policy Management. *International Journal of Computer Science and Applications, Year 2009: Volume VI Issue II - Special Issue on Networking Mobile Virtual Knowledge*, Feb. 2009

Matzen, B.; Ebert, J. P.; and Karl, H. (2003) Electromagnetic emission reduction for radio communication networks by using a multi hop ad hoc approach. *TKN Technical Report TKN-03-004*, Berlin, February 2003.

Veichtlbauer, A.; and Dorfinger, P. (2007) Modeling of Energy Efficient Wireless Communication. *Proceedings of the 2007 International Conference on Software, Telecommunications and Computer Networks (SoftCom 2007)*, September 2007

Veichtlbauer, A.; and Dorfinger, P. (2008) Energy Efficient Communication in a Skiing Environment. *Proceedings of the 7th International Conference Communications 2008 (COMM 2008)*, Bukarest, June 2008

Veichtlbauer, A.; and Dorfinger, P. (2009) Modeling and Simulation of Energy Efficient Communication in a Skiing Environment. *MTA Review Vol. XIX, No. 1*, pp. 55-76, March 2009

Zyren, J.; and Petrik, A. (1998) Tutorial on Basic Link Budget Analysis. *Application Note AN9804*, Harris Semiconductor, April 1998.

15

Towards Performance Enhancement of Short Range Wireless Communications in Reliability- and Delay-Critical Applications

Yang Liu and Ye Liu
Department of Computer Science, University of Vaasa
PL 700, 65101 Vaasa, Finland
Email: yang.liu@uwasa.fi

1. Introduction

More and more applications demand highly reliable and low latency short range wireless communications nowadays, one extreme example of which is the wireless communication used in RoboCup Small Size League (SSL) robots (Liu et al, 2007). RoboCup is the world's top level international robotics competition held every year, and SSL is for a team of multiple fast-moving robots under a dynamic environment to autonomously play soccer game against another team. Due to the highly dynamic nature of the competition, the requirements and constraints for the wireless communication are extremely tight. The challenge is that wireless communication is involved in the control loop and therefore the reliability and propagation delay are vital factors which directly affect the team performance. Beside, various interferences with known and unknown frequency / transmission power usually present at the competition site, which is hazardous environment to achieve reliable and low latency performance for wireless communication. This study investigates the performance strengths and weaknesses of various short range wireless communications e.g. RadioMetrix, IEEE 802.11a/b, IEEE 802.15.4, DECT, Linx, etc, which are commonly used nowadays in different RoboCup SSL wireless communication implementations. Unfortunately most of these commercial solutions are not able to provide satisfactory performance to such kind of reliability- and delay-critical applications especially under interferences. In the case study, a typical commercial short range wireless communication module which has weak immunity to interference has been tested and its performance has been evaluated with test bed. An adaptive error correction and frequency hopping scheme (Liu, 2008) has been proposed to improve its immunity to interference and therefore enhance the wireless communication performance for reliability and delay-critical applications. Such scheme can be easily adopted to similar applications using short range wireless communications.

2. Communication System Design and Testing

2.1 Choosing wireless technologies

Many different wireless technologies have been considered for use in the robots. The main ones are: RadioMetrix 433 MHz and 869 MHz RF, IEEE 802.11a/b, IEEE 802.15.4, DECT, Linx, etc. While making decision which technology to choose, we also need to keep in mind about RoboCup's rules and regulations and also compliance with regulations of the country hosting the competition. Before the competition all teams should notify the local organizing committee of the wireless communication technology, power and frequency. To avoid direct interference, each team should be able to select between at least two carrier frequency bands before the match.

After experienced unsatisfied performance from the RF modules, the prospective choices are from IEEE 802.11a/b, IEEE 802.15.4, DECT and Linx. Among these wireless technologies, IEEE 802.11a/b is based on CSMA/CA and therefore considered not to be optimum solution for real-time applications. According to Tse et al (2005), the performance of IEEE 802.15.4 drops significantly where there are many 802.11 terminals connected to access points, which is the case at the competition site, and therefore this is not considered as optimum solution either. Both DECT and Linx are designed to support voice transmission capability and optimized for real-time performance, so the communication system for the new generation robot design will use these two and choose the one which will perform better during the competition, according to the opponent team's radio to be used.

2.2 Designing the test bed

The purpose of this testing is to observe and study wireless communication performance of Linx modules such as round trip delay, bit error, packet error, RSSI (Received Signal Strength Indication), and how they are affected by interference. A test bed has been built to carry out the tests and collect data to a PC. Both wireless transmitter and receiver modules are connected to an ARM7 microcontroller UART port. The packets which have been transmitted over wireless link are compared by ARM7 with the packages that have been received. The testing data are sent to PC for further processing. The timer feature is used to record the transmission time per each byte, and the result are also read by the microprocessor and sent to PC. Linx HP3 RF modules also provide a RSSI function which is connected to ADC so that a digital RSSI value can be read to indicate each byte's signal strength. The test bed simulates a full-duplex wireless transmission. Linx transmitter A will send data through channel A to Linx receiver B. Transmitter B will send what receiver B received through channel B to receiver A. In such way, we could measure the time delay for the round trip, RSSI for both channel A and B, error rate of the data, etc.

Much attention has been put to design PCB carefully following the standard industrial practices and choosing high quality components.

For testing purposes a test bed has been designed and implemented which will be able to work with both DECT and Linx modules. Here we present briefly some board design issues with some theoretical background when needed. The testing board is made of two parts: the mother board and the daughter board for LCD display and buttons, which will be mounted on top of the mother board.

The board can support both serial communication (through its DB9 female connector) and Ethernet connection (using RJ-45 XPort jack). Serial connection is left for downwards compatibility with the old server. It's connected directly to DECT and Linx modules. In the near future only Ethernet interface will be used to communicate with the server, and therefore in the testing board we use two XPort devices. One is connected directly to communication modules. The other one is connected to microcontroller to be able to set and control communication parameters, like RSSI values and communication channel number for monitoring and data logging to be used for further analysis.

Linx HP3 Series transmitter and receiver are high-performance RF modules (Linx 2007a, Linx 2007b) commonly used in wireless data transfer and industrial automation. They use popular 902-928 MHz frequency band. On the PCB we have possibility for serial channel selection of 100 channels by using microcontroller's I/O ports. Anyway in case of problems with microcontroller's operation during the competition we have provided also parallel selection of 8 channels using octal rotary DIP switches. For digital data transmission, Linx uses FSK modulation to ensure reliable performance. Equation (1) shows the FSK modulation formula. In FSK the modulating signal shifts the output frequency between two discrete values so called mark and space frequencies:

$$s(t) = \begin{cases} A_c \cos(2\pi f_1 t), \; m(nT_b) = 1 \\ A_c \cos(2\pi f_2 t), \; m(nT_b) = -1 \end{cases} \tag{1}$$

Therefore comparing to On-Off Keying modulation, FSK has increased noise immunity and ability to capture in the presence of multiple signals which is really helpful especially in such crowded band. Using those SIP style Linx modules we don't need any additional RF components (except of the antenna of course). Receiver has an exceptional sensitivity of -100 dBm typical.

