ETSI TR 102 662 V1.1.1 (2010-03)

Technical Report

Satellite Earth Stations and Systems (SES); Advanced satellite based scenarios and architectures for beyond 3G systems

Reference
DTR/SES-00288

Keywords
3G, architecture, MSS, satellite

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Satellite Earth Stations and Systems (SES).

Introduction

The analysis contained in this Technical Report is intended to assist ETSI in defining future standardisation activities; specifically standardisation for the medium-term evolution of current SatCom "2G" and "3G" standards, and for the long-term definition of future "4G" SatCom standards.

The material presented in this Technical Report represents the efforts of many research facilities which include ETRI, University of Surrey, University of Bologna, ESA and CNES.

1 Scope

The present document addresses the role of satellite communications as terrestrial communication systems begin to evolve towards beyond 3G and 4G architectures.

The present document identifies the possible roles of satellites in beyond 3G and 4G networks and how to make the best use of innovative technologies in order to achieve these roles. The present document makes a contribution in these directions, by identifying possible future system architectures and roles for satellites in this evolving context. It reviews and analyzes some of the latest communication technologies that would enable satellite systems to realize cost-effectively these architectures and claim these roles.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI EN 302 583: "Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz".

[i.2] ETSI TS 101 851: "Satellite Earth Stations and Systems (SES); Satellite Component of UMTS/IMT-2000; Part 1: Physical channels and mapping of transport channels into physical channels; Sub-part 1: G-family (S-UMTS-G 25.211)".

[i.3] Kanatas et al, "Land Mobile Satellite Channel Measurements in Athens City Center at 1 800 MHz", IMSC'97, Pasadena, CA, June 1997.

[i.4] J. V. Evans, "Satellite systems for personal communications", Proc. of the IEEE, vol. 86, no.7, 1998, pp. 1325-1341.

[i.5] J. Goldhirsh, W. J. Vogel, "Earth-satellite tree attenuation at 20 GHz: foliage effects", IEE Electr. Letters, vol. 29, no. 18, pp. 1640-1641, September 1993.

[i.6] W. J. Vogel, U. Hong, "Measurement and modeling of land mobile satellite propagation at UHF and L-band", IEEE Trans. on Antennas and Propagation, vol. 36, no. 5, pp. 707-719, May 1988.

[i.7] W. J. Vogel, J. Goldhirsh, "Mobile satellite propagation measurements at L-band using MARECS-B2", IEEE Trans. on Ant. and Prop., vol. 38, no. 2, pp. 259-264, February 1990.

[i.8] W. J. Vogel, J. Goldhirsh, "Multipath Fading at L Band and Low Elevation Angle, Land Mobile Satellite Scenarios", IEEE Journ. on Sel. Areas in Comm., vol. 13, no. 2, pp. 259-264, February 1995.

[i.9] International Telecommunication Union, ITU-R Recommendation P.681, "Propagation data required for the design of earth-space land mobile telecommunication systems", Geneva, 1997.

[i.10] S. R. Saunders, A.A. Zavala. "Antennas and Propagation for Wireless Communication Systems", 2nd edition, Wiley&Sons, 2007.

[i.11] G. E. Corazza, F. Vatalaro, "A statistical model for land mobile satellite channels and its applications to nongeostationary orbit systems", IEEE Trans. on Veh. Tech., vol. 43, no. 3, pp. 738-742, August 1994.

[i.12] C. Loo, "A statistical model for land mobile satellite link", IEEE Trans. on Veh. Tech. vol. 34, no. 3, pp. 122-127, August 1985.

[i.13] S-H Hwang, K-J Kim, J-Y Ahn, K-C Whang, "A channel model for nongeostationary orbiting satellite system", 47th IEEE Veh. Tech. Conf., Vol. 1, Issue 4-7, pp. 41-45, May 1997.

[i.14] M. J. Gans, "A Power-Spectral Theory of Propagation in the Mobile-Radio Environment", IEEE Trans. on Veh. Technology, Vo. 21, No. 1, February 1972.

[i.15] C.Q. Xu, C. L. Law, S. Yoshida, "On the Doppler Power Spectrum at the Mobile Unit Employing a Directional Antenna", IEEE Comm. Letters, Vol. 5, No. 1, January 2001.

[i.16] W. T Ng, V.K. Dubey, "Comments on "On the Doppler Spectrum at the Mobile Unit Employing a Directional Antenna", IEEE Comm. Letters, Vol. 6, No. 11, November 2002.

[i.17] R. H. Clarke, "A statistical theory of mobile radio reception", Bell Syst. Tech. J., vol. 43, pp. 957-1000, July 1968.

[i.18] T. Aulin, "A Modified Model for the Fading Signal at a Mobile Radio Channel", IEEE Trans. on Veh. Tech., VT-28, No. 3, August 1979.

[i.19] J. Goldhirsh, W. J. Vogel, "Fade-durations derived from the land mobile satellite measurements in Australia", IEEE. Trans. on Comm., vol. 39, no. 5, pp. 664-668, May 1999.

[i.20] R. Tafazolli, P. Taaghol, "Correlation model for shadow fading in land mobile satellite systems", IEE Electr. Letters, vol. 33, no. 15, pp. 1287-1289, July 1997.

[i.21] M. Patzold, Y, Li, F. Laue, "A study of a Land Mobile Satellite Channel Model with Asymmetrical Dopller Spectrum and Lognormally Distributed Line-of Sight", IEEE Trans. on Veh. Tech., Vol. 47, No. 1, February 1998.

[i.22] B. Vucetic, J. Du, "Channel modeling and simulation in satellite mobile communication systems", IEEE J. Select. Areas Commun., vol. 10, no. 8, pp. 1209-1218, 1992.

[i.23] E. Lutz, et al., "The land mobile satellite communication channel - recording statistics and channel model", IEEE Trans. on Veh. Technologu, vol. 40, no. 2, pp. 375-386, May 1991.

[i.24] E. Lutz, "A Markov Model for Correlated Land Mobile Satellite Channels", Int. Jour. Of Sat. Comm., vol. 14, pp. 333-339, 1996.

[i.25] E. Lutz, "Issues in satellite personal communication systems", Wireless Networks, vol. 4, pp. 109-124, 1998.

[i.26] E. Lutz, M. Werner, A. Jahn, "Satellite Systems for personal and broadband communications", Springer 2000.

[i.27] F. Perez-Fontan, et al. "A versatile framework for narrow- and wide-band statistical propagation model for the LMS channel", IEEE Trans. on Broadcasting, vol. 43, no. 4, pp. 431-458, December 1997.

[i.28] F. Perez-Fontan, et al. "Complex envelope threestate Markov model based simulator for the narrow-band LMS channel", Int. J. on Sat. Comm., vol. 15, pp. 1-15, 1997.

[i.29] F. Perez-Fontan, et al. "S-Band LMS propagation channel behavior for different environments, degrees of shadowing and elevation angles", IEEE Trans. On Broadcasting, vol. 44, no. 1, pp. 40-76, March 1998.

[i.30] F. Perez-Fontan, et al. "Statistical modeling of the LMS channel", IEEE Trans. on Veh. Tech., vol. 50, no. 6, pp.1549-1567, November 2001.

[i.31] B. Vucetic, J. Du, "Channel modeling and simulation in satellite mobile communication systems", in Proc. Int. Conf. Satel. Mobile Commun., Adelaide, Australia, pp. 1-6. August 1990.

[i.32] B. Belloul, S. Saunders, M.A.N. Parks, B. G. Evans, "Measurement and modeling of the wideband propagation at L and S band applicable to the LMS channel", IEEE Proc. Microwaves, Antennas and Prop., vol. 147, no. 2, pp. 116-121, April 2000.

[i.33] A. Jahn, H. Bischl, G. Heiss, "Channel characterization for spread spectrum satellite communications", IEEE 4th Inter. Symp., on Spread Spectrum Techniques and Applications, vol. 3, 22-25, pp. 1221-1226, September 1996.

[i.34] A Jahn, "Propagation considerations and fading countermeasures for mobile multimedia services", Inter. Journal of Satel. Comm., vol. 19, pp. 223-250, 2001.

[i.35] M. Dottling, A. Jahn, D. Didascalou, W. Wiesbeck, "Two- and three- dimensional ray tracing applied to the land mobile satellite (LMS) propagation channel", IEEE Antennas and Propagation Magazine, vol. 43, no. 6, pp. 27-37, December 2001.

[i.36] S. R. Saunders, C. Tzaras, B. G. Evans, "Physical-statistical methods for determining state transition probabilities in mobile-satellite channel models", Int. Jour. Of Satellite Communications, vol. 19, pp. 207-222, 2001.

[i.37] C. Oestges, D. Vanhoenacker-Janvier, "Physical-statistical predication of performance for land mobile satellite communications systems", IEEE Proc. on Microwaves and Prop., vol. 146, issue 5, pp. 362-368, October 1999.

[i.38] C. Oestges, S. R. Saunders, D. Vanhoenacker-Janvier, "Physical-statistical modeling of the land mobile satellite channel based on ray tracing", IEE Proc. Microwave Antennas and Propagation, vol. 146, no. 1, pp. 45-49, February 1999.

[i.39] C. Tzaras, B. G. Evans, S. R. Saunders, "Physical-statistical analysis of land mobile-satellite channel", IEE Electronics Letters, vo. 34, no. 13, pp. 1355-1357, June 1998.

[i.40] P.R. King, B. G. Evans, S. Savrou, "Physical-statistical model for the land mobile-satellite channel applied to satellite/HAP-MIMO", 11th European Wireless Conference, vol. 1, April 2005.

[i.41] P.R. King, S. Savrou, "Land mobile-satellite MIMO capacity predictions", IEE Electr. Letters, vol. 41, no. 13, pp. 749-751, June 2005.

[i.42] P.R. King, S. Savrou, "Low Elevation Wideband Land Mobile Satellite MIMO Channel Characteristics", IEEE Trans. Wireless Comm., vol. 6, no. 7, pp.2712-2720, July 2007.

[i.43] P.R. King, S. Savrou, "Capacity improvements for a land mobile single satellite MIMO system", IEEE Antennas and Wireless Prop. Letters, vol. 5, pp. 98-100, December 2006.

[i.44] D2-1.3-Part1-ASC-MAESTRO-V3.3.pdf.

[i.45] D8-3-SPH-MAESTRO-V3.1.pdf.

[i.46] A. J. Viterbi, CDMA: "Principles of Spread Spectrum Communications", Addison-Weley.

[i.47] S. Verdu, "Multiuser Detection", Cambridge University Press, New York 1998.

[i.48] A. Nordio, M. Hernandez, G.Caire, "Low-Complexity Turbo Equalization and Multiuser Decoding for TD-CDMA", IEEE Trans. on Wireless Comm., Vol. 3, No. 2, pp. 454-465, March 2004.

[i.49] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-devision multiple-access channels", IEEE Trans. on Info. Theory, Vol. 35, pp. 123-136, January 1989.

[i.50] C. SanKaran, A. Ephremides, "Solving a Class of Optimum Multiuser Detection Problems with Polynomial Complexity", IEEE Trans. on Info. Theory, Vol. 44, pp. 1958-1961, September 1998.

[i.51] S. Ulukus, R. Yates, "Optimum Multiuser Detection Is Tractable for Synchronous CDMA Systems Using M-sequences", IEEE Comm. Lett. Vol. 2, pp. 89-91, 1998.

[i.52] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problem using semi-definite programming", J. ACM, vol. 42, pp. 1115-1145, 1995.

