

Telecomunicações

Revista do Instituto Nacional de Telecomunicações

Revista Científica Periódica

Volume 09

Número 02

Julho de 2007

ISSN 1516-2338

Editorial

Tutorial

Channel Capacity Estimation for a Multi-Carrier Code-Division Multiple Access System

Dayan Adionel Guimarães - Inatel 1

Engenharia Aplicada

Análise de Desempenho da Esfera Integradora para Calibração de Medidores de Potência Óptica

Hilton Henrique Bertan and Yuzo Iano - UNICAMP 7

Propriedades Direcionais de Irradiação de Antenas Rádio-Base

Marco Antonio Brasil Terada - UnB 13

Projeto, Simulação e Implementação em Hardware de um Sistema de Sincronismo para o MI-SBTVD

Marcelo A. C. Fernandes - UNICAMP

Cristiano Panazio - USP

Sandro A. Fasolo - Inatel 19

Codificação LDPC em Sistemas de Televisão Digital

Tarciano F. Pegoraro, Fábio A. L. Gomes, Fábio Lumertz, Renato R. Lopes, Fabrício A. Oliveira and Roberto Gallo - UNICAMP

Marcelo C. Paiva and José S. G. Panaro - Inatel 25

Revista Científica/Periódica publicada pelo Inatel - Instituto Nacional de Telecomunicações

Diretor: Wander Wilson Chaves

Vice-diretor: Carlos Roberto dos Santos

Editor

Carlos Alberto Ynoguti - Inatel

Conselho Editorial

Antônio Marcos Alberti - Inatel

Dayan Adionel Guimarães - Inatel

José Marcos Câmara Brito - Inatel

Corpo de Revisores

Anderson S. L. Gomes - UFPE

Anilton Salles Garcia - UFES

Antônio Alves Ferreira Júnior - Inatel

Antônio Marcos Alberti - Inatel

Arismar Cerqueira Sodré Júnior - UNICAMP

César Kyn d'Ávila - CEDET

Diego Grosz - BELL Laboratórios USA

Dilson Frota de Moraes - Leucotron Equipamentos Ltda.

Edson Moschim - UNICAMP

Eduardo César Grizendi - Rail Com

Francisco José Fraga da Silva - Inatel

Franco Callegati - DEIS

Geraldo Gil R. Gomes - Inatel

Guilherme Augusto Barucke Marcondes - Inatel

Hani Yehia - UFMG

Helio Waldman - UNICAMP

Ivanil S. Bonatti - UNICAMP

Jaime Portugheis - UNICAMP

João César Moura Mota - UFC

José Antônio Justino Ribeiro - Inatel

José de Souza Lima - LINEAR

José Edimar Barbosa Oliveira - ITA

José Marcos da Silva Nogueira - UFMG

Júlio César Tibúrcio - Inatel

Luciano Leonel Mendes - Inatel

Luiz Geraldo Pedroso Meloni - UNICAMP

Márcio Lourival Xavier dos Santos - UNITAU

Marcos R. Salvador - CTIT

Maria Regina Campos Caputo - Inatel

Marlene Sabino Pontes - CETUC

Martin Zieher - FHTE (Alemanha)

Maurício Silveira - Inatel

Nelson Soares Wisnik - N. Wisnik Consultoria

Omar Carvalho Branquinho - CPqD

Paulo Gomide Cohn - Embassy Systems

Pierre Kaufmann - Mackenzie/INPE/UNICAMP

Rainer Doster - FHTE (Alemanha)

Renato Baldini Filho - UNICAMP

Sandro Adriano Fasolo - Inatel

Silvio Ernesto Barbin - EPUSP

Shusaburo Motoyama - UNICAMP

Wilton Ney do Amaral Pereira - Inatel

Yuzo Iano - UNICAMP

Expediente

Assessoria de Comunicação & Marketing - ASCOM
e-mail: ascom@inatel.br

Diagramação
Centro de Informações Científicas e Tecnológicas
Setor de Editoração Eletrônica

Tiragem: 3.500 exemplares
Distribuição Gratuita

Instituto Nacional de Telecomunicações
Av. João de Camargo, 510
Caixa Postal: 05
Santa Rita do Sapucaí - MG - BRASIL
CEP 37540-000
Tel: (35) 3471.9200 Fax: (35) 3471.9314
e-mail: inatel@inatel.br
<http://www.inatel.br>

EDITORIAL

A área das telecomunicações é possivelmente aquela em que as mudanças tecnológicas ocorrem com a maior velocidade. Novas formas de uso destas tecnologias também vêm se verificando, com a explosão do comércio eletrônico e das comunicações sem fio, só para citar as mais evidentes.

Neste cenário, a Revista Telecomunicações contribui com a divulgação de trabalhos de pesquisa científica e aplicada nesta área, permitindo que profissionais e pesquisadores mantenham-se atualizados com as tecnologias atuais e aquelas que ainda estão por vir.

Com este número termina a série especial sobre o Sistema Brasileiro de Televisão Digital (SBTVD), com artigos sobre os problemas de sincronismo e codificação LDPC. Temos ainda um excelente tutorial sobre estimação de canais para um sistema de múltiplo acesso multiportadora, e mais dois artigos de engenharia aplicada, nas áreas de medidores de potência óptica e irradiação de antenas rádio base.

Carlos Alberto Ynoguti
Editor

Channel Capacity Estimation for a Multi-Carrier Code-Division Multiple Access System[♦]

Dayan Adionel Guimarães

Abstract—This paper presents some results concerning the investigation of the Shannon channel capacity for a Multi-Carrier, Direct-Sequence, Code-Division Multiple Access (MC-DS-CDMA) system operating in a multi-path, frequency-selective, Rayleigh-fading channel. The channel capacity estimation uses a Monte Carlo technique. The results reveal that, for the considered system, it is preferable to explore the maximum order of diversity instead of the maximum parallelism of the data stream. The results also indicate that one possible solution for a channel coding scheme to be used along with the system is a concatenation of a powerful low-rate outer code with the inherent inner repetition code of the system.

Index Terms—Channel Capacity, Multi-Carrier CDMA.

Resumo—Neste artigo são apresentados alguns resultados referentes à capacidade de canal de Shannon de um sistema CDMA multiportadora operando em um canal Rayleigh seletivo em frequência. A estimativa da capacidade de canal é realizada por meio do método de Monte Carlo. Os resultados apresentados revelam que, para o sistema sob análise, é preferível explorar a máxima ordem de diversidade em vez de explorar o máximo paralelismo do feixe de dados. Tais resultados também indicam que uma possível solução para codificação de canal para o sistema em questão corresponde à concatenação de um código externo potente e de taxa baixa com o código de repetição inerente à estrutura do sistema MC-DS-CDMA.

Palavras chave—Capacidade de canal, sistemas CDMA multiportadora.

I. INTRODUCTION

It is a fairly collective opinion that multi-carrier modulation, especially when combined with the code-division multiple access (CDMA) technique, is a potential candidate for the next generation (4G) of wireless communication systems. But the scarce of spectrum and the multitude of foreseen services will demand new supporting technologies for these systems that will make their real utilization feasible. Among these technologies, efficient channel coding schemes can be highlighted. Nevertheless, after the conception of the system, it is advantageous if one knows its limits of

Manuscript received on November 9, 2006.

D. A. Guimarães (dayan@inatel.br) is with INATEL - Instituto Nacional de Telecomunicações. Av. João de Camargo, 510 - Santa Rita do Sapucaí - MG - Brazil - 37540-000.

[♦]Part of this work was presented during the 4th International Workshop on Multi-Carrier Spread-Spectrum, MC-SS 2003, Oberpfaffenhofen, Germany, September 17-19, 2003, and published in its proceedings: *Multi-Carrier Spread-Spectrum*, FAZEL, Khaled and S. KAISER (editors), pp. 115-122, ISBN: 1-4020-1837-1, Kluwer Academic Publishers, The Netherlands, 2004, under the title: *Turbo Product Codes for an Orthogonal Multicarrier DS-CDMA System*.

transmission of information in order to support the code construction and to evaluate its performance.

After the remarkable contribution from Shannon [1], several authors have extended his studies in order to estimate the channel capacity for fading channels. Among them we can mention [2] [3] [4] [5] [6] [7], for unlimited input alphabet, and [8] [9] for BPSK signaling. It also can be found some studies related to channel capacity calculations for multi-carrier fading channels [10].

In this paper, the channel capacity for the Multi-Carrier Direct-Sequence Code Division Multiple Access (MC-DS-CDMA) system of [11], is estimated using a Monte Carlo technique, and the results for some choice of the system parameters are presented. The Monte Carlo technique is applied here to make possible an alternative calculation of capacity formulae.

The results lead to important conclusions about the choice of the parameters of the considered system, unveiling that it is preferable to explore the maximum order of diversity instead of exploring the maximum parallelism of the data stream, and giving an insight about some characteristics of an error correcting code scheme to be designed for this system.

II. SYSTEM DESCRIPTION

Fig. 1 and Fig. 2 [11] show, respectively, the block diagrams for the transmitter and receiver of the MC-DS-CDMA system under consideration.

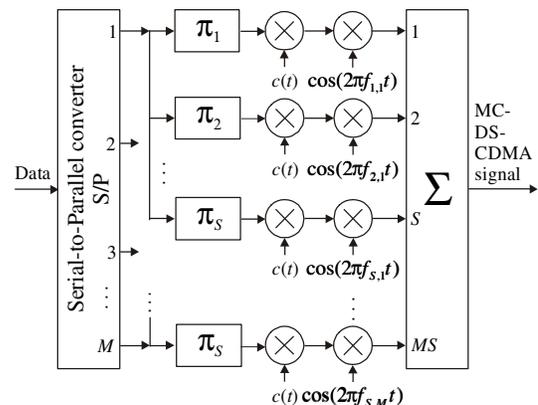


Fig. 1. MC-DS-CDMA Transmitter.

In the system suggested in [11], transmitted data bits are serial to parallel converted to M parallel branches. On each branch, each bit is repeated S times and the replicas feed

different block interleavers (π). Then, these identical bits are direct-sequence spread spectrum BPSK modulated and transmitted using orthogonal carries. Hence, there are a total of MS carriers and time-frequency diversity is obtained. At the receiver, the matched filters outputs of the S identical-bit carriers are combined prior to decoding. The system of [11] can be viewed as a combination or generalization of the *copy-type* and *S/P-type* configurations described in [12]. The main attributes of this system are: 1) the possibility of overcoming the performance of the conventional single-carrier CDMA system, and 2) the reduction of complexity through the use of one matched filter per carrier, instead of a RAKE receiver, situation that is achieved if the number of carriers satisfies [11]

$$MS \geq 2L_1 - 2 \quad (1)$$

where L_1 is the number of resolvable propagation paths for a single-carrier CDMA system with the same total bandwidth as for the MC-DS-CDMA.

In [11] it is allowed a 50% overlap of adjacent modulated carriers, still maintaining orthogonality. Fig. 3 illustrates the main lobe spectrums of the transmitted signal for $M = 4$ and $S = 3$. In this figure, the modulated carriers identified with the same pattern transport the replicas of bits and can be viewed as diversity carriers.

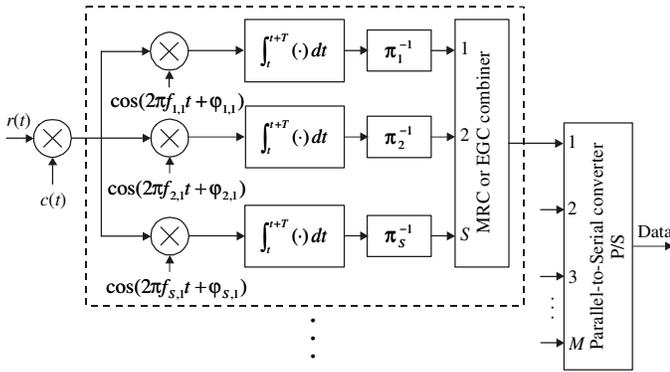


Fig. 2. MC-DS-CDMA receiver.

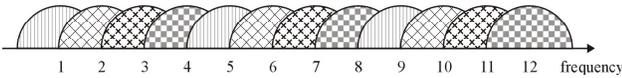


Fig. 3. MC-DS-CDMA spectrum for $M = 4$, $S = 3$.

III. CHANNEL CAPACITY ESTIMATION

In this section, the method for estimating the channel capacity for the MC-DS-CDMA system is described. Initially, a short review of some previous published results is presented, followed by a description of the Monte Carlo technique used for solving the capacity formulas.

A. Background

According to [5], whose credit was given to Wolfowitz [13], let $g[i]$ be a stationary and ergodic stochastic process associated to the channel state information, assuming values on a finite set U of discrete memoryless channels. Let $C(u)$ denote the capacity of a channel $u \in U$, and $p(u)$ denote the

probability, or fraction of time, that the channel is in state u . The capacity for this time-varying channel is given by [5]

$$C^* = \sum_{u \in U} C(u)p(u) \quad (2)$$

Now, consider an AWGN-fading channel and assume that the channel gains $g[i]$ can be considered as a stationary and ergodic stochastic process, known by the transmitter and receiver. A time-invariant AWGN channel with average received signal-to-noise ratio (SNR) γ and unconstrained input alphabet has capacity, in bits per second, given by

$$C_\gamma = B \log_2(1 + \gamma) \quad (3)$$

Let $p(\gamma) = p(\gamma[i] = \gamma)$ be the probability density function (p.d.f.) of γ . Then, based on (2) and (3), the capacity of the time-varying channel is [5]

$$C = \int_\gamma B \log_2(1 + \gamma) p(\gamma) d\gamma \quad (4)$$

Still following [5], if the *compatibility constraint* is satisfied, then the capacity of the channel with *side information at the receiver only* is also given by the average capacity formula (4). The compatibility constraint is satisfied if the channel sequence $g[i]$ is independent and identically distributed (i.i.d.), and if the input distribution that maximizes the mutual information is the same, regardless the channel state. The result (4) will be considered later on in this paper.

Now, consider a BPSK modulation for which the symbols, with energy E_s , are sent through an AWGN channel with equal probability. In this case, it can be shown that the channel capacity, in bits per channel use, can be calculated by [14]

$$C_{BPSK} = \int_{-\infty}^{\infty} p(y | \sqrt{E_s}) \log_2 \frac{p(y | \sqrt{E_s})}{p(y)} dy \quad (5)$$

where

$$p(y | \sqrt{E_s}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y - \sqrt{E_s})^2}{2\sigma^2} \right] \quad (6)$$

$$p(y) = \frac{1}{\sqrt{8\pi\sigma^2}} \left\{ \exp \left[-\frac{(y - \sqrt{E_s})^2}{2\sigma^2} \right] + \exp \left[-\frac{(y + \sqrt{E_s})^2}{2\sigma^2} \right] \right\} \quad (7)$$

and σ^2 is the noise power of the samples at the output of the optimum receiver (matched filter or correlator).

If the receiver knows the channel state information, the capacity of a memoryless Rayleigh fading channel with BPSK signaling can be calculated as [8]

$$C_{BPSK} = - \int_g \int_y p(g)p(y | \sqrt{E_s}, g) \log_2 [f(g, y)] dy dg \quad (8)$$

where $p(g)$ is the p.d.f. of the fading amplitude,

$$f(g, y) = \frac{1}{2} \left\{ 1 + \exp[-(4/N_0)yg\sqrt{E_s}] \right\} \quad (9)$$

and $N_0/2 = \sigma^2$ is the bilateral noise power spectral density at the receiver input.

The results (5) and (8) will also be considered latter on in this paper.

Finally, it is known [15, p. 344] that the capacity of a set of U parallel and independent AWGN channels is equal to the sum of the individual capacities. So, if the multi-carrier, frequency-selective, fading channel is approximately considered as a set of independent parallel and conditioned AWGN channels (conditioned on the fading amplitudes), the total capacity can be estimated as the sum of U capacities calculated via (4), if the input alphabet is unconstrained, and via (8) for BPSK signaling.

The inherent problem with (4) and (8), even if numerical integration is used, is that the probability density functions operated into each of these expressions are not always easily obtained, specially in the case of (8). For example, if it is of interest to consider the fading statistics at the output of an EGC combiner with S inputs, the task of obtaining the p.d.f. of the sum of Rayleigh envelopes becomes more and more complex as the value of S increases. In this context, we suggest a Monte Carlo method for channel capacity estimation, without the need for operating such density functions. The method is described in the next subsection.

B. Applying the Monte Carlo Method

From this point on, it is assumed that the receiver has perfect knowledge of the channel state information and that the transmit power is constant (no power adaptation scheme). It is further assumed that the compatibility constraint described in Section II is satisfied.

Let $g[i]$ represent the channel state information at the discrete-time instant i , and assume that it is possible to generate by computer, directly or indirectly, a sufficient large number X of values for g , based on a specific and known probability distribution. Then, it is claimed here that the capacity expressed by (4) can be estimated using the discrete-time average

$$C = \frac{1}{X} \sum_{i=1}^X B \log_2 (1 + \gamma g^2[i]) \quad (10)$$

By a sufficient large value of X it is meant a value enough for convergence in (10).

Then, for BPSK signaling in a fading channel, the capacity expressed by (5) can be adapted to this Monte Carlo technique, leading to

$$C_{BPSK} = \frac{1}{X} \sum_{i=1}^X \int_{\Lambda[i]}^{\Lambda[i]} p(y | \psi[i]) \log_2 \frac{p(y | \psi[i])}{p(y)} dy \quad (11)$$

where $\psi[i] = g[i]\sqrt{E_s}$, $\Lambda[i] = \psi[i] + 6\sigma$,

$$p(y | \psi[i]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - \psi[i])^2}{2\sigma^2}\right] \text{ and } \quad (12)$$

$$p(y) = \frac{\exp\left[-\frac{(y - \psi[i])^2}{2\sigma^2}\right] + \exp\left[-\frac{(y + \psi[i])^2}{2\sigma^2}\right]}{\sqrt{8\pi\sigma^2}} \quad (13)$$

Obviously, for an AWGN channel $g[i] = 1$. The results obtained through (10) and (11) demonstrate perfect agreement [14] with those obtained through their counterparts (4) and (8), respectively, showing the applicability of the method for both unconstrained and constrained input alphabets. Fig. 4 shows some results concerning this comparison.

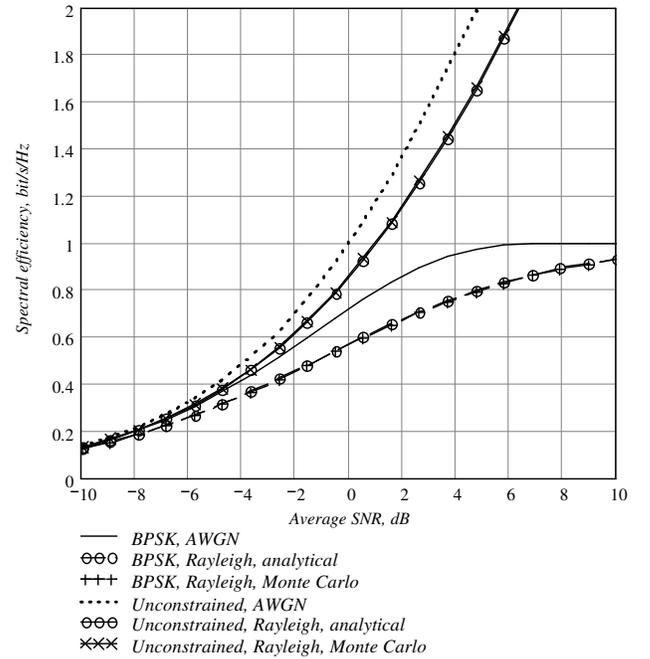


Fig. 4. Comparison between *analytical* and *Monte Carlo* methods used for channel capacity estimation.

The reverse link of the MC-DS-CDMA system for a user of reference can be interpreted, in one hand, as a set of M parallel channels with BPSK signaling. These channels are defined from each of the M S/P converter outputs at the transmitter to each of the M combiner's outputs at the receiver. On the other hand, this link can be interpreted as a set of MS parallel channels defined from each of the MS modulator inputs at the transmitter to each of the MS matched filters outputs at the receiver. Then, according to previous statements (see Section II), the channel capacity for the MC-DS-CDMA system can be estimated as the sum of M or MS individual capacities, depending on the case under consideration. This sum is possible if it is presumed independence among the M or MS channels, a reasonable assumption when the bandwidth occupied by each modulated carrier is smaller than the *coherence bandwidth* of the channel and a frequency interleaver is applied.

Then, if the sum of the interference terms at the receiver input is modeled as Gaussian, the capacity for *each* of the M or MS channels of the MC-DS-CDMA system can be approximately estimated using (11). Furthermore, the value of $\sqrt{E_s}$ in this expression should be replaced by [11] $T\sqrt{P/2}$, where P is the average transmitted power per carrier and T is the BPSK symbol duration.

Table I shows the values for $g[i]$ and for the variances of the interference terms plus noise, σ^2 , to be operated in (11), according to each case taken into consideration here. The values of the average signal-to-noise ratio γ , used to calculate the average transmitted power per carrier, P , in each situation are also given in Table I. The value of $\beta[i]_\nu$, $i = 1, 2, \dots, X$, corresponds to the i -th value of the computer generated Rayleigh random variable, i.i.d. for all i and ν , and J_ν accounts for the interference variances at the output of each matched filter, at the receiver. In fact, the values of J_ν are different for different combiner's outputs [11], but if this difference is not taken into consideration, the channel capacity results are not significantly affected [14].

Table I – Values operated directly or indirectly in (11), (12) and (13).

	γ	$g[i]$	σ^2
M ch. EGC	$\frac{PT}{N_0 S} E \left[\left(\sum_{\nu=1}^S \beta_\nu \right)^2 \right]$	$\sum_{\nu=1}^S \beta[i]_\nu$	$\sum_{\nu=1}^S J_\nu + \frac{N_0 T S}{4}$
M ch. MRC	$\frac{PT}{N_0} E \left(\sum_{\nu=1}^S \beta_\nu^2 \right)$	$\sum_{\nu=1}^S \beta[i]_\nu^2$	$g[i] \left[\frac{1}{S} \sum_{\nu=1}^S J_\nu + \frac{N_0 T}{4} \right]$
MS ch.	$\frac{PT}{N_0} E (\beta^2)$	$\beta[i]$	$\frac{1}{S} \sum_{\nu=1}^S J_\nu + \frac{N_0 T}{4}$

IV. RESULTS AND DISCUSSION

In this paper we consider three situations for the MC-DS-CDMA system analyzed: Equal Gain Combining (EGC), Maximal Ratio Combining (MRC) and no combining. The first two situations are interpreted as M parallel channels. The last one is interpreted as MS parallel channels. In all cases, the total capacity was calculated as the sum of M or MS capacities, according to the case under consideration, versus the average SNR at the output of each channel. The results were converted to spectral efficiency versus the Shannon minimum average SNR per information bit for error-free transmission.

Fig. 5 shows capacity results for $MS = 6$ and variable M and S . For M channels with diversity, the results shown are for MRC combining. The systems with EGC combining have capacities identical to those of systems with MRC combining, but, for a given SNR, the necessary transmitted power per carrier with EGC is greater than that necessary for MRC.