The critical requirement for both modules performance was big ground plane on lower layer of PCB. To minimize losses we had to stick to the rules of designing microstrip as shown in Figure 1, the trace running between module and antenna.

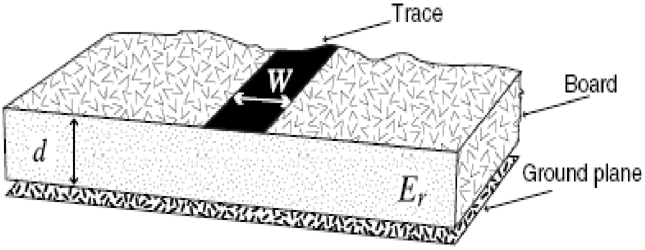


Fig. 1. Microstrip (Linx 2007b)

Microstrip design is based on Equation (2) and (3):

$$E_e = \frac{E_r + 1}{2} + \frac{E_r - 1}{2} \cdot \frac{1}{\sqrt{1 + 12d/W}} \tag{2}$$

$$Z_0 = \begin{cases} \frac{60}{\sqrt{E_e}} \cdot \ln\left(\frac{8d}{W} + \frac{W}{4d}\right) & \text{For } \frac{W}{d} \leq 1 \\ \frac{120\pi}{\sqrt{E_e} \cdot \left(\frac{W}{d} + 1.393 + 0.667 \cdot \ln\left(\frac{W}{d} + 1.444\right)\right)} & \text{For } \frac{W}{d} \geq 1 \end{cases} \tag{3}$$

E_r =Dielectric constant of PCB material

Another important thing is antenna's length. We have decided to use whip style antenna. Its optimal length can be calculated using Equation (4):

$$L = \frac{234}{f_{MHz}} \tag{4}$$

L is length in feet of quarter-wave length and f_{MHz} is the operating frequency in MHz.

Therefore in our case we will need two 8 cm long whip style antennas. The antennas are located as far as possible from each other and in the same line facing to the field with playing robots.

2.3 EMC and wireless performance testing

We have measured the testing board for radiated E-field emissions using European Union standard EN 55022 in EMC laboratory. This test measures unintentional E-field emissions from product in normal operating mode. Linx modules including a transmitter and a receiver have been tested. Simple application written in C and running on ATMega8535L microcontroller displays on LCD screen actual RSSI values. The board is connected with laptop computer outside EMC room with cross-over CAT5e cable. Client-server application written in Java constantly sends a packet which is passed to Linx transmitter and waits for response from the receiver, to check if the received value is correct. At the end of the measurement there is a table showing in percentage how many errors occurred at which particular frequencies.

The channel is set to 69. For these settings transmitter frequency should be 919.87 MHz. In EMC laboratory, jamming signals are feed at fine steps from 915 to 925 MHz. We were observing error rate occurring for different frequencies and in the same time we were recording Received Signal Strength Indicator (RSSI) value. The results from the measurements are shown in Table 1 and plotted in Figure 2.

F [MHz]	Errors [%]	RSSI
915	0.00	140
916	0.05	142
917	0.17	139
918	0.69	135
919	96.71	120
920	0.56	125
921	0.14	130
922	0.00	135
923	0.13	129
924	0.23	127
925	1.32	130

Table 1. Error rate and RSSI measurements

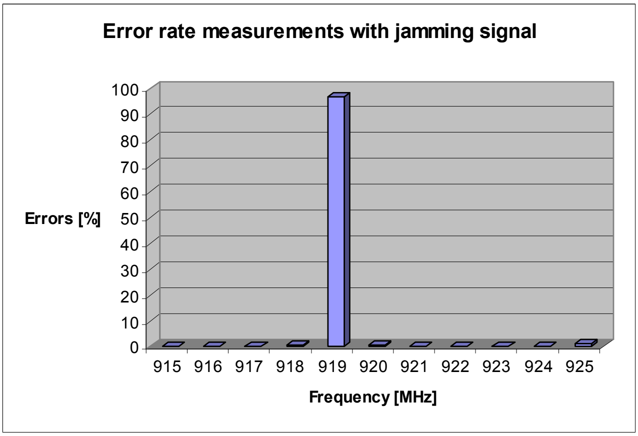


Fig. 2. Error rate measurements with jamming signal

As expected, most errors have occurred on transmitter frequency 919 MHz which is better visible in Figure 2 showing percentage of errors for given frequency. The value of RSSI was much lower around 919 and 920 MHz and it was varying rapidly around the values shown in Table 1, but the value has never gone below 120. The adjacent channels, although affected by the jamming frequencies too, have significantly lower error rates. This indicates that frequency hopping can be used as one very effective solution to improve the immunity to jamming signals.

Signal to Noise Ratio (SNR) variable in the formula below is defined as the signal to noise ratio. The probability of error can be calculated for a receiver system using non-coherent FSK modulation. The system is modelled with two matched filters which are centred at f1 and f2 with envelope detectors summed to a decision circuit. Equation (5) can be used to calculate the probability of error for non-coherent FSK modulation:

$$P_{\varepsilon} fsk_{SNR} = \frac{1}{2}\exp(-\frac{SNR}{4}) \tag{5}$$

Real-life performance of Linx wireless communication has been also intensively investigated in terms of round trip delay, bit error rate and packet error rate, etc, by interfering with simulated jamming signals, and results are shown in Figure 3.

The typical round trip time delay for Linx wireless transmission is around 209 µs to 809 µs after deducting the processing time. Under heavy interference (10 V/m) the most common bit error rate is 3 bits per byte. Hence, a 3-bit error correction scheme should be sufficient to significantly improve the reliability of wireless performance. Besides error correction, frequency hopping is also employed to resist interferences. The RSSI levels of all channels are scanned and a black list is generated so that all the bad channels are excluded from the random frequency hopping sequence to ensure the hopping channel quality, and therefore improve the communication performance even further.

Figure 4 shows a detailed snapshot of good quality channels. The sequences of the plot are RSSI, Delay, Bit Error Rate, and Packet Error Rates. The data cursor tells that while jam frequency are located among channel 40 to channel 48, the transmission give a better performance with strong RSSI, evenly distributed time delay, low bit error rate and low packet loss rate. Among channel 53 to channel 74, the transmission gives an even better performance. It can be concluded that, if the transmission is allocated from 1 to 10 channels below or above the jamming frequency, the wireless performance is significantly better. This has proved again that frequency hopping scheme will be quite effective to improve the existing wireless transmission.