[i.53] W. Ma, T.N. Davidson, K.M. Wong, Z.Q. Luo, P.C. Ching, "Quasi-Maximum-Likelihood Multiuser Detection using Semi-Definite Relaxation with Application to Synchronous CDMA", IEEE Trans. Signal Processing, Vol. 50, No. 4, pp. 912-922, April 2002.

[i.54] Y. Nesterov and A. Nemirovskii, "Interior-Point Polynomial Algorithms in Convex Programming", Philadelphia, USA: SIAM, 1994.

[i.55] E.S. Levitin and B. T. Polyak, "Constrained minimization methods", USSR Computational Mathematics and Mathematical Physics, vol. 6, pp. 1-50, 1966.

[i.56] P.H. Tan, L.K. Rasmussen, T.J. Lim, "Constraint Maximum-Likelihood Detection in CDMA", IEEE Trans. on Comm., vol. 49, no. 1, pp. 142-153, January 2001.

[i.57] A. Yener, R.D. Yates, S. Ulukus, "A nonlinear programming approach to CDMA multiuser detection", in Proc. Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA, October 1999, pp/ 1579-1583.

[i.58] B. H. Ahn, "Iterative methods for linear complementary problems with upper bounds on primary variables", Math. Program., vol. 26, no. 3, pp. 295-315, 1983.

[i.59] L. Brunel, J.J. Boutros, "Lattice Decoding for Joint Detection in Direct-Sequence CDMA Systems", IEEE Trans on Inf. Th., Vol. 49, No. 4, April 2003.

[i.60] E. Viterbi, J. Boutros, "A Universal Lattice Code Decoder for Fading Channels", IEEE Trans on Inf. Th., Vol. 45, No. 5, July 1999.

[i.61] J.R. Shewchuck (1994): "An Introduction to the Conjugate Direction Method Without the Agonizing Pain".

NOTE: See http://www.icos.ethz.ch/.

[i.62] M.J.Juntii, B.Aazhang, J.O.Lilleberg, "Iterative Implementation of Linear Multiuser Detection for Dynamic Asynchronous CDMA systems", IEEE Trans. On Comm., Vol. 46, No. 4, pp. 503-508, April 1998.

[i.63] R.I. Hernadez, R. Baghaie, K. Kettunen, "Implementation of Gram-Schmidt Conjugate Direction and Conjugate Gradient Algorithms", in Proceedings IEEE Finnish Signal Processing Symposium, FINSIG'99, Oulu, Finland, pp. 165-169, May 1999.

[i.64] R. Lupas and S. Verdu, "Near-Far Resistance of Multiuser Dectors in Asynchronous Channels", IEEE Trans. on Comm., Vol. 38, No. 4, pp. 496-508, April 1990.

[i.65] M. Honig, U. Madhow and S. Verdu, "Blind adaptive multiuser detection", IEEE Trans. Inform. Theory, vol. 41, pp. 944-960, July 1995.

[i.66] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in DS/CDMA system", IEEE Journal on Selected Areas in Communications, vol. 12, no. 5, pp. 796-807, June 1994.

[i.67] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code division multiple-access communications", IEEE Transactions on Communications, vol. 38, no. 4, pp. 509-519, April 1990.

[i.68] M. K. Varanasi and B. Aazhang, "Near-optimum detection in synchronous code-division multiple-access communications", IEEE Transactions on Communications, vol. 39, no. 5, pp. 725-736, May 1991.

[i.69] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels", IEEE Trans. Commun., vol. 43, pp. 421-433, February-April 1995.

[i.70] M. K. Varanasi, "Decision feedback multiuser detection: A systematic approach", IEEE Trans. Inform. Theory, vol. 45, pp. 219-240, January 1999.

[i.71] M. K. Varanasi, "Group detection for synchronous Gaussian code-division multiple-access channels", IEEE Trans. Inform. Theory, vol. 41, pp. 1083-1096, July 1995.

[i.72] D. Divsalar, M. K. Simon and D. Raphaeli, "Improved Parallel Interference Cancellation for CDMA", IEEE Trans. on Comm., Vol. 46, No.2, pp. 258-268, February 1998.

[i.73] L. K. Rasmussen, T. J. Lim, A. Johansson, "A Matrix-Algebraic Approach to Successive Interference Cancellation in CDMA", IEEE Transactions on Communications, Vol. 48, No. 1, January 2000, pp. 145-151.

[i.74] L. B. Nelson, H. V. Poor, "Iterative Multiuser Receivers for CDMA Channels: An EM-Based Approach", IEEE 2 Transactions on Communications, Vol. 44, No. 12, December 1996, pp. 1700-1710.

[i.75] X. Zhang, D. Brady, "Asymptotic Multiuser Efficiencies for Decision-Directed Multiuser Detectors", IEEE Transactions on Information Theory, Vol. 44, No. 2, Mar. 1998, pp. 502-515.

[i.76] J. Luo, K. Pattipati, P. Willet, "Optimal Grouping Algorithm for a Group Decision Feedback Detector in Synchronous CDMA Communications", IEEE Trans. Comm., Vol. 51, No. 3, pp. 341-346, March 2003.

[i.77] V. J. Rayward-Smith, I. H. Osman, C. R. Reeves, and G. D. Smith, editors, Modern Heuristic Search Methods. John Wiley and Sons, 1996.

[i.78] J. Luo, G. Levchuk, K. Pattipati, and P. Whillett, "A class of coordinate descent methods for multiuser detection", in Proc. 2000 IEEE Int. Conf. Acoust., Speech, and Signal Processing, pp. 2853-2856, 2000.

[i.79] Y. Sun, "A family of linear-complexity likelihood ascent search detectors for CDMA multiuser detection", in Proc. IEEE Int. Symp. Spread Spectrum Techniques and Application, (New Jersey, USA), pp. 713-717, September 2000.

[i.80] S. Abedi and R. Tafazolli, "Genetic multiuser receiver for code division multiple access communications", IEE Electron. Lett., vol. 36, pp. 1957-1958, November 2000.

[i.81] K. Yen and L. Hanzo, "Hybrid genetic algorithm based detection schemes for synchronous CDMA systems", in Proc. IEEE VTC (Tokyo, Japan), pp. 1400-1404, May 2000.

[i.82] F. Glover, "Future Paths for Integer Programming and Links to Artificial Intelligence", Computers and Operations Research, Vol. 13, pp. 533-549, 1986.

[i.83] R. Golden, "Mathematical Methods for Neural Network Analysis and Design", The MIT Press, Cambridge, MA, 1996.

[i.84] Y. Bar-Shalom and X. R. Li, "Estimation and Tracking: Principles, Techniques and Software". Dedham, MA: Artech House, 1993.

[i.85] J. Luo, K. Pattipati, P. Willett, F. Hasegawa, "Near-Optimal Multiuser Detection in Synchronous CDMA using Probabilistic Data Association", IEEE Comm. Letters, Vol. 5, No. 9, pp. 361-363, September 2001.

[i.86] Y. Jia, C. Andrieu, R. P. Piechocki, M. Sandell, "Gaussian Approximation Based Mixture Reduction for Near Optimum Detection in MIMO Systems", IEEE Comm. Letters, Vol. 9, No. 11, November 2005.

[i.87] X. Wang, H.V. Poor, "Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA", IEEE Trans. on Comm. Vol. 47, No. 7, July 1999.

[i.88] Tan, L. Rasmussen, J. Luo, "Iterative Multiuser Decoding Based on Probabilistic Data Association", IEEE International Symposium on Information Theory, Yokohama, Japan, June 2003.

NOTE: See Iterative Multiuser Decoding Based on Probabilistic Data Association.

[i.89] A. Sanderovich, M. Peleg, and S. Shamai (Shiz), "LDPC Coded MIMO Multiple Access with Iterative Joint Decoding", IEEE Trans. on Inform. Theory, vol. 51, No. 4, April 2005.

[i.90] M. Sellathurai, P. Guinand, J. Lodge, "Space-Time Coding in Mobile Satellite Communications Using Dual-Polarized Channels", IEEE Trans. on Veh. Technology, vol. 55, January 2006.

[i.91] D. Zheng, P. Karabinis, "Adaptive Beam-Forming with Interference Suppression in MSS with ATC", 23rd International Communications Satellite Systems Conference (ICSSC), Rome, September 27, 2005.

[i.92] P. Karabinis, S. Dutta, W.W. Chapman, "Interference Potential to MSS due to Terrestrial Reuse of Satellite Band Frequencies", 23rd International Communications Satellite Systems Conference (ICSSC) in Rome, September 27, 2005.

[i.93] A.F. Naguib, N. Seshadri and A.R. Calderbank, "Increasing Data Rate over Wireless Channels. Space Time Coding and Signal Processing for High Data Rate Wireless Communications", IEEE Signal Processing Magazine, May 2000.

[i.94] J.D. Parsons. "The Mobile Radio Propagation Channel". Wiley, 2nd edition, 2000.

[i.95] S.M. Alamouti, "A simple transmit diversity technique for wireless communications", IEEE Journal on Selected Areas in Communications, 16:1451-1458, October 1998.

[i.96] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-time block codes from orthogonal designs", IEEE Transactions on Information Theory, 45:1456-1467, July 1999.

[i.97] V. Tarokh, N. Sheshadri, and A.R. Calderbank, "Space-time codes for high-data rate wireless communication: Performance criteria and code construction", IEEE Transactions on Information Theory, 44(3):744-765, March 1998.

[i.98] H. Wang and X.-G. Xia, "Upper bounds of rates of space-time block codes from complex orthogonal designs", in Proc. IEEE Int. Symp. Inform. Theory, Lausanne, Switzerland, June 30-July 5 2002, p. 303.

[i.99] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding", IEEE Trans. Wireless Commun., vol. 2, pp. 294-309, March 2003.

[i.100] X. Ma, G.B. Giannakis, "Full-Diversity Full-Rate Complex-Field Space-Time Coding", IEEE Trans. on Signal Processing, Vol. 51, No. 11, pp. 2917- 2930, November 2003.

[i.101] E. Lingskog, A.Paulraj, "A transmit diversity scheme for channels with Intersymbol interference", in Proc. ICC, vol.1, pp. 307-311, 2000.

[i.102] S. Zhou, G.B.Giannakis, "Space-time coding with maximum diversity gains over frequency-selective fading channels", IEEE Signal Processing Letters, vol. 8, pp. 269-272, October 2001.

[i.103] W. H. Gerstacker, F. Obernosterer, R. Schober, A. Lehmann, A. Lampe, P. Gunreben, "Widely Linear Equalization for Space-Time Block-Coded Transmission over Fading ISI Channels", IEEE 56th Veh. Tech. Conf., Vol. 1, pp 238-242, 24-28 September 2002.

[i.104] W. H. Gerstacker, R. Schober, A. Lampe, "Equalization with Widely Linear Filtering", in Proc. of Int. Symposium on Inf. Theory, Washington D.C., June 2001, p. 265.

[i.105] J. Hudson, "Adaptive Array Principles", Peter Peregrinus Ltd, IEEE Electromagnetic Wave Series 11, 1981.

[i.106] J. Winters, "Optimum combining in digital mobile radio with co-channel interference", IEEE Jour. Selected Areas in Comm., (4):528-539, 1984.

[i.107] S. Haykin, "Adaptive Filter Theory", Prentice Hall, 1996.

[i.108] A.B. Gershman, "Robust adaptive beamforming in sensor arrays", Int. Journal Electron. Comm.,vol 53, pp.305-314, December 1999.