It is possible to note from Fig. 5 that it is more advantageous to explore the maximum order of diversity instead of exploring the maximum parallelism of the data stream. This indicates that one possible coding scheme is a concatenation of a powerful outer code with the inherent repetition inner code of the system. At the receiver, the outputs

of the combiners can be viewed as soft-inputs for the outer decoder.

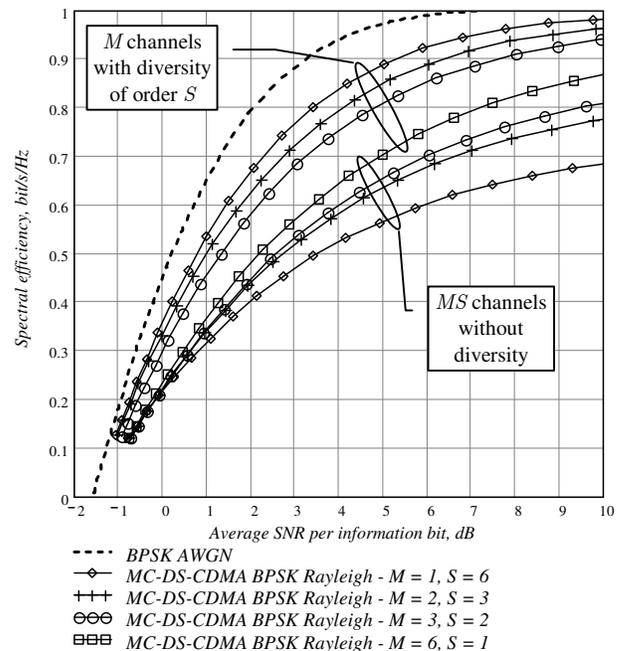


Fig. 5. Spectral efficiency for the MC-DS-CDMA system on the Rayleigh channel; $MS = 6$, variable M and S .

The length N_1 of the spreading code for a single-carrier CDMA system taken for reference was made equal to 60, and the number of resolvable paths L_1 for this single-carrier case was made equal to 4. In this case (1) is satisfied and the number L of resolvable paths per carrier reduces to 1.

The multi-path intensity profile was considered uniformly distributed and the number of active users, K , was made equal to 10. The expressions that permit the estimation of the variances J_ν in Table I can be obtained from the Appendix.

It can also be seen from Fig. 5 that the use of diversity is critical to system performance improvement and that, as the diversity order S increases, the channel capacity approximates the AWGN one. The observation of MS channels by the receiver, without diversity, significantly reduces the capacity, especially for high information rates.

Still referring to the results presented in Fig. 5, it can be seen that the best choice for the system parameters is $M = 1$ and $S = 6$. However, in real channels it is almost impossible to guarantee low correlation between diversity carriers if they are adjacent and, furthermore, overlapping. This fact leads us to the conclusion that the best choice for M is the minimum value such that the diversity carriers experience low correlation. The value of S is then readily obtained.

The use of the concatenated coding scheme suggested as an option for this MC-DS-CDMA system not necessarily increases the transmission bandwidth or reduces the information data rate. The length of the spreading code per carrier, N , can be adjusted [14] to compensate for the reduced coded symbols duration $R_c T$ due to coding of rate R_c , keeping unchanged both the bandwidth and the information rate, relative to the uncoded system. In this case, however, the

channel capacity is reduced, since the total variance of the interference terms in the decision variable is increased.

Fig. 6 shows some results for the spectral efficiency of the MC-DS-CDMA system for $M = 1$, $S = 6$ and variable number of active users, K . As expected, the capacity reduces as the number of users is increased, all other parameters unchanged. The results shown in Fig. 6, and also those presented in Fig. 5, indicate that, referring to the concatenated coding scheme already mentioned, it is preferable to use a low-rate outer code in order to aim the best performance, since the capacity is changed less than 1 dB for code rates below 0.2. In [14] it is suggested a low-rate product code with turbo-decoding as an outer coding/decoding scheme for the MC-DS-CDMA system and it is shown that good performance results can be obtained, even if the length of the spreading code per carrier is changed to maintain the information rate and the occupied bandwidth.

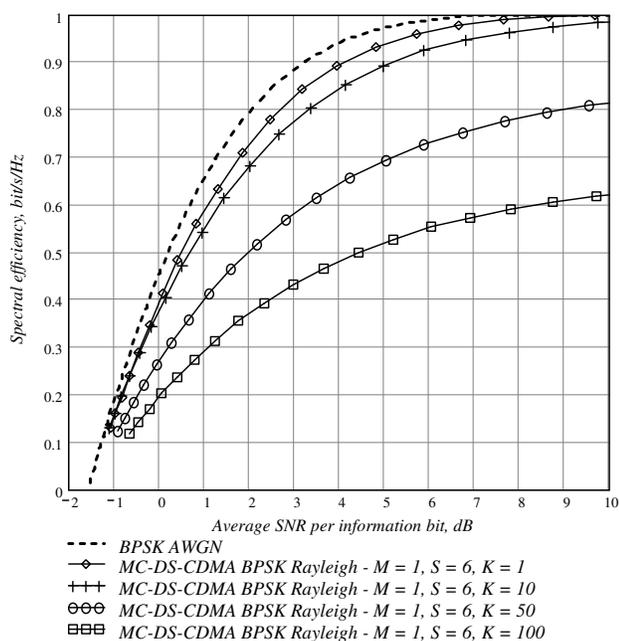


Fig. 6. Spectral efficiency for the MC-DS-CDMA system on the Rayleigh channel; $M = 1$, $S = 6$, variable K .

V. CONCLUSIONS

This paper presented some results of the investigation about the Shannon channel capacity for the MC-DS-CDMA system proposed in [11], considering that the fading amplitudes are independent and identically distributed in time and frequency, and that the channel state information is available at the receiver. The channel capacities were estimated through a Monte Carlo technique, which can avoid the characteristic manipulation of probability density functions typically found in channel capacity expressions. One drawback of this method is that it can take a long time of computer processing to demonstrate convergence, obviously depending on the complexity of the capacity formula.

It was concluded that, for the MC-DS-CDMA system under consideration, it is preferable to explore the maximum order of diversity instead of exploring the maximum parallelism of the data stream. It was also concluded that one possible solution

for a channel coding scheme to be used along with this system is a concatenation of a powerful outer code with the inherent inner repetition code of the system. At the receiver, the combiner's outputs serve as soft-inputs for the outer channel decoder. It was also pointed out that it would be preferable to choose an outer low-rate code in order to aim better performance, from the point of view of capacity achieving results.

REFERENCES

- [1] C. E. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 1948, 379-423, 623-656.
- [2] T. Huschka, M. Reinhardt & J. Lindner, Channel Capacities of Fading Radio Channels. in *7th IEEE Int. Symp. Personal, Indoor, and Mobile Radio Communication, PIMRC'96*, Taipei, Taiwan, 1996, 467-471.
- [3] M. S. Alouini & A. Goldsmith, Capacity of Nakagami Multipath Fading Channels, in *Proc. IEEE Veh. Technol. Conf. VTC'97*, Phoenix, AZ, 1997, 358-362.
- [4] C. G. Günther, Comment on "Estimate of Channel Capacity in Rayleigh Fading Environment", *IEEE Transactions on Vehicular Technology*, 45(2), 1996, 401-403.
- [5] A. J. Goldsmith & P. P. Varaiya, Capacity of Fading Channels with Channel Side Information, *IEEE Transactions on Information Theory*, 43(6), 1997, 1986-1992.
- [6] L. H. Ozarow, S. Shamai & A. D. Wyner, Information Theoretic Considerations for Cellular Mobile Radio, *IEEE Transactions on Vehicular Technology*, 43(2), 1994, 359-378.
- [7] E. Biglieri, J. Proakis & S. Shamai, Fading Channels: Information-Theoretic and Communications Aspects, *IEEE Transactions on Information Theory*, 44(6), 1998, 2619-2692.
- [8] E. K. Hall & S. G. Wilson, Design and Analysis of Turbo Codes on Rayleigh Fading Channels, *IEEE JSAC*, 16(2), 1998, 160-174.
- [9] T. F. Wong, Numerical Calculation of Symmetric Capacity of Rayleigh Fading Channel with BPSK/QPSK, *IEEE Communications Letters*, 5(8), 2001, 328-330.
- [10] L. Goldfeld & V. Lyandres, Capacity of the Multicarrier Channel with Frequency-Selective Nakagami Fading, *IEEE Transactions on Communication*, E83-B(3), 2000, 697-702.
- [11] E. Sourour & M. Nakagawa, Performance of Orthogonal Multicarrier CDMA in a Multipath Fading Channel, *IEEE Trans. on Comm.*, 44(3), 1996, 356-367.
- [12] S. Kaiser, *Multi-carrier CDMA mobile radio systems - analysis and optimisation of detection, decoding and channel estimation* (Ph.D. Thesis, VDI Verlag GmbH. Düsseldorf, 1998).
- [13] J. Wolfowitz, *Coding Theorems of Information Theory* (New York: Springer-Verlag, 2nd ed., 1964).
- [14] D. A. Guimarães, *A Class of Product Codes and Its Turbo Decoding Applied to a Multicarrier CDMA System*, Ph.D. thesis, Unicamp, SP, Brazil, 2003 (in Portuguese).
- [15] R. G. Gallager, *Information Theory and Reliable Communication* (John Wiley & Sons, 1968).

VI. APPENDIX

The following expressions permit the calculation of the variances J_v shown in Table I and used for the channel capacity estimations considered in this paper. A more complete mathematical analysis, as well as generic expressions for interference calculation for the MC-DS-CDMA system can be obtained from [11].

$$J_\nu = \frac{PT^2}{6N^3}r + \frac{1}{M} \sum_{p=1}^M \frac{PT^2}{2\pi^2 N^3} \mu Q_{p,\nu} \quad (14)$$

where

$$N = \frac{2M}{MS+1} N_1 \quad (15)$$

$$r = 2(K-1)N^2 \quad (16)$$

$$\mu = (K-1)N^2 \quad (17)$$

$$Q_{p,\nu} = \sum_{\substack{m=1 \\ \neq p+(\nu-1)M}}^{MS} \frac{1}{[m-p-(\nu-1)M]^2} \quad (18)$$



Dayan Adionel Guimarães was Born in Carrancas, MG, Brazil, on March 01, 1969. He holds the titles: Electronics Technician (ETE “FMC”, 1987), Electrical Engineer (Inatel, 1994), Specialist in Data Communication Engineering (Inatel, 2003), Specialist in Human Resources Management (FAI, 1996), Master in Electrical Engineering (Unicamp, 1998) and Doctor in Electrical Engineering (Unicamp, 2003).

From 1988 to 1993 he developed equipment for Industrial Instrumentation and Control, and was also Manufacturing Supervisor and Product Engineering Supervisor at *SENSE Sensores e Instrumentos*. Since January 1995 he is Professor at Inatel where, for eight years, he was responsible for the structure that supports practical teaching activities for the Electrical Engineering undergraduate course. His research includes the general aspects about Digital Transmission and Mobile Communications, specifically Multi-Carrier CDMA systems, and coding for fading channels, specifically Block Turbo Codes.

Dr. Dayan is member of the *Telecomunicações* magazine’s Editorial Board, member of the Inatel’s Master Degree Counseling Board and of the IEICE (*Institute of Electronics, Information and Communication Engineers*).

Análise de Desempenho da Esfera Integradora para Calibração de Medidores de Potência Óptica

Hilton Henrique Bertan & Yuzo Iano

Abstract—We have studied the integrating sphere for application as a standard in optical power measurement. This work was developed using resources and equipment commonly available in a laboratory. We have compared the sphere with the standard optical power meter of calibration laboratory and the results indicate that the measurement uncertainty of the integrating sphere is smaller.

Index Terms— Integrating Sphere, Measurement Uncertainty, Optical Power.

Resumo—A esfera integradora foi estudada para sua aplicação como um padrão em medição de potência óptica. Este trabalho foi desenvolvido utilizando recursos e equipamentos limitados disponíveis no laboratório. A esfera foi comparada com o medidor de potência óptica padrão do laboratório de calibração e os resultados indicaram que a incerteza de medição da esfera integradora é menor.

Palavras chave—Esfera Integradora, Incerteza de Medição, Potência Óptica.

I. INTRODUÇÃO

A evolução da tecnologia está permitindo o desenvolvimento de técnicas mais precisas e dispositivos mais confiáveis a custos menores com o objetivo constante de aumentar o desempenho dos sistemas de telecomunicações quanto à sua velocidade, capacidade e confiabilidade. Portanto, esses sistemas vêm exigindo instrumentos cada vez mais precisos para a medição de diversas grandezas.

Uma das grandezas cuja medição requer extrema precisão é a potência óptica, cuja importância vem aumentando à medida que equipamentos mais complexos são projetados e utilizados nos laboratórios do mundo todo.

Esses equipamentos necessitam de calibração com incertezas de medição muito baixas, o que é conseguido utilizando uma cadeia de rastreabilidade formada por equipamentos confiáveis e cujas propriedades garantem um ótimo desempenho.

Os métodos de medição de potência óptica evoluíram com o tempo, chegando a equipamentos como o radiômetro

criogênico.

A. Radiômetro Criogênico

O funcionamento do radiômetro criogênico é baseado no método da substituição elétrica, onde são comparadas a potência da radiação incidente no dispositivo e a potência elétrica de referência. O detector do radiômetro sofre um aquecimento devido à incidência da radiação e essa diferença de temperatura é medida. Quando é interrompida a radiação, uma corrente elétrica de referência, com intensidade precisamente medida, aquece o detector a mesma temperatura que a causada pela radiação, permitindo que sejam comparadas as potências óptica e elétrica.

Diversos institutos nacionais de metrologia, que são os responsáveis pela regulamentação e manutenção da rastreabilidade em medições, possuem o radiômetro criogênico como padrão primário, pois este possui incerteza de medição extremamente baixa [1]. O Inmetro, Instituto Nacional de Metrologia, Normalização e Qualidade Industrial, possui um radiômetro desse tipo com o objetivo de garantir a rastreabilidade dos laboratórios brasileiros ao que há de melhor em medição óptica.

Embora detentor dessas vantagens relativas à precisão, o radiômetro possui uma operação muito complexa que dificulta seu uso em operações rotineiras, tornando impraticável sua utilização como padrão de trabalho em um laboratório que exige um uso constante.

Devido à complexidade na operação do radiômetro criogênico, são utilizados padrões de transferência, que são padrões utilizados como intermediários para comparar padrões [2], dentre eles destacando-se os detectores do tipo armadilha óptica, também chamados de detectores *trap* ou detectores de armadilha óptica [3].

B. Detectores de Armadilha Óptica

Os detectores ópticos que utilizam o conceito de armadilha óptica possuem sensores instalados em seu interior posicionados para formar uma armadilha para a radiação incidente. Dessa maneira, as reflexões que fatalmente ocorrem na superfície de um fotodetector podem ser absorvidas e medidas por outro, o que diminui os erros na medição.

Esta configuração tipo armadilha possui as vantagens de reduzir os efeitos do coeficiente de temperatura, da

Manuscrito recebido em 29 de novembro de 2006; revisado em 26 de fevereiro de 2007.

H. H. Bertan (hiltonbertan@yahoo.com.br) e Y. Iano (yuzo@decom.fee.unicamp.br) pertencem ao Departamento de Comunicações da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas – FEEC/UNICAMP. Av. Albert Einstein, 400 - Cidade Universitária Zeferino Vaz – Campinas – SP – Brasil - 13083-852.

reflectância e da influência da não-uniformidade dos detectores [1]. Laboratórios de diferentes institutos do mundo utilizam esses detectores como padrões de transferência para comparações de entre seus radiômetros criogênicos [4].

O radiômetro criogênico funciona segundo o princípio da medição de diferença de temperatura ocasionada pela absorção da radiação incidente. No caso dos detectores do tipo armadilha, os fotodetectores instalados em seu interior são construídos de forma que a energia dos fótons que incidem no semiconductor é usada para forçar a retirada dos elétrons da banda de valência levando-os para a banda de condução. Através desta fotoionização, os fótons dão origem a portadores de carga que, antes que se recombinem no interior do dispositivo, são movimentados por um campo elétrico gerando a corrente elétrica que será medida por um equipamento apropriado [5].

Os detectores tipo armadilha são muitas vezes utilizados como padrões de transferência na calibração de outros padrões, que por sua vez serão utilizados rotineiramente nos laboratórios de calibração. Nos laboratórios de calibração que trabalham na caracterização de equipamentos para telecomunicações, os medidores usados são otimizados para os comprimentos de onda de 1300nm e 1550nm, principalmente. O Silício (Si) é o material geralmente utilizado na primeira janela de transmissão (850nm), mas nas segunda e terceira janelas, 1300nm e 1550nm, respectivamente, são utilizados o Germânio (Ge) ou InGaAs.

II. A ESFERA INTEGRADORA

Os radiômetros de Germânio ou InGaAs citados são utilizados sem uma esfera integradora acoplada, ou seja, a radiação incidente atinge diretamente o fotodetector. Estes são muito empregados como padrões de referência em laboratórios de calibração, mas um radiômetro usando a esfera integradora pode ser utilizado e trazer algumas vantagens.

A. Estrutura

A esfera integradora é uma estrutura de forma esférica com no mínimo duas aberturas, sendo uma para a entrada da radiação e outra para a instalação do fotodetector. Este fotodetector recebe a luz refletida pelo interior da esfera e origina a corrente que pode ser medida por um equipamento apropriado. A estrutura esférica possui o interior revestido com um material especial que reflete e difunde a luz que entra por sua abertura de maneira uniforme em todas as direções, representando dessa maneira uma superfície com reflexão difusa perfeita. A constante de tempo do dispositivo, ou seja, o tempo necessário para que as reflexões ocorram de forma que desapareça o transiente, varia de alguns nanossegundos até dezenas de nanossegundos [6].

A esfera integradora pode ser instalada na entrada de um radiômetro de forma a dar novas características ao conjunto, diminuindo a incerteza de medição e possibilitando que ela seja utilizada como padrão de trabalho ou até mesmo como padrão de referência do laboratório. Padrão de referência é o

padrão, geralmente tendo a mais alta qualidade metrológica disponível em um dado local ou em uma dada organização, a partir do qual as medições lá executadas são derivadas [2].

B. Teoria

A esfera integradora tem seu funcionamento baseado no princípio da troca de radiação entre superfícies difusoras [6]. Vamos considerar a troca de radiação entre dois elementos diferenciais de uma superfície difusora denominados de dA_1 e dA_2 , separados por uma distância S . A fração da energia que parte de dA_1 e incide sobre dA_2 é denominada fator de troca dF_{d1-d2} , dado por (1), onde θ_1 e θ_2 são ângulos medidos em relação as normais da superfície.

$$dF_{d1-d2} = \frac{\cos\theta_1 \cos\theta_2}{\pi S^2} dA_2 \quad (1)$$

Consideremos esses dois elementos diferenciais situados dentro de uma esfera com superfície interna difusora, conforme Figura 1.

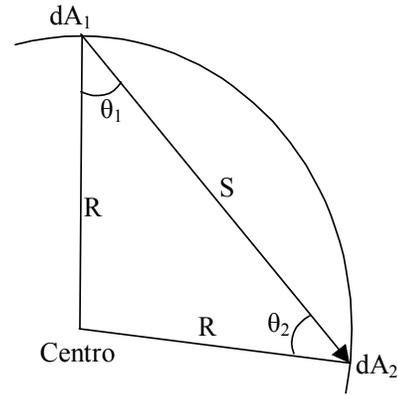


Fig. 1. Esfera com dois elementos trocando radiação

Como a distância S entre os elementos é dada por (2), então o fator de troca dF_{d1-d2} é dado por (3).

$$S = 2R \cos\theta_1 = 2R \cos\theta_2 \quad (2)$$

$$dF_{d1-d2} = \frac{dA_2}{4\pi R^2} \quad (3)$$

Pelas equações (2) e (3) notamos que o resultado independe dos ângulos θ_1 e θ_2 e das distâncias entre os elementos dA_1 e dA_2 , fazendo com que a fração do fluxo luminoso recebido por dA_2 é sempre o mesmo para qualquer radiação emitida de qualquer ponto da superfície da esfera. Se a área infinitesimal dA_1 emitir radiação agora para uma área finita igual a A_2 , teremos o calculado em (4).

$$\int dF_{d1-d2} = \frac{1}{4\pi R^2} \int_{A_2} dA_2 = \frac{A_2}{4\pi R^2} \quad (4)$$

Como o resultado é independente de dA_1 , temos a relação descrita em (5), onde A_S representa a área interna total da esfera e a fração do fluxo radiante recebido por A_2 é a razão entre A_2 e A_S .

$$F_{1-2} = \frac{A_2}{4\pi R^2} = \frac{A_2}{A_S} \quad (5)$$

A radiação incidente na superfície interna da esfera integradora cria uma fonte de luz virtual por reflexão, sendo essa fonte que emite radiação para o fotodetector, possibilitando sua medição.

O fluxo de radiação incidente é perfeitamente difundido na primeira reflexão, e depois de várias reflexões no interior da cavidade, o fluxo total incidente na superfície da esfera é maior que o fluxo de entrada devido a múltiplas reflexões dentro da estrutura, resultando em uma radiância (densidade de fluxo por unidade de ângulo sólido) na superfície dada por (6), onde Φ_i é o fluxo de entrada, A_S é a área atingida pela radiação, ρ é a reflectância da superfície interna da esfera e π é o ângulo sólido total projetado da superfície [6]. A esfera possui uma entrada para o feixe e uma saída para o detector, sendo que essas aberturas, além de não refletirem radiação, absorvem a luz que nelas incidem, fazendo com que em (6) tenhamos a variável f que representa a relação entre a soma das áreas das aberturas da estrutura e a área total.

$$L = \frac{\Phi_i}{\pi A_S} \cdot \frac{\rho}{1 - \rho(1 - f)} \quad (6)$$

III. ESTUDOS REALIZADOS

Definimos um sistema de medição formado por um picoamperímetro do fabricante Keithley modelo 6485, utilizado para as medições de corrente da ordem de nanoampéres, e uma esfera integradora do fabricante Labsphere modelo LPM-040-SF de 4 polegadas de diâmetro e duas aberturas circulares, sendo uma delas para o fotodetector de germânio.

A. Verificações

Aproveitando a boa infra-estrutura do Laboratório de Radiometria (Larad) do Inmetro, realizamos alguns ensaios para verificar o comportamento do fotodetector trabalhando isoladamente e também acoplado à esfera, sendo obtidos resultados muito satisfatórios que comprovaram suas características principais e motivaram a continuação dos trabalhos. Utilizamos primeiramente um detector tipo armadilha óptica calibrado em 632,8nm para calibrar um detector piroelétrico, tendo como fonte óptica um *laser* HeNe estabilizado em 632,8nm montado em uma bancada antivibratória especial para seu perfeito alinhamento. Como o detector piroelétrico possui intrinsecamente responsividade espectral uniforme, e este utilizado é especificado pelo fabricante para medições com erro menor que 1% em comprimentos de onda que abrangem a faixa de 600nm à 1550nm, o mesmo tornou-se nosso padrão de calibração também para os comprimentos de onda de interesse, ou seja, 1300nm e 1550nm. Portanto, este detector foi utilizado como padrão para verificar o desempenho da esfera integradora.