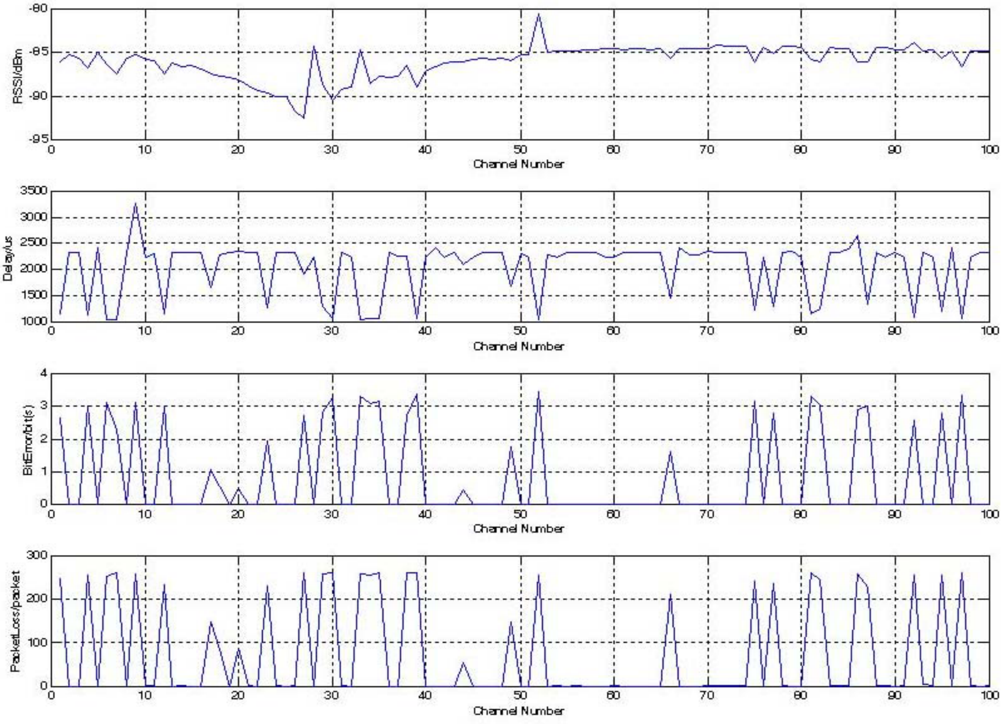


Fig. 3. Performance evaluation of RSSI, delay, bit error rate, packet loss rate

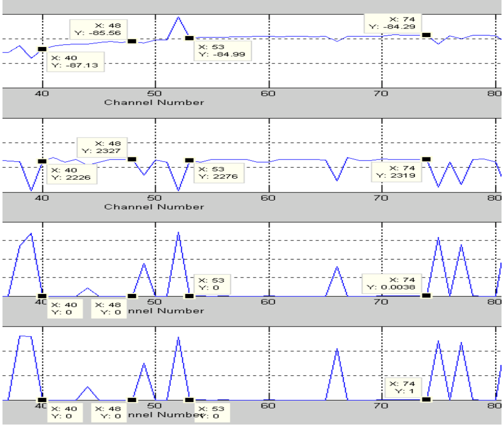


Fig. 4. Performance of RSSI, delay, bit error rate, packet loss rate in good quality channels

3. Analysis and Enhancement of Wireless Communication Performance

3.1 General analysis of improvements proposals

Reliable communication is critical for success in the competition. Most intelligence and all tactics are processed on the centralized server so if there is excessive data error or delay in the communication with the server, robots cannot achieve reliable real-time performance. In case of this application, propagation delay in the communication is much more critical than the throughput. Specific data transmission doesn't require high throughput but working with real-time systems means that information which will come to receiver too late is useless, and will result in foul or loosing a goal during competition.

During the tests it has been noticed that 96.71% of errors occur on the same frequency as the one set on Linx module. Any concurrent transmission on that frequency causes excessive errors and decreases reliability. In case of very small SNR the system doesn't work at all.

Because the described communication system has to work in very noisy environment we have to consider use of additional error detection and correction algorithms to improve communication's reliability. One of the improvements could be dynamic checking for channel with the highest SNR and choosing the best channel for communication purpose.

Other idea is to use two antennas inside the robots making use of antenna diversity, the fact that fading is space dependant and one of the antennas will receive stronger signal than the other. This is important especially that the robots move fast and can change position and orientation rapidly.

Because the data is sent to robots using broadcast, not point to point, so there are more than one channel which could be used. Therefore there is possibility to use frequency diversity, using two channels to communicate with robots (to broadcast the same data). The channel which is more probable to have correct data is the one with the highest SNR.

Finally, from the measurement results obtained from the test bed, it can be concluded that utilizing frequency hopping is probably the most effective solution to improve the wireless communication performance without changing much existing communication system structure.

3.2 Adaptive frequency hopping

An adaptive frequency hopping scheme with duplex link is proposed as an improvement solution. Frequency hopping is a powerful solution towards to interference and multi-path fading. As normal frequency hopping system, a hop-sequence generator has to be determined, so that the frequency hopping system will avoid the congested channel and transmit the data in clear channel. The hop-sequence generator is determined based on the link quality testing results as well as the analysis of wireless communication theory.

Frequency hopping is categorized into slow hopping and fast hopping. By slow hopping, more than one data symbol is transmitted in same channel. Fast hopping change frequency several times during one symbol (Rappaport, 2001). The pattern of channel usage is called the hopping sequence (Jochen, 2001). Explain the notion of hopping sequences in a more practical way is that how to determine the next channel to hop. There are two types of hopping sequence: random hopping sequences and the deterministic hopping sequences. In this study a hopping system with slow deterministic frequency hopping sequence is focused. Compare with other kinds of wireless communications, HF communication selectively fading because of the multi-path propagation and abundance interference from others.

Hence, a channel with sustaining stable signal-to-interference ratio, SIR is normally used for a narrowband communication, and new channel is used according to the same SIR criteria when the previous channel quality changes. Narrowband schemes will not work as expected when there is noise burst although it is very efficient while slowing changing interference and fading environment is engaged. For wideband communication, a random frequency hopping schemes together with forward error correction, FEC method, the short noise burst will not effect wideband transmission; it is protected from jamming transmission and has low probability of interception.

The improvement proposal is meant to combine the advantage in of both narrowband and wideband schemes. Additionally, from previews EMC test results in RSSI tells each channel's quality so that a "bad channel" list can be generated. Therefore, the improvement proposal is determined, using a wideband frequency hopping scheme that would avoid known interference: the "bad channels". This can be done by functionalized one of the duplex channel as the feedback channel. The feedback information contains the channel numbers which are in use. The selection of active channels is based from EMC RSSI testing results, the channel RSSI which is lower than -95 dBm is considered as "bad channel". This system is an adapted frequency hopping system which can utilize the existing hardware functionalities to achieve the optimized wireless performance.

For this duplex communication system, the selection of which frequencies to be used is based on the feedback from uplink. As motioned before, the downlink transmitter A transmit the data, at the receiver B side the RSSI value of downlink which in this case is equivalent as SIR, is measured. At the receiver side a link quality analysis scheme is implemented, once the SIR is below the criterion, for instance, -95 dBm, LQA will determine that the channel needs to be switched. After the performance of the hopping sequence generator, the new channel number is send to the transmitter over the uplink.