[i.109] R.J.Piechocki, "Space-time Techniques for W-CDMA and OFDM Wireless Networks", Ph.D. thesis, University of Bristol, March 2002.

[i.110] H.Cox, R.M. Zeskind, M.H. Owen, "Robust adaptive beamforming", IEEE Trans. Acoust., Speech, Signal Processing, vol 35, pp.1365-1376, October 1987.

[i.111] D.D. Deldman, L.J. Griffiths, "A projection approach to robust adaptive beamforming", IEEE Trans. Signal Processing, vol. 42, pp. 867-876, April 1994.

[i.112] L. Chang, C.C. yeh, "Performance of DMI and eigenspace-based beamformers", IEEE Trans. Antennas Propagation, vol. 40, pp. 1336-1347, November 1992.

[i.113] S.A. Vorobyov, A.B. Gershman, Z-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization via second-order cone programming", in Proc. ICASSP2002, Orlando, FL, May 2002.

[i.114] A.B. Gershman, S.A. Vorobyov, Z-Q. Luo, "An application of second-order cone programming to robust adaptive beamforming", in Proceedings ICOTA2001, Hong Kong, December 2001, volume1, pp. 308-315.

[i.115] L.Häring, A.Gershman, T.Kaiser "Robust Adaptive Beamforming in TD-SCDMA", Second Joint Symposium on Optoelectronic Devices and Circuits, 10-16 March 2002, Stuttgart, Germany.

[i.116] H.Krim, M.Viberg, "Two decades of arrays signal processing research", IEEE Trans on Comm., 11:1214-1224, 1972.

[i.117] J.Capon, "High resolution frequency-wavenumber spectrum analysis", Proc. Of the IEEE, August 1969.

[i.118] A.Swindlehurst and P.Stoica, "Maximum Likelihood methods in radar array signal processing", Proc. of the IEEE, 1998.

[i.119] A.Paulraj, T.Kailath, "A subspace rotation approach to signal parameter estimation", Proc. of the IEEE, 1986.

[i.120] M.Viberg, B. Ottersten, T.Kailath, "Detection and estimation in sensor arrays based on weighted subspace fitting", IEEE Trans. SP, 39(11):2436-2449, 1991.

[i.121] A.Belouchrani, M.G.Amin, "Jammer mitigation in spread spectrum communications using blind source separation", Signal Processing, 80, pp. 723-729, 2000.

[i.122] T. Ristaniemi, K. Raju, J. Karhunen, "Jammer mitigation in DS-CDMA Array system using Independent Component Analysis", IEEE ICC2002.

[i.123] C. Berrou, A. Glaveux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes", Proc. Int. Conf. Comm., pp. 1064-1070, 1993.

[i.124] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. IEEE Transactions on Information Theory, IT-20:284--287, March 1974.

[i.125] S. Benedetto, G. Montorsi, "Unveiling Turbo Codes: some results on parallel concatenated coding schemes", IEEE Transactions Information Theory, vol 42, no.2, pp. 409-428, March 1996.

[i.126] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding", IEEE Transactions on Information Theory_44_3:906-926, May 1998.

[i.127] Robertson, P., Villebrun, E., and Hoeher, P.: "A comparison of optimaland sub-optimal MAP decoding algorithms operating in the log domain", Proc. IEEE Int. Conf. Communications, Seattle, WA, USA, June 1995,pp. 1009-1013.

[i.128] Tan, J., and Stuber, G.L.: "New SISO decoding algorithms", IEEE Trans. Commun., 2003, 51, (6), pp. 845-848.

[i.129] M. Arzel, et al., "Semi-Iterative Analog Turbo Decoding", IEEE Trans. on Circuits and Systems-1: Regular Papers, vol. 54, no. 6, June 2007.

[i.130] R.G. Gallager, "Low-density parity-check codes", IRE Trans. Inf. Theory, vol. 8, pp.21-28, January 1962.

[i.131] D. J. C. MacKay and R. M. Neal, "Near-Shannon-limit performance of low-density parity-check codes", Electr. Lett., vol. 32, pp. 1645-1646, August 1996.

[i.132] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices", IEEE Trans. Inform. Th., vol. 45, pp. 399-432, March 1999.

[i.133] Y. Kou, et al., "Low density parity check codes based on finite geometries: a rediscovery and new results", IEEE Trans. Inform. Theory, vol. 47, no. 7, November 2001.

[i.134] Z. Liu, D. A. Pados, "A decoding algorithm for finite-geometry LDPC codes", IEEE Trans. Commun., Vol. 53, No. 3, March 2005.

[i.135] Wiberg, N.,Loeliger, H.A., and Koetter, R.: "Codes and iterative decoding on general graphs", Eur. Trans. Telecommun., 1995, 6, pp. 513-526 .

[i.136] S. Hemati, A. Banihashemi, "Dynamics and Performance Analysis of Analog Iterative Decoding for Low-Density Parity-Check (LDPC) Codes", IEEE Trans. on Comm., vol. 54, no.1, January 2006.

[i.137] S. Lin and D.J. Costello, Jr., "Error Control Coding", Prentice-Hall, 2nd ed., May 2004.

[i.138] B. Li, "Reliable Multicast Transmissions using Forward Error Correction and Automatic Retransmission Requests", Canadian Conference on Electrical and Computer Engineering, Vol. 2, pp. 1145-1150, May 2001.

[i.139] J. Nonnenmacher, E.W. Biersack, and D. Towsley, "Parity-Based Loss Recovery for Reliable Multicast Transmission", IEEE/ACM Transactions on Networking, Vol. 6, Issue 4, pp. 349-361, August 1998.

[i.140] T. Lestayo, M. Fernandez, and C. Lopez, "Adaptive Approach for FEC Reliable Multicast", IEEE Electronics Letters, Vol. 37, Issue. 22, pp. 1333-1335, October 2001.

[i.141] J.W. Byers, M. Luby, and M. Mitzenmacher, "A Digital Fountain Approach to Asynchronous Reliable Multicast", IEEE Journal on Selected Areas in Communications, Vol. 20, Issue 5, pp. 1528-1540, October 2002.

[i.142] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols", ACM Computer Communication Review, Vol. 27, Issue 2, pp. 24-36, April 1997.

[i.143] L. Ping, X, Huang, and N. Phamdo, "Zigzag Codes and Concatenated Zigzag Codes", IEEE Transactions on Information Theory, Vol. 47, Issue 2, pp. 800-807, February 2001.

[i.144] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor Codes for Reliable Download Delivery in Wireless Broadcast Systems", Proceedings of IEEE Consumer Communications and Networking Conference, Las Vegas, Nevada, USA, pp. 192-197, January 2006.

[i.145] A. Shokrollahi, "Raptor codes", IEEE Transactions on Information Theory, Vol. 52, Issue 6, pp. 2551-2567, June 2006.

[i.146] M. Luby, "LT codes", Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, pp. 271-280, November 2002.

[i.147] M. Luby et al., "The use of FEC in Reliable Multicast", IETF RFC 3453, December 2002.

[i.148] P. Maymounkov and D. Mazieres, "Rateless Codes and Big Downloads", Proceedings of the 2nd International Workshop on Peer-to-Peer Systems, pp. 247-255, February 2003.

[i.149] A. Yamaguchi, H. Kurosu, S. Fukumoto, and K. Iwasaki, "Fault-Tolerance Design for Multicast Using Convolutional-Code-Based FEC and its Analytical Evaluation", Proceedings of IEICE Transactions on Information and Systems, Vol. E85-D, No. 5, pp. 864-873, May 2002.

[i.150] S.B. Wicker, "Error Control Systems for Digital Communication and Storage", Prentice Hall, January 1995.

[i.151] P. Basu and K. Kanchanasut, "A Reliable Multicast Protocol for Unidirectional Satellite Link", Proceedings of the 2003 Symposium on Applications and the Internet, pp. 390-393, January 2003.

[i.152] C. Wang and V.C.M. Leung, "Performance Evaluations of SRMTP for Reliable Multicasting over Satellite Networks", Proceedings of the IEEE Wireless Communications and Networking Conference, Vol. 3, pp. 1813-1818, March 2003.

[i.153] G. Fairhurst, "SAT-RMTP: A Protocol for Reliable Multicast Content Distribution via VSAT", Proceedings of the IEE Seminar on End-to-End Challenges of Broadband via Satellite, pp. 37-44, January 2005.

[i.154] WiMAX Forum, "Mobile WiMAX - Part I: A Technical Overview and Performance Evaluations," August 2006.

NOTE: Available at: http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part1_Overview_and_Performance.pdf

[i.155] WiMAX Forum, "Mobile WiMAX - Part II: A comparative Analysis", May 2006.

NOTE: Available at: http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part2_Comparative_Analysis.pdf

[i.156] ETSI TS 136 211: "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211, Release 8)".

[i.157] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group", November 2002.

NOTE: Available online: http://www.fcc.gov/sptf/reports.html.

[i.158] Q. Zhao and B. M. Sadler, "A Survey of Dynamic Spectrum Access: Signal processing, networking, and regulatory policy", IEEE Signal Processing Magazine, May 2007.

[i.159] R. Coase, "The federal communications commission", J. Law and Economics, pp. 1-40, 1959.

[i.160] D. Hatfield and P. Weiser, "Property rights in spectrum: Taking the next step", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks, November 2005.

[i.161] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, "DRiVE-ing to the Internet: Dynamic radio for IP services in vehicular environments", in Proc. 25th Annual IEEE Conf. Local Computer Networks, November 2000.

[i.162] Y. Benkler, "Overcoming agoraphobia: Building the commons of the digitally networked environment", Harv. J. Law Tech, vol. 11, no. 2, 1998.

[i.163] W. Lehr and J. Crowcroft, "Managing shared access to a spectrum commons", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks, November 2005, pp. 420-444.

[i.164] C. Raman, R. Yates, and N. Mandayam, "Scheduling variable rate links via a spectrum server", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks, 2005.

[i.165] O. Ileri, D. Samardzija, and N. Mandayam, "Demand Responsive Pricing and Competitive Spectrum Allocation via a Spectrum Server", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks - 2005.

[i.166] S. Chung, S. Kim, J. Lee, and J. Cioffi, "A game-theoretic approach to power allocation in frequency-selective Gaussian interference channels," in Proc. IEEE Int. Symp. Information Theory, June 2003.

[i.167] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks, - 2005.

[i.168] J. Huang, R. Berry, and M. Honig, "Spectrum sharing with distributed interference compensation", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks 2005.

[i.169] J. Mitola, "Cognitive radio for flexible mobile multimedia communications", Proc. IEEE Int. Workshop Mobile Multimedia Communications, 1999.

[i.170] Report and Order and Notice of Proposed Rulemaking, FCC 03-15, Flexibility for Delivery of Communications by Mobile Satellite Service Providers in the 2 GHz Band, the L-Band, and the 1.6/2.4 Bands, IB Docket No. 01-185, Adopted: January 29, 2003, Released: February 10, 2003.

[i.171] P. D. Karabinis, S. Dutta, W. Chapman, "Interference Potential to MSS due to Terrestrial Reuse of Satellite Band Frequencies", presented at the 23rd International Communications Satellite Systems Conference (ICSSC) in Rome, September 27, 2005.

[i.172] D. Zheng, P. D. Karabinis, "Adaptive Beam-Forming with Interference Suppression in MSS with ATC", presented at the 23rd International Communications Satellite Systems Conference (ICSSC) in Rome, September 27, 2005.