Foi realizada a verificação do detector de germânio da esfera integradora isoladamente e também acoplado ao sistema, incluindo o picoamperímetro para as leituras de corrente elétrica gerada. Esta verificação foi realizada utilizando-se *laser* nos comprimentos de onda de 1300nm e 1550nm. Os valores encontrados estão coerentes com os indicados na literatura para os detectores de germânio, indicando que o método utilizado estava correto e o detector em bom funcionamento, o que motivou a calibração do sistema de medição no Laboratório de Fibras Ópticas do Instituto Electrotécnico Português (IEP), localizado na cidade do Porto (Portugal).

B. Calibração do Sistema

O sistema formado pelo picoamperímetro, detector de germânio e esfera integradora foi calibrado no laboratório do IEP utilizando-se *laser* nos comprimentos de onda de 1300nm e 1550nm. Foi utilizado como padrão uma esfera integradora calibrada com rastreabilidade a organismos europeus reconhecidos pelo Inmetro.

O certificado de calibração emitido forneceu os valores de responsividade do sistema e as respectivas incertezas de medição para cada comprimento de onda calibrado. Todo o sistema foi calibrado como uma unidade, ou seja, os valores de incerteza pertencem à esfera, seu detector e picoamperímetro funcionando em conjunto.

No mesmo período também foi calibrado o medidor de potência óptica padrão do Laboratório de Calibração da Fundação CPqD. Este equipamento, denominado HP81521B, também utiliza um detector de germânio, mas não possui esfera integradora acoplada em sua estrutura, isto é, a radiação incide diretamente no elemento sensor. A calibração desse equipamento foi importante porque puderam ser comparados os valores de sua incerteza com a esfera calibrada no mesmo período e laboratório, e também com os mesmos padrões, ou seja, utilizando-se a mesma cadeia de rastreabilidade.

C. Experimentos Finais

A conclusão dos estudos foi realizada no Laboratório de Calibração do Centro de Pesquisa e Desenvolvimento em Telecomunicações – Fundação CPqD. Este laboratório possui uma ótima estrutura para os experimentos finais, sendo montado em um ambiente especialmente planejado para evitar interferências externas. Seu teto, paredes, bancadas e armários são todos revestidos com material especial para sala limpa visando a não retenção de partículas, além de possuir uma ante-sala e um tapete especial para retirada de sujeira do calçado. O laboratório possui temperatura e umidade controladas por equipamentos calibrados.

A montagem dos equipamentos é realizada sobre bancada antivibratória com tampo de granito, que está localizada no interior de um fluxo laminar formado por dois estágios de filtros e cortina especial antiestática, constituindo dessa forma um ambiente de sala limpa Classe 100. A sala possui uma linha de nitrogênio para a limpeza dos dispositivos e os

experimentos foram realizados usando-se vestuário apropriado constituído de aventais, capuz e luvas antiestáticos e sapato especial.

Toda e qualquer limpeza é realizada com materiais próprios para esses ambientes, como panos de limpeza e cotonetes para sala limpa. Um microscópio estava disponível para qualquer análise mais detalhada de algum dispositivo. As anotações foram realizadas diretamente em um microcomputador, ou na ausência deste utilizou-se papel vegetal, que possui fibras longas contribuindo para a eliminação de partículas no ambiente.

Foi realizada no laboratório do CPqD a calibração de um medidor de potência óptica do fabricante Anritsu modelo MA9301A, que possui características similares à maioria dos equipamentos encontrados no mercado e não possui esfera integradora acoplada. Esta calibração foi realizada utilizando-se como padrão o HP81521B, equipamento já consolidado como padrão de medição do Laboratório, e a Esfera Integradora Labsphere LPM-040-SF, com a finalidade de comparar diretamente os resultados. O experimento foi montado conforme esquemas das Figuras 2 e 3.

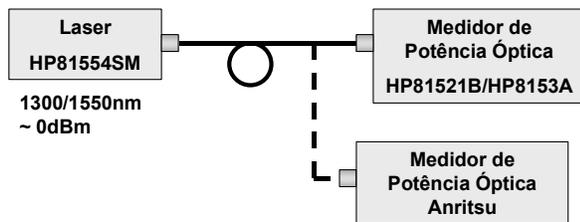


Fig. 2. Montagem para calibração do medidor Anritsu com o HP81521B.

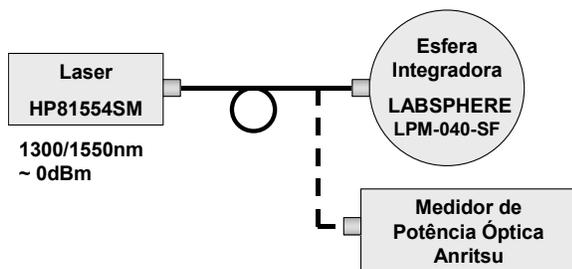


Fig. 3. Montagem para calibração do medidor Anritsu com a Esfera.

IV. AS COMPONENTES DE INCERTEZA

Algumas pesquisas foram realizadas para determinar quais são as componentes de incerteza de medição que precisam ser consideradas durante as medições de ambos os medidores. As informações da norma internacional IEC 1315 [7] e o Guia para a Expressão da Incerteza de Medição [8], também conhecido como GUM, foram seguidos.

A. Temperatura

Uma das componentes relevantes que precisou ser determinada foi aquela devido à variação da potência do laser

com a mudança de temperatura e/ou umidade relativa do ambiente. Foi projetado um experimento realizado em uma câmara climática do CPqD, com temperatura e umidade relativa controladas, onde os equipamentos permaneceram sob testes controlados via GPIB por software (LABView/National Instruments).

Os equipamentos utilizados no experimento foram aqueles calibrados no IEP: o HP81521B e o *laser*, que foram interligados com fibras ópticas monomodo conectorizadas. Foi notado que a variação de umidade relativa não influencia consideravelmente as medições, mas a variação de temperatura tem efeito considerável, principalmente porque a fonte óptica utilizada já não é um equipamento novo, o que já havia sido notado em estudos anteriores [9], onde uma pequena variação em seu comprimento de onda ao longo do tempo foi verificada, mas cuja influência é desprezível. Considerando uma distribuição retangular para o dado obtido experimentalmente, temos, para um nível de confiança de 68%, uma incerteza de medição devido à variação da temperatura de 1,77%.

B. Conectores

Outra componente de incerteza considerável e que foi pesquisada foi a dependência quanto ao tipo de conector óptico utilizado. Para isto foram utilizados os resultados obtidos no programa internacional de comparação interlaboratorial realizado entre o CPqD e o IEP, desenvolvido no mesmo período em que foram realizadas as calibrações dos equipamentos HP e Labsphere. Nessa comparação interlaboratorial, dentre outros resultados importantes, foram obtidos os que mostram o comportamento das medições quando são utilizados conectores ópticos de vários tipos, como FC/PC, FC/APC, construídos em material cerâmico ou metálico, e a combinação destes. Considerando uma distribuição retangular para o dado obtido experimentalmente, temos, para um nível de confiança de 68%, os valores de incerteza apresentadas nas tabelas correspondentes, sendo a maior delas de 0,17%.

Também ficou evidenciado nesse experimento que as medidas com a esfera variam menos quando se substitui um conector por um de outro tipo ou material, o que constitui uma vantagem em relação ao medidor HP, que não possui esfera em sua estrutura.

C. Análise dos resultados

A Tabela I mostra as componentes de incerteza de medição calculadas, já excluídas as consideradas desprezíveis, para o medidor HP81521B, que não utiliza esfera integradora em sua estrutura. A Tabela II lista as componentes referentes à esfera integradora Labsphere.

Calculada conforme o GUM, a incerteza expandida mostrada nas tabelas é o resultado da incerteza combinada das componentes de incerteza multiplicada pelo fator de abrangência (k), cujo valor é de 1,960 para um nível de

confiança de 95%. Observamos que as incertezas expandidas referentes à esfera são realmente menores.

O motivo de encontrarmos incertezas quanto à dependência do tipo de conector maiores para a esfera integradora não foi estudado neste trabalho, já que os baixos valores das outras componentes contribuem para a redução da incerteza da esfera.

TABELA I
INCERTEZAS DE MEDIÇÃO PARA O DETECTOR HP

COMPONENTE DE INCERTEZA	[%]	
	1310nm	1550nm
Certificado do Padrão	0,60	0,58
Linearidade do Padrão (Certificado)	0,23	0,23
Estabilidade Laser – Tempo (Certificado)	0,03	0,03
Não-uniformidade do detector (Dependência da Geometria/Alinhamento do Feixe)	0,12	0,12
Varição do Laser com temperatura	1,77	1,77
Dependência do tipo de Conector	0,09	0,14
Dependência quanto ao Lambda do Laser	0,06	0,58
Dependência da Polarização	0,29	0,29
Incerteza Expandida $k=1,960$ (95%)	3,74	3,90

TABELA II
INCERTEZAS DE MEDIÇÃO PARA A ESFERA LABSPHERE

COMPONENTE DE INCERTEZA	[%]	
	1310nm	1550nm
Certificado do Padrão	0,55	0,56
Linearidade do Padrão (Certificado)	0,23	0,23
Estabilidade Laser - Tempo (Certificado)	0,03	0,03
Não-uniformidade do detector (Dependência da Geometria/Alinhamento do Feixe)	0,00	0,00
Varição do Laser com temperatura	1,77	1,77
Dependência do tipo de Conector	0,12	0,17
Dependência quanto ao Lambda do Laser	0,06	0,58
Dependência da Polarização	0,00	0,00
Incerteza Expandida $k=1,960$ (95%)	3,67	3,85

Nas Tabela III e IV, que listam os resultados obtidos na calibração do equipamento Anritsu utilizando o padrão HP e utilizando o padrão esfera integradora, observamos que a maior componente é aquela herdada do padrão, como já era esperado. Esta calibração do medidor Anritsu com os dois padrões foi realizada seguindo o mesmo procedimento, inclusive com a mesma fonte óptica configurada para aproximadamente 0dBm, pois esta esfera utilizada no experimento não é apropriada para medições de baixa potência.

Notamos que, embora o erro na calibração utilizando-se a esfera integradora em 1300nm seja o dobro do mesmo erro utilizando-se o padrão HP, e praticamente a metade em 1550nm, este fato não é muito relevante neste caso, já que estes valores são menores que as incertezas associadas, permanecendo os resultados de ambos os padrões dentro da faixa de valores determinada pela incerteza de medição.

TABELA III
CALIBRAÇÃO DO MEDIDOR ANRITSU UTILIZANDO O PADRÃO HP

Parâmetro	[%]	
	1310nm	1550nm
Erro	1,11	3,44
Incerteza Tipo A (Leituras)	0,00	0,05
Incerteza Resolução	0,07	0,07
Incerteza Combinada (Padrão)	1,91	1,99
Incerteza Expandida $k=1,960$ (95%)	3,74	3,90

TABELA IV
CALIBRAÇÃO DO MEDIDOR ANRITSU UTILIZANDO O PADRÃO ESFERA

Parâmetro	[%]	
	1310nm	1550nm
Erro	2,22	1,76
Incerteza Tipo A (Leituras)	0,00	0,05
Incerteza Resolução	0,07	0,07
Incerteza Combinada (Padrão)	1,87	1,96
Incerteza Expandida $k=1,960$ (95%)	3,67	3,85

V. CONCLUSÕES

Esse trabalho analisou o desempenho do radiômetro baseado na esfera integradora, mostrando que suas características são ótimas para a calibração de medidores de potência óptica. Foram estudadas as propriedades principais do dispositivo e determinadas as incertezas de medição relevantes para sua aplicação. Os experimentos foram realizados utilizando-se equipamentos relativamente comuns, demonstrando que mesmo com recursos limitados é possível um trabalho prático e de utilidade. Obviamente, a incerteza referente à esfera ainda pode ser diminuída caso forem utilizados equipamentos novos e de qualidade superior ou demais instrumentos para realização de experimentos com melhor precisão [10], que não estavam disponibilizados para a realização desse trabalho devido ao custo elevado.

A esfera integradora, por não apresentar problemas quanto ao alinhamento e variações angulares do feixe, e dependência quanto a não-uniformidade do detector [6], e por possuir incerteza de medição inferior que o medidor HP, pode inclusive ser utilizada como padrão de referência de um laboratório. Foi observado também que os erros encontrados na calibração do medidor Anritsu, medidos usando o padrão HP ou o padrão esfera, estão muito próximos, considerando as incertezas envolvidas, concluindo uma comparação fiel dos dois equipamentos.

VI. AGRADECIMENTOS

Agradecemos as instituições e a todos que contribuíram no desenvolvimento desse estudo, destacando o Eng. Celso P. Saraiva (Laboratório de Calibração/CPqD), o Tecnologista Augusto P. da Cunha (Larad/Inmetro) e o Eng. Modesto de Moraes (Laboratório de Fibras Ópticas/IEP).

REFERÊNCIAS

- [1] A. Cunha et al. “Radiômetro criogênico: referência primária do Inmetro para medidas de radiação óptica”. *Metrologia*, 2003.
- [2] INMETRO. “VIM-Vocabulário internacional de termos fundamentais e gerais de metrologia”. Duque de Caxias, 1995. 52 p.
- [3] M. Simões et al. “Montagem e caracterização de detectores padrão do tipo armadilha óptica”. *Metrologia*, 2003.
- [4] A. Lassila et al. “Intercomparison of cryogenic radiometers using silicon trap detectors”. *Meas. Sci. Technol.*, 1997. v. 8 p. 123–127.
- [5] W. F. Giozza, E. Conforti, H. Waldman. “Fibras Ópticas: tecnologia e projeto de sistemas”. 1. ed. São Paulo: Makron, McGraw-Hill, 1991.
- [6] Labsphere, Inc. “A guide to integrating sphere radiometry and photometry”. North Sutton, 2003. 26 p.
- [7] Comissão Eletrotécnica Internacional. “IEC 1315 - Calibração de medidores de potência em fibra óptica”. 1ª ed. Genebra, 1995-04. 60 p.
- [8] ABNT, INMETRO. “Guia Para a Expressão da Incerteza de Medição (GUM)”. Rio de Janeiro, 3ª ed Brasileira, 2003. 120 p.
- [9] C. P. Saraiva, H. H. Bertan, M. Morais, Y. Iano. “Um estudo sobre a reprodutibilidade de técnica espectrométrica de alta resolução para sistemas DWDM”. ENQUALAB - Congresso e Feira da Qualidade em Metrologia. São Paulo, 2006.
- [10] J. Envall, P. Kärhä, E. Ikonen. “Measurements of fibre optic power using photodiodes with and without an integrating sphere”. *Metrologia*, 2004. v. 41, p. 353–358.

Hilton Henrique Bertan recebeu os títulos de Técnico em Eletro-Eletrônica pelo Colégio Técnico de Campinas - COTUCA/Unicamp em 1997 e Engenheiro Eletricista pela Unicamp em 2004.

De 2000 a 2006 foi Metrologista do Laboratório de Calibração da Fundação CPqD - Centro de Pesquisa e Desenvolvimento em Telecomunicações. Atualmente é aluno de Mestrado da FEEC/UNICAMP e Engenheiro Eletrônico do Centro de Manutenção de Equipamentos da Unicamp.

Yuzo Iano recebeu os títulos de Engenheiro Eletrônico, Mestre e Doutor em Engenharia Elétrica através da UNICAMP. Atualmente é Professor Adjunto do DECOM/FEEC/UNICAMP. É o responsável pelo Laboratório de Comunicações Visuais do DECOM.

Seus interesses estão em processamento digital de sinais de áudio e vídeo. Trabalha com pesquisas em televisão e em especial com a HDTV desde 1986.

Propriedades Direcionais de Irradiação de Antenas Rádio-Base

Marco Antonio Brasil Terada
Departamento de Engenharia Elétrica
Universidade de Brasília
Caixa Postal 4386 – Brasília/DF 70919-970

Abstract—This work discusses in detail the directional properties of the radiation of radio base antennas in cellular systems with respect to various variables, in addition to the usually employed variable of the distance from the bottom of the radio base tower. It is demonstrated that the electric field intensity peaks at different distances from the bottom of the radio base tower depending on these variables, leading to the conclusion that fixing a minimum distance from populated areas for the installation of radio bases is not the proper way to ensure safety.

Index Terms—Applied Electromagnetics, Antennas and Propagation, Mobile Communications, Biological Effects.

Resumo—Este trabalho¹ investiga em detalhe as propriedades direcionais de irradiação de antenas rádio-base em sistemas de comunicações celulares em função de diversas variáveis, além da usual distância da base da estação rádio-base. É mostrado que o máximo da intensidade do campo elétrico ocorre a distâncias diferentes da base da ERB dependendo destas outras variáveis, levando à conclusão que se fixar uma distância mínima de regiões povoadas para a instalação de ERBs não é a maneira adequada para se assegurar a segurança e o bem-estar da população.

Palavras chave—Eletromagnetismo Aplicado, Antenas e Propagação, Comunicações Móveis, Efeitos Biológicos.

I. INTRODUÇÃO

O explosivo recrudescimento de serviços baseados em sistemas de comunicações sem-fio representa um mercado que diretamente suporta e viabiliza o processo de globalização através de projetos de valor agregado superior a centenas de bilhões de reais. Correntemente existem cerca de 2 bilhões e 740 milhões de usuários de comunicações móveis celulares e PCS no mundo [1]. Esses dados não incluem usuários que estarão exclusivamente imersos em redes locais e pessoais, apesar de um certo grau de compartilhamento seja esperado. A proliferação de sistemas de comunicações sem-fio com alto grau de sofisticação necessitará do projeto e implementação de novas configurações, a serem usadas no atendimento à demandas técnicas cada vez mais exigentes derivadas de novos serviços e aplicações.

O mercado Brasileiro de comunicações celulares segue um caminho similar, com quase 100 milhões de usuários de serviços de comunicações sem-fio e de voz no final de

2006. No entanto, o desenvolvimento técnico e a implementação de novos serviços têm sofrido atrasos consideráveis nestes últimos anos. Isto se deve ao desconhecimento da população em geral quanto aos fundamentos de funcionamento das comunicações celulares e seus possíveis efeitos em nossas vidas. Estas dúvidas, ainda que justificáveis do ponto de vista da proteção de nosso bem-estar e saúde, levaram à elaboração de diversas leis que tentam restringir a instalação de Estações Rádio-Base (ERBs) perto de áreas povoadas. Estações Rádio-Base são estações de radiocomunicações de base do Serviço Móvel Pessoal (a telefonia móvel), usadas para radiocomunicação com estações móveis, ou seja, com os aparelhos terminais da telefonia móvel (os telefones celulares)².

Este trabalho tem por objetivo investigar as propriedades direcionais de radiação da antenas de ERBs. Será também evidenciado que as leis anteriormente mencionadas são arbitrárias e sem consistência científica, podendo inclusive introduzir problemas ao invés de solucioná-los. A maioria destas leis requer que a instalação de ERBs ocorra à uma distância mínima de escolas e unidades imobiliárias, e obrigam o re-posicionamento de ERBs que já estejam instaladas e em operação em distâncias inferiores a estes mínimos. Na realidade, a distância da base da ERB não é a única variável que deve ser levada em consideração, e o re-posicionamento da ERB para uma distância de 50 m [2], por exemplo, pode aumentar a radiação eletromagnética nas unidades imobiliárias que as leis estão tentando proteger. É importante se ressaltar que todas as ERBs conhecidas do autor atendem às especificações da Agência Brasileira de Telecomunicações (ANATEL), as quais ao contrário de leis como a [2] não obrigam que as ERBs sejam instaladas à uma distância mínima de unidades imobiliárias, mas requerem que a irradiação eletromagnética em todas as regiões povoadas esteja abaixo de valores mínimos de acordo com a frequência de operação [3].

II. CONSIDERAÇÕES PRELIMINARES

Inicialmente, esclarece-se que a distância da base da antena não é a única nem a mais importante variável a partir da qual se define qual a intensidade do campo eletromagnético proporcionado por essa mesma antena [4,5]. Há diversas outras variáveis ao lado da distância da base da antena que são

¹Este trabalho foi desenvolvido com suporte da Associação Nacional das Operadoras Celulares (ACEL), através de um contrato entre a ACEL e a Universidade de Brasília.

² Cf. Art. 3º, XIV, do Regulamento do Serviço Móvel Pessoal, aprovado pela Resolução nº. 316, de 27/09/2002, da Anatel.

relevantes para se apurar qual a intensidade do campo eletromagnético produzido pela antena. Sem o intuito de esgotar o rol dessas variáveis, o que seria desnecessário para o escopo desse trabalho, pode-se destacar dentre elas: (1) a altura de instalação da antena; (2) a altura, em relação ao solo, do ponto em que se deseja medir a intensidade do campo eletromagnético; (3) o ângulo de inclinação da antena; e (4) a potência efetivamente irradiada pela antena [3].

A representação gráfica que segue (Figura 1) apresenta a geometria dessa questão relativamente às duas primeiras variáveis destacadas:

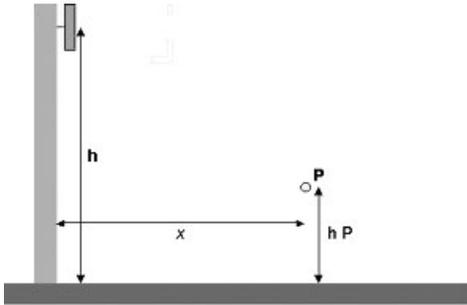


Figura 1 - Distâncias relevantes para apuração do campo eletromagnético.

Considerado um ponto qualquer no espaço (ponto P, na Figura 1), a intensidade do Campo Eletromagnético nesse ponto em decorrência da antena ali representada variará não apenas em função da distância desse ponto até a torre onde está instalada a antena (x), mas também em função da distância desse ponto ao chão (hP), e da altura de instalação da antena (h).

O ângulo de inclinação da antena também é relevante, pois as antenas de telefonia móvel são direcionais⁴. As Figuras 2 e 3 objetivam apontar que ângulo seria este.

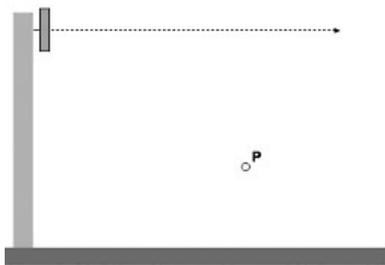


Figura 2 – Antena sem inclinação.

³ A potência efetivamente irradiada, *e.r.p.*, é definida como o produto da potência de entrada da antena pelo ganho linear da antena com relação à antena do tipo dipolo. Nesse sentido, também, a bibliografia já citada [4,5].

⁴ Antenas direcionais apresentam uma *e.r.p.* que é função da direção, ou ângulo, entre a antena e o ponto de observação. Conforme podemos ver das Figs. 2 e 3, o ângulo entre a seta de linha pontilhada e o ponto P varia de acordo com o ângulo de inclinação da antena, o que resulta em campos eletromagnéticos diferentes no ponto P se a antena for direcional. Uma antena omnidirecional produz campos uniformes ao seu redor, independente da direção. Desta forma, os campos produzidos por uma antena omnidirecional no ponto P das Figs. 2 e 3 seriam os mesmos independente do ângulo de inclinação da antena, em contraste com as antenas direcionais que dependem do mesmo.