In this proposed system, the uplink can be immobile and used for only sending feedback to downlink. If the downlink sends a packet with chips with each chip contain one channel symbol, and the uplink will send chips as feedback. Since the uplink's feedback is not totally reliable, therefore assume that the transmission with one bit per chip in Equation (6):

$$C_f = N_a \log_2 N + C_{OH} + R_\tau \tag{6}$$

$N = TotalAvialableChannels = 100$

$N_a = ActiveChannels$

$C_f = ChipsOnFeedback$

$C_{OH} = FeedbackOverhead$

$R = ChipRate$

$\tau = propogationTime + LOADelay$

During the EMC test, an RSSI testing for each channel is also measured. This can indicate each channel's quality from SIR aspect. The test is done by setting the jamming frequency at channel 27 and let the transmission carry out from channel 1 to channel 100, and then the same procedure while jamming frequency is set at channel 75. The result is shown in Figure 5. The green curve in Figure 3 presents the RSSI curve based on 100 channels' transmission while jamming frequency is set at channel 75, and the red curve in Figure 3 presents the RSSI curve while jamming frequency is set at channel 27. From this data, a "bad channel" black list can be generated.

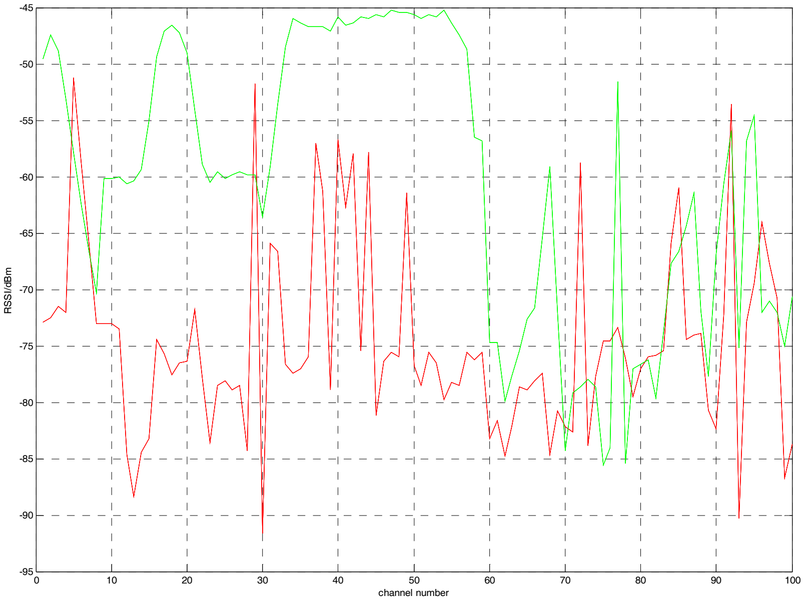


Fig. 5. Channel quality measurement: RSSI

3.3 Interference model and performance analysis

Interference always exists to any wireless system, in the improved system RSSI and the bit error rate is still important in aspect to determine whether if proposal is an improvement.
In this system, channel hops due to the SIR changes. Some of the channels are crowed while others are clear. Therefore, in the analyses, a simple two state channel model, "Gilbert-Elliot Model" (Gilbert, 1960 and Elliot, 1963) is used. "Gilbert-Elliot Model" is a two-state Markov chain (Wang et al, 1995) with states named "Good" and "Bad". According to this model, each of the channels may either be in a congested condition or clear. The classification is made due to received signal quality which means that except the interference, the two-state model also includes other transmission aspects, such as the signal strength. The bad state will be described as a channel (BSC) with a high bit error probability while the good state corresponds to a channel with a low bit error probability. Assume that all the transitions between the two states may be modelled by a Markov chain as shown in Figure 6.

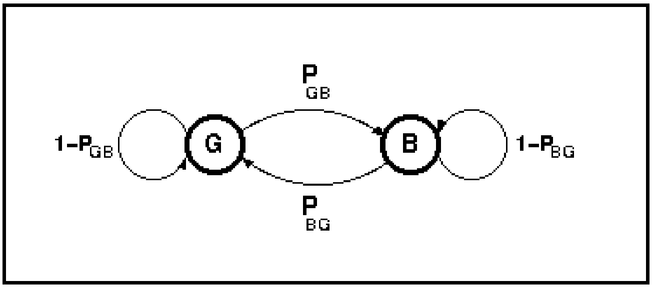


Fig. 6. A two-state Gilbert-Elliot model (Kanal et al, 1978)

The probability of leaving bad or good states will denote as *n*, *m*, and the occupancy of the frequency band by *Q*. Also assume that the average duration the transmission stay in the bad state is *t* (Andersson et al 1991 and Bröms 1991). As we know that the congestion of the channel is based on the SIR, the noise level and also the transmit power, higher the transmission power the lower the occupancy. According to Zander et al (1995) the probability of using a bad state channel is:

$$Q = P_B = \frac{n}{n+m}; \tag{7}$$

$$m = \frac{1}{t}; \tag{8}$$

$$n = \frac{1}{t} \cdot \frac{Q}{1-Q}; \tag{9}$$

Estimating the performance of the proposed system, the bit error probability is time-dependent, so assume that *M* channels being in a bad state at chip *i*, and the probability of error in chip *i* is P_i:

$$E\left[M_i \middle| M_0 = x\right] = x\left(1-(n+m)\right)^i + N_a \cdot n \frac{1-\left(1-(n+m)\right)^i}{n+m} \tag{10}$$

Assuming random hopping schemes select each of the active channels with equal probability, therefore:

$$P_{i(x)} = \frac{N_a - E\left[M_i \middle| M_0 = x\right]}{N_a} P_G + \frac{E\left[M_i \middle| M_0 = x\right]}{N_a} P_B \approx P_G + \frac{E\left[M_i \middle| M_0 = x\right]}{N_a} P_B \tag{11}$$

And assume that:

$$P_G << P_B ; E\left[M_i \middle| M_0 = x\right] << N_a$$

Combining Equation (10) and (11) yields:

$$\begin{aligned} P_{i(x)} &\approx P_G + \left(\frac{x}{N_a}\left(1-(m+n)\right)^i + n\frac{1-\left(1-(m+n)\right)^i}{m+n} \right) P_B \\ &= P_G + P_B \frac{n}{n+m} - P_B\left(\frac{n}{n+m} - \frac{x}{N_a} \right)\left(1-(n+m)\right)^i \\ &= P_G + P_B Q - P_B\left(Q - \frac{x}{N_a} \right)\left(1 - \frac{1}{(1-Q)t} \right)^i \end{aligned}$$