[i.173] J. Mitola, Software Radios: Wireless Architecture for the 21st Century. New York: Wiley, 2000.

[i.174] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications", IEEE Journal on Selected Areas in Communications, vol. 23, No. 2, February 2005.

[i.175] M. Milliger et al., Eds., "Software Defined Radio: Architecture, Systems and Functions"Wiley, New York, 2003.

[i.176] J. Neel, R. Menon, J. H. Reed and A. B. MacKenzie, "Using Game Theory to Analyze Physical Layer Cognitive Radio Algorithms", in Proc. Research Conf. On the Econ., Tech. and Policy of Unlicensed Spectrum, Michigan State University, East Lansing, Michigan, 16 - 17th May 2005.

[i.177] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework", IEEE J. Selected Areas in Commun.: Special Issue on Adaptive, Spectrum Agile and Cognitive Wireless Networks, vol. 25, no. 3, April 2007.

[i.178] Q. Zhao, L. Tong, and A. Swami, "Decentralized cognitive MAC for dynamic spectrum access", in Proc. IEEE Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN), November 2005.

[i.179] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors", in Proc. IEEE Asilomar Conf. Signals, Systems, Computers, 2006.

[i.180] U. Berthold and F.K. Jondral, "Guidelines for designing OFDM overlay systems", in Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks, November 2005.

[i.181] Cabric, D., Mishra, S.M., Brodersen, R.W. "Implementation issues in spectrum sensing for cognitive radios",Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference Vol 1, 7-10 November 2004.

[i.182] D.Cabric, M.S.W.Chen, D.A.Sobel, J.Yang, and R.W.Broadersen, "Future wireless systems: UWB, 60GHz, and Cognitive radios", Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005, September 2005.

[i.183] J.O.Juntunen, "Think ahead Stay ahead-strategic visions on future research directions in the wireless field", WWRF 15 Meeting, Paris, France, December 8, 2005.

[i.184] N. Hoven, R.Tandra, and A.Sahai, "Some Fundamental Limits on Cognitive Radio", Wireless Foundations, EECS, University of California at Berkeley, February 11, 2004.

[i.185] M.P.Olivieri, G.Barnett, A.Lackpour, A.Davis, and P.Ngo, "A scalable dynamic spectrum allocation system with interference mitigation for teams of spectrally agile software defined radios", New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on, November 2005.

[i.186] H.Tang, "Some physical layer issues of wide-band cognitive radio systems", New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on, November 8-11, 2005.

[i.187] M. Gandetto, C. Regazzoni, "Spectrum Sensing: A Distributed Approach for Cognitive Terminals", IEEE Journal on Selected Areas in Comm., vol. 25, No. 3, April 2007.

[i.188] Y. Hur, J. Park, W. Woo, K. Lim, C.-H. Lee, H.S. Kim, J. Laskar, "A Wideband Analog Multi-Resolution Spectrum Sensing (MRSS) Technique for Cognitive Radio (CR) Systems", IEEE Intern. Symp. On Circuits and Systems (ISCAS) 2006, May 2006.

[i.189] Z. Tian, G. B. Giannakis, "A Wavelet Approach to Wideband Spectrum Sensing for Cognitive Radios", 1st Intern. Conf. on Cognitive Radio Oriented Wireless Networks and Communications\ (CROWNCOM), 8-10 June 2006.

[i.190] S.M.Mishra, A.Sahai, and R.W.Brodersen, "Cooperative sensing among cognitive radio", ICC, 2006.

[i.191] N. Devroye, P. Mitran, V. Tarokh, "Limits on Communications in Cognitive Radio Channel", IEEE Comm. Magazine, June 2006.

[i.192] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA", IEEE Trans. Commun., vol. 42, pp. 3178-3188, December 1994.

[i.193] X. Wang and H. V. Poor, "Group-blind multiuser detection for uplink CDMA", IEEE J. Select. Areas Commun., vol. 17, pp. 1971-1984, November 1999.

[i.194] Y. Hongbo, R. Sumit, "Parallel Interference Cancellation for Uplink Multirate Overlay CDMA Channels", IEEE Trans. on Comm., Vol. 53, No. 1, January 2005.

[i.195] Max H. M. Costa, "Writing on dirty paper", IEEE Transactions on Information Theory, vol. 29, no. 3, pp. 439-441, May 1983.

[i.196] R. Hoshyar, F. Wathan, R. Tafazolli, "Novel Low-Density Signature Structure for Synchronous DS-CDMA Systems", IEEE GLOBECOM, 2006.

[i.197] Lo, T.K.Y., "Maximum ratio transmission", Communications, IEEE Transactions on, vol.47, no.10, pp.1458-1461, October 1999.

[i.198] M. Schubert, H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints", Vehicular Technology, IEEE Transactions on, vol 53, issue 1, pp. 18-28, January 2004.

[i.199] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for cancelling known interference", IEEE Trans. Inf. Theory, vol. 51, no. 11, pp. 3820-3833, November 2005.

[i.200] T. Philosof, U. Erez, and R. Zamir, "Combined shaping and precoding for interference cancellation at low SNR", in Proc. IEEE Int. Symp. Information Theory, Yokohama, Japan, June/July 2003.

[i.201] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme", IEEE Trans. on Inform. Theory, October 2005.

[i.202] A. Bennatan, D. Burshtein, G. Caire, S. Shamai (Shitz), "Superposition coding for side-information channels", Information Theory, IEEE Transactions on, Vol. 52, No. 5.pp. 1872-1889, 2006.

[i.203] M. Uppal, V. Stankovic, Z. Xiong, "Code Designs for MIMO Broadcast Channels", IEEE Intern. Symp. on Inf. Theory, July 2006.

[i.204] A. D. Dabbagh, D. J. Love, "Precoding for Multiple Antenna Gaussian Broadcast Channels With Successive Zero-Forcing", IEEE Trans. on Signal Processing, vol. 55, no. 7, July 2007.

[i.205] P. Piantanida, P. Duhamel, "Dirty-paper Coding without Channel Information at the Transmitter and Imperfect Estimation at the Receiver", IEEE Intern. Conf. on Com, Glasgow, 24-28 June 2007.

[i.206] N. Devroye, P. Mitran, and V. Tarokh, "Cognitive Multiple Access Networks", Proc. IEEE Int'l. Symp. Info. Theory, September 2005.

[i.207] D. Cabric and R. W. Brodersen, "Physical Layer Design issues Unique to Cognitive Radio Systems", 16th IEEE International Symposium on Personal Indoor and Mobile Radio Communications, (PIMRC 2005), September 11-14, 2005.

[i.208] H.Nie and P.T.Mathiopoulos, "Adaptive Prediction and Cancellation Digitization Method for Wideband Multistandard Software Radio Base-Station Receivers", IEEE Transactions on Vehicular Technology, Vol. 55, issue. 3, pp 887-902, May 2006.

[i.209] A.Mahajan, M.Agarwal, and A.K.Chaturvedi, "A Novel Method for Down-Conversion of Multiple Bandpass Signals", IEEE Transactions on Wireless Communications, Vol. 5, No. 2, pp 427-434, February 2006.

[i.210] Wireless World Research Forum, Technologies for the Wireless Future, edit by R.Tafazolli, Wiley, 2005.

[i.211] J. Wang, H .Zhai, Y. Fang, and M.C. Yuang "Opportunistic media access control and rate adaptation for wireless ad hoc networks", IEEE International Conference on Communications, 2004.

[i.212] F. Capar, I. Martoyo, T. Weiss, F. Jondral "Analysis of Coexistence Strategies for Cellular and Wireless Local Area Networks", IEEE Vehicular Technology Conference 2003.

[i.213] S.M. Mishra, D. Cabric, C. Chang, D. Willkomm, B. Van Schewick, A. Wolisz and R.W. Brodersen "A real time cognitive radio testbed for physical and link layer experiments", First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, November 2005.

[i.214] P. Subrahmanyam, M. Cummings "PERSPECTIVES ON A METALANGUAGE FOR RECONFIGURABLE WIRELESS SYSTEMS", Software Defined Radio Forum SDRF-05-I-0003.

[i.215] S. Mangold, Z. Zhong, K. Challapali and C. Chun-Ting "Spectrum agile radio: radio resource measurements for opportunistic spectrum usage", IEEE Global Telecom. Conf., 2004.

[i.216] T.A. Weiss, F.K. Jondral "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency", IEEE Communications Magazine, Vol. 42, No. 3 pp. 8-14, March 2004.

[i.217] L. Berlemann, S. Mangold, B. H. Walke, "Policy-based Reasoning for Spectrum Sharing in Cognitive Radio Networks", Proc. IEEE DySPAN, 8th - 11th Nov. 2005, Baltimore, MD, USA.

[i.218] A.Sahai, D. Cabric, R. W. Brodersen, N.Hoven, S. M. Mishra, and R.Tandra, "Spectrum sensing-fundamental limits and practical challenges", First IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), November 2005.

[i.219] H. Yanikomeroglu, "Fixed and mobile relaying technologies for cellular network", 2nd Workshop on Application and Services in Wireless Networks, Paris, France, July 2002.

[i.220] Cover and A. El Gamal, "Capacity Theorems for the Relay Channel", IEEE Transactions on Information Theory, Vol. IT-25, No. 5, pp. 572-584, September 1979.

NOTE: Available at: Capacity Theorems for the Relay Channel,

[i.221] J.N. Laneman, D.N.C. Tse, and G.W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior", IEEE Trans. Inform. Theory, vol. 50, pp.3062-3080, 2004.

[i.222] M.A. Khojastepour, A. Sabharwal, and B. Aazhang, "On capacity of Gaussian "cheap" relay channel", GLOBECOM, 2003.

[i.223] Y. Liang, and V.V. Veeravalli, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity". IEEE Trans. Inform. Theory, vol 51, pp.3284-3289, 2005.

[i.224] R.B. Nabar, H. Bolcskei, F.W. Kneubuhler, "Fading Relay Channels: Performance Limits and Space-Time Signal Design", IEEE J. Select. Areas Commun., vol. 22, no.6, pp1099-1109, August 2004.

[i.225] K. Azarian, H. El Gamal, Hesham, P. Schniter, Philip, "On the Achievable Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels", IEEE Trans. Inform. Theory, vol. 51, no. 12, pp.4152-4172, 2005.

[i.226] P. Mitran, H. Ochia, V. Tarokh, "Space-Time Diversity Enhancements using Cooperative Communications", IEEE Trans. Inform. Theory, volume 51, no. 6, pp.2041-2057, 2005.

[i.227] Slepian and Wolf, "Noiseless Coding of Information Sources", IEEE Trans. on Inform. Theory, vol. IT-19, no. 4, pp. 471-480, July 1973.

[i.228] Wyder and Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", IEEE Trans. on Inf. Theory, vol. IT-22, no. 1, pp. 1-10, January 1976.

[i.229] A. Sendonaris, E. Erkip, B. Aazhang, "User cooperation diversity Part I. System description", IEEE Trans. Commun., vol.51, pp.1927-1938, 2003.

[i.230] G.F. Pedersen, M. Tartiere, M.B. Knudsen, "Radiation efficiency of handheld phones", Proc. IEEE 51st Vehicular Technology Conference (VTC), 2000.

[i.231] R.G. Vaughan, J. A. Bach, "Channels, Propagation and Antennas for Mobile Communications", IEE Press, London, UK, 2003.