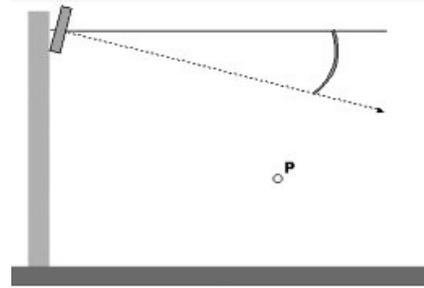


Figura 3 – Antena com inclinação.

Como se constata pela comparação entre as figuras, tem-se na Figura 2 um ângulo de 0° quando a antena é paralela à torre em que está instalada. Na Figura 3 a antena é inclinada em direção ao solo, daí surgindo um ângulo. Trata-se do ângulo de inclinação do eixo da antena, que faz com que esse eixo se aproxime do ponto P, no exemplo citado, fazendo com que seja maior a intensidade do campo eletromagnético no mesmo. Para um correto entendimento do funcionamento dos sistemas de comunicações celulares, é portanto necessário que se esclareça não só como varia a intensidade do campo eletromagnético em função da distância x como também como varia em função das demais variáveis apontadas (altura da antena, altura do ponto P e ângulo de inclinação da antena).

Esclarece-se, ainda, que seja nos regulamentos tratando da matéria, seja na legislação, ora se fala em *campo elétrico*, ora em *campo eletromagnético*. Os dois conceitos se distinguem, dado que o campo eletromagnético é a co-existência dos campos elétrico e magnético variando no tempo simultaneamente (um gera o outro), mas para o que interessa ao presente trabalho, é importante deixar destacado que os valores da intensidade do campo elétrico e do campo magnético são sempre diretamente proporcionais um ao outro. No escopo deste trabalho, a intensidade do campo magnético é simplesmente a intensidade do campo elétrico dividida por 377 [4,5]. As unidades são diferentes, no entanto, sendo que o campo elétrico é dado em V/m (Volts por metro) e o campo magnético em A/m (Amperes por metro). Não faz sentido se falar em intensidade de campo eletromagnético, a menos que esteja implícita as intensidades dos campos elétrico e magnético separadamente. Na continuidade deste trabalho, serão utilizados sempre os dados do campo elétrico, ficando, porém, claro que a toda e qualquer variação do campo elétrico varia em igual proporção o campo magnético.

Apresentamos inicialmente um gráfico que aponta a variação da intensidade do campo elétrico gerado por uma dada antena à medida que se varia a distância à base da ERB. Para tanto, utilizando técnicas analíticas de análise de antenas, bem como utilizando o programa computacional GRADMAX [6,7], desenvolvido pelo autor, é apresentado o gráfico que segue (Figura 4), calculado a partir de uma antena instalada a dez metros de altura ($h = 10\text{m}$), distância do solo ao ponto analisado de um metro e meio ($hP = 1,5\text{m}$) e com uma

potência efetivamente irradiada de 2512 Watts^[5] e um ângulo de inclinação da antena de 8°.

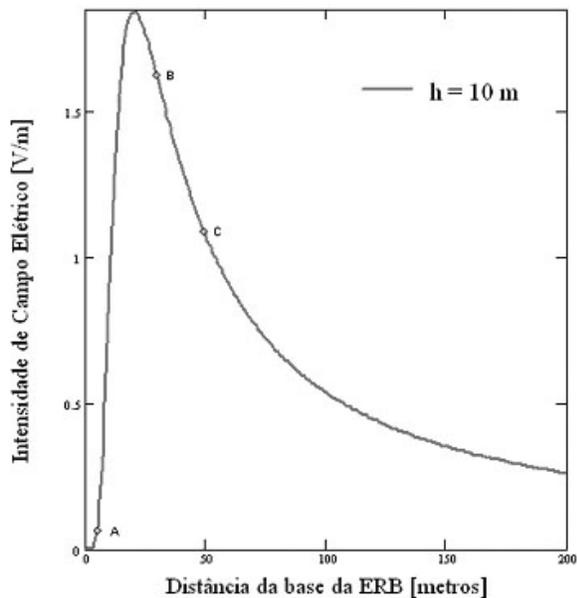


Figura 4 - Intensidade de campo elétrico em função da distância x da base da ERB. A análise foi feita para uma antena instalada a 10 metros e com uma inclinação de 8°.

No gráfico em questão (Figura 4), o eixo horizontal mostra a distância x em metros da base da ERB e o eixo vertical mostra a intensidade de campo elétrico em V/m (Volts por metro). Da Figura 4, conclui-se que:

- a intensidade do campo elétrico é a mais baixa exatamente quando são menores as distâncias do ponto até a torre da antena (valores de x próximos a zero);
- os valores da intensidade do campo elétrico vão aumentando progressivamente, atingindo o valor máximo de 1,850 V/m quando a distância x é de 20 m;

A TABELA I mostra os valores da intensidade do campo elétrico para as distâncias $x = 5$ m, 30 m e 50 m. Assim, quando se está a 1,5 m de altura e 5,0 metros de distância de uma torre de antena com as especificações dadas, a intensidade do campo elétrico é de 0,077 Volts por metro. À medida que a distância x vai aumentando (à medida que se afasta da base da antena), esse valor aumenta, até atingir 1,850 V/m a 20 metros da base da antena. Começa a diminuir, chegando a 1,615 V/m

a uma distância de 30 metros e a 1,076 V/m a cinquenta metros, prosseguindo em queda.

TABELA I
INTENSIDADE DO CAMPO ELÉTRICO.

	x [metros]	Intensidade do Campo Elétrico [V/m]
A	5	0,077
B	30	1,615
C	50	1,076

III. ALTERAÇÕES DA INTENSIDADE DO CAMPO ELÉTRICO EM FUNÇÃO DAS DEMAIS VARIÁVEIS

A. Variação da intensidade do campo elétrico em função da variação da altura da antena

O gráfico seguinte (Figura 5) mostra o comportamento da intensidade do campo elétrico em função da distância x da base da ERB para três diferentes alturas de antena, quais sejam, 10 m, 20 m e 30 m, mantendo-se as demais variáveis constantes^[6].

A Tabela II lista os valores da intensidade do campo elétrico para as distâncias $x = 5$ m, 30 m e 50 m, bem como para o pico do campo em cada uma das torres.

A partir da comparação das três curvas da Fig. 5 e dos valores listados na Tabela II, constata-se que a intensidade do campo elétrico varia sensivelmente em função da variação da altura de instalação da antena, levando às seguintes conclusões:

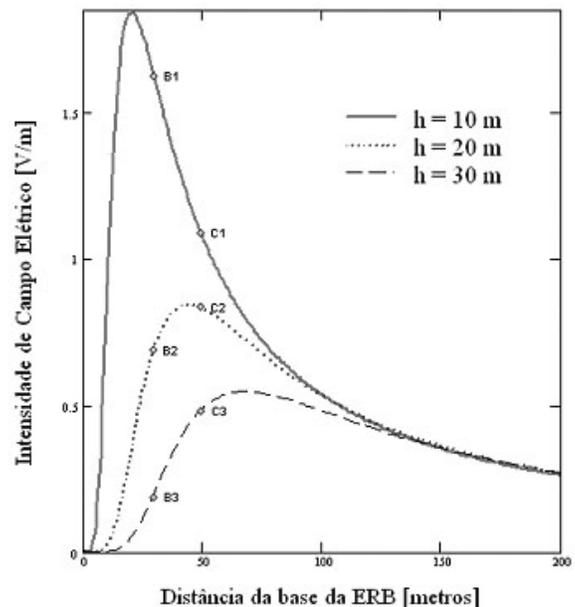


Figura 5 - Intensidade de campo elétrico em função da distância x da base da ERB considerando três diferentes alturas de instalação da antena.

⁵ Essa especificação corresponde ao valor máximo admitido na regulamentação editada pela Anatel para o Serviço Móvel Pessoal [8], ou seja, corresponde à especificação de *e.r.p.* (potência efetivamente irradiada) de 64 dBm para antenas que operam na frequência de 869 a 894 MHz (Banda A da telefonia móvel), com uma potência de entrada de 100 W. Por definição, $e.r.p.$ [dBm] = $10 \log(e.r.p. \text{ em mW})$, daí resultando que 64 dBm correspondem a 2512W (Nesse sentido, também, a bibliografia já citada nas notas [4,5]).

⁶ ponto de observação (ponto "P") distante 1,5 m do nível do solo, inclinação da antena de 8°, potência efetivamente irradiada de 2512 W.

TABELA II
INTENSIDADE DE CAMPO ELÉTRICO EM FUNÇÃO DA
DISTÂNCIA X DA BASE DA ERB PARA DIVERSAS ALTURAS DA
ANTENA E UMA INCLINAÇÃO DA ANTENA DE 8°.

	Distância da torre	Diferentes alturas de instalação da Antena		
		h = 10 m	h = 20 m	h = 30 m
		Intensidade do Campo Elétrico em Volts por metro		
A	5m	0,077	0,000635	0,00004749
B	30m	1,615	0,705	0,207
C	50m	1,076	0,833	0,492
Pico	variável	1,850	0,847	0,550

- ◆ Um ponto situado a 50 metros de uma antena instalada a 10 metros de altura está sujeito a uma intensidade do campo elétrico superior ao pico do campo que pode ser produzido por antenas semelhantes instaladas a 20 ou 30 metros de altura.
- ◆ O valor máximo da intensidade do campo elétrico (pico de cada curva) é menor à medida que se aumenta a altura de instalação da antena:
 - para uma antena instalada a 10 metros de altura, com as demais especificações já apontadas [7], o pico é de 1,850 Volts por metro;
 - Caso essa mesma antena fosse instalada a 20 metros de altura, o pico seria reduzido a 0,847 Volts por metro;
 - Caso a antena fosse instalada a 30 metros de altura, o pico passaria a ser de 0,550 Volts por metro.
- ◆ Os valores máximos (o pico de cada curva), a medida em que a altura de instalação da antena é maior, são atingidos em pontos mais distantes da torre:
 - para a antena instalada a 10 metros de altura, o ponto a 1,5m de altura do solo (ponto P, de análise) em que se observa a maior intensidade do campo elétrico (pico) está a 20 metros de distância da torre;
 - para a mesma antena instalada a 20 metros de altura, o pico passa a ser registrado a 44 metros de distância da torre;
 - para a antena instalada a 30 metros de altura, o pico está a 68 metros de distância da torre.
- ◆ Tomando-se em consideração uma distância específica da torre de instalação da antena (de 50m, por exemplo), constata-se que a intensidade do campo elétrico não só tem valores diferentes como está em momentos distintos da curva:
 - para a antena instalada a 10 metros de altura, a intensidade do campo elétrico a 50 metros da antena é de 1,076 V/m e está em momento de curva descendente acentuada;

⁷ ponto de observação (ponto “P”) distante 1,5 m do nível do solo, inclinação da antena de 8°, potência efetivamente irradiada de 2512 W.

- para a antena instalada a 20 metros de altura, a intensidade do campo elétrico a 50 metros da antena é de 0,833 V/m e está praticamente no pico da curva, que é apenas 1,7% superior a ele (0,847 V/m) e foi atingido seis metros antes (44 m);
- para a antena instalada a 30 metros de altura, a intensidade do campo elétrico a 50 metros da antena é de 0,492 V/m e ainda não atingiu o pico, que se verificará a 18 metros adiante (68 m).

B. Variação da intensidade do campo elétrico em função da variação do ângulo de inclinação da antena

O próximo gráfico (Figura 6) mostra o comportamento da intensidade do campo elétrico em função da distância x da base da ERB para três diferentes ângulos de inclinação da antena, quais sejam, 8°, 4° e 2°, mantendo-se as demais variáveis constantes [8]. Em seguida, a Tabela III aponta os valores da intensidade do campo elétrico para as distâncias $x = 5$ m, 30 m e 50 m da torre da antena, bem como para o pico da intensidade do campo elétrico correspondente a cada uma das alturas de instalação das antenas.

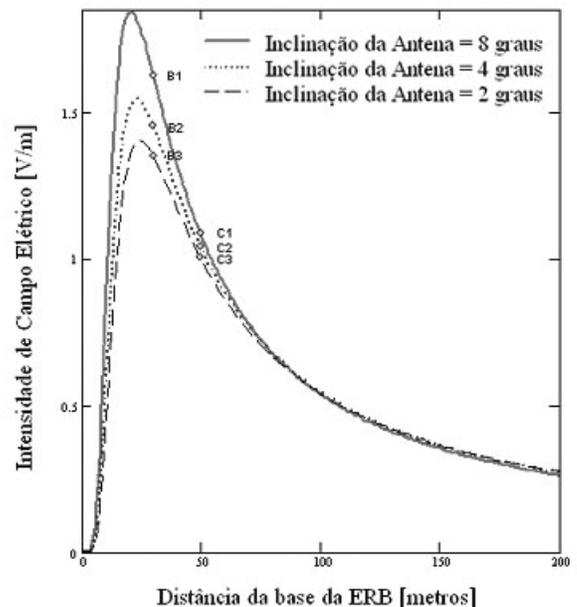


Figura 6 - Intensidade de campo elétrico em função da distância x da base da ERB considerando três diferentes ângulos de inclinação da antena.

Aos dados mostrados, acrescenta-se que ainda que na representação gráfica isso não transpareça de modo tão claro, também há variação do pico de cada uma das curvas (ponto em que há maior intensidade do campo elétrico proporcionado pela antena) quanto à distância em que o mesmo se verifica. Assim, os valores máximos para as inclinações de 8°, 4° e 2° são verificados respectivamente à distância de $x = 20$ m, 23 m e 24 m.

⁸ ponto de observação (ponto “P”) distante 1,5 m do nível do solo, altura da antena de 10m, potência efetivamente irradiada de 2512 W.

Novamente se pode constatar que à medida que a inclinação é menor, menores são os picos das curvas, o que significa que são menores as intensidades dos campos elétricos ao nível do solo ou, como nos dados apurados, a 1,5 metros de altura do solo e tais picos são atingidos a distâncias mais distantes da torre.

TABELA III
INTENSIDADE DE CAMPO ELÉTRICO EM FUNÇÃO DA DISTÂNCIA X DA BASE DA ERB PARA UMA ALTURA DA ANTENA DE 10 M E DIVERSAS INCLINAÇÕES DA ANTENA.

	Distância da torre	Diferentes ângulos de inclinação da Antena		
		8°	4°	2°
		Intensidade do Campo Elétrico em Volts por metro		
A	5m	0,077	0,033	0,021
B	30m	1,615	1,448	1,349
C	50m	1,076	1,034	0,997
Pico	variável	1,850	1,545	1,404

C. *Variação da intensidade do campo elétrico em função da variação da potência efetivamente irradiada da antena*

O gráfico apresentado a seguir (Figura 7) mostra o comportamento da intensidade do campo elétrico em função da distância x da base da ERB para três diferentes potências efetivamente irradiadas da antena, quais sejam, 2512 W, 1256 W e 754 W, mantendo-se as demais variáveis constantes ^[9]. Em seguida, a Tabela IV aponta as intensidades dos valores do campo elétrico para as distâncias x = 5 m, 30 m e 50 m da torre da antena, bem como para o pico do campo em cada uma das torres.

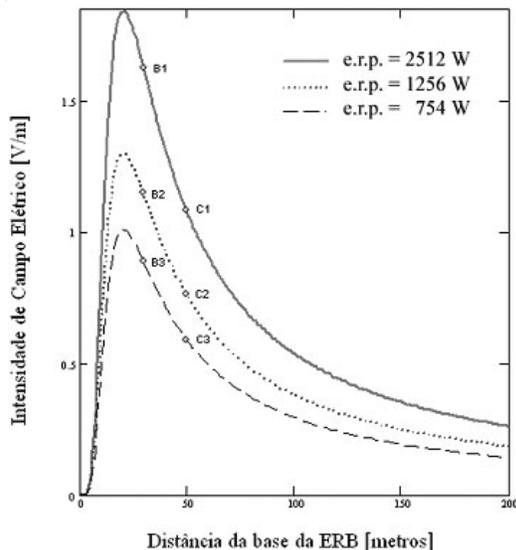


Figura 7 - Intensidade de campo elétrico em função da distância x da base da ERB considerando três diferentes potências efetivamente irradiadas da antena.

⁹ ponto de observação (ponto “P”) distante 1,5 m do nível do solo, altura da antena de 10m, ângulo de inclinação de 8°

TABELA IV
INTENSIDADE DE CAMPO ELÉTRICO EM FUNÇÃO DA DISTÂNCIA X DA BASE DA ERB PARA UMA ALTURA DA ANTENA DE 10 M E DIVERSAS POTÊNCIAS DE ENTRADA.

	Distância da torre	Diferentes potências efetivamente irradiadas		
		2512 W	1256 W	754 W
		Intensidade do Campo Elétrico em Volts por metro		
A	5m	0,077	0,018	0,014
B	30m	1,615	1,142	0,885
C	50m	1,076	0,761	0,589
Pico	20m	1,850	1,303	1,009

Nota-se, a partir da Figura 7 que o pico das curvas, representando o máximo de intensidade do campo elétrico, ocorre sempre a uma mesma distância da base da ERB independente da potência da antena.

- As três curvas têm esse pico a 20 m de distância da torre.

Conclui-se, também, que quanto menor é a potência da antena, menor é o valor da intensidade do campo elétrico a uma mesma distância da base da ERB:

- para uma antena com potência efetivamente irradiada de 2512 W o pico é de 1,850 Volts por metro;
- para uma antena com potência efetivamente irradiada de 1256 W o pico é de 1,303 Volts por metro;
- para uma antena com potência efetivamente irradiada de 754 W o pico é de 1,009 Volts por metro.

IV. CONCLUSÕES

As propriedades direcionais de radiação de antenas rádio-base foram apresentadas e discutidas em detalhe neste trabalho. Foram consideradas diversas variáveis além da usual distância à base da torre da estação rádio-base (ERB). Foi demonstrado que o máximo da intensidade do campo elétrico ocorre a distâncias diferentes da base da ERB dependendo destas outras variáveis, levando à conclusão que se fixar uma distância mínima de regiões povoadas para a instalação de ERBs não é a maneira adequada para se assegurar a segurança e o bem-estar da população. O re-posicionamento de uma ERB para uma distância de 50 m, por exemplo, pode na realidade aumentar a intensidade do campo elétrico em mais de mil vezes na área que se deseja proteger. Isto demonstra a arbitrariedade e a falta de consistência técnica na imposição de distâncias mínimas de proteção. O critério científico correto e seguro é o de se limitar o valor da intensidade do campo elétrico máximo resultante de todas as antenas em operação na faixa de frequências considerada [3].

REFERÊNCIAS

- [1] GSM World e Wireless Intelligence (<http://www.gsmworld.com> e <https://www.wirelessintelligence.com>) obtido em 18 de Janeiro de 2007.
- [2] Lei nº. 3.446, de 23 de setembro de 2004, do Distrito Federal, publicada no Diário Oficial do Distrito Federal de 7 de outubro de 2004.
- [3] “Regulamento sobre limitação da exposição a campos elétricos, magnéticos e eletromagnéticos na faixa de radiofrequências entre 9 kHz e 300 GHz” aprovado pela Resolução Anatel nº. 303, de 2 de julho de 2002.
- [4] L.C. Godara (ed.), *Handbook of Antennas in Wireless Communications*, CRC Press, 2002.
- [5] W.L. Stutzman e G.A. Thiele, *Antenna Theory and Design*, 2ª Edição, John Wiley & Sons, 1998.
- [6] M.A.B. Terada, A.J.M. Soares, F.C. Silva, e S.B.A. Fonseca, “GRADMAX: Um Programa Para Análise e Otimização de Antenas de Fios”, *TELEMO92 (Software Educacional): X Simpósio Brasileiro de Telecomunicações e V Simpósio Brasileiro de Microondas*, Brasília - DF, pp. 15-18, Julho de 1992.
- [7] M.A.B. Terada, A.J.M. Soares e S.B.A. Fonseca, “Análise de Estruturas de Linhas de Transmissão Através do Método dos Elementos Finitos”, *IV Simpósio Brasileiro de Microondas: SBMO - EESC/USP*, São Carlos – SP, pp. 199-204, Julho de 1990.
- [8] Resolução nº. 315 da ANATEL, de 19 de Setembro de 2002.



Marco Antonio Brasil Terada nasceu em São Paulo, SP, em Novembro de 1966. Recebeu os títulos de Engenheiro Eletricista e Mestre em Engenharia Elétrica pelo Departamento de Engenharia Elétrica da Universidade de Brasília (UnB) em 1989 e 1991, respectivamente. Em 1995 recebeu o título de Ph.D em Engenharia Elétrica pela Universidade Virginia Tech - EUA.

Desde Agosto de 1995 é professor da UnB, onde atua nas áreas de Antenas e Sistemas de Comunicações Sem-Fio. Foi Engenheiro Sênior e

Gerente de Antenas do programa INTELSAT IX (sete satélites operando nas bandas C e Ku) na empresa INTELSAT (1998 a 2001), Washington – EUA, e professor da Universidade Estadual do Novo México – EUA. (2001 a 2004). Seus interesses correntes incluem a pesquisa e desenvolvimento de antenas e sistemas de comunicações com alta taxa de rejeição sistêmica (isolamento). É bolsista do CNPq em desenvolvimento tecnológico e extensão inovadora.

Projeto, Simulação e Implementação em Hardware de um Sistema de Sincronismo para o MI-SBTVD

Marcelo A. C. Fernandes, Cristiano Panazio e Sandro A. Fasolo

Abstract—In 2005 the Brazilian government supported many research consortia in order to develop a Digital Television System employing advanced technologies for multimedia broadcasting. One of the proposals for the physical layer of this system was entitled Innovative Modulation for the Brazilian Digital TV System (MI-SBTVD). The MI-SBTVD Project includes high performance error-correcting codes, transmit spatial diversity and multi-carrier modulation. The objective of this article is to present the synchronization techniques that were used in the first proof-of-concept for the MI-SBTVD System. We present the simulated performance results for both the AWGN and multipath channels. Finally, we also comment on the implementation details of the synchronization scheme using Field Programmable Gate Array (FPGA) devices.

Index Terms— OFDM Synchronization, Digital Television, SBTVD.

Resumo—Em 2005 o governo brasileiro financiou um grupo de consórcios formado por vários pesquisadores para desenvolver um Sistema de Televisão Digital incorporando o estado da arte das tecnologias atuais. Uma das propostas para a camada física deste sistema foi chamada de Modulação Inovadora para o Sistema Brasileiro de TV Digital (MI-SBTVD). O projeto MI-SBTVD utiliza uma codificação de canal de alto desempenho, diversidade espacial e modulação com múltiplas portadoras. O objetivo deste artigo é apresentar a técnica de sincronização utilizada na primeira prova de conceito do protótipo do MI-SBTVD. O desempenho do receptor empregando a técnica de sincronização proposta é avaliada, tanto em canais AWGN como em canais com múltiplos percursos. Finalmente, são apresentados os detalhes da implementação do protótipo em FPGA - *Field Programmable Gate Array*.

Palavras chave— Sincronização, OFDM, SBTVD.