Removing the condition *x* and can see that the probabilities of error in chip *i*:

$$P_i = P_G + P_B \cdot Q - P_B\left(Q - \frac{E[M_0]}{N_a} \right) \cdot \left(1 - \frac{1}{(1-Q)\cdot t} \right) \tag{12}$$

If *i* approximately is linear, than error probability would be:

$$P = \frac{1}{L} \cdot \sum_{i\_1}^{L} P_i \approx P_G + P_B \cdot \left[\frac{E[M_0]}{N_a} + \frac{L}{2}\left(n - \frac{E[M_0]\cdot(n+m)}{N_a} \right) \right] \tag{13}$$

Equation (13) is dependent on the value of M_0 and if assume that *B* is the number of bad channels in active channels, and L_f is the duration of the bad states may change, then:

$$E[M_0] = QN_a - \left(QN_a - E[B]\right) \cdot \left(1 - \frac{1}{(1-Q)\cdot t} \right)^{L_f} \approx QN_a - \left(QN_a - E[B]\right)\left(1 - \alpha L_f\right) \tag{14}$$

where:

$$\alpha = \frac{1}{(1-Q)\cdot t}$$

Because bad channels are binomially distributed so that:

$$E[B] = \sum_{i=0}^{N_a} (N_a - i)\binom{N}{i}(1-Q)^i \cdot Q^{N-i} \qquad (15)$$

Imagine more complex case which there is a probability P_f that a feedback message may be lost and no changes will be made until the last packet is send. Therefore an extended feedback time is assumed as $L + 2L_f$, this gives:

$$\begin{aligned} E[M_0] &\approx 1 - P_f\left[QN_a - (QN_a - E[B])(1-\alpha L_f)\right] + P_f\left[QN_a - (QN_a - E[B])(1-\alpha(L+2L_f))\right] \\ &= QN_a - (QN_a - E[B]) \times \{1 - \alpha(L_f + P_f(L + L_f))\} \end{aligned} \qquad (16)$$

From the equations above, the frequency hopping has been implemented and simulated with Matlab. Figure 7 shows the performance comparison with and without frequency hopping. With the proposed frequency hopping solution, the BER can be significantly reduced especially in bad channel conditions with low RSSI, compared to the performance without frequency hopping. This has proved the proposed system can well perform the transmission, and the best performance will be obtained if using fixed rate block codes with ideal interleaving and a soft decoder. Also the smaller the packet size gets, the lower the probability of error rate.

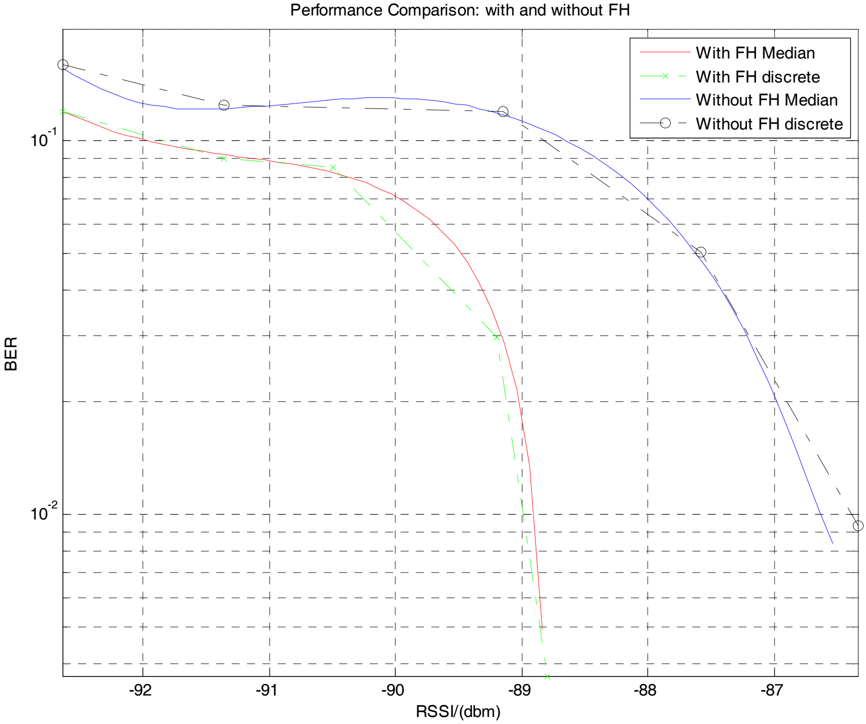


Fig. 7. Performance comparison: with and without frequency hopping

4. Conclusion

In this chapter, we have investigated different wireless technologies in order to optimize the wireless communication performance in application of RoboCup. Linx wireless communication technology is intensively tested using self developed test bed and EMC measurements, and its performance has been analyzed.

Reliable and low latency wireless communication ensures smooth and accurate control of fast dynamic process and is critical for success not only in RoboCup competition but also in many industrial automation applications where short range wireless communication technologies are used to replace existing cables. The prototype test bed with EMC measurements along with RoboCup application forms an ideal real world testing and analytical environment for various wireless communication technologies, to find out how different jamming signals affect their performances and therefore to make further improvements possible.

An adaptive frequency hopping scheme has been proposed to improve the immunity of interference of commercial short range wireless communication and therefore enhance the wireless communication performance. With the test bed to evaluate the wireless communication performance, the adaptive frequency hopping scheme can significantly reduce the bit error rate and packet loss rate. In future research, the proposed adaptive frequency hopping scheme will be adopted to other wireless links to be evaluated further.

5. References

Andersson, A. & Bergzen, H. (1991). "HF-interference measurements", *TELUB internal report*, TR914109, 1991

Anthes, J. (2007). "OOK, ASK and FSK Modulation in the Presence of an Interfering signal", *RF Monolithics*, Dallas, Texas

Bröms, M. (1991). "Some results from measurements of interference in Sweden", *IEE conference publication 339*, pp. 337-342

Elliot, E. O. (1963). "Estimates of Error Rates for Codes On Burst-Noise Channels", *Bell Syst. Tech. J*, pp. 1977-1997

Gilbert, E. N. (1960). "Capacity of a burst-noise channel", *Bell Syst. Tech. J*, pp. 1253-1265

Jochen, S. (2001). *Mobile Communication*, Addison Wesley

Kanal, L. N. & Sastry, A. R. K. (1978). "Models for channels with memory and their application to error and control", *IEEE, Proceedings,* Vol. 66, July 1978, pp. 724-744

Linx transmitter Data Guide (2007a). Linx Technologies, Inc., www.linxtechnologies.com/documents/TXM-900-HP3-xxx_Data_Guide.pdf

Linx receiver Data Guide (2007b). Linx Technologies, Inc., www.linxtechnologies.com/documents/RXM-900-HP3-xxx_Data_Guide.pdf

Liu, Y.; Mazurkiewicz, M. & Kwitek, M. (2007). "A Study towards Reliability- and Delay-Critical Wireless Communication for RoboCup Robotic Soccer Application", *Proceedings of 3rd IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China