[i.232] P. Gupta, P.R. Kumar, "The capacity of wireless networks", IEEE Trans. Inform. Theory, vol. 46, no. 2, pp. 388-404, 2002.

[i.233] M. Grossglauser, D. Tse, "Mobility increases the capacity of ad hoc wireless networks", IEEE ACM Trans. on Networking, vol. 10, no.4, 2002.

[i.234] M. Dohler, E. Lefranc, H. Aghvami, "Space-Time Block Codes for Virtual Antenna Arrays", IEEE PIMRC 2002, Lisbon, Portugal, 2002.

[i.235] M. Dohler, E. Lefranc, H. Aghvami, "Virtual Antenna Arrays for Future Wireless Mobile Communication Systems", IEEE ICT 2002, Beijing, China, 2002.

[i.236] M. Dohler, Z. Zeng, H. Aghvami, "System Performance of a W-CDMA based Network with deployed Virtual Antenna Arrays", ICT2002, Beijing, China, June 2002.

[i.237] M. Dohler, J. Dominguez, H. Aghvami, "Link Capacity Analysis for Virtual Antenna Arrays", VTC Fall, Vancouver, Canada, September 2002.

[i.238] M. Dohler, H. Aghvami, "A step towards MIMO: Virtual Antenna Arrays", COST273, Barcelona, Catalonia, January 2003.

[i.239] M. Dohler, B. Rassool, H. Aghvami, "Performance Evaluation of STTCs for Virtual Antenna Arrays", VTC Spring 2003, Jeju, Korea.

[i.240] A. Kastrisios, M. Dohler, H. Agvami, "Influence of Channel Characteristics on the Performance of VAA with deployed STBCS", IEEE VTC Spring 2003, Jeju, Korea.

[i.241] T. Miyano, H. Murata, K. Araki, "Cooperative Relaying Scheme with Space Time Code for Multi-hop Communications among Single Antenna Terminals", IEEE GLOBECOM "04, November-3 December 2004.

[i.242] J. N. Laneman, G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks",IEEE Trans. on Information Theory, Vol. 49, no. 10, pp.2415-2425, October 2003.

[i.243] A. Sendonaris, E. Erkip, B. Aazhang, "User cooperation diversity Part II. Implementation aspects and performance analysis", IEEE Trans. Commun., vol.51, pp.1939-1948, 2003.

[i.244] M.A. Khojastepour, N. Ahmed, B. Aazhang, "Code design for the relay channel and factor graph decoding", in Proc. of Asilomar Conference, 2004.

[i.245] A. Chakrabarti, A. de Baynast, A. Sabharwal, A., B. Aazhang, "LDPC code design for half-duplex decode-and-forward relaying", In Proc. of the Allerton Conference, Monticello, 2005.

[i.246] O.-S. Shin, A. Chan, H.T. Kung, V. Tarokh, "Design of an OFDM cooperative space-time diversity system", IEEE Trans. on Vehicular Technology, vol. 56, no. 4, July 2007.

[i.247] S. Hanes, H. Yanikomeroglu, B. Hashem, "Diversity and AMC (adaptive modulation and coding) aware routing in TDMA multi-hop networks", IEEE GLOBECOM, October 2003.

[i.248] H. Viswanathan, S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling", IEEE VTC, October 2003.

[i.249] M. Lott, M. Weckerle, W. Zirwas, H. Li, and E Schulz, "Hierarchical cellulr multi-hop networks", 5th European Personal Mobile Comm. Conference (EPMCC 2003), April 2003.

[i.250] V. Sreng, H. Yanikomeroglu, D.D. Falconer, "Relayer selection strategies in cellular networks with peer-to-peer relaying", IEEE VTC, Fall 2003.

[i.251] W. Zirwas, M. Lampe, H. Li, M. Lott, M. Weckerle, E. Schulz, "Radio Resource Management in Cellular Multi-hop Networks", MoMuC, 2003.

[i.252] B. Walke, R. Pabst, et. al., "Relay-based Deployment Concepts for Wirless and Broadband Cellular Radio", Wireless World Research Forum (WWRF), 2003.

NOTE: Available at http://www.wireless-world-research.org/.

[i.253] K. Gosse, et. al., "Short Range Communication with New Radio Air Interfaces based on Next Genetation of WLANs, WPANs and WBANs", WWRF, 2003.

NOTE: Available at http://www.wireless-world-research.org/.

[i.254] C.E. Perkins, "Ah-hoc Networking". Addison Wesley, December 2000.

[i.255] M. Frodigh, P. Johansson, P. Larsson, "Wireless Ad Hoc Networking - The Art of Networking without a Network", Ericsson Review, vol. 4, pp. 248-263, 2000.

[i.256] I.F. Akyildiz, W. Su, et. al., "Wireless sensor networks: a survey", Computer Networks, vol. 38, 2002.

[i.257] T. Clausen, P. Jacquet, "Optimized Link State Routing Protocol", IETF January 2003.

[i.258] C. Perkins, E.M. Royer, "Ad-hoc On-Demand Distance Vector Routing", on Proc. of WMCSA '99, 1999.

[i.259] D.B. Johanson, D.A. Maltz, and J. Broch, "DSR: The Dinamic Source Routing Protocol", chapter Chapter 5, pages 139-172. Addison-Wesley, 2001.

[i.260] C. Perkins, "IP Mobility Support", IETF RFC 2002, 2996.

[i.261] J. Xi, C. Bettstetter, "Wireless Multi-hop Internet Access: Gateway Discovery, Routing, and Addressing", in Proceedings of the World Wireless Congress, San Francisco, USA 2002.

[i.262] K. Weniger, M. Zitterbat, "IPv6 Autoconfiguration in Large Scale Mobile Ad-hoc Networks", in Proc. of European Wireless 2002, Florence, Italy, February 2002.

[i.263] T. Clausen, et. al. "Optimized Link State Routing Protocol", Proc. of INMIC '01, 2001.

[i.264] M. Mauve, A. Widmer, and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks", IEEE Network, vol. 15, no. 6 pp.30-39, 2001.

[i.265] X. Hong, K. Xu, M. Gerla, "Scalable routing protocols for mobile ad-hoc networks", IEEE Network, vol. 16, no.4, pp.11-21, 2002.

[i.266] A. Tsirios, Z.J. Haas, "Multipath routing in the presence of frequent topological changes. IEEE Communications Magazine", vol.39, no. 11, pp.132-138, 2004.

[i.267] Y. Zhenzhen, Y. Hua, "Networking by parallel relays: diversity, lifetime and routing overhead", in Proc. of Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1302-1306, 2004.

[i.268] D.S.J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high throughput path metric for multi-hop wireless routing", Proc. of MobiCom '03, 2003.

[i.269] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "Protocol for IEEE 802.11 Wireless Networks", Proc. of BROADNETS 2004, San José, 2004.

[i.270] R. Draves, J. Padhye, B. Zill, "Comparison of routing metrics for static multi-hop ireless networks". In SIGCOMM '04: Proceedings of the 2004 conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, NewYork, 2004.

[i.271] I. Stojmenovic, "Handbook of Wireless Networks and Mobile Computing", Wiley-Interscience, 2002.

[i.272] IEEE 802.11: "Draft Supplement to Standard for Telecommunication and Information Exchange Between Systems - LAN/MAN Specific Requirements - Parti I: Wireless Medium Access Control (MAC) and physical layer (PHY) specification: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)", Draft Supplement to Standard IEEE 802.11, IEEE, New York, November 2001.

[i.273] S. Mangold, et. al. "IEEE 802.1 le Wireless LAN for Quality of Service", European Wireless 2002.

[i.274] IEEE Draft P802.15.3: "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPAN)", February 2003.

[i.275] S. Buchegger, J.-Y. Le Boudec, "The selfish node: Increasing routing security in mobile ad-hoc networks", Research Report RZ 3354, IBM Research, Zurich Research Laboratory, 8803 Ruschlikon, Switzerland, May 2001.

[i.276] L. Buttyan, and J.-P. Hubaux, "Nuglets: a virtual currency to stimulate cooperation in self-organized mobile ad-hoc networks" Technical Report DSC/2001/001, Institute for Computer Communications and Applications, Department of Communication Systems, Swiss Federal Institute of Technology, January 2001.

[i.277] J.P. Hubaux, L. Buttyan, S. Capkun, "The quest for security in mobile ad-hoc networks", Proceedings of the ACM Symposium on Mobile Ad-hoc Networking and Computing (MobiHOC), Long Beach, CA, ACM, October 2001.

[i.278] A. Perrig, et.al. "SPINS: Security protocols for sensor networks", 7th ACM International Conference on Mobile Computing and Networking, Rome, Italy, 2001.

[i.279] IEEE15web (2006). IEEE 802.15.

NOTE: Standard Web Site. http://www.ieee802.org/15/.

[i.280] IEEE 802.16-2004: "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems".

[i.281] Y.K. Kim, R. and Prasad, "4G Roadmap and Emerging Communication Technologies", Artech House, 2006.

[i.282] S. Frattasi, et. Al. "Defining 4G Technology from the User Perspective", IEEE Network Magazine, 2006.

[i.283] M. Katz, F.H.P. Fitzek, "On the Definition of the Fourth Generation Wireless Communications Networks: The Challenges Ahead", In International Workshop on Convergent Technology (IWCT) 2005, Oulu, Finland, 2005.

[i.284] A.H. Kupetz, K.T. Brown, "4G - A Look Into the Future of Wireless Communications". Rollings Business Journal, 2003.

[i.285] ITU ITU-R Recommendation M.1645 (2003): "Framework and overall objectives of the future development of imt-2000 and systems beyond imt-2000".

[i.286] IrDA (2006). Infrared data association (irda).

NOTE: Available at: http://www.irda.org.

[i.287] D.C. O'Brien, M. Katz, "Optical wireless communications within fourth-generation wireless systems", Journal of Optical Networking, vol. 4, no. 6, pp. 312-322, 2005.

[i.288] D.C. O'Brien, M. Katz, "White paper: Short-range optical wireless communications", WWRF 14th Meeting, pages 1-22, Chicago, IL, 2005.

[i.289] Y. Tanaka,T. Komine, S. Haruyama, M. Nakagawa, "Indoor visible light data transmission system utilizing white led lights", IEICE Trans. on Commun., vol. 86, no. 8, pp.2440-2454, 2003.

[i.290] Taylor Nelson Sofres (TNS), "Two-day battery life tops wish list for future all-in-one phone device. Technical report", 2005.

[i.291] S. Frattasi, F.H.P. Fitzek, A. Mitseva, R. Prasad, "A Vision on Services and Architectures for 4G", In 1st CTIF B3G/4GWorkshop, Aalborg, Denmark, 2005.

[i.292] S. Frattasi, R.L. Olsen, M. de Sanctis, F.H.P. Fitzek, R. Prasad, "Innovative Services and Architectures for 4G", Wireless Mobile Communication Systems. In IEEE ISWCS, Siena, Italy, 2005.

[i.293] ETSI TS 102 584: "Digital Video Broadcasting (DVB); DVB-SH Implementation Guidelines".

[i.294] ETSI TS 136 211: "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211 Release 8)".

[i.295] ETSI TS 136 212: "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding (3GPP TS 36.212 Release 8)".

[i.296] 3GPP TSG-RAN WG1 #48-bis, St. Julian's, Malta, March 26th - 30th, 2007.

[i.297] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks", Proceedings of the IEEE, vol. 95, no. 12, pp. 2410-2431, December 2007.

[i.298] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas", IEEE Trans. Inf. Theory, vol. 48, no. 6, pp. 1277-1294, June 2002.