I. INTRODUÇÃO

Desde os anos 80, a transição do sistema analógico de TV para o digital tem sido discutida e implementada em diferentes países. Basicamente, há três padrões que disputam o panorama

internacional: o americano ATSC [2], o europeu DVB-T [3], e o padrão japonês ISDB-T [4]. Estes três sistemas fazem uso da mesma tecnologia de compressão de vídeo, o MPEG-2. Por outro lado, eles diferem com respeito às técnicas de modulação e codificação de canal.

O padrão ATSC é baseado em uma única portadora com modulação 8-VSB (*Vestigial Side Band* com 8 níveis)[5], enquanto o DVB-T e ISDB-T são baseados na técnica de transmissão OFDM (*Orthogonal Frequency Division Multiplexing*) [6] com vários níveis de modulação QAM. No ano de 2005 o governo brasileiro contratou vinte consórcios, compostos por um grande número de pesquisadores, para dar início ao desenvolvimento do Sistema Brasileiro de Televisão Digital (SBTVD). O desenvolvimento deveria incorporar o estado da arte das técnicas de transmissão para sinais digitais, com o objetivo de prover melhorias consideráveis frente aos padrões mais antigos.

O Consórcio MI-SBTVD foi composto pelo Instituto Nacional de Telecomunicações (INATEL), a Universidade Estadual de Campinas (UNICAMP), a Universidade Federal de Santa Catarina (UFSC), a Universidade Federal do Paraná (UTFPR) e a empresa Linear Equipamentos Eletrônicos. O MI-SBTVD utiliza código LDPC para codificação de canal e sistema OFDM combinado com técnicas de codificação espaço temporal (STC) [7] no processo de modulação. O sistema STC-OFDM [8] com diversidade espacial na transmissão é uma técnica muito eficiente para situações onde é necessária uma alta taxa de transmissão em canais móveis e em ambientes fortemente seletivos no tempo e na frequência.

Para garantir uma demodulação correta no receptor, o sistema precisa compensar os atrasos de propagação, o efeito Doppler e as variações na frequência do oscilador local. Estes objetivos só podem ser alcançados pelo uso de uma técnica de sincronização corretamente projetada, dado que o sincronismo representa um papel muito importante em sistemas OFDM [9] [10]. O início do tempo do símbolo OFDM e a frequência do oscilador local devem ser estimados pelo circuito de sincronização para que o processo de demodulação tenha sucesso. Pois somente após a sincronização é que o restante dos processos - FFT, estimação da resposta em frequência do canal, decodificações e detecção - pode ser efetivado. Os diagramas em blocos do transmissor e do receptor do sistema STC-OFDM proposto são mostrados nas Figuras 1 e 2,

Manuscrito recebido em 15 de maio de 2005; revisado em 18 de novembro de 2005.

Marcelo A. C. Fernandes é aluno de doutorado no Departamento de Comunicações da Universidade Estadual de Campinas - Unicamp, Campinas, SP (maugusto@decom.fee.unicamp.br). Cristiano Panazio é Professor Doutor da Escola Politécnica da Universidade de São Paulo - USP, São Paulo (cpanazio@lcs.poli.usp.br). Sandro Adriano Fasolo é Professor Adjunto do Instituto Nacional de Telecomunicações - Inatel, Santa Rita do Sapucaí, MG (sandro.fasolo@inatel.br). Este trabalho foi financiado pela FINEP - Financiadora de Estudos e Projetos no âmbito do Sistema Brasileiro de TV Digital - SBTVD.

respectivamente. Um fato importante a enfatizar é que as técnicas de sincronização convencionais para sistema com uma única antena de transmissão não são aplicadas diretamente ao STC-OFDM. Características particulares da modulação STC-OFDM devem ser levadas em conta no projeto do sistema de sincronização.

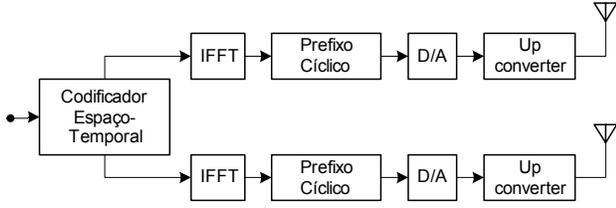


Fig. 1. Diagrama de blocos do Transmissor STC-OFDM.

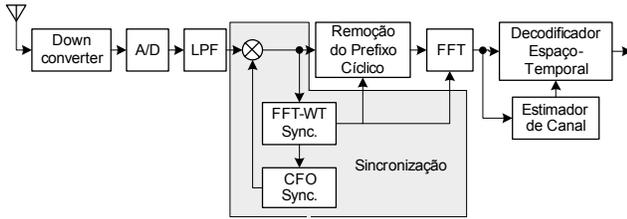


Fig. 2. Diagrama de blocos do Receptor STC-OFDM.

O principal objetivo deste artigo é apresentar os detalhes do esquema de sincronização proposto no receptor de TV digital para o sistema MI-SBTVD. Os aspectos teóricos e práticos de implementação em hardware, alguns resultados interessantes para canais AWGN e com múltiplos percursos também são apresentados e discutidos.

II. SINCRONIZAÇÃO PARA RECEPTORES OFDM COM CÓDIGOS ESPAÇO-TEMPORAIS

O sincronismo é um das partes mais importantes de um receptor para OFDM. É um processo indispensável que é usado para compensar a presença de imperfeições de sistema, tais como *desvios* do oscilador local, efeito Doppler e atrasos na propagação do sinal.

Chamando de f_{erro} o *desvio* de frequência entre o transmissor e o receptor, θ o *desvio* de fase e τ_{erro} o *desvio* de tempo, pode-se então caracterizar o sinal recebido $r(t)$ como:

$$r(t) = s(t - \tau) e^{j2\pi f_{erro} t + \theta} + n(t), \quad (1)$$

onde $s(t)$ é o sinal transmitido e $n(t)$ é um ruído branco gaussiano aditivo. Neste artigo não será levada em conta a sincronização de fase, dado que isto pode ser facilmente realizado pelo estimador de canal sem qualquer perda de desempenho. Então, daqui em diante não será considerada a fase θ nas equações.

Como mostrado em [9] e [10], é possível observar que o

desvio de frequência entre a frequência do sinal recebido e do oscilador local no receptor pode destruir a ortogonalidade entre as subportadoras, causando ICI (*Inter-Carrier Interference*) no sinal recebido. Se o *desvio* de tempo não é compensado no caso em que é maior que o prefixo cíclico, T_{CP} , então ICI e ISI (*Interferência entre Símbolos OFDM*) aparecem juntas [9] [10].

As técnicas de sincronização podem ser divididas em duas categorias, quais sejam, as técnicas cegas, que não utilizam um sinal piloto, chamadas na literatura de *non-pilot aided* [11]-[13] e as técnicas supervisionadas. Nestas últimas o receptor tem um conhecimento *a priori* de algumas informações dos sinais pilotos, que podem ser multiplexadas no tempo ou em frequência [14]-[16]. No caso específico das técnicas cegas, alguns algoritmos exploram propriedades de correlação induzidas pela inserção de prefixo cíclico.

Como apresentado na Figura 1 e de acordo com [9] e [10], o processo de sincronismo pode ser dividido em duas partes, quais sejam, *Carrier Frequency Offset* (CFO), que corrige f_{erro} , e a *FFT (Fast Fourier Transform) Window Timing* (FFT-WT) que minimiza τ_{erro} e obtém f_{erro} .

A sincronização de tempo e frequência pode ser feita em duas fases, uma chamada de sincronização grosseria e outra chamada de sincronização fina. Enquanto a sincronização grosseria é feita no domínio do tempo, a fina é realizada no domínio da frequência [10]. A sincronização fina permite obter uma estimação mais precisa, i.e., mais robusta aos efeitos do ruído. Contudo, neste trabalho não foram utilizadas as técnicas de sincronização fina, o que significa que os resultados aqui obtidos podem ser ainda melhores.

A. Perdas Causadas pelos Erros de Sincronização

Valores elevados de f_{erro} aumentam a ICI e valores elevados de τ_{erro} aumentam a ICI e ISI, resultando assim em uma degradação do desempenho. De acordo com [10] é possível descrever a degradação do desempenho em canais AWGN em termos da SNR (Relação Sinal/Ruído), como uma função de f_{erro} e τ_{erro} que é determinada pelas seguintes equações:

$$SNR_m \approx \frac{E_S(m) \text{sinc}^2(f_{erro} T_S)}{N_0 + E_S(m) + \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \text{sinc}^2(n - m + f_{erro} T_S)} \quad (2)$$

$$SNR_m \approx \frac{\left(1 - \frac{\tau_{erro}}{T_S}\right)^2}{\frac{N_0}{E_S(m)} + \frac{\tau_{erro}^2}{T_S^2} + \frac{2\tau_{erro}^2}{T_S^2} + \sum_{n=0}^{N-1} \text{sinc}^2\left(\frac{(n-m)\tau_{erro}}{T_S}\right)} \quad (3)$$

A Equação (2) ilustra a degradação da SNR na m -ésima portadora causada pelo *desvio* de frequência e a Equação (3) apresenta a degradação da SNR na m -ésima portadora quando a ISI está presente na janela da FFT causada pelo *desvio* de tempo. Nestas equações, T_S representa o período de um símbolo OFDM, N representa o número total de portadoras, $E_S(m)$ é a energia média da m -ésima portadora e N_0 representa a densidade espectral de potência do ruído AWGN no canal. Como pode ser analisado a partir das Equações (2) e (3), as degradações causadas por f_{erro} e τ_{erro} são consideráveis. O importante é ressaltar que quando $|\tau_{erro}| \leq T_{CP}$, o erro de tempo resulta apenas em uma distorção de fase que pode ser facilmente corrigida pelo estimador de canal.

III. PARÂMETROS DO PROCESSO DE SINCRONIZAÇÃO PARA A IMPLEMENTAÇÃO

Para implementar a sincronização de tempo e correção do CFO, foi escolhida uma técnica de máxima verossimilhança apresentada em [13]. Esta técnica permite estimar de forma conjunta o desvio de tempo e de frequência por meio da propriedade de correlação do prefixo cíclico, como mostrado na Figura 3.

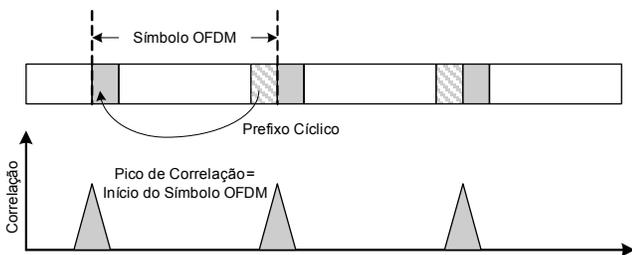


Fig 3. Técnica de Correlação do Prefixo Cíclico.

As técnicas clássicas de sincronização usam o critério de máxima verossimilhança para calcular f_{erro} e τ_{erro} , como apresentado em [10] e [13]. Em [13], o *desvio* de frequência estimado, \hat{f}_{erro} , e o *desvio* de tempo estimado $\hat{\tau}_{erro}$ são obtidos pelas seguintes equações:

$$\gamma(n) = \sum_{k=n}^{n+M-1} v(k)v^*(k+N), \quad (4)$$

$$\phi(n) = \sum_{k=n}^{n+M-1} |v(k)|^2 + |v(k+N)|^2, \quad (5)$$

$$\hat{f}_{erro} = -\frac{1}{2\pi} \angle \gamma(\hat{\tau}_{erro}), \quad (6)$$

$$\hat{\tau}_{erro} = \arg\left(\max\left\{|\gamma(\tau_{erro})|^2 - \rho\phi(\tau_{erro})\right\}\right), \quad (7)$$

onde \angle representa o argumento de um número complexo, ρ é uma constante que depende da SNR e representa a magnitude da correlação entre as amostras $v(k)$ e $v(k+N)$, N é a duração da FFT e M é o tamanho do prefixo cíclico.

A escolha da técnica de sincronização deste projeto tem como base um compromisso que envolve a complexidade de implementação, robustez e o desempenho do sistema. Tomando como base que o sistema proposto, o MI-SBTVTD, não transmite uma seqüência de treinamento, uma técnica cega é obrigatória para a sincronização de tempo e CFO. Os algoritmos implementados, com base nas Equações (4), (5), (6) e (7), são mostrados nas Figuras 4 e 5.

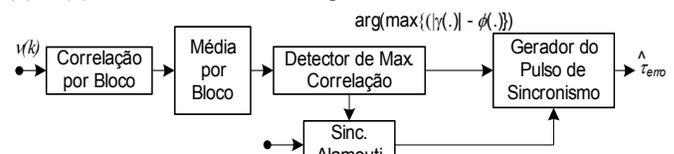


Fig. 4. Esquema de Sincronismo da FFT-WT.

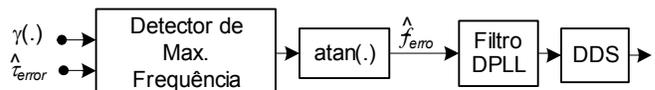


Fig. 5. Esquema de Sincronismo do CFO.

Em termos de complexidade, esta técnica tem um baixo custo computacional devido à utilização de somas móveis para a estimativa dos valores das correlações necessárias. Em canais com múltiplos percursos, a perda de desempenho parece ser desprezível. Em [13], para um canal com um perfil de decaimento exponencial com um comprimento de 15 amostras, a perda de desempenho situa-se ao redor de 1,2dB para uma SNR de 25dB.

Após a aquisição, quando os símbolos pilotos no domínio da frequência e a estimação de canal já estão disponíveis, pode-se então adotar técnicas supervisionadas (por exemplo, [17]) para refinar a sincronização de tempo e CFO. Porém, tal aproximação aumentaria a complexidade do sistema, dado que seriam incluídos mais algoritmos para a sincronização e novos circuitos para controle.

A. Detalhes da Implementação

Devido a algumas características de operação do sistema proposto, pode-se reduzir a complexidade da técnica apresentada em [13]. Em primeiro lugar, pode-se notar que o

coeficiente de correlação ρ é uma função da SNR. Conseqüentemente, para seu cálculo é necessário calcular a SNR e fazer uma divisão que não é uma tarefa computacional simples. Porém, o valor de ρ é quase igual a 1 para valores pequenos de SNR (por exemplo, para 12dB $\rho = 0,94$). Assim, seu impacto no desempenho pode ser desprezível.

A segunda simplificação é feita no cálculo do $\hat{\tau}_{erro}$. Dado que o valor da correlação dado por $\gamma(n)$ é um número complexo, então o uso de uma função de módulo como definido em (7) requeria uma função de raiz-quadrada. Assim, para reduzir a complexidade, esta operação foi aproximada pelo módulo da parte real mais o módulo da parte imaginária, como mostrada na Figura 6. No pior caso para tal aproximação, o erro relativo não é maior que $\sqrt{2}$. Contudo, ao se implementar a correção do CFO em malha fechada, temos $\gamma(n)$ tende para um número real quando o CFO é corretamente estimado. Deste modo, a parte imaginária tende a zero, fazendo com que o erro resultante da aproximação tenda também a zero.

O valor de correlação dado por $\gamma(n)$ é igual a uma amostra do símbolo de OFDM, que é igual a $T = 63/512\mu s$. Qualquer valor para o *desvio* de tempo menor que o período de uma amostra pode ser compensado pelo estimador de canal. O processo de correção no CFO é ilustrado na Figura 6.

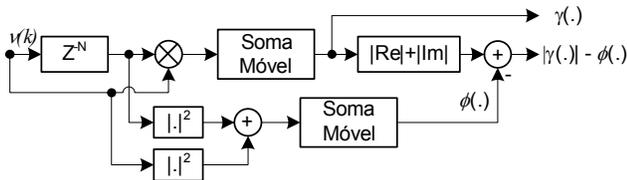


Fig. 6. Proposta da Técnica de Correlação.

Com o objetivo de melhorar a robustez do valor máximo estimado em (7), foi implementado um circuito de média por blocos em quatro sucessivos símbolos OFDM, como apresentado na Figura 4. Os resultados simulados para esta técnica são mostrados na Figura 8.

Neste projeto também foi desenvolvido um esquema de sincronização levando em consideração a atualização das palavras códigos do esquema STC de Alamouti. Isto deve ser feito porque a estimação de canal é executada a uma taxa especificada pelo esquema de Alamouti, e qualquer atualização dentro da palavra código do Alamouti implica em uma quebra de ortogonalidade do esquema de codificação.

A frequência estimada pelo CFO, \hat{f}_{erro} , é alimentada em um filtro, que é composto de um integrador digital escalonado. A saída é inserida então em um DDS (*Direct Digital Synthesizer*) que possui um funcionamento semelhante a um VCO (*Voltage Controlled Oscillator*). O DDS gera os sinais de co-seno e seno que são aplicados aos componentes em fase e quadratura do sinal recebido, corrigindo assim o CFO. A malha composta da estimação do CFO, filtro, DDS e Correção de CFO é chamada de *Digital Phase Locked Loop* (DPLL).

Nos tópicos seguintes serão apresentadas algumas curvas de desempenho do esquema de sincronização proposto. Os resultados foram obtidos para alguns dos canais usados nos testes para TV digital realizados no Brasil [1] (Tabela I).

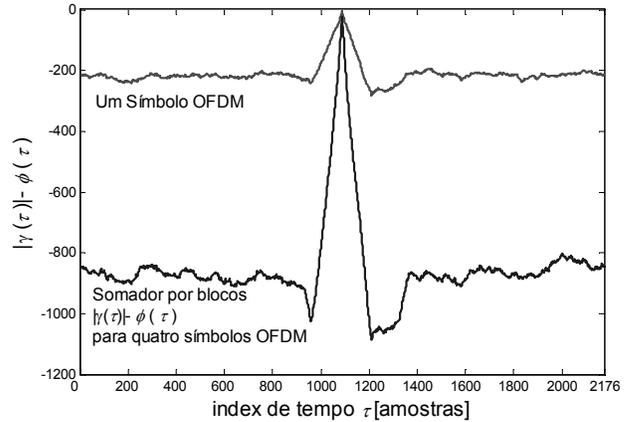


Fig. 7. Índice de tempo detectado para um símbolo OFDM, utilizando o somador de blocos com quatro símbolos OFDM.

TABELA I
PERFIL DE POTÊNCIA DOS CANAIS SIMULADOS.

Nome	Desc.	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
Brasil	Atraso (μs)	0	0,15	2,22	3,05	5,86	5,93
A	Atenuação (dB)	0	13,8	16,2	14,9	13,6	16,4
Brasil	Atraso (μs)	0	0,3	3,5	4,4	9,5	12,7
B	Atenuação (dB)	0	12	4	7	15	22

IV. TESTES DE DESEMPENHO

Nas Figuras 8 e 9 são apresentadas as variâncias normalizadas do tempo e da frequência estimados em função da SNR. Foram feitas simulações para vários canais com múltiplos percursos. Foi utilizado um período da amostra, T , para a normalização da variância do tempo e utilizada a banda da sub-portadora (3968Hz) para a normalização da variância da frequência.

Os resultados mostram um pouco de degradação de desempenho para canais com grande dispersão temporais, em particular para o canal B que possui múltiplos percursos severos, apresentando o pior desempenho. Todavia, a variância do tempo estimado é menor que o período T e a variância da frequência é menor que a banda da sub-portadora. Assim, pode-se concluir que este resultado não deverá degradar consideravelmente a taxa de erro de bit.

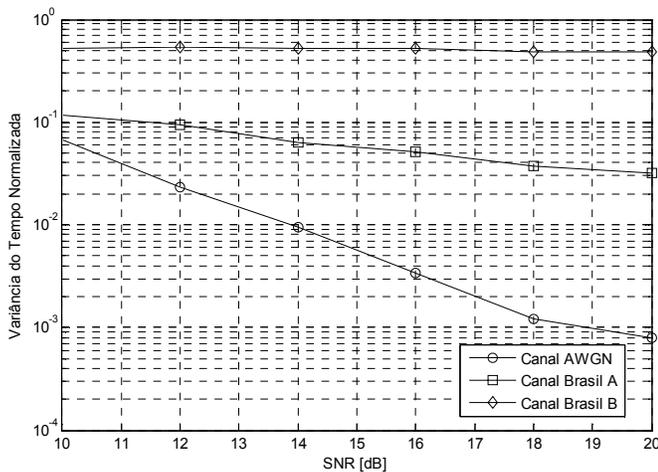


Fig. 8. Variância do tempo estimado.

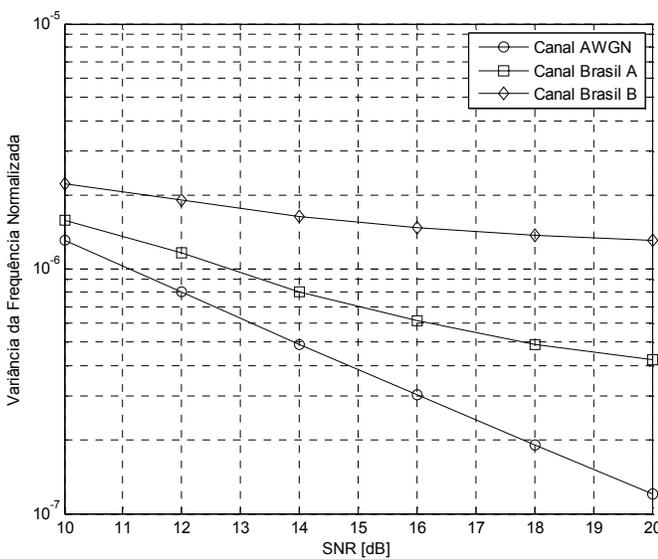


Fig. 9. Variância da frequência estimada.

V. ASPECTOS DA IMPLEMENTAÇÃO EM FPGA

O sistema proposto foi implementado em uma placa de desenvolvimento, Nallatech toolkit [18], que é composta de uma placa com dois conversores A/D e dois conversores D/A, várias interfaces digitais, e uma FPGA Xilinx Virtex 4 SX35 [19].

O protótipo desenvolvido é formado de um estágio de FI (Frequência Intermediária) com um *Digital Down-converter* (DDC), um estágio de sincronização de tempo e frequência, estimador de canal, um estágio de demodulação OFDM e um decodificador espaço temporal utilizando a técnica de Alamouti. Para uma prototipação rápida foi utilizada a ferramenta Xilinx de DSP *Design Flow* que auxilia no desenvolvimento de sistemas complexos através de Matlab/Simulink.

O resumo de percentuais de utilização do protótipo, em termos de hardware, é apresentado na Tabela II. A maioria dos dispositivos de BRAMs da FPGA, apresentadas na Tabela II (58 de 94 BRAMs), é usada pelos algoritmos de

sincronização, enquanto a maioria dos elementos de lógica (11.132 de 13.063 *slices*) é usada pelo restante do protótipo.

TABELA II
RECURSOS DE FPGA UTILIZADOS PELO PROTÓTIPO DESENVOLVIDO.