Liu, Y. (2008). "Enhancement of Short Range Wireless Communication Performance Using Adaptive Frequency Hopping", *Proceedings of 4th IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, Dalian, China

Rappaport, T. S. (2001). *Wireless Communication: Principles and Practice, 2nd ed.*, Prentice Hall

Tse, D. & Viswanath, P. (2005). *Fundamentals of Wireless Communication*, Cambridge University Press

Wang, H. S. & Moayeri, N. (1995). "Finite State Markov Channel – A Useful Model for radio communication channels", *Proceedings of IEEE Transactions on Vehicular Technology*, pp. 163-171

Zander, J. & Malmgren, G. (1995). "Adaptive frequency hopping in HF communications", *IEE Proceedings Communications*, Vol. 142, Apr 1995, pp. 99-105

16

Capacity Dimensioning for Wireless Communications System

Xinsheng Zhao and Hao Liang
Southeast University
P.R.China

1. Introduction

With the development of mobile communications system, lots of novel techniques in wireless link had been invented by research institutes and adopted by different kinds of mobile communications system standards[1-2]. Meanwhile, there are still many new technologies emerging for standardization of next generation mobile communications system [3].

Capacity dimensioning methods have been continuously exploited for the system capacity performance evaluation and wireless network planning in development process of any generation of mobile communications system. All of the capacity dimensioning approaches can be divided into two main groups, i.e. the analytical methods and the simulation methods. Analytical methods give the close-form expression of system capacity with system parameters, transmission techniques and wireless network architecture, user distribution and wireless services by establishing mathematical models and get the capacity solutions by computer calculations. The simulation methods simulate the system process by precisely establishing simulation platform and analyze the system capacity through Monte Carlo simulation procedure. Performance comparisons of the two kinds of methods are shown in table X.1.1.

	simulation method	analytical method
Precision	Precise under any condition	Accuracy of the result depends on the mathematical methods
Time cost	Much time	Less time
Complexity	Cost a lot of computations. Powerful computer is necessary	Cost less computations
Convenience	It is not easy to establish simulation platform	If using accurate mathematical models, it will be convenient to solve practical problems

Table X.1.1 Comparisons of simulation method and analytical method

2 Analytical Methods of Capacity Dimensioning

Analytical capacity dimensioning methods include approaches of limited system resources-based capacity dimensioning, system interference model-based capacity dimensioning and queue-based capacity dimensioning.

2.1 Limited System Resources-Based Analytical Approaches

Ignoring the factors of self-interference and supposing all of the radio resources be orthogonal, limited system resources-based analytical method considers radio resources like spectrum bandwidth in FDMA system, time slots in TDMA system, code channels in CDMA system, subcarriers in OFDMA system, antennas in MIMO system or mixture of the radio resources in multi-dimensions radio resource mobile communications system. By taking the total number of radio resources occupied by system as system states and constructing Markov chain, the system capacity can be calculated based on the stationary state distribution of Markov chain. If system resources are modeled properly, this approach is applicable for any kind of mobile communications system. But, it is not accurate for CDMA system due to its self-interference.

[4] extended the Erlang B/C law to multi-service as multi-service Erlang B law and hybrid multi-service Erlang B/C law for the performance evaluation of UMTS with multi-services. By using Markov chain-based system resources modeling method for estimation of Erlang capacity of CS domain voice service and recursive algorithm proposed by Kauffman and Roberts for calculating the stationary state distribution of the Markov chain, it obtained a closed-form expression to calculate the blocking rate for each service in a multi-service context. And also, it got a UMTS radio dimensioning flow chart for offering more accurate dimensioning.

[5] presented a multi-service dimensioning procedure for evaluation of the number of CDMA cells needed in a given area based on the analysis of an original CDMA multi-service uplink capacity estimation model and the multi-service Erlang-B formula.

[6] made a contribution to the generic problem of having simple and accurate models to estimate radio cells with data traffic on a GPRS or EDGE network for avoiding the derivation of any multi-dimensional Markovian (or semi-Markovian) model. On the basis of ON/OFF modeled service, the number of time slots occupied by system could be modeled as first-order Markov chain. Capacity of GPRS/EDGE was figured out by calculating stationary state distribution of the Markov chain.

2.2 System Interference Model-Based Analytical Approaches

System interference model-based analytical method calculates the system capacity based on uplink/downlink interference model in the system and Shannon information capacity theorem. Markov chain can be used to model the number of users in the system. This method is applicable for any kind of mobile communications system with consideration of system interference and especially for CDMA system as the number of users in CDMA system is restricted by the interference of uplink and the transmit power of downlink.

[7] evaluated the cell capacity of the HS-DSCH channel in HSDPA in the presence of circuit switched (CS) services based on description of the SIR expression of an HSDPA user in the presence of CS users and power consumption and code tree resources consumption of CS services. Hybrid Automatic Repeat Request (HARQ), Fast Cell Selection, Adaptive

Modulation and Coding (AMC) were revisited for the capacity calculation of the HS-DSCH channel. Monte Carlo simulation was performed for the verification of the correction of the analytical expressions.

[8] analytically estimated the cell throughput of HSDPA based on the analysis of distribution of SNR and influence of AMC, HARQ and different scheduling algorithms, analyzed the TCP model and effect of TCP on wireless network on condition of different scheduling algorithm.

[9] presented an analytical model for the evaluation of UMTS HSDPA capacity based on the analytical estimation of SINR for quality of service guarantee, used the model for the capacity performance comparison of different MIMO transmission scheme (space time coding and BLAST) and proposed the use of adaptive MIMO in the HSDPA system with the aim of increasing cell capacity.

[10] estimated the MIMO/HSDPA cell capacity for UMTS by a proposed semi analytical model and analytical framework. The model included the effect of fast fading, mobility and coverage by establishing interference model with fast fading, mobility model and coverage model. The performances of MIMO with different transmission scheme were also investigated.

[11] compared the cell capacity of HSDPA with adaptive modulation and coding, Hybrid ARQ, fast scheduling, fast cell selection and different MIMO transmission schemes (space time coding and Blast) by introducing an analytical method based on the interference model establishment.

[12] compared the cell capacity of a proposed scheduling algorithm to the cell capacity of other scheduling algorithms in HSDPA by introducing an analytical interference model with the assumption of a dense multipath frequency selective channel with uncorrelated signal envelope following a Rayleigh distribution and Wide-sense Stationary Uncorrelated Scattering (WSSUS) radio channel.

[13] evaluated the reverse link capacity of a CDMA cellular voice system by analysis of system interference and blocking probability.

[14] defined downlink multiservice loading as performance parameter of system capacity, set up a comprehensive downlink dimensioning framework for capacity and coverage analysis of downlink in WCDMA system, analyzed the influence of offered bandwidth, non-orthogonal factor, soft handover and diversity gain, output power, shadowing to system capacity.