[i.299] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks, Part I: Theoretical framework", IEEE Trans. Wireless Commun., vol. 4, no. 2, pp. 614-624, February 2005.

[i.300] Z. Shen, J. G. Andrews and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints", IEEE Trans. On Wireless Commun., vol. 4, pp. 2726-2736, November 2005.

[i.301] C. Wang, Z. Shen, B. L. Evans and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems", in Proceedings, IEEE Signal Processing Workshop, pp. 1-6, Austin TX, October 2004.

[i.302] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems", IEEE Journal on Selected Areas Commun., vol. 21, pp. 171-178, February 2003.

[i.303] R. Wang and V. K. N. Lau, "Cross layer design of downlink multi-antenna OFDMA systems with imperfect CSI for slow fading channels", IEEE Trans. Wireless Commun., vol. 6, pp. 2417-2421, July 2007.

[i.304] Y. Yao and G. B. Giannakis, "Rate-maximizing power allocation in OFDM based on partial channel knowledge", IEEE Trans. Wireless Commun.,vol. 4, pp. 1073-1083, May 2005.

[i.305] C. Wong and B. L. Evans, "OFDMA Resource Allocation for Ergodic Capacity Maximization with Imperfect Channel Knowledge", Proc. IEEE Int. Global Communications Conf., Washington, DC USA, November 26-30, 2007.

[i.306] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation"Proc., IEEE Veh. Technology Conf., pages 1085-89, Tokyo, May 2000.

[i.307] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks"IEEE J. Sel. Areas Commun., vol. 23, pp. 1056-1065, May 2005.

[i.308] Romain Masson, "E-UTRA RACH within the LTE system", Master's Degree Project Stockholm, Sweden 2006-02-03, XR-EE-KT 2006:002.

[i.309] Panasonic, NTT DoCoMo, "Zadoff-Chu sequence allocation on RACH for complexity reduction", TSG-RAN WG1 meeting #47bis, Sorrento, Italy Jan 15-19, 2007, R1-070189, Discussion and Decision.

[i.310] D. Tse and P. Viswanath, Fundamentals of Wireless Communications. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[i.311] C. Wong and B. L. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates", IEEE Transactions on Wireless Commun., vol. 7, no. 3, pp. 962-971, March 2008.

[i.312] Y. Yao and G. B. Giannakis, "Rate-maximizing power allocation in OFDM based on partial channel knowledge", IEEE Trans. Wireless Commun.,volume 4, pp. 1073-1083, May 2005.

[i.313] J. G. Andrews, A. Ghosh, and R. Muhamed, Fundamentals of WiMAX, Prentice Hall, 2007.

[i.314] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view", IEEE Trans. Commun., vol. 49, pp. 1561-1571, September 2001.

[i.315] IEEE 802.16m-08/004r1 Evaluation Methodology document.

NOTE: see: http://www.wirelessman.org/tgm/.

[i.316] M. Shen, G. Li, and Hui Liu, "Effect of traffic channel configuration on the orthogonal frequency division multiple access downlink performance", IEEE Trans. Wireless Commun., vol. 4, pp. 1901-1913, July 2005.

[i.317] C.Ciochina, F. Buda, and H. Sari, "An Analysis of OFDM Peak Power Reduction Techniques for WiMAX Systems", Proceedings of ICC'06, IEEE International Conference of Communications 2006, vol. 10, pp. 4676-4681.

[i.318] Technical Note 2 (TN2), "Detailed analysis of physical and access layer and cross layer techniques", deliverable of the "Study of Satellite Role in 4G Mobile Networks", ESA ITT AO/1-5115/06/NL/US.

[i.319] M. Papaleo, R. Firrincieli, G. E. Corazza, A. Vanelli-Coralli, "On The Application of MPE-FEC to Mobile DVB-S2: Performance Evaluation in Deep Fading Conditions", Proceedings of the International Workshop on Satellite and Space Communications 2007, Salzburg, Austria, September 2007, pp. 223-227.

[i.320] M. Papaleo, R. Firrincieli, S. Cioni, A. Vanelli-Coralli, G. E. Corazza, H.J.Lee, and P.S.Kim, "Link Layer FEC in DVB-RCS: Performance Evaluation in nLoS Conditions", Proceedings of the IEEE 67th Vehicular Technology Conference, VTC 2008 Spring, Marina Bay, Singapore, May, 2008.

[i.321] ETSI TR 102 377 (V1.2.1): "Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines".

[i.322] DVB Bluebook A120: "DVB-SH Implementation Guidelines".

NOTE: Available at http://www.dvb.org/technology/standards/.

[i.323] "Physical and Access Layer Architecture (PALA)", deliverable of the "Study of Satellite Role in 4G Mobile Networks", ESA ITT AO/1-5115/06/NL/US.

[i.324] Technical Note 2 (TN2), "Detailed analysis of physical and access layer and cross layer techniques", deliverable of the "Study of Satellite Role in 4G Mobile Networks", ESA ITT AO/1-5115/06/NL/US.

[i.325] Heuberger, H. Stadali, B. Matuz, A. Del Bianco, R. Hoppe, O. Pulvirenti, A. Bolea Alamanac, O. Smeyers, R. De Gaudenzi "Experimental Validation of Advanced Mobile Broadcasting Waveform in S-Band", 4th ASMS 2008 Conference, 26-28 August 2008, Bologna, Italy.

[i.326] Lee. S, "Satellite DMB in Korea", International Workshop for B3G/4G satellite Communications, Korea, 18 November 2004.

[i.327] ETSI TR 102 604 (V.1.2.1):"Satellite Earth Stations and Systems (SES);Satellite Digital Radio (SDR) Systems; Guidelines for the use of the physical layer standards".

[i.328] ETSI TS 102 550 (V1.3.1): "Satellite Earth Stations and Systems (SES); Satellite Digital Radio (SDR) Systems;Outer Physical Layer of the Radio Interface".

[i.329] ITU-R REPORT M.2077: "Traffic forecasts and estimated spectrum requirements for the satellite component of IMT-2000 and systems beyond IMT-2000 for the period 2010 to 2020" (ITU-R-REP-M.2077-2006-MSW-E.doc).

[i.330] ETSI TS 122 105: " Universal Mobile Telecommunications System (UMTS); Service aspects; Services and Service Capabilities (3G TS 22.105)".

[i.331] CENELEC EN 50478: "Functional receiver specification of satellite digital interactive television with a low data rate return channel via satellite - Modem layer specification".

[i.332] ITU-T Recommendation Y.1540 (2002): "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[i.333] IEEE 802.16: "Air Interface for Fixed and Mobile Broadband Wireless Access System".

[i.334] IEEE 802.16d: "Maintenance and System profiles for 2-11 GHz".

[i.335] IEEE 802.16e: "Mobile Broadband Wireless Access System".

[i.336] ITU-R Report ITU-R Recommendation M.2023: "Spectrum requirements for International Mobile Telecommunications-2000 (IMT-2000)".

[i.337] ITU-R Recommendation M.1391: "Methodology for the Calculation of IMT-2000 Satellite Sprectrum Requirements".

[i.338] ETSI TS 123 203: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Policy and charging control architecture (3GPP TS 23.203)".

[i.339] ITU-T Recommendation Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[i.340] IEEE 802.15: "Wireless Personal Area Networks".

[i.341] IEEE 802 11s: "Extended Service Set Mesh Networking Standard".

[i.342] IEEE 802.15.4. "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specification for Low-Rate Wireless Personal Area Networks (WPANs)".

[i.343] IEEE 802.15.5: "Mesh Topology Capability in Wireless Personal Area Networks (WPANs)".

[i.344] IEEE 802.16a: "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz ".

[i.345] IEEE 802.16c: "IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 1: Detailed System Profiles for 10-66 GHz".

[i.346] IEEE 802.22: "Information Technology- Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks -Specific Requirements - Part 22.1: Standard to Enhance Harmful Interference Protection For Low Power Licensed Devices Operating".

[i.347] IEEE 802.16j: "Local and Metropolitan Area Networks - Part 16: Air Interface for Broadband Wireless Access Systems ".

[i.348] IEEE 802.11k: " Local and Metropolitan Area Networks - Specific Requirements Part 11Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 1: Radio Resource Measurement of Wireless LANs".

[i.349] Void.

[i.350] IEEE 802.11a: "IEEE Standard for Information technology - Telecommunications and information exchange between systems—Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - Amendment 1: High-speed Physical Layer in the 5 GHz band ".

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

1D	one-Dimensional
2D	two-Dimensional
ACE	Active Constellation Extension
ACK	Acknowledgment
ACM	Adaptive Coding and Modulation
ADC	Analog to Digital Converter
AES	Advanced Encryption Standard
AF	Amplify and Forward
AM	Acknowledged Mode
AMC	Adaptive Modulation and Coding
AMS	Adaptive MIMO Switching
AODV	Ad hoc On-Demand Distance Vector routing
APP	A-Posteriori-Probability
ARQ	Automatic Repeat-reQuest
ASN	Access Service Network
ASP	Application Service Provider
ATC	Auxiliary Terrestrial Component
ATN	Auxiliary Terrestrial Network
AWGN	Additive White Gaussian Noise
BAN	Body Area Network
BCJR	Bahl, Cocke, Jelinek and Raviv algorithm
BEC	Binary Erasure Channel
BER	Bit Error Rate
BF	Bit-Flipping
BLER	Block Error Rate

BP	Belief Propagation
BS	Base Station
BS/AP	Base-Station/Access-Point
BSS	Blind Source Separation
CA	Collision Avoidance
CAGR	Cumulative Annual Growth Rate
CC	Convex Constraint
CCI	Co-Channel Interference
cdf	cumulative density function
CDMA	Code Division Multiple Access
CF	Compress and Forward
CGC	Complementary Ground Component
CN	Core Network
CO-OFDM	COoperative-OFDM
CP	Cyclic Prefix
CQI	Channel Quality Indicator
CQICH	Fast Channel Feedback
CR	Cognitive Radio
CRC	Cyclic Redundancy Check
CSI	Channel State Information
CSMA	Carrier Sense Multiple Access
CSN	Connectivity Servjice Network
CTC	Convolutional Turbo Codes
CTS	Clear To Send
DAAP	Dynamic Address Allocation Protocol
DAB	Digital Audio Broadcasting
DAD	Duplicated Address Detection
DA-OFDM	Double-Antenna OFDM
DC	Direct Current
DCF	Distributed Coordination Function
DF	Decode and Forward
DFT	Discrete Fourier Transform
DL	Down-Link
DMB	Digital Multimedia Broadcasting
DoA	Direction of Arrival
DPC	Dirty Paper Coding
DPS	Dynamic Frequency Selection
DRCP	Dynamic Registration and Configuration Protocol
DSP	Digital Signal Processing
DSR	Dynamic Resource Printing
DVB	Digital Video Broadcasting
DVB-RCS	DVB Return Channel via Satellite
DVB-S	DVB via Satellite
DVB-SH	DVB Satellite Handheld
DVB-T	DVB Terrestrial
EAP	Extensible Authentication Protocol
EIRP	Effective Isoptorpic Radiated Power
eNB	eNodeB
ePDG	evolved Packet Data Gateway
ERS	Empirical Roadside Shadowing
ESNR	Effective SNR
ESR	Erroneous Seconds Ratio
ETX	Expected Transmission Count
FBSS	Fast Base Station Switching
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPGA	Field Programme Gate Array
FSS	Fixed Satellite Service