	Usado	Disponível	Utilização
Número de <i>slices</i> ocupados	13063	15360	85%
Número de BRAMs	94	192	48%
Número de DSP48s	47	192	24%

O algoritmo de sincronização é responsável pela utilização de um grande número de BRAMs [18], devido aos algoritmos de correlação entre *N* amostras separadamente e a média por blocos feita a cada quatro símbolos OFDM. Esta estratégia de implementação de hardware foi bastante interessante devido ao grande número de BRAMs na FPGA utilizada. Se uma grande quantidade de BRAMs não estivesse disponível, um número grande de *slices* (medida de dispositivos lógicos para as FPGAs da Xilinx) seria consumido para executar funcionalidades de memória. Este tipo de abordagem não seria possível devido ao grande número de *slices* consumidos pelo protótipo, algo em torno de 73% do total da FPGA.

A primeira prova de conceito para o MI-SBTVD foi implementada em janeiro de 2006. Na transmissão um módulo de RF foi anexado ao *front-end* do transmissor utilizando misturadores, amplificadores de potência e osciladores para transmitir um sinal de DTV na portadora de 512,8137MHz. No caso da recepção, o módulo de RF, desenvolvido pela empresa Linear, foi anexado à parte de FI do protótipo. O sinal analógico da saída do módulo de RF, centrado em 8,126984MHz, é amostrado pelo *down-converter* do protótipo a uma taxa de amostragem quatro vezes superior. A referência de relógio externa para o protótipo foi de 65,015873MHz. Nesta primeira prova de conceito o esquema de sincronismo implementado teve um desempenho satisfatório.

VI. CONCLUSÕES

Este artigo apresentou o esquema de sincronismo projetado, simulado e implementado para o MI-SBTVD. Foram mostrados os aspectos de projeto, resultados de simulação e detalhes de implementação em hardware usando FPGA. Melhorias adicionais estão sendo investigadas com o objetivo de refinar os parâmetros calculados e melhorar as condições de quantização das variáveis no FPGA. Além disso, outras técnicas estão sendo estudadas para melhorar a robustez do esquema de sincronização desenvolvido.

REFERÊNCIAS

[1] José M. C. Brito *et al.*, "Design, Simulation and Hardware Implementation of a Digital Television System: System Overview", The 9th International Symposium on Spread Spectrum Techniques and Applications - ISSSTA-2006, Manaus, Brazil, 2006.
 [2] ATSC Standard A/53, ATSC Digital Television Standard.

- [3] European Broadcasting Union, "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television", ETSI EN 300 744 V1.4.1, 2001.
- [4] International Telecommunication Union, "Channel Coding, Frame Structure and Modulation Scheme for Terrestrial Integrated Services Digital Broadcasting (ISDB-T)", ITU-R 205/11, 1999.
- [5] Fernando C. C. Castro, Maria C. F. Castro, Marcelo A. C. Fernandes and Dalton S. Arantes, "8-VSB Channel Coding Analysis for DTV Broadcast," in IEEE Transactions on Consumer Electronics, vol. 46, 3, pp. 539-547, August, 2000.
- [6] K. Welling and M. Rice, "Coded Orthogonal Frequency Division Multiplexing for the Multipath Fading Channel," In Proceedings of the International Telemetering Conference, Las Vegas, NV, October 1999.
- [7] S. Alamouti, "A simple transmit diversity technique for wireless communications", IEEE J. Select. Areas Comm., vol 16, no 8, pp. 1451-1458, 1998.
- [8] K. F. Lee and D. B. Williams, "A Space-Time Coded Transmit Diversity Technique for Frequency Selective Fading Channels," IEEE Sensor Array and Multichannel Signal Processing Workshop, pp. 149-152 Cambridge, MA, March 2000.
- [9] S. Hara and R. Prasad, Multicarrier Techniques for 4G Mobile Communications, Artech House, 2003.
- [10] K. Fazel and S. Kaiser, Multi-carrier and Spread Spectrum Systems, Wiley, 2003.
- [11] Daffara F. and Chouly A., "Maximum likelihood frequency detectors for orthogonal multi-carrier systems", in Proc. IEEE International Conference on Communications (ICC'93), Geneva, Switzerland, pp. 766-771, May 1993.
- [12] Daffara F. and Adami O., "A new frequency detector for orthogonal multi-carrier transmission techniques," in Proc. IEEE Vehicular Technology Conference, Chicago, USA, pp. 804-809, July 1995.
- [13] J. J. van de Beek, M. Sandell, P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," in IEEE Transactions on Signal Processing, vol. 45, 7, pp. 1800-1805, July, 1997.
- [14] Fazel K., Kaiser S. and Robertson P., "OFDM: A key component for terrestrial broadcasting and cellular mobile radio," in Proc. International Conference on Telecommunication (ICT'96), Istanbul, Turkey, pp. 576- 583, April 1996.
- [15] Fazel K., Kaiser S., Robertson P. and Ruf M.J., "A concept of digital terrestrial television broadcasting," Wireless Personal Communications, vol. 2, nos. 1 & 2, pp. 9-27, 1995.
- [16] Moose P.H., "A technique for orthogonal frequency division multiplexing frequency offset correction", Transactions on Communications, vol. 42, pp. 2908-2914, Oct. 1994.
- [17] Baoguo Yang, K. B. Letaief, R. S. Cheng, Zhigang Cao, "Timing recovery for OFDM transmission," in IEEE Journal on Selected Areas in Communications, vol. 18, 11, pp. 2278-2291, Nov., 2000.
- [18] Nallatech, Inc., "Virtex-4, Dual 105MSPS ADC, Dual 160MSPS DAC DIME-II Module User Guide," Reference Manual, 2005.
- [19] Xilinx Inc., "Virtex-4 User Guide," Reference Manual, March, 2006.

interesses incluem filtragem adaptativa, recuperação de sincronismo, técnicas multiportadora e receptores espaço-temporais.

Sandro Adriano Fasolo graduou-se em Licenciatura em Física pela Pontifícia Universidade Católica do Rio Grande do Sul (1989), graduou-se em Engenharia Elétrica pela Universidade Federal de Santa Maria (1993), obteve o mestrado em Engenharia Elétrica pela Universidade Estadual de Campinas (1996) e o doutorado em Engenharia Elétrica também pela Unicamp (2001). Atualmente é Professor Adjunto da Fundação Instituto Nacional de Telecomunicações. Tem experiência em diversas áreas da Engenharia Elétrica, com ênfase em Telecomunicações. Seus principais temas de interesse são televisão digital, transmissão digital, equalização adaptativa, equalização cega, rede de frequência única e simulação de sistemas de comunicações.

Marcelo Augusto C. Fernandes graduou-se em Engenharia Elétrica pela Universidade Federal do Rio Grande do Norte, em 1997, e obteve o mestrado em Engenharia Elétrica também pela UFRN, em 1998. Atualmente é Professor do Centro Universitário Salesiano de São Paulo (Unisal) e Pesquisador do ComLab/UNICAMP, onde está em fase final de conclusão do doutorado no Departamento de Comunicações (DECOM) da Faculdade de Engenharia Elétrica da Unicamp (FEEC/UNICAMP). Tem experiência em diversas áreas da Engenharia Elétrica, com ênfase em Telecomunicações. Suas principais áreas de interesse são equalização adaptativa, antenas inteligentes, sincronização, TV-Digital, sistemas CDMA, sistemas 3G, desenvolvimento de sistemas em Matlab/Simulink e programação FPGA.

Cristiano Panazio é professor da Escola Politécnica da Universidade de São Paulo desde 2006. Recebeu seus títulos de bacharelado e mestrado em Engenharia Elétrica pela Universidade Estadual de Campinas (UNICAMP), em 1999 e 2001, respectivamente. Em 2005 obteve o título de doutor pelo Conservatoire National des Arts et Métiers (CNAM), Paris, França. Seus

Codificação LDPC em Sistemas de Televisão Digital

Tarciano F. Pegoraro, Fábio A. L. Gomes, Fábio Lumertz, Renato R. Lopes, Fabrício A. Oliveira, Roberto Gallo, Marcelo C. Paiva e José S. G. Panaro

Abstract—In this paper, we describe the design methodology and hardware implementation of a low-density parity-check (LDPC) code for a digital television (DTV) system. We begin the paper describing LDPC codes and the design strategies we used. We also provide some simulation results that show that the proposed code greatly outperforms codes used by other DTV standards. Finally, we provide details of the hardware implementation of the code.

Index Terms—LDPC codes, Digital Television, FPGA, SBTVD.

Resumo—Neste artigo descreve-se a metodologia de projeto e implementação em hardware de um código de matriz de paridade de baixa densidade (LDPC) aplicado a um sistema de televisão digital. Começamos o artigo descrevendo códigos LDPC e as estratégias de projeto empregadas. Também fornecemos alguns resultados de simulação que mostram que o código proposto fornece um desempenho significativamente superior ao empregado em outros padrões de televisão digital. Finalmente, fornecemos detalhes sobre a implementação em hardware do código proposto.

Palavras chave—Códigos LDPC, FPGA, SBTVD, Televisão Digital.

I. INTRODUÇÃO

Códigos de matriz de paridade de baixa densidade, ou códigos LDPC (do inglês, *low-density parity-check codes*), como são mais conhecidos, foram introduzidos por Gallager em 1962 [1] [2], quando o próprio Gallager também introduziu o seu algoritmo de decodificação, conhecido hoje como algoritmo Soma-Produto (SPA) ou *belief propagation*. Embora introduzido há mais de 3 décadas, o tema permaneceu praticamente estático, sem receber maior atenção da comunidade científica, até a sua “redescoberta” por Mackay [3], no fim dos anos 90. O grande impulso dessa redescoberta foi o progresso dos códigos turbo, outra família de códigos cuja decodificação se baseia em um algoritmo iterativo e que apresenta desempenho próximo ao limite de Shannon. Observou-se desde então que códigos LDPC eram capazes de atingir o mesmo nível de desempenho dos códigos turbo, com praticamente a mesma complexidade, porém com a vantagem de permitir um ajuste mais fino do compromisso entre complexidade e desempenho. Além disso, progressos mais recentes introduziram estruturas especiais de matrizes de paridade, como é o caso dos códigos eIRA (do inglês, *extended Irregular Repeat-Accumulate*) [9], utilizados no projeto aqui apresentado, que permitem reduzir não só a complexidade do processo de decodificação, mas também do processo de codificação.

Tarciano F. Pegoraro, Fábio A. L. Gomes, Fábio Lumertz, Renato R. Lopes, Fabrício A. Oliveira e Roberto Gallo estão vinculados à Universidade Estadual de Campinas - UNICAMP, Campinas, SP, Brasil, 13083-852. (tarciano, adriano, lumertz, rlopes, fabricio@decom.fee.unicamp.br), (gallo@ic.unicamp.br). Marcelo C. Paiva e José S. G. Panaro estão vinculados ao Instituto Nacional de Telecomunicações - INATEL, Santa Rita do Sapucaí, MG, Brasil, 37540-000.

Outra área onde vários progressos vêm sendo produzidos recentemente é a dos algoritmos de decodificação. A generalização do algoritmo SPA forma a classe dos algoritmos de passagem de mensagens, que, entre outros, inclui o algoritmo Min-Sum (ou Mínimo-Soma, em português). O Min-Sum pode ser visto como uma aproximação do SPA de menor complexidade e com desempenho um pouco pior, porém com grande utilidade em casos onde os recursos computacionais são mais restritos. Uma grande parcela das pesquisas recentes relacionadas a LDPC se concentra no tema dos algoritmos de passagem de mensagens, principalmente no desenvolvimento de versões de menor complexidade, utilizando mensagens compostas por um número reduzido de bits, e que possam ser implementados em hardware, na forma de ASICs (do inglês, *Application-Specific Integrated Circuit*) e FPGAs (do inglês, *Field-Programmable Gate Arrays*). Entre os aspectos que devem ser considerados pelo projetista estão as limitações de memória, área de silício e tempo de processamento.

Outro desafio do projeto de sistemas práticos usando códigos LDPC é a implementação do codificador. Várias propostas [5] [6] foram desenvolvidas com o intuito de reduzir sua complexidade, que a princípio é proporcional ao quadrado do comprimento do código, que pode chegar à ordem de dezenas de milhares de bits por bloco. Entre as propostas mais satisfatórias está o uso da classe estendida de códigos de repetição e acumulação irregulares (eIRA), que é usada neste projeto e é descrita na Seção II.

Neste artigo, detalhamos a implementação de um sistema prático de codificação e decodificação LDPC, desenvolvido como parte de um sistema de transmissão para televisão digital. Oferecemos uma descrição completa do processo de desenvolvimento, partindo do projeto do código até a implementação em hardware do codificador e do decodificador. Como ponto de partida, nos baseamos na classe de códigos eIRA, parcialmente inspirados pelos códigos LDPC utilizados na proposta para o padrão DVB-S2 [7]. Como é mostrado no restante do artigo, muitas das restrições impostas ao projeto do código estão diretamente relacionadas às limitações do hardware onde implementamos nossa prova de conceito. Por isso, o processo de desenvolvimento seguiu vários ciclos entre a fase de projeto da matriz de paridade, as simulações de desempenho, e o projeto do hardware, até que pudéssemos chegar a um resultado satisfatório, que pudesse ser testado em um ambiente real.

Os conceitos básicos dos códigos LDPC são discutidos em mais detalhes nas Seções II and III. A Seção IV descreve o método utilizado para projetar códigos eIRA estruturados, com desempenho otimizado para as restrições impostas. Os resultados de desempenho dos códigos projetados são mostrados na Seção V. A Seção VI descreve os detalhes da

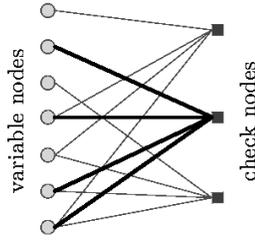


Fig. 1. Grafo de Tanner para o código de Hamming (7, 4).

implementação em hardware do decodificador. Finalmente, a Seção VII apresenta nossas conclusões e comentários sobre o projeto.

II. CÓDIGOS LDPC

Nesta seção descreveremos os códigos LDPC em mais detalhes. Começaremos com uma definição genérica e algumas propriedades destes códigos. Então, discutiremos o processo de codificação e decodificação.

Códigos LDPC (n, k) são códigos de bloco lineares e binários. Desta forma, estes códigos podem ser vistos como um conjunto de vetores binários \mathbf{c} que satisfazem $\mathbf{c}\mathbf{H}^T = \mathbf{0}$, onde \mathbf{H} é uma matriz binária $(n - k) \times n$ chamada *matriz de paridade* e as operações de soma e produto são binárias. A característica que define os códigos LDPC é o fato de que sua matriz de paridade é esparsa, ou seja, o número de valores não nulos é muito menor que o número total de valores na matriz \mathbf{H} .

Quanto à regularidade, existem dois tipos de códigos LDPC: regulares e irregulares. Em códigos regulares, todas as linhas e colunas de \mathbf{H} têm o mesmo número de uns, embora este número possa ser diferente para ambas. Se o número de uns muda entre colunas e/ou linhas de \mathbf{H} , o código é dito irregular.

A decodificação de códigos LDPC assim como algumas com suas características, são descritas em termos de seus grafos de Tanner [8]. Um grafo de Tanner é um grafo bipartido que contém dois tipos de nó: nós de cheque e nós de variável. Qualquer código de bloco linear binário tem um grafo de Tanner correspondente, o qual é construído a partir de sua matriz de paridade \mathbf{H} . Cada bit na palavra-código corresponde a um nó de variável, e cada equação de paridade corresponde a um nó de cheque. Um nó de variável é conectado a um nó de cheque no grafo de Tanner se, e somente se, o correspondente bit da palavra-código faz parte da correspondente equação de paridade. A Figura 1 mostra o grafo de Tanner associado ao código de Hamming (7, 4), cuja matriz de paridade é dada por

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}. \quad (1)$$

Na Figura 1, as linhas mais grossas correspondem à segunda linha de \mathbf{H} . Em [4], é mostrado que o desempenho de códigos LDPC é determinado pela chamada *distribuição de densidades*, que são polinômios que provêm uma descrição da estrutura do grafo de Tanner. De fato, define-se o grau de um nó como o número de ramos conectados a ele. Assim, o grau de um nó de variável é o número de equações de paridade do

qual o bit correspondente faz parte, enquanto que o grau de um nó de cheque é o número de bits envolvidos na correspondente equação de paridade. Seja λ_i (ρ_i) a fração de ramos conectados a nós de variável (nós de cheque) de grau i . Então,

$$\lambda(x) = \sum_i \lambda_i x^{i-1} \quad (2)$$

é a distribuição de graus dos nós de variável, e

$$\rho(x) = \sum_i \rho_i x^{i-1} \quad (3)$$

é a distribuição de graus dos nós de cheque. Já que estes polinômios são determinantes no desempenho de um código LDPC, vários algoritmos foram propostos para determinar a distribuição ótima. Alguns destes algoritmos serão discutidos na Seção III.

A. Decodificação de códigos LDPC

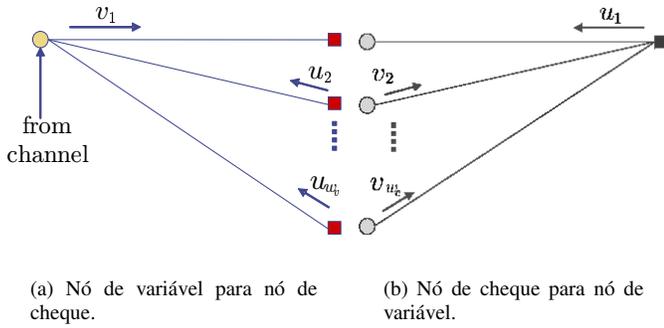
O principal objetivo em qualquer algoritmo de decodificação é determinar uma estimativa da probabilidade máxima a posteriori (MAP) dos bits transmitidos, já que ela minimiza a probabilidade de erro de bit. Para sistemas binários, a estimação MAP de c_i , o i -ésimo bit da palavra-código transmitida, pode ser feita a partir de

$$L_i = \log \frac{Pr(c_i = 0 | \mathbf{r})}{Pr(c_i = 1 | \mathbf{r})}, \quad (4)$$

onde \mathbf{r} é a sequência recebida. Assim, a estimação MAP é $c_i = 0$ se $L_i > 0$, e $c_i = 1$ caso contrário. Na literatura, L_i é chamada razão do log das verossimilhanças (LLR, do inglês *log likelihood ratio*). O cômputo exato da LLR é um problema NP completo e, assim, impraticável. Entretanto, baseado no grafo de Tanner, é possível derivar-se algoritmos iterativos capazes de obter uma boa aproximação da LLR.

Os algoritmos iterativos são baseados numa troca de mensagens entre nós de variável e nós de cheque, como ilustrado na Figura 2. Desta forma, a Figura 2(a) mostra o procedimento para um nó de variável de grau w_v . Nesta Figura, nós vemos que o nó de variável combina a informação vinda dos nós de cheque, u_i , $i = 2, \dots, w_v$ com a observação do canal, $-2r_i/\sigma^2$ (considerando uma modulação onde $0 \rightarrow -1$ e $1 \rightarrow 1$), para formar uma mensagem que é enviada novamente para o primeiro nó de cheque, v_1 . A Figura 2(b) mostra a operação equivalente em um nó de cheque de grau w_c . A principal diferença entre as figuras está na forma pela qual as mensagens são combinadas, e no fato de que os nós de cheque não tem acesso à informação do canal. A Figura 2 também ressalta o importante conceito de informação extrínseca. Como visto nestas figuras, para ambos os tipos de nó, a mensagem de saída de um dado ramo não é função da mensagem de entrada do mesmo ramo. Isto significa que todas as mensagens correspondem a informação extrínseca, o que impede realimentação positiva de informação.

Talvez o principal algoritmo de decodificação para códigos LDPC seja o algoritmo soma-produto (SPA) [3] [2]. Este algoritmo considera que as mensagens em todos os ramos são independentes, e suas mensagens são os valores “exatos” da informação extrínseca. Esta consideração de independência


 Fig. 2. Ilustração do algoritmo *message passing*.

é sempre verdadeira em grafos sem ciclos, para os quais o SPA produz valores exatos para a LLR em um número finito de iterações. Infelizmente, grafos correspondentes a códigos práticos costumam ter ciclos. Mesmo nestes casos, os valores aproximados obtidos pelo SPA são consideravelmente bons, embora ciclos pequenos piorem seu desempenho.

Para um nó de variável j , o SPA computa a mensagem deste nó para um nó de cheque i como

$$v_i = \sum_{j \neq i} u_j. \quad (5)$$

A mensagem proveniente de um nó de cheque para um nó de variável é dada por

$$u_i = 2 \tanh^{-1} \left(\prod_{j \neq i} \tanh \left(\frac{v_j}{2} \right) \right). \quad (6)$$

Claramente, a complexidade do SPA, e de outros algoritmos iterativos baseados em grafos, depende do número de ramos incidente nos nós: poucos ramos significa baixa complexidade. Entretanto, poucos ramos também significa que a matriz de paridade é esparsa. Isto explica porque códigos LDPC podem ser decodificados por algoritmos quase-MAP de baixa complexidade.

B. Codificação de códigos LDPC

Como visto anteriormente, a decodificação de códigos LDPC é relativamente simples. Infelizmente, a codificação de códigos LDPC em geral não é assim tão simples. De fato, considere que queremos obter a matriz geradora sistemática \mathbf{G} correspondente a \mathbf{H} . Começamos decompondo \mathbf{H} como

$$\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2], \quad (7)$$

onde \mathbf{H}_1 e \mathbf{H}_2 são matrizes binárias de dimensões $(n-k) \times k$ e $(n-k) \times (n-k)$. Então, \mathbf{G} é dada por

$$\mathbf{G} = [\mathbf{I}_k \ \mathbf{H}_1^T \mathbf{H}_2^{-T}]^T. \quad (8)$$

Infelizmente, a inversa de uma matriz esparsa não é esparsa, fazendo com que a codificação por força-bruta empregando \mathbf{G} seja uma tarefa de alta complexidade.

Desta forma, vários códigos LDPC com baixa complexidade de codificação têm sido propostos. Normalmente, a baixa

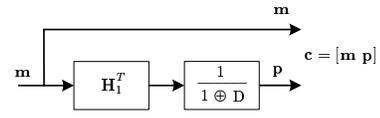


Fig. 3. Codificador para códigos LDPC do tipo eIRA.

complexidade é obtida impondo alguma estrutura à matriz de paridade. Em nosso trabalho, usamos os códigos *extended irregular repeat-accumulate* (eIRA) de [9]. Para estes códigos, a sub-matriz \mathbf{H}_1 ainda é aleatoriamente gerada de acordo com uma distribuição de graus pré-determinada. Já a matriz \mathbf{H}_2 , por outro lado, é dada por

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}. \quad (9)$$

Em outras palavras, a diagonal principal e a diagonal abaixo dela contém uns e os elementos restantes são zeros. Empregando esta definição, a multiplicação de um vetor por \mathbf{H}_2^{-T} pode ser implementada com um acumulador simples. O codificador resultante para um código LDPC do tipo eIRA é mostrado na Figura 3.