[15][16][17] derived the close-form expression of outage probability and the forward-link Erlang capacity for CDMA system forward link with mixed multi-rate sources based on the establishment of data traffic model, system interference model, introduction of the forward-link power factor and impact of soft handoff.

[18] designed a recursive procedure for calculation of system throughput and blocking probability in the presence of power control error in CDMA system based on the establishment of a three-dimensional Markov chain for the integrated CS/PS system model, intercell and intracell interference model in different conditions of perfect power control and imperfect power control.

[19] extended the notion of capacity for the consideration of wireless service quality of service requirements influence, evaluated the uplink capacity for single cell scenario and multi-cell scenario of a multi-service WCDMA system based on the analysis of system interference model, real time service model and best effort model.

[20] presented a capacity dimensioning flow chart of for WCDMA with HSDPA and HSUPA based on the analysis of link level capacity and system level capacity. Supposing that HSDPA system work in a TDM mode and each user has sufficient services to transmit, only one user be scheduled during each time slot in system level capacity analysis, interference model was established directly according to the interference model, PDF of user's receiving SNR used in capacity dimensioning was derived by using different scheduling algorithms. Then, the system performances including the sector throughput and user satisfied percentage of WCDMA with HSDPA and capacity with HSUPA were calculated.

2.3 Queue-Based Analytical Approaches

Queue-based analytical method combines Markov chain modeling of the wireless transmission channel states in air interface and Markov chain modeling of the wireless service transmission buffer states into a multi-dimensional Markov chain. The system capacity can be calculated by solving the stationary state distribution of the multi-dimensional Markov chain. Theoretically, this method is qualified to any system if the mathematical model of Markov chain can be properly constructed for the mobile communications system and wireless service transmission buffers.

[21] extended RIO (random early drop routers with in/out bit) algorithm to 5 class wireless services by continuous-time 2-D Markov chain including wireless link state and wireless services states, derived the state transition probability for the calculation of the stationary state distribution of Markov chain by matrix analysis, obtained system performance like system throughput and queuing delay time.

[22] compared the system performance like relative system throughput and average queued packets number of a proposed dynamic rate allocation technique with traditional static rate allocation technique in HSDPA system by derivation of a discrete time Markov model of the system and establishment of three kinds of traffic models, verified the analytical method by simulation process.

[23] [24] established system model for wireless system with adaptive modulation and coding (AMC), analyzed system performance by Markov chain diagram describing the transition of the pair containing both service queues and channel states, and given a cross layer design example.

[25] developed two discrete time Markov chain for the quality of service (QoS) evaluation of TCP-Friendly Rate Control protocol (TFRC) in hybrid wireless/wired networks, one is for analysis of link utilization and packet loss rate, and the other is for delay performance analysis.

[26] proposed a four-dimensional discrete time Markov chain (DTMC) model to describe the system behavior of HSDPA system with different PS wireless services, investigated the effect of AMC, HARQ and the packet multimedia traffics by a discrete Markov modulated Poisson process (D-MMPP), derived closed-form expressions for the system performance metrics such as throughput, delay and packet loss rate based on the stationary state distribution of the DTMC, provided analytical and simulation results for the validation of the analytical method.

3 Simulation Methods of Capacity Dimensioning

Simulation methods of capacity dimensioning can provide more accurate results and involve the approaches of event driven-based simulation platform and time driven-based simulation platform. Both of the simulation methods can be applied to any wireless communications system with time consuming work.

3.1 Event Driven-Based Approaches

Event driven-based simulation method describes the system's behavior by state machine mechanism in simulation environment like OPNET or C/C++. The communications procedures in the wireless communications system are simulated by the event drive transitions among the established states. The average throughput, packet loss rate, custom satisfaction and service transmission delay can be calculated in the simulation process.

[27] carried out extensive system level simulation for WCDMA system with various real time wireless service and mixed wireless services of real time service and non-real time service by event-based simulation implemented in OPNET platform, studied the single service capacity, average session throughput of mixed wireless services, custom satisfaction and session outage.

3.2 Time Driven-Based Approaches

Time driven-based simulation method simulates system behavior based on time-shift. According to the continuous of different simulation scenarios, time driven-based simulation methods can be divided into three groups, i.e. static, dynamic and hybrid simulation-based capacity dimensioning.

In static simulation method, system behavior is modeled as discrete actions by a series of irrelevant snapshot procedures.

[28] described statistic relations among the system and environment parameters (propagation, terminal distribution within the cells, base station distribution) by the linear equations for wireless network modeling, got the characteristics results of transmitting/receiving power and system capacity (number of users in a cell) for UMTS WCDMA system by simulation method.

In dynamic simulation method, the communications scenarios are relative with each other. System parameters are treated as the functions of time which can be used to describe the communications behavior.

[29] described a detailed simulation methodology consideration to HSDPA system based on the analysis of key techniques in HSDPA, gave the simulation results of transmitted cell power distribution, cell throughput, per HSDPA-user throughput by dynamic simulation in a standard three sector network topology, and discussed environment sensitivity of simulation platform.

[30] compared the achievable throughput performance of three different scheduling algorithms, analyzed the influence of the number of access users and traffic on throughput by establishing a dynamic simulation platform for HSDPA system.

[31] presented physical layer performance and system performance of HSDPA by dynamic simulation method.

[32] established a comprehensive simulation tools for evaluation of URTRA FDD system with mixed services of circuit switching domain and packet domain by using pipeline mechanism for modeling of wireless channel to dynamic simulation procedure.

Hybrid simulation method is a mixture of static and dynamic simulation method. This method usually uses static simulation method as main simulation process and embeds dynamic simulation method into part of simulation platform. It can get more detail simulation results than static simulation method and save time than dynamic simulation method.

[33] presented a novel simulation method by using short-term dynamic (STD) simulation concept, divided simulation process into two stages i.e. snapshot stage for static simulation of system and establishment of HSDPA components for dynamic simulation in order to make a fast UMTS HSDPA planning simulation.

[34] evaluated various performance indicators (capacity, coverage, call blocking and dropout probabilities, packets delay and performance of burst admission) of forward link in CDMA cellular systems with mixed services of voice and high speed packet-switched service by using static simulation for static system capacity and cell coverage and dynamic simulations for the influence of shadow fading in an urban environment, mobility and soft handoff events.

4 Conclusions and Discussions

Capacity dimensioning methods can be divided into analytical approaches and simulation approaches for wireless network in mobile communications system. Previous contributions showed that analytical methods establish mobile communications system model by using mathematical expressions of limited radio resources, system interference and transmission queues for fast capacity dimensioning and simulation methods set up simulation platform by using static, dynamic and hybrid ways to simulate the action of mobile communications system for accurate capacity dimensioning. Both methods can be used for correctness verification of each other.

With the development of mobile communications system and the appearance of novel technologies in physical layer (like MIMO, OFDM and etc.), wireless network architecture (such as flat wireless network, mesh wireless network, relay wireless network, cooperative wireless network) and wireless services, capacity dimensioning is still an open issue for further investigations.