FTP	File Transfer Protocol
FUSC	Fully Used Sub-Carrier
GA	Genetic Algorithm
GBN	Go-Back-N
GDF	Group Decision Feedback
GEO	Geostationary Earth Orbit
GF	Galois Field
GMR	Geo Mobile Radio
GPM	Gradient Projection Method
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSO	Geostationary Satellite Orbit
GTP-U	GPRS Tunnelling Protocol-User plane
HARQ	Hybrid ARQ
HHO	Hard HandOff
HMCN	Hierarchical Multi-hop Cellular Network
HoM	High order Modulation
HPA	High Power Amplifier
HSDPA	High Speed Downlink Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
IBO	Input Back Off
IC	Interference Cancellation
ICA	Independent Component Analysis
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electronic and Electrical Engineers
IETF	Internet Engineering Task Force
IF	Intermediate Frequency
IFFT	Inverse Fast Fourier Transform
IMR	Intermediate Module Repeater
IMS	Industrial Medical Scientific
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telecommunications
IP	Internet Protocol
IrDA	Infrared Data Association
ITU	International Telecommunications Union
LAN	Local Area Network
LDGM	Low Density Generator Matrix
LDPC	Low Density Parity Check
LEO	Low Earth orbit
LHCP	Left Hand Circular Polarisation
LMS	Land Mobile Satellite
LO	Local Oscillator
LOS	Line Of Sight
LS	Least Square
LSA	Local Search Algorithm
LSMI	Loaded Simple Matrix Inversion
LT	Luby Transform
LTE	Long Term Evolution
MAC	Medium Access Control
MACA	Multiple Access with Collision Avoidance
MAN	Metropolitan Area Network
MANET	Mobile Ad-hoc NETworks
MAP	Maximum-A-Posterior
MBMS	Multimedia Broadcasting Multicast Service
MBS	Multicast and Broadcast Service
MBSFN	MBMS Single Frequency Networ
MCE	MBMS Coordination Entity
MDHO	Macro Diversity Handover
MDS	Maximum Distance Separable
MED	Modified Exponential Decay
MF	Maximum Fairness

MFN	Multi Frequency Diversity Network
MHN	Multi-hop-Capable Node
MIC	Mean Mutual Information per Bit
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
ML	Maximum Likelihood
MM	MultiMedia
MME	Mobility Management Entity
MMIB	Mean Mutual Information per Bit
MMR-SG	Mobile Multi-hop Relay Study Group
MMSE	Minimum Mean Square Error
MN	Mobile Node
MPA	Message Passing Algorithm
MPE-FEC	Multi Protocol Encapsulation Forward Error Correction Technique
MPLS	Multi-Protocol Label Switching
MRC	Maximal Ratio Combining
MRT	Maximum Ratio Transmission
MS	Mobile Station
MSD	Maximum Distance Separable
MSE	Mean Square Error
MSR	Maximum Sum Rate
MSS	Mobile Satellite System
MSV	Mobile Satellite Ventures
MTBL	Maximum Tolerable Burst Length
MUD	Multi-User Detection
MVDR	Minimum Variance Distortionless Response
NACK	Negative ACK
NAS	Non-Access Stratum
NGSO	Non-Geostationary Satellite Orbit
NLOS	Non Line Of Sight
NRM	Network Reference Model
NRT	Non-RT
NSP	Network Service Provider
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLSR	Optimized Link State Routing
OSA	Opportunistic Spectrum Access
OSAR	Opportunistic Packet Scheduling and Auto Rate
OSI	Open Systems Interconnection
PAN	Personal Area Network
PAPR	Peak to Average Power Ratio
PBCH	Physical Broadcast Channel
PCEF	Policy and Charging Enforcement Function
PCFICH	Physical Control Format Indicator Channel
PCRF	Policy and Charging Rules Function
PDA	Probabilistic Data Association
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Control Protocol
Pdf	Probability density function
PDN GW	Packet Data Network Gateway
PDP	Power Delay Profile
PDSCH	Physical Downlink Shared Channel
PER	Packet Error Rate
PF	Proportional Fairness
PHICH	Physical Hybrid ARQ Indicator Channel
PHY	PHysical Layer
PIC	Parallel Interference Canceller
PKMv2	Privacy and Key Management Protocol Version 2
PL-FEC	Packet Level FEC
PLMN	Public Land Mobile Network
PLR	Packet Loss Ratio
PMCH	Physical Multicast Channel

PMR	Private Mobile Radio
PNC	PicoNet Coordinator
POCS	Projection Onto Convex Set
POMDP	Partially Observable Markov Decision Processes
PRACH	Physical Random Access Channel
PRB	hysical Resource Block
PUCCH	Physical Uplink Control Channel
PUSC	Partially Used Sub-Carrier
PUSCH	Physical Uplink Shared Channel
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RA	Resource Allocation
RAN	Evolved Radio Access Network
RBIR	Received Bit Mutual Information Rate
RF	Radio Frequency
RHCP	Right Hand Circular Polarisation
RLC	Radio Link Control
RM	Rate Matching
RMUS	Reliable Unicast over Unidirectional Satellite link
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Reed Solomon
RT	Real Time
RTS	Request To Send
RTT	Round Trip Time
RUIM	Removable User Identify Module
SAA	Stateless Address Autoconfiguration
SAC	Spectrum Access Control
SA-OFDM	Single-Antenna OFDM
SAT-RMTP	SATellite Reliable Multicast Transport Protocol
SCARI	Software Communications Architecture Reference Implementation
SC-FDMA	Single Carrier FDMA
SDE	Sphere DEcoding
S-DMB	Satellite Digital Multimedia Broadcasting
SDR	Software Defined Radio
SDU	Service Data Unit
SF	Stretch Factor
SFN	Single Frequency Network
SGW	Serving GateWay
SIC	Successive Interference Canceller
SIM	Subscriber Identify Module
SINR	Signal-to-Interference plus Noise-Ratio
SIR	Signal to Interference Ratio
SISO	Single Input Single Output
SM	Spatial Multiplexing
SMI	Sample Matrix Inversion
SNIR	Signal to Noise plus Interference Ratio
SNR	Signal to Noise Ratio
SOC	Second Order Cone
SOFDM	Scalable OFDM
SOFDMA	Scalable OFDMA
SPA	Sum-Product Algorithm
SR	Selective-Repeat
SS	Subscriber Station
ST	Space-Time
STBC	ST Block Codes
STC	Space-Time Coding
STICS	Satellite/ Terrestrial Integrated mobile Communication System
STTC	ST Trellis Codes
S-UMTS	Satellite Universal Mobile Telecomunication Systems
SW	Sliding Window

TA	Tracking Area
TC	Turbo Codes
TCP	Transmission Control Protocol
TD	Transmit Diversity
TDD	Time Division Duplex
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
T-DMB	Terrestrial Digital Multimedia Broadcasting
TETRA	Terrestrial Trunked RAdio
TM	Transparent Mode
TTI	Time Transmission Interval
UE	User Equipment
UL	Upper Layer (annex I) or Up-Link
UM	Unacknowledged Mode
USIM	Universal SIM
UWB	Ultra Wide-Band
VAA	Virtual Antenna Array
VCO	Voltage Controlled Oscillator
VINR	Carrier to Interference Noise Ratio
VoIP	Voice over IP
W-CDMA	Wideband CDMA
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless LAN
WRAN	Wireless Regional Area Network
WSN	Wireless Sensor Network
ZF	Zero-Forcing

4 Overview on future Mobile Satellite Systems

4.1 Aims

The present document aims to provide a detailed review of future Satellite Communications (SatCom) system architectures and the associated technologies for the delivery of mobile services.

The identification of future satellite architectures has been based on the following main factors:

- Services that are foreseen to be driving the mobile communications market over the coming years (see clause 7.1.2 for some examples).
- An assessment of the possible roles of satellites in providing services in a cost-effective manner in the context of emerging and future terrestrial networks.
- The foreseen deployment of integrated or hybrid architectures for the delivery of mobile broadcasting, interactive services and or telecommunications services.
- The potential impact of state-of-the-art satellite and communication technologies in enhancing the cost-effectiveness of satellites for the delivery of services.
- Spectrally efficient approaches including dynamic spectrum management.

Based on the above factors, two main approaches to satellite architectures for beyond 3G have been identified. These are:

a) Evolution of existing and emerging radio interfaces for satellite services (e.g. GMR, S-UMTS Family SL, DVB-RCS, Satmode, DVB-SH and ETSI SDR). This is also referred to within this document as the "Medium-Term Architecture". Alternatively, it could be possible to adapt the emerging terrestrial mobile radio interfaces for satellite services. In this regard, this document focuses on the WiMAX and LTE standards. Consideration of such an adapted radio interface has also been incorporated in the "Medium-Term Architecture".

b) The potential roles of satellites in beyond 3G converged networks and in the context of related-future communication technologies on a **bottom-up basis**. This is also referred to within this document as the "Long-Term Architecture".

4.2 Future Mobile Satellite System Architectures

The overall system can be classified in an integrated system type and a hybrid system type.

An integrated Mobile Satellite System (MSS) is a system employing space and ground components where the ground component is complementary to and operates as a part of the MSS and, together with the satellite component, provides an integrated service offering. In such system the ground component is controlled by the satellite resource and network management system. Further, the ground component uses the same designated portions of the frequency band as the associated operational MSS.

On the other hand, a hybrid satellite and terrestrial system is a system employing satellite and terrestrial components where the satellite and terrestrial components are interconnected and in specific cases these systems may share the same core network, but they can operate independently of each other. In such systems the satellite and terrestrial components can have separate network management systems and do not operate necessarily in the same frequency band.

NOTE: It is also possible to have a combination of both integrated and hybrid systems. To illustrate this, let us consider an integrated MSS based on a satellite component and a CGC operating in the same frequency band. This integrated MSS would cooperate with a cellular system operating in a different frequency band. The "triple mode" terminals would then select the best signal according to their environment and the service availability of the cellular system, the satellite component or the CGC of the integrated MSS.

4.2.1 Medium-Term Architecture

The first level of investigation has been defined through consideration of the following factors:

- The assumption that emerging hybrid mobile broadcasting architectures will be commercially successful and this will induce interest and confidence in operators for introducing new services.
- The assumption that integrated mobile communications systems will be commercially successful and this will induce interest and confidence in operators for introducing new services.
- Expanding the service range on mobile broadcasting terminals implies introduction of bi-directional data transmission capability.
- Examination of WiMAX and LTE air interfaces and architectures with regard to their adaption to satellite operation.

The business validity of introducing bi-directional mobile data services over the hybrid network (on an integrated multimedia terminal platform) can be considered separately for urban and rural (i.e. lacking low-cost terrestrial network coverage) regions. In urban areas, a multiplicity of terrestrial networks (GPRS, 3G/LTE, WiMAX, Wi-Fi) will be dominating the scene. A critical question arising is whether the terrestrial component could allow satellite operators to offer competitive services. This complex question is too difficult to be addressed quickly in a systematic way, but the answer would possibly depend on factors such as:

- Spectrum availability and related regulatory/licensing constraints.
- The exact architecture of the ground network (terrestrially interconnections and backbone? possibilities for a cellular architecture for re-using the interactive band? and cell sizes?).

- The impact of advanced communications techniques (fixed-mobile relaying, cooperative transmission techniques, advanced signal processing - e.g. MIMO, MUD).

Despite the requirement of a more in-depth system-commercial analysis, it could be commented that the foreseen deployment of ground infrastructures for supporting mobile broadcasting services in densely populated urban areas, can potentially leave satellite operators in a favourable position for getting into the mobile data comms mass-market. Recent support of this view can be drawn from recent developments in the USA, where several SatCom operators have announced provision of bidirectional services over hybrid architectures (e.g. ICO-G, MSV, Terrestar). These new systems, however, adopt a direct approach towards bidirectional services over hybrid architectures, compared to the considered more gradual and targeted extension of mobile broadcasting services.

On the other hand, in rural environments the provision of low-medium data services (including VoIP) over the satellite component could potentially make business sense, depending on the cost of services and the cost-size-quality of terminals (which are expected to be handheld multi-service, multi-application platforms). The above assertion is justified to some extend by the example of the Thuraya system, which has shown that the market size of such services is not insignificant, provided that charging of services is comparable to that of terrestrial networks, and these are provided on relatively low cost multi-network-multiservice "cellular-style" devices.

Some other factors for supporting optimism about the potential size of the rural SatCom market include:

- Basic data services (e-mail, browsing, messaging, etc.) are becoming increasingly indispensable for larger percentages of the population (in an increasingly globalized world).
- Increasing demand for multi-network handheld devices that support multiple parallel/integrated services (e.g. 2G/3G, DVB-SH, Wi-Fi, Bluetooth... plus SatCom).
- The expected impact of state-of-the art physical and higher layer techniques (e.g. diversity, MIMO, multi-user detection, co-operative transmissions) in reducing the cost services and enhancing service availability.
- The impact of advances in payload design, antenna and battery technologies in reducing the cost of services.

Also, from the service operator's point of view, extending services in this direction (i.e. low-medium speed SatCom), would not necessarily present a critical business risk, since coverage could be targeted and commitment of bandwidth gradual. Moreover business failure would not be catastrophic, since backtracking to mobile broadcasting services is possible.

Some possible usage scenarios for such type of services are:

- Business-tourist traveller uses his integrated mobile-TV/GPS/... device to check e-mails, browse, etc. (blackberry type of services)
- Intercontinental traveller avoids expensive roaming or solves terminal incompatibility problems.
- Mariner or hiker uses a distress service or makes a VoIP call

With respect to the investigation on the medium-term architecture, this study has focused on the "satellite-only" coverage domain, and in particular the following topics have been reviewed and analyzed:

- Service QoS requirements.
- Applicable propagation models.
- Suitable anti-fading features.
- Benefits of physical and upper-layer state-of-the art techniques.
- Possible standardization paths based on existing and emerging satellite and terrestrial standards.

Consideration of the service and performance requirements outlined by 3GPP in TS 22 105 [i.330] is also appropriate here and the reader is referred to reference.

There are several options for the satellite radio interface:

- Start from existing MSS satellite radio interface (e.g. GMR, S-UMTS Family SL).
- Start from existing FSS satellite radio interface (e.g. DVB-RCS and Satmode).
- Start from envisaged terrestrial mobile radio interface (e.g. LTE and WiMAX).
- In each case, the broadcast mode could be integrated or separated using e.g. DVB-SH.

In each case forward link of an interactive radio interface could be replaced by a mobile broadcast radio interface (e.g. DVB-SH).

4.2.2 Long-Term Architecture

The current prevailing visions and technological paths for 4G systems are reviewed in clause 6.1, 4G systems will very likely consist of the convergence of heterogeneous networks and services, which will be transparently accessible on personal mobile devices. Consolidation and optimization of "3G technologies" (as these are reviewed to a large extent within clause 5), will be a necessary but not a sufficient factor for achieving the very ambitious spectrum efficiency and performance objectives of 4G systems. Instead, **cooperation** appears to be the key technological enabler in 4G, as this will be pursued on the networking level, on the spectrum access level, and on the physical-layer transmission techniques.

Having misjudged their significance (relative to terrestrial radio systems) and possible roles within 3G systems, satellite systems need to make a careful assessment of their capabilities and potentials for playing some significant roles in 4G systems, with a key objective being to become an indispensable component of the union of sub-networks which will constitute the 4G networks. Obvious potential roles of satellites include provision of positioning services, mobile broadcasting and provision of bidirectional data services in remote areas (including sea and air). However new potential roles can be identified in the context of cooperative networking, for example in organising and administrating infrastructure-less ad-hoc networks.

Realising all these potential roles will on one hand depend on the maturity of terminal technologies in accommodating satellite transreceivers (within the multi-network 4G terminals), but perhaps more importantly in satellite systems taking the outmost advantage of state-of-the-art and emerging technologies in order to achieve low-cost service provision and overall competitiveness relative to alternative technological options. Already there are several examples where satellite systems begin to embrace new concepts and architectures in order to bring competitiveness to their services. For instance, fixed relaying in hybrid mobile broadcasting architectures, implemented within S-DMB and supported by the recently developed DVB-SH standard (which also specifies optimised physical and link-layer techniques), gives a competitive edge against purely terrestrial broadcasting networks (DVB-H). Furthermore, recent FCC regulations, which allow the terrestrial re-use of satellite bands in USA, is a very significant development which has stimulated satellite operators in designing optimised hybrid architectures and achieving competitiveness relative to purely terrestrial networks.

With the above in mind, the emphasis in the analysis of the Long-Term Architecture (clause 6) is given to reviewing and analyzing some of the latest communications technologies that are currently being investigated in the terrestrial arena for enabling 4G networks, namely:

- Spectrum sharing techniques (through cognitive radio and cross-system interference suppression).
- Cooperative transmission techniques (relaying and virtual antenna arrays).
- Ad-hoc networking.

The reviews of these topics are included in clauses 6.2, 6.3 and 6.4, respectively. The end objective of reviewing the above techniques (and the most recent related research results) is to draw conclusions about their applicability in satellite systems, and conversely to identify possible scenarios where satellites could play a supporting role in their implementation. These conclusions are provided at the ends of each of these clauses.

No specific service and performance requirements are outlined for the Long-Term Architecture as this was considered inappropriate.

4.3 Land Mobile Satellite Channel Modelling

The availability (and thus QoS) of land mobile satellite systems operating at frequencies below 3GHz, is limited predominantly by shadowing and blocking effects typically caused by vegetation, and man-made and physical structures. Unlike terrestrial systems, where large fade margins can be "afforded" for providing reliable non-LOS coverage, the power-limited satellite systems would have to make a totally uneconomical use of the available spectrum if they were to support non-LOS coverage; since this would require fade margins in the order of 30 dB. Even in LEO systems, where larger fade margins can be allowed (for example the Iridium system specified a figure of 16,5 dB), achieving large percentages of availability in a cost-effective manner, particularly in urban environments, is challenging (at least with a single satellite). In densely populated urban areas the deployment of Complementary Ground Components (CGC) might present a viable solution for achieving cost-effective provision of services, but still rural and sub-urban regions will need to rely on the satellite coverage alone.

Based on the above it is deduced that achieving adequate QoS with the minimum possible fade margins is a key system design objective within the medium-term architecture under investigation. Given any specific satellite system, pushing the boundaries of this optimization target relies on making best possible use of advances in communication techniques, such as:

- Diversity techniques; terminal antenna diversity, satellite diversity, polarization and time/frequency diversity.
- Sate-of-the-art physical layer and upper layer coding.
- Time-interleaving.
- Adaptive Coding and Modulation.
- Space-time coding, multi-user detection, inter-spot interference cancellation.
- Efficient upper layer protocols (e.g. HARQ).
- Mobile relaying, co-operative MIMO.

The design and optimization of these techniques is heavily dependent on the mobile satellite channel. Therefore accurate channel characterization and modelling, is not only crucial for predicting the performance and availability of the system, but also for optimizing its design in order to minimize the required fade margin.

This clause provides a review of Land Mobile Satellite (LMS) channel models that have been proposed up to date, focusing on those applicable in L and S bands. The review also includes recent results on the characterization, modelling and capacity of the multi-satellite and dual-polarized MIMO satellite channels ([i.40] to [i.43]).

The different classes of LMS channel models are reviewed in detail in annex A.

4.3.1 Summary of Models and Conclusions

There exist several different types of channel models that could be used to predict the performance and availability of land mobile satellite systems. The main model categorisations that have been reviewed (see annex A) are:

- **Empirical models**: which predict the channel attenuation based on empirical formulas, derived through extensive measurements in specific types of environments and frequencies. They are only very accurate for predicting the attenuation for environments that are similar to the ones where measurements were carried out. Moreover empirical models do not allow to synthesise a time-series of the channel coefficient, which is required for simulation evaluations.

- **Statistical models**: which make use of parameterised statistical distributions and Markov state transition models, in order to describe the variability of the signal's envelope and phase. Different parameters need to be specified, through best fits on measured data, for different environment types. They are well developed and well studied and various different methods are available for synthesizing the channel coefficient for simulation purposes.

- **Physical Models**: which are based on deterministic modelling of the physical phenomena that govern electromagnetic wave propagation. They are very accurate and useful for predicting the coverage in specific layouts (a specific office, or city), but their application becomes difficult for large coverage areas.

- **Physical-Statistical Models**: they combine physical and statistical modelling in order to characterise accurately and efficiently the channel variability in specific environment layouts, even large-scale ones.

For evaluating a system's performance within medium-term architecture scenarios, the 3-state Fontan model would provide a generic enough framework for capturing different types of environments (such as sub-urban, tree shadowing) and different propagation conditions (LOS, shadowing and heavy shadowing).

For evaluating the performance benefits of a dual-polarised MIMO configuration, only physical-statistical and empirical-statistical models have been identified (reviewed in annex A).

5 Analysis of Medium-Term MSS Architecture

5.1 QoS Requirements

The investigation in the medium-term architecture focuses on the "satellite-only" coverage scenario, where it is assumed that bidirectional data services are supported on different classes of mobile/portable devices (as these are described in [i.1] and [i.2]). Since different terminal classes are characterised by different transmit power levels, antenna gains, G/T and possibilities to utilize more than a single antenna, the average and maximum bit-rates for delivering data services could be made dependent on the terminal class (e.g. through the use of ACM, or different spreading factors in CDMA based system). However a minimum data rate needs to be defined, per type of service to be provided, that will be supportable even by cellular-size handheld terminals, which represent the most challenging case. The choice of these minimum data rates is a crucial system design parameter, since it defines the average user capacity of any particular system.

This clause reviews the QoS requirements of different application types, as these appear in ITU recommendations. The definition of QoS requirements will give an indication of the bit-rates (and other performance parameters) the system should aim for in order to support applications that are relevant to the medium-term architecture. QoS requirements can also impose restrictions on physical layer design aspects, such as the interleaver depths (as they introduce additional delays in the system) for countering shadowing and blockage in mobile satellite environments.

According to ITU-T Recommendation Y.1540 [i.332] the following protocol independent performance measures are defined for IP-based data communication services:

- Throughput.
- PER.
- Transfer Delay.
- Jitter (that occurs mainly due to processing and queuing in satellite links).
- Packet Loss Ratio (PLR).

Different applications are characterised by different levels of sensitivity with respect to the above parameters, and these are typically characterised in qualitative terms (low, medium, high). Specific numbers for the above parameters can however be derived from ITU guidelines and other standardization bodies. References [i.3] and [i.4] provide typical QoS requirements, on a qualitative and quantitative level (as specified by ITU guidelines), respectively.

Table 5.1: Qualitative Characterization of QoS Requirements for popular IP-based applications

Application	Throughput	Loss	Delay	Jitter
Web-Browsing	