III. CAPACIDADE DE CÓDIGOS LDPC

Para muitos canais e tipos de decodificadores iterativos, os códigos LDPC apresentam um limiar de operação, ou seja, a medida que o tamanho do bloco tende ao infinito, uma probabilidade de erro de bit arbitrariamente pequena pode ser obtida se o nível de ruído for menor que um certo limiar. Por outro lado, para um nível de ruído acima deste limiar a probabilidade de erro de bit é maior que uma constante não nula. Gallager foi o primeiro a observar este fenômeno para canais simétricos binários (BSC, do inglês *Binary Symmetric Channel*) quando introduziu códigos LDPC regulares [1] [2]. Luby *et al.* generalizaram esta idéia para códigos LDPC irregulares de construção aleatória [10]. Mais recentemente em [4], Richardson e Urbanke generalizaram estas observações para um grande número de canais de entrada binária, incluindo o AWGN, e vários algoritmos de decodificação, como o SPA. Nesta seção, descreveremos com mais detalhes alguns métodos para se obter este limiar de convergência de códigos LDPC que, como veremos no decorrer deste trabalho, é fundamental para o projeto destes códigos.

O primeiro método a ser descrito é a *evolução de densidades* [4], o qual calcula iterativamente a função densidade de probabilidade (fdp) das mensagens no SPA. Chung *et al.* melhoraram o esforço computacional ao publicarem a versão discreta do algoritmo de evolução de densidades [11] [12], o qual permitiu se chegar a um limiar 0,0045 dB acima do limite de Shannon para canal AWGN e a um código com comprimento de bloco 10^7 e desempenho 0,04 dB acima do limite de Shannon a uma probabilidade de erro de bit (BER, do inglês *Bit Error Rate*) de 10^{-6} .

Ainda assim, o cálculo do limiar e a otimização das distribuições de graus empregando evolução de densidades são tarefas computacionalmente intensas, o que pode inviabilizar seu uso para canais práticos. No entanto, existem alternativas computacionalmente menos complexas e que, dependendo do caso, podem ser suficientemente precisas. Um destes métodos é baseado na aproximação das fdp das mensagens por Gaussianas (no caso de códigos regulares) ou misturas de Gaussianas (para códigos irregulares). Nesta *aproximação gaussiana*, sem sacrificar exageradamente a precisão, a média das densidades Gaussianas (uma quantidade uni-dimensional) pode representar a densidade de probabilidade das mensagens (uma quantidade cuja dimensão é infinita).

Embora a fdp das mensagens enviadas por nós de variável possa ser adequadamente aproximada por distribuições Gaussianas, o mesmo não acontece com a fdp das mensagens enviadas por nós de cheque. Assim, de forma a ter um método computacionalmente menos complexo que o evolução de densidades e mais preciso que com aproximação Gaussiana, Ardakani & Kschischang propuseram o método de *aproximação semi-Gaussiana* [13]. Neste método as fdp das mensagens enviadas por nós de variável continuam sendo aproximadas por Gaussianas ou misturas de Gaussianas. Já as fdp das mensagens enviadas por nós de cheque são obtidas empregando-se apenas uma iteração do algoritmo de evolução de densidades. Curvas de transferência de informação extrínseca (EXIT, do inglês *Extrinsic Information Transfer*) [14] são empregadas então para verificar o maior nível de ruído para o qual o algoritmo iterativo ainda converge, e assim determinar o limiar.

Embora existam outros métodos para a determinação do limiar, descreveremos em mais detalhes os três métodos discutidos anteriormente.

A. Evolução de Densidades Discretas

Considere novamente o valor de LLR v como sendo a mensagem de um nó de variável de grau d_v para um nó de cheque. No decodificador soma-produto, v é igual a soma de todas as LLRs, ou seja,

$$v = \sum_{i=0}^{d_v-1} u_i, \quad (10)$$

onde u_i , $i = 1, \dots, d_v - 1$ são as LLRs que chegam dos vizinhos dos nós de variável, e u_0 é a mensagem inicial do nó de variável proveniente do canal.

Já mensagem de saída de um nó de cheque pode ser obtida através da relação

$$\tanh \frac{u}{2} = \prod_{j=1}^{d_c-1} \tanh \frac{v}{2}, \quad (11)$$

onde v , $j = 1, \dots, d_c - 1$ são as LLRs provenientes dos $d_c - 1$ nós de cheque vizinhos, e u é a mensagem de saída do nó de cheque em questão.

De agora em diante, assumimos um código irregular com distribuições de grau, $\lambda(x)$ e $\rho(x)$.

Agora, seja $\mathcal{Q}(w)$ a mensagem w quantizada, isto é,

$$\mathcal{Q}(w) = \begin{cases} \left\lfloor \frac{w}{\Delta} + \frac{1}{2} \right\rfloor \Delta, & \text{se } w \geq \frac{\Delta}{2} \\ \left\lceil \frac{w}{\Delta} - \frac{1}{2} \right\rceil \Delta, & \text{se } w \leq -\frac{\Delta}{2} \\ 0, & \text{c.c.} \end{cases} \quad (12)$$

onde $\mathcal{Q}(w)$ é o operador quantização; Δ é o intervalo de quantização; $\lfloor x \rfloor$ é o maior inteiro menor ou igual a x ; e $\lceil x \rceil$ é o menor inteiro maior ou igual a x .

O decodificador soma-produto quantizado é definido como aquele no qual todas as mensagens de entrada e saídas são quantizadas de acordo com (12). Desta forma, no nó de variável temos que $\bar{v} = \sum_{i=0}^{d_v-1} \bar{u}_i$, onde $\bar{v} = \mathcal{Q}(v)$ e $\bar{u}_i = \mathcal{Q}(u_i)$ para $i = 0, \dots, d_v - 1$. Denomina-se a função massa de probabilidade (fmp) da mensagem \bar{w} por $p_w[k] = \Pr(\bar{w} = k\Delta)$ para $k \in \mathbb{Z}$. Assim, p_v relaciona-se com sua fmp de entrada por

$$p_v = \bigotimes_{i=0}^{d_v-1} p_{u_i} \quad (13)$$

onde p_v é a fmp de \bar{v} ; p_{u_i} é a fmp de \bar{u}_i ; e \bigotimes é a operação de convolução discreta. Desde que os \bar{u}_i sejam identicamente distribuídos para $1 \leq i < d_v$, a equação anterior pode ser reescrita como

$$p_v = p_{u_0} * \left(\bigotimes_{i=1}^{d_v-1} p_{u_i} \right) \quad (14)$$

onde $p_u = p_{u_i}$ e $*$ é a convolução discreta. Isto pode ser calculado eficientemente através da FFT.

No nó de cheque, define-se o operador \mathcal{R} como

$$\mathcal{R}(a, b) = \mathcal{Q} \left(2 \tanh^{-1} \left(\tanh \frac{a}{2} \tanh \frac{b}{2} \right) \right) \quad (15)$$

onde a e b são mensagens quantizadas. Note que esta operação pode ser feita empregando uma tabela de valores pré-computados, o que faz a evolução de densidades discretas ser computacionalmente mais eficiente. Usando este operador, calcula-se a mensagem \bar{u} de (11) como:

$$\bar{u} = \mathcal{R}(\bar{v}_1, \mathcal{R}(\bar{v}_2, \dots, \mathcal{R}(\bar{v}_{d_c-2}, \bar{v}_{d_c-1}))) \quad (16)$$

onde assumimos que o processo de decodificação soma-produto discreta nos nós de cheque é feita par-a-par.

Considere $c = \mathcal{R}(a, b)$. A fmp p_c de c é dada por

$$p_c[k] = \sum_{(i,j):k\Delta=\mathcal{R}(i\Delta,j\Delta)} p_a[i]p_b[j]. \quad (17)$$

Abusando da notação, escrevemos isto como $p_c = \mathcal{R}(p_a, p_b)$.

Desde que os p_{v_i} sejam todos iguais, definimos $p_v = p_{v_i}$, para $1 \leq i < d_c$, e escrevemos $p_u = \mathcal{R}(p_v, \mathcal{R}(p_v, \dots, \mathcal{R}(p_v, p_v), \dots))$ como $p_u = \mathcal{R}^{d_c-1} p_v$. Definindo $\lambda(p) = \sum_{i=2}^{d_l} \lambda_i \bigotimes^{i-1} p$ e $\rho(p) = \sum_{j=2}^{d_r} \rho_j \mathcal{R}^{j-1} p$ para qualquer fmp p podemos escrever a evolução de densidades discreta como:

Teorema: A evolução de densidades discreta é descrita como

$$p_u^{(l+1)} = \rho(p_{u_0} * \lambda(p_u^{(l)})) \quad (18)$$

onde a fmp inicial $p_u^{(0)}$ tem toda a massa de probabilidade concentrada em zero e l é o número da iteração.

Para executar este algoritmo assume-se inicialmente que a palavra-código toda nula foi enviada. Assim, fixam-se os parâmetros do canal e executa-se o algoritmo iterativamente até que a densidade u tenda a um ponto de massa no infinito (equivalente a uma probabilidade de erro tender a zero) ou o algoritmo convirja para uma densidade tendo uma probabilidade de erro finita.

Definição 1: O limiar de convergência para um código LDPC de distribuição de graus $\lambda(x)$ e $\rho(x)$ é definido como o nível máximo de ruído no qual a probabilidade de erro ainda tende a zero a medida que o número de iterações tende a infinito.

B. Aproximação Gaussiana

Considere um código LDPC irregular com distribuições $\lambda(x)$ e $\rho(x)$. Aproxima-se as fdp de u , u_i , v e v por Gaussianas ou misturas de Gaussianas. Como uma Gaussiana pode ser completamente representada por sua média e variância, podemos manter apenas média e variância durante as iterações. Existe uma condição importante, chamada *condição de simetria* [15], que é preservada durante a evolução de densidades para todas as mensagens, o qual pode ser expressa como $f(x) = f(-x)e^x$, onde $f(x)$ é a fdp de uma mensagem. Aplicando esta condição à aproximação Gaussiana a cada iteração, podemos melhorar a precisão da aproximação. Para uma Gaussiana de média m e variância σ^2 , esta condição se reduz a $\sigma^2 = 2m$ e, assim, podemos manter apenas a média.

Assim, a partir da equação (11), a média $m_{v,i}^{(l)}$ da mensagem de saída do nó de variável de grau i na l -ésima iteração é dada por

$$m_{v,i}^{(l)} = m_{u_0} + (i-1)m_u^{(l-1)} \quad (19)$$

onde m_{u_0} é a média de u_0 e $m_u^{(l-1)}$ é a média de u na iteração $(l-1)$. A variância da densidade de saída é dada por $2m_{v,i}^{(l)}$. Além disso, na l -ésima iteração, uma mensagem de entrada em nó de cheque v terá a seguinte mistura de Gaussianas $f_v^{(l)}$ como fdp:

$$f_v^{(l)} = \sum_{i=2}^{d_l} \lambda_i \mathcal{N}(m_{v,i}^{(l)}, 2m_{v,i}^{(l)}). \quad (20)$$

Já em relação aos nós de cheque, com alguma manipulação adicional (veja [11] [12]) chegamos a

$$m_u^l = \sum_{j=2}^{d_r} \rho_j \phi^{-1} \left(1 - \left[1 - \sum_{i=2}^{d_l} \lambda_i \phi \left(m_{u_0} + (i-1)m_u^{l-1} \right) \right]^{j-1} \right) \quad (21)$$

onde

$$\phi(x) = \begin{cases} 1 - \frac{1}{\sqrt{4\pi x}} \int_{\mathbb{R}} \tanh \frac{u}{2} e^{-\frac{(u-x)^2}{4x}} du & \text{se } x > 0 \\ 1, & \text{se } x = 0. \end{cases} \quad (22)$$

A mensagem inicial proveniente de qualquer nó de cheque é nula, logo $m_u^{(0)} = 0$. Assim, podemos chegar a seguinte definição:

Definição 2: O limiar de um código LDPC irregular com distribuições de grau $\lambda(x)$ e $\rho(x)$ é o maior nível de ruído (equivalente ao maior valor absoluto de m_{u_0}) em que a média dos valores de LLR das mensagens enviadas de nós de cheque para nós de variável u ainda convirja para ∞ a medida que o número de iterações $l \rightarrow \infty$.

C. Aproximação Semi-Gaussiana

No método de aproximação semi-Gaussiana, assim como no método de aproximação Gaussiana convencional, as fdp das mensagens enviadas de nós de variável para nós de cheque, v , é aproximada por uma somatória de gaussianas, ou seja, as equações (19) e (20) ainda são válidas. No entanto, as fdp das mensagens enviadas por nós de cheque para nós de variável u não são mais consideradas como misturas de gaussianas. Agora, a fdp de u é calculada para uma iteração da treliça da Figura 4. Isto pode ser feito através de simulações de Monte Carlo ou através de uma iteração do método de evolução de densidades. Um decodificador LDPC

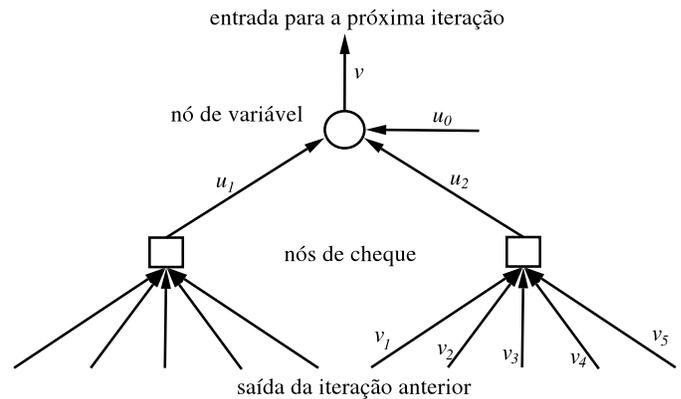


Fig. 4. Treliça da aproximação semi-Gaussiana.

pode ser visto como um bloco que, a cada iteração, usa duas fontes de informação sobre a palavra-código transmitida: a informação proveniente do canal (informação intrínseca I_0) e a informação vinda da iteração anterior (informação extrínseca I_{in}). A partir destas duas fontes de informação, o algoritmo de decodificação tenta produzir uma informação mais confiável sobre a palavra-código transmitida, gerando assim uma nova informação extrínseca I_{out} para a próxima iteração.

Este conceito de evolução da confiabilidade da informação a cada iteração é empregado para a confecção de EXIT charts. Embora empregue-se geralmente a entropia como esta grandeza de informação, Ardakani & Kschischang empregaram, sem perda de desempenho, probabilidade de erro de mensagem, como pode ser visto na EXIT chart na Figura 5. Assim, para cada iteração do algoritmo de decodificação temos que I_{in} , I_0 e I_{out} são associadas, respectivamente, as probabilidades de erro de mensagem p_{in} , p_0 e p_{out} . Assim, a cada iteração, temos $p_{out} = g(p_{in}, p_0)$. A EXIT chart pode ser representada por um gráfico com a função g e sua inversa g^{-1} , como na Figura 5. Cada seta representa uma iteração do decodificador. Muitas das características do decodificador, como número de iterações necessárias para alcançar-se uma

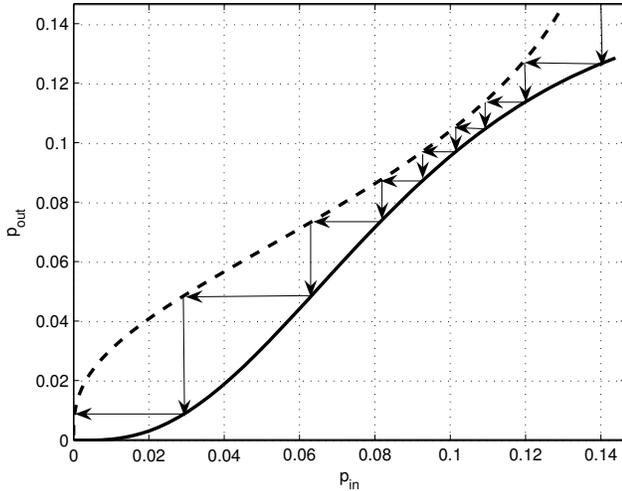


Fig. 5. EXIT Chart.

dada p_{out} e o limiar de convergência podem ser obtidas através de EXIT charts. Note que se o túnel de decodificação (espaço entre g e g^{-1}) estiver fechado, o algoritmo de decodificação converge para uma probabilidade de erro alta. Assim, podemos ter uma outra definição para o limiar de convergência de um código LDPC irregular:

Definição 3: O limiar de convergência para um código LDPC de distribuição de graus $\lambda(x)$ e $\rho(x)$ pode ser definido como o maior nível de ruído no qual o EXIT chart ainda encontra-se aberto.

Para manter a EXIT chart aberta a taxa de erro de mensagem na entrada do decodificador p_{in} deve ser maior que a taxa de erro de mensagem na saída, dada por

$$p_{out} = \sum_{i=2}^{d_l} \lambda_i g_i(x), \quad (23)$$

onde $g_i(x)$ é uma EXIT chart elementar (EXIT chart para nó de variável de grau i).

IV. METODOLOGIA DE PROJETO PARA CÓDIGOS eIRA ESTRUTURADOS

O projeto de um código LDPC do tipo eIRA estruturado [9] [6] consiste em duas etapas: a otimização das distribuições de grau $\lambda(x)$ e $\rho(x)$, e a construção da parte \mathbf{H}_1 da matriz de paridade. Na primeira etapa, o objetivo é maximizar a capacidade do código. Já na construção de \mathbf{H}_1 , distribui-se os "1"s respeitando-se as distribuições de grau, minimizando o número de ciclos de pequeno grau no grafo de Tanner e facilitando a representação e armazenamento da matriz de paridade. Este processo leva a códigos LDPC com bom desempenho e com codificadores e decodificadores com complexidade computacional reduzida.

A. Otimização de $\lambda(x)$ e $\rho(x)$

A otimização conjunta de $\lambda(x)$ e $\rho(x)$ é uma tarefa de alta complexidade computacional. Em [9], por exemplo,

o autor emprega o algoritmo de evolução de densidades com otimização via evolução diferencial. Embora complexo, este método permite obter códigos para qualquer taxa de codificação.

Uma forma de reduzir a complexidade é utilizar uma distribuição concentrada para $\rho(x)$ [12]. Dessa forma temos

$$\rho(x) = \rho x^{d_r-1} + (1-\rho)x^{d_r}, \quad (24)$$

onde seria necessário apenas estimar o valor escalar de ρ . Experimentos realizados em [11] [12] mostraram que não houve degradação de desempenho com o emprego desta forma concentrada de $\rho(x)$.

Outra alternativa para a redução da complexidade da otimização é a substituição do método de evolução de densidade por métodos como a aproximação Gaussiana e semi-Gaussiana. Nestes dois métodos o problema se resume a uma otimização via programação linear. Dentre os métodos estudados, o com aproximação semi-Gaussiana é aquele que fornece a melhor relação entre desempenho e complexidade computacional.

Richardson *et al.* [15] mostraram que a taxa de codificação pode ser expressa em termos de $\lambda(x)$ e $\rho(x)$ da seguinte forma

$$r(\lambda, \rho) = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx} = 1 - \frac{\sum_{i=2}^{d_r} \frac{\rho_i}{i}}{\sum_{i=1}^{d_l} \frac{\lambda_i}{i}}. \quad (25)$$

Eles também mostraram que o número de nós de variável de grau i , $N_v(i)$, pode ser expresso por

$$N_v(i) = \frac{n\lambda_i}{i \int_0^1 \lambda(x) dx} = \frac{n\lambda_i}{i \sum_{j=1}^{d_l} \frac{\lambda_j}{j}} \quad (26)$$

Assim, empregando a EXIT chart e a equação (26), e levando em consideração que para códigos do tipo eIRA $N_v(1) = 1$ e $N_v(2) = (n-k-1)$, o problema de otimização pode ser formulado pelo seguinte programa linear

$$\begin{aligned} & \text{maximizar} && \sum_{i=2}^{d_l} \frac{\lambda_i}{i} \\ & && \lambda_i \geq 0 \\ & \text{sujeito a} && \sum_{i=1}^{d_l} \lambda_i = 1 \\ & && \lambda_1 = \frac{1}{(n-1)} \sum_{i=2}^{d_l} \frac{\lambda_i}{i} \\ & && \lambda_2 = \frac{2(n-k-1)}{(k+1)} \sum_{i \neq 2}^{d_l} \frac{\lambda_i}{i} \\ & && \sum_{i=2}^{d_l} \lambda_i g_i(p_{in}) < p_{in} \quad \forall p_{in} \in (0, p_0] \end{aligned} \quad (27)$$

As entradas deste programa linear são o modelo do canal, o nível de ruído (limiar), o comprimento da palavra-código n , d_l e $\rho(x) = x^{d_r-1}$. O valor inicial para o nível de ruído pode ser ligeiramente superior ao limite de Shannon para canal AWGN. Caso a taxa de codificação resultante da maximização (27) seja diferente da especificada para o sistema (r), o nível de ruído deve ser incrementado e o processo de otimização repetido. O nível de ruído para o qual a maximização (27) resulta em uma taxa de codificação igual a r é considerado como o limiar de convergência do código.

Este método de otimização das distribuições de graus é uma das contribuições deste trabalho, já que é o primeiro trabalho onde é empregado aproximação semi-Gaussiana no projeto de códigos LDPC do tipo eIRA.

B. Construção da Matriz de Paridade

A matriz \mathbf{H}_1 poderia ser contruída aleatoriamente respeitando-se as distribuições de grau $\lambda(x)$ e $\rho(x)$ para nós de grau maior que 2. Ou então, diretamente através do algoritmo *progressive edge-growth* (PEG) [16], de forma a minimizar a ocorrência de pequenos ciclos no grafo de Tanner. No entanto, a memória necessária para armazenar-se \mathbf{H}_1 seria significativamente grande. Uma forma de diminuir a quantidade de memória para representar \mathbf{H}_1 é acrescentar a ela alguma regularidade.

O primeiro passo para construir a matriz \mathbf{H}_1 de dimensão $(n - k) \times k$ é criar uma matriz \mathbf{A} de dimensão $a \times b$, M vezes menor que a dimensão de \mathbf{H}_1 . M é o número de ramos que podem ser processados em paralelo no decodificador, e deve ser projetado levando-se em consideração o compromisso entre a taxa efetiva de dados e a área em hardware ocupada pelo decodificador. A matriz \mathbf{A} pode ser construída com o algoritmo PEG.

A seguir, constrói-se uma matriz \mathbf{B} com as mesmas dimensões de \mathbf{H}_1 substituindo cada “0” em \mathbf{A} por matrizes nulas quadradas de ordem M e cada “1” por \mathbf{J}^{ab} , onde \mathbf{J} é uma permutação da matriz identidade de ordem M , como na equação (28) para $M = 5$. Matrizes \mathbf{J} e suas permutações são empregadas em códigos LDPC do tipo *Array Codes* [17].

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (28)$$

Assim, a matriz \mathbf{H}_1 pode ser obtida através da seguinte permutação de linhas da matriz \mathbf{B}

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_{M+1} \\ \vdots \\ \mathbf{b}_{(a-1)M+1} \\ \mathbf{b}_2 \\ \mathbf{b}_{M+2} \\ \mathbf{b}_{(a-1)M+2} \\ \vdots \\ \mathbf{b}_{n-k} \end{bmatrix} \quad (29)$$

onde \mathbf{b}_i é a i -ésima linha da matriz \mathbf{B} .

Este processo reduz o número de pequenos ciclos no grafo de Tanner, embora não os elimine por completo. Além disso, a regularidade na sub-matriz \mathbf{H}_1 permite uma representação compacta da matriz de paridade \mathbf{H} e reduz a complexidade do codec.

Como um sistema de televisão digital requer uma recepção quase livre de erros ($\text{BER}=10^{-11}$), nosso objetivo para o projeto de códigos LDPC é obter o código que fornece o melhor desempenho em termos de E_b/N_0 para uma $\text{BER}=10^{-5}$. No sistema como um todo, um código externo como o Reed-Solomon eliminaria por completo um possível *error-floor* e levaria o sistema a uma situação quase livre de erros na saída do receptor.

V. RESULTADOS DE SIMULAÇÃO

O receptor de Televisão Digital deve ser robusto a vários tipos de canal. Para recepção fixa em uma área rural o canal pode ser puramente AWGN. Por outro lado, para um receptor móvel o canal pode ser Rayleigh. Assim, o esquema de codificação de canal deve ser capaz de se adaptar a todas estas condições. Como visto em [18], códigos LDPC irregulares projetados para canal AWGN são também robustos para outros tipos de canal. Assim, o canal AWGN foi escolhido para o projeto dos códigos.

Para o Sistema de Televisão Digital proposto foram escolhidas cinco taxas de codificação, 1/2, 2/3, 3/4, 5/6 and 7/8, e palavras-código de 9792 bits. Empregando evolução de densidades com aproximação semi-Gaussiana como descrito anteriormente resultou nas distribuições de graus e limiares de convergência da Tabela I. A capacidade de códigos LDPC com estas distribuições de graus ficou a uma distância de 0,13 dB a 0,37 dB do limite de Shannon para canais AWGN binários.

TABELA I
DISTRIBUIÇÕES DE GRAU

	1/2	2/3	3/4	5/6	7/8
$\rho(x)$	x^6	x^{11}	x^{16}	x^{23}	x^{25}
λ_1	0,00003	0,00003	0,00002	0,00003	0,00003
λ_2	0,2857	0,1666	0,1176	0,0833	0,0769
λ_3	0,2544	0,3779	0,4118	0,5000	0,6923
λ_5	0,1223				
λ_6					0,2308
λ_9		0,0989			
λ_{10}	0,3197			0,4167	
λ_{11}	0,0180				
λ_{12}		0,3566	0,4767		
limiar (dB)	0,428	1,236	1,804	2,582	3,211
gap (dB)	0,24	0,13	0,18	0,27	0,37

Simulações de Monte Carlo foram feitas com os códigos gerados a partir das distribuições de grau da Tabela I. A Figura 6 mostra o desempenho em termos de taxa de erro de bit. Os códigos obtidos ficam de 0,7 dB a 1 dB do limite de Shannon para canais AWGN binários (considerando uma BER de 10^{-5}). Como pode ser visto na Figura 7, o código LDPC (9792,4896) projetado pela nossa metodologia é 3 dB melhor (a $\text{BER}=10^{-5}$) que o código convolucional empregado nos sistemas DVB-T e ISDB-T. Para estas simulações considerou-se modulação BPSK, canal AWGN e $M = 51$ ramos em paralelo no decodificador. A BER foi computada depois de 50 palavras-código erradas, considerando-se um máximo de 50 iterações para o decodificador soma-produto.

VI. IMPLEMENTAÇÃO EM HARDWARE

Dentre as diferentes taxas de codificação consideradas para o sistema de televisão digital proposto, para a implementação em hardware optou-se pela taxa 3/4 com palavras-código de 9792 bits. Assim, cada palavra-código consiste em 7344 bits de informação e 2448 bits de paridade.

O codificador e o decodificador LDPC foram implementados em *Field Programmable Gate Arrays* (FPGA). O codificador foi desenvolvido em *Very High Speed Integrated Circuit*

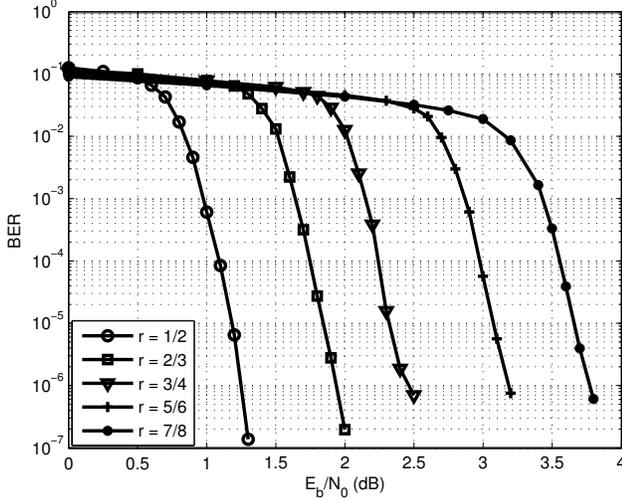


Fig. 6. Desempenho dos códigos LDPC para o STVD.

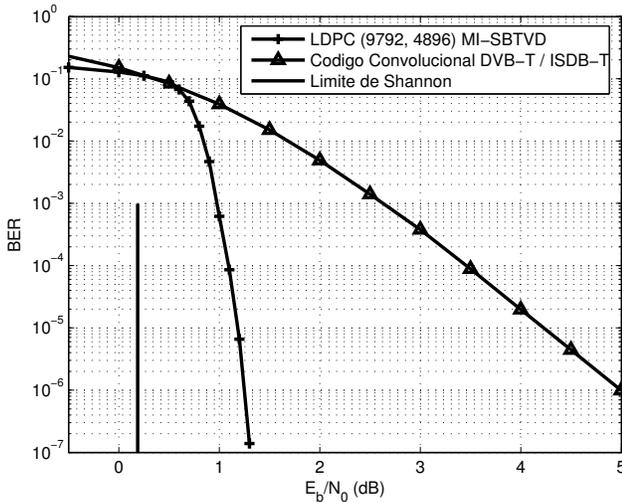


Fig. 7. Comparação entre o código LDPC (9792,4896) e o código convolucional dos sistemas DVB-T e ISDB-T.

Hardware Description Language (VHDL – IEEE 1164) utilizando a ferramenta de desenvolvimento Quartus, da Altera. O decodificador LDPC foi implementado em VHDL e System Generator, da Xilinx.

No decodificador, as operações em nós de cheque podem ser feitas de forma serial com uma estrutura em treliça, ou em paralelo na forma de uma estrutura em ramos. A estrutura em treliça foi adotada devido ao fato de consumir significativamente menos recursos da FPGA que a implementação com estrutura em ramos.

Talvez a questão mais importante na implementação em hardware seja o compromisso entre a área ocupada no dispositivo e a velocidade de processamento. O taxa efetiva mínima requerida pelo sistema proposto é de 19,33Mbps. Por outro lado, existe a limitação física do dispositivo FPGA em área de silício, que limita o grau de paralelismo da implementação.

Os detalhes da implementação do codificador e decodifica-

dor LDPC serão discutidos a seguir.

A. Codificador LDPC

Conforme visto na Figura 3, o codificador LDPC multiplica o bloco da mensagem por \mathbf{H}_1^T e envia o resultado para um acumulador. O acumulador é obtido através da simples operação OU-Exclusivo (XOR), sendo assim facilmente implementado. Por outro lado, um grande problema a ser considerado é como armazenar a matriz \mathbf{H}_1 , de dimensões 7344×2448 . Felizmente, conforme discutido na Subseção IV-B, é necessário armazenar apenas os índices dos nós de variável conectados aos 144 nós de cheque indexados por múltiplos de m , que, no nosso caso, é igual a 51. Os índices dos nós de variável são armazenados em uma estrutura que possui 112 linhas de 3 elementos e 32 linhas de 12 elementos, aqui denominada de estrutura \mathbf{T} . Uma vez que cada elemento desta estrutura \mathbf{T} possui 12 bits, \mathbf{H}_1 pode ser representada usando-se apenas 8.6 Kbits, uma redução de aproximadamente 2000 vezes, se comparado com a representação original completa.

Assim, a estrutura \mathbf{T} foi armazenada em 15 blocos de memória independentes (BRAMs), sendo que 3 armazenam os elementos das primeiras 112 linhas e 12 armazenam os elementos das últimas 32. Desta forma, torna-se possível acessar simultaneamente todos os dados necessários para cada iteração, permitindo a saída serial dos bits de paridade imediatamente após a chegada do último bit da mensagem.

B. Decodificador LDPC

Para uma implementação em hardware do decodificador, as expressões (5) e (6) devem ser reescritas de modo mais eficiente. A informação total disponível no nó de variável v é dada por $v = \sum_j u_j$. Assim, as mensagens provenientes de nós de variável, dadas pela equação (5), podem ser computadas de forma eficiente por

$$v_i = v - u_i, \quad (30)$$

o que requer apenas $2w_v + 1$ adições. Para as operações nos nós de cheque em (6) é possível evitar o cômputo direto da função transcendental \tanh . Assumindo um nó de cheque com grau dois e sendo U e V as mensagens que chegam a este nó, (6) pode ser reescrita como

$$\begin{aligned} L(U \oplus V) &= 2 \tanh^{-1} \left(\tanh \frac{U}{2} \tanh \frac{V}{2} \right) \\ &= \min\{|U|, |V|\} + z(U, V), \end{aligned} \quad (31)$$

onde $z(U, V) = \log \left(\frac{1 - \exp^{-|U+V|}}{1 - \exp^{-|U-V|}} \right)$ é chamado de *fator de correção*.

Aproximando o fator de correção por $z(U, V) \equiv 0$ obtemos uma operação com reduzida complexidade computacional, dando origem ao algoritmo *min-sum*. Da mesma forma, uma expressão com um melhor compromisso entre custo computacional, precisão e quantização é a *aproximação constante* [19], dada por

$$z(x, y) = \begin{cases} 0.5 & \text{if } |x| \leq 2, |y| > 2|x| \\ -0.5 & \text{if } |y| \leq 2, |x| > 2|y| \\ 0 & \text{caso contrário.} \end{cases} \quad (32)$$

Para a operação nos nós de cheque como um todo, diversas topologias podem ser empregadas [20]. A topologia mais complexa é a *força bruta*, a qual computa o valor da equação (6). Já a *implementação paralela* permite a maior taxa efetiva de dados ao custo de uma maior ocupação em hardware, além de apresentar grande sensibilidade aos efeitos de quantização. Por outro lado, a *implementação serial* é muito robusta em termos de efeitos de quantização e requer um menos recursos de hardware quando sintetizada em uma FPGA.

Para a implementação serial, (6) é reescrita como

$$u_1 = L(b_2) \quad (33)$$

$$u_i = L(f_{i-1} \oplus b_{i+1}), \quad i = 2, \dots, w_c - 1 \quad (34)$$

$$u_{w_c} = L(f_{w_c-1}) \quad (35)$$

onde $f_1 = c_1$ e $f_2 = f_1 \oplus c_2, \dots, f_{w_v} = f_{w_v-1} \oplus c_{w_v}, b_{w_v} = c_{w_v}, b_{w_v-1} = b_{w_v} \oplus c_{w_v-1}, \dots, b_1 = b_2 \oplus c_1$ são dois conjuntos de variáveis aleatórias auxiliares que representam o fluxo de mensagens diretas e reversas entre nós de cheque. O cômputo da mensagem u_i pela equação (34) requer a obtenção dos parâmetros $L(f_1), \dots, L(f_{w_v}), L(b_1), \dots, L(b_{w_v})$. Estes podem ser obtidos recursivamente substituindo-se os valores conhecidos por $L(c_1), \dots, L(c_{w_v})$ na equação (31). A complexidade computacional resultante é de $3 \cdot (w_v - 2)$ operações.

A arquitetura empregada em nosso projeto é similar a empregada em [21]. Como visto na Figura 8, esta implementação tem sete subcomponentes principais: buffer de entrada (inBuffer), buffer de saída (outBuffer), processadores dos nós de variável (vnp), processadores dos nós de cheque (cnp), entrelaçador e desentrelaçador, e uma unidade de controle (CR). As últimas cinco unidades formam o kernel do decodificador.

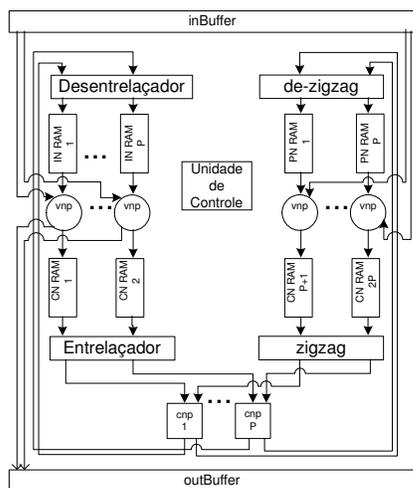


Fig. 8. Arquitetura do decodificador LDPC

O inBuffer é responsável por manter todos os blocos de dados disponíveis e acessíveis ao kernel. De forma a minimizar atrasos no kernel, ele consiste em dois blocos de buffer circulares de escrita serial e leitura paralela. vnp e cnp computam, respectivamente, (30) e (31). Ambos entrelaçador e desentrelaçador executam endereçamento aleatório em linhas e deslocamento em colunas. Um bloco completo do decodificador LDPC é enviado em paralelo entre vnp e cnp através

do entrelaçador e são armazenados e lidos dos dois blocos de memórias internas com os valores das mensagens de LLR.

C. Resultados

O decodificador LDPC foi implementado em um dispositivo Altera Stratix II 2S60 e desenvolvido em VHDL. Com a estrutura da Figura 3 e a regularidade na matriz H_1 , o codificador teve um baixo custo computacional. De fato, somente 4% da área do dispositivo e 5% dos blocos internos de memória RAM (BRAMs) foram ocupados.

Para o decodificador empregou-se aproximação constante para o fator de correção e uma implementação serial. O decodificador foi implementado em um dispositivo Xilinx Virtex II XC2V3000 empregando VHDL e System Generator, também da Xilinx. A frequência do relógio para o dispositivo FPGA foi de 97 MHz. Em recursos da FPGA foram empregados 186 BRAMs (96%) e 9478 slices (61%). As mensagens foram armazenadas em ponto fixo com 5 bits (1 para o sinal, 3 para a parte inteira e 1 para a parte fracionária) e, de forma a respeitar a taxa efetiva de dados mínima, o número máximo de iterações foi definido como 13.

VII. CONCLUSÕES

Neste artigo descrevemos a metodologia de projeto de um código LDPC e sua implementação em hardware. Vimos que o código LDPC do tipo eIRA empregado neste trabalho permite a implementação de um codificador de baixa complexidade computacional, empregando menos de 5% de uma FPGA do tipo Altera Stratix II. A implementação do decodificador é mais complexa, ocupando quase que a totalidade de uma FPGA do tipo Xilinx Virtex IV. Os resultados de simulação mostraram que, para uma BER de 10^{-5} , o código proposto opera em torno de 1 dB do limite de Shannon para canais AWGN, e fornece uma vantagem de 3 dB em relação ao código convolucional empregado em outros sistemas de televisão digital.

AGRADECIMENTOS

Este trabalho foi financiado pela Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e pela Linear Equipamentos Eletrônicos S/A.

REFERÊNCIAS

- [1] R. G. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. IT-8, pp. 21–28, Jan. 1962.
- [2] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.
- [3] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Information Theory*, pp. 399–431, Mar. 1999.
- [4] T. J. Richardson e R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Information Theory*, vol. 47, pp. 599–618, Fev. 2001.
- [5] T. J. Richardson e R. L. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 638–656, Fev. 2001.
- [6] Y. Zhang, W. Ryan e Y. Li, "Structured eIRA codes with low floors," *Proceedings of the International Symposium on Information Theory - ISIT2005*, pp. 174–178, Set. 2005.

- [7] DVB-S.2 Standard Specification, ETSI EN 302 307 V1.1.1 (2005-03). http://webapp.etsi.org/action/PU/20050322/en_302307v010101p.pdf
- [8] R. M. Tanner, "A recursive approach to low-complexity codes," *IEEE Trans. Information Theory*, pp. 533–547, Set. 1981.
- [9] M. Yang, W. Ryan e Y. Li, "Design of efficiently encodable moderate-length high-rate irregular LDPC codes," *IEEE Trans. Communications*, vol. 52, no. 4, pp. 564–571, Abr. 2004.
- [10] M. Luby, M. Mitzenmacher, A. Shokrollahi e D. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proc. 30th Annual ACM Symp. Theory on Computing*, pp. 249–258, 1998.
- [11] S.-Y. Chung, G. D. Forney Jr., T. J. Richardson e R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon Limit," *IEEE Communications Letters*, vol. 5, no. 2, pp. 58–60, Fev. 2001.
- [12] S.-Y. Chung, T. J. Richardson e R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Information Theory*, vol. 47, pp. 657–669, Fev. 2001.
- [13] M. Ardakani e F. R. Kschischang, "A more accurate one-dimensional analysis and design of irregular LDPC codes," *IEEE Trans. Communications*, vol. 52, no. 12, pp. 2106–2114, Dez. 2004.
- [14] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Communications*, vol. 49, pp. 1727–1737, Out. 2001.
- [15] T. J. Richardson, A. Shokrollahi e R. Urbanke, "Design of capacity-approaching low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Fev. 2001.
- [16] X. Y. Hu, E. Eleftheriou e D. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *IEEE Trans. Communications*, vol. 52, pp. 386–398, Fev. 2005.
- [17] J. L. Fan, "Array codes as low density parity check codes," in *Proc. of the 2nd International Symposium on Turbo Codes and Related Topics*, pp. 543–546, Set. 2000.
- [18] J. Hou, P. Siegel e L. Milstein, "Performance analysis and code optimization of low density parity-check codes on Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 924–934, Maio 2001.
- [19] M. Shen, H. Niu, H. Liu e J. A. Ritcey, "Finite precision implementation of LDPC coded M-ary modulation over wireless channels," *Conference on Signals, Systems and Computers - ACSSC2003*, v. 1, pp. 114–118, Nov. 2003.
- [20] X.-Y. Hu, E. Eleftheriou, D. Arnold e A. Dholakia, "Efficient Implementations of Sum-Product Algorithm for Decoding LDPC Codes." *IEEE GLOBECOM2001*, v.2, pp. 1036–1036, 2001.
- [21] F. Kienle, T. Brack e N. Wehn, "A synthesizable IP Core for DVB-S2 LDPC code decoding" *Proc. of Design, Automation and Test in Europe*, v. 3, pp. 100–105, 2005.



Tarciano Facco Pegoraro é doutorando da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas (UNICAMP) e engenheiro de software da Motorola Industrial Ltda, em Jaguariúna, SP. Recebeu os títulos de Engenheiro Eletricista pela Universidade Federal de Santa Maria (UFSM) em 1998 e Mestre em Telecomunicações e Telemática pela UNICAMP em 2000. De 2000 a 2004 atuou como engenheiro de software para redes CDMA na Ericsson Telecomunicações do Brasil e Nortel Networks. Tem

interesse nas áreas de processamento digital de sinais e comunicações digitais em geral, especialmente códigos corretores de erros e codificação espaço-temporal.

Marcelo Carneiro de Paiva nasceu em Ipameri, GO, em março de 1979. Formou-se Técnico em Eletrônica pela ETE "FMC" (Escola Técnica de Eletrônica Francisco Moreira da Costa) em 1998 e Engenheiro Eletricista com ênfase em Telecomunicações pelo INATEL (Instituto Nacional de Telecomunicações) em 2003. Atualmente, cursa o Mestrado no INATEL. Trabalha com pesquisa e desenvolvimento de sistemas de transmissão em TV Digital desde 2003.

Fábio Adriano Lisboa Gomes nasceu em Picos, PI, em 07 de março de 1975. Recebeu os títulos de Engenheiro Eletricista e Mestre em Engenharia Elétrica, ênfase em Computação, pelo Departamento de Engenharia Elétrica da Universidade Federal do Rio Grande do Norte em 1998 e 2000, respectivamente. Desde 2003 é professor do Departamento de Informática das Faculdades Hoyer, campus de Hortolândia. Desde 2004 coordena o desenvolvimento de Aplicativos para WEB e para Dispositivos Móveis da empresa TiSoft, nas plataformas J2EE e J2ME. De 2005 a 2006 foi coordenador científico de projeto PIPE/FAPESP relacionado ao desenvolvimento de Sistemas para Dispositivos Móveis. Tem interesse nas áreas de Processamento digital de sinais, Engenharia de software, Tecnologias Orientadas a Objetos, Plataformas de desenvolvimento para WEB 2.0, Linguagens de programação e Automação de testes de software.



(SBTVD) na proposta de modulação inovadora.

Fábio Lumertz é doutorando do departamento de comunicações (DECOM) da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas (UNICAMP) e membro do ComLab, laboratório de pesquisas em comunicações digitais e WebLabs vinculado ao mesmo departamento. Graduou-se em engenharia elétrica com ênfase em telecomunicações na Pontifícia Universidade Católica do Rio Grande do Sul no ano de 2003. Como aluno e pesquisador da Unicamp, participou do projeto do Sistema Brasileiro de Televisão Digital



estimação de canal, códigos corretores de erros, e receptores iterativos.

Renato R. Lopes recebeu os títulos de Engenheiro Eletricista e Mestre em Engenharia Elétrica pela Universidade de Campinas (UNICAMP) em 1995 e 1997, respectivamente. Ele é PhD em Engenharia Elétrica pelo Instituto de Tecnologia da Geórgia (GeorgiaTech), EUA, onde também concluiu, em 2001, um mestrado em matemática. De 2003 a 2005 fez pós-doutorado na Faculdade de Engenharia Elétrica e de Computação da UNICAMP, onde é professor assistente desde 2006. Tem interesse na grande área de teoria das comunicações, incluindo equalização e

Fabrizio C. de A. Oliveira nasceu em São Paulo, SP, em 06 de setembro de 1976. Possui os títulos de: Engenheiro Eletricista, modalidade Eletrônica, pela Universidade Federal de Pernambuco (2000), Mestre em Engenharia Elétrica, com ênfase em Telecomunicações, pela Universidade Estadual de Campinas (2002). Atualmente cursa o doutorado na Universidade Estadual de Campinas. É pesquisador nas áreas de teoria da informação, compressão de imagens e vídeo e codificação de canal para sistemas de comunicação. Atuou nos projetos de compressão de vídeo usando H.264, transcodificação de vídeo digital e modulação durante o desenvolvimento do sistema brasileiro de televisão digital. Seus principais interesses incluem o modelamento teórico de sistemas de comunicação e as aplicações práticas da teoria da informação.

Roberto Gallo nasceu em Campinas, SP, em 17 de maio de 1978. É Engenheiro de Computação (IC - UNICAMP, 2001) e Mestre em Ciência da Computação (IC - UNICAMP, 2003). É especialista no projeto e desenvolvimento de arquiteturas para hardware reconfigurável (FPGA), em SoCs. Atualmente é aluno de Doutorado em Ciência da Computação (IC - UNICAMP), na área de processadores criptográficos. De 2004 a 2006 foi coordenador científico de projeto PIPE FAPESP na área de hardware criptográfico baseado em sistemas reconfiguráveis. Possui experiência de mais de seis anos no desenvolvimento de sistemas digitais, tendo produzido diversos *IP Cores* de processadores e aceleradores criptográficos. Áreas de interesse: hardware reconfigurável (FPGA), sistemas digitais de larga escala, criptografia.