5. Reference

[1] Subramanya, S.R.; Yi, B.K., ‘Mobile communications - an overview’, Potentials, IEEE Volume 24, Issue 5, Dec. 2005 Page(s): 36–40.

[2] Erik Dahlman, Stefan Parkvall, Johan Skold and Per Beming, 3G Evolution: HSPA and LTE for Mobile Broadband, Academic Press, October 3, 2008.

[3] http://www.itu.int/ITU-R

[4] Adiego D, Cordier C. ‘Multi-service radio dimensioning for UMTS circuit-switched services’, IEEE 54th VTC, 2001, pp. 2745–2749.

[5] Brandin H. 'A multiservice dimensioning procedure for 3G CDMA', 3G Mobile Communication Technologies, 2000. First International Conference on, 27-29 March 2000 Page(s): 406-410
[6] Georges N, Bruno B, Pierre E., 'An analytical model for the dimensioning of a GPRS/EDGE network with a capacity constraint on a group of cells', International Conference on Mobile Computing and Networking, 2005.
[7] Assaad M, Zeghlache D., 'Effect of circuit switched services on the capacity of HSDPA', IEEE Transactions on Wireless Communications, 2006, 5: 1044-1054.
[8] Assaad M, Zeghlache D., 'Cross-Layer design in HSDPA system to reduce the TCP effect', IEEE Journal on Selected Areas in Communications, 2006, 24: 614-625.
[9] Assaad M, Zeghlache D., 'Comparison between MIMO techniques in UMTS-HSDPA system', IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications, 2004, pp 874-878, Sep 2004.
[10] Assaad M, Zeghlache D. 'MIMO/HSDPA with fast fading and mobility: capacity and coverage study', 15th IEEE International Symposium on Personal Indoor and Mobile Radio Communications, 2004, 3: 2181–2186.
[11] Assaad M, Zeghlache D., 'On the capacity of HSDPA' IEEE GlOBECOM, 2003, 1: 60-64.
[12] Assaad M, Zeghlache D., 'Scheduling study in HSDPA system', IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005, 3:1890-1894.
[13] Viterbi A M, Viterbi A J., 'Erlang Capacity of a Power Controlled CDMA System', IEEE Journal on Selected Areas in Communications, 1993, 11(6): 892-900.
[14] Cordier C, Ortega S., 'On WCDMA downlink multiservice coverage and capacity', IEEE 54th Vehicular Technology Conference, 2001, 4:2754 - 2758.
[15] Wan Choi, Jin Young Kim, 'Forward-link capacity of a DS/CDMA system with mixed multirate sources', IEEE Transactions on Vehicular Technology, 2001, 50(3): 737–749.
[16] Wan Choi, Jin Young Kim, 'Forward link capacity of 3G wideband CDMA system with mixed traffic sources', Vehicular Technology Conference IEEE 53rd VTC, 2001, 4:2620–2624.
[17] Wan Choi, Byung Shik Kang, Jun Cheol Lee, et al. 'Forward link Erlang capacity of 3G CDMA system', 3G International Conference on Mobile Communication Technologies, 2000, pp. 213–217.
[18] Lin Wang, Aghvami A H, Chambers W, et al. 'Performance Analysis of an Integrated CS/PS Services CDMA System', IEEE Transactions on Vehicular Technology, 2005, 54:1488–1499.
[19] Nidhi Hegde, Eitan Altman, 'Capacity of Multiservice WCDMA Networks with Variable GoS', Wireless Networks, 2006, 12(2): 241-253.
[20] Zhao X, Qi J, Liang H, 'An Analytical Method for Capacity Dimensioning of WCDMA with High Speed Wireless Link', Proc. IEEE Wireless Communications and Network Conference, pp. 4172-4176, Mar. 2007.
[21] Gyasi-Agyei A., 'Performance analysis of differentiated services over wireless links', 5th IEEE International Conference on High Speed Networks and Multimedia Communications, 2002, pp. 86–90.
[22] Caponi L, Chiti F, Fantacci, 'A dynamic rate allocation technique for wireless communication systems', IEEE International Conference on Communications, 2004, 7:4263–4267.

[23] Qingwen Liu, Shengli Zhou, Georgios B Giannakis, 'Queuing with adaptive modulation and coding over wireless links: Cross-layer analysis and design', IEEE transactions on wireless communications, 2005, 4:1142-1153.

[24] Xin Wang, Qingwen Liu, Georgios B Giannakis, 'Analyzing and Optimizing Adaptive Modulation Coding Jointly With ARQ for QoS-Guaranteed Traffic', IEEE Transactions on Vehicular Technology, 2007, 56: 710-720.

[25] Shen H, Cai L, Shen X., 'Performance analysis of TFRC over wireless link with truncated link-level ARQ', IEEE Transactions on Wireless Communications, 2006, 5: 1479-1487.

[26] Xinsheng Zhao; Hao Liang; Zhenjie Gu; "A Markov Chain-Based Capacity Dimensioning Method for Wireless Communications System with AMC, HARQ and Packet Multimedia Traffic Source" Communications, 2008. ICC' 08. IEEE International Conference on 19-23 May 2008 Page(s):429–434

[27] Malik S A, Zeghlache D., 'Downlink capacity and performance issues in mixed services UMTS WCDMA networks', IEEE 55th Vehicular Technology Conference, 2002, 4:1824-1828.

[28] Bem D J, Nawrocki M J, Wieckowski T W, et al. 'Modeling methods for WCDMA network planning', IEEE 53rd Vehicular Technology Conference, 2001, 2: 962–966.

[29] Pedersen K I, Lootsma T F, Stottrup M, et al. 'Network performance of mixed traffic on high speed downlink packet access and dedicated channels in WCDMA', IEEE 60th Vehicular Technology Conference, 2004, 6: 4496–4500.

[30] Ofuji Y, Morimoto A, Abeta S, et al., 'Comparison of packet scheduling algorithms focusing on user throughput in high speed downlink packet access', IEEE 13th International Symposium on Personal, Indoor and Mobile Radio Communications, 2002, 3:1462–1466.

[31] Love R, Ghosh A, Weimin Xiao, et al., 'Performance of 3GPP high speed downlink packet access (HSDPA)', IEEE 60th Vehicular Technology Conference, 2004, 5: 3359–3363.

[32] Adiego and Cordier C., 'Multi-service Radio Dimensioning for UMTS Packet-Switched Services', IEEE Proceeding of PIMRC, 2002, 5: 2409–2413.

[33] Turke U, Koonert M, Schelb R, et al., 'HSDPA performance analysis in UMTS radio network planning simulations', IEEE 59th Vehicular Technology Conference, 2004, 5:2555–2559.

[34] Abarca-Reyna J P, Lara-Rodriguez D., 'Performance evaluation of the downlink CDMA cellular system supporting integrated voice/data traffic', IEEE 61st Vehicular Technology Conference, 2005, 5:2741–2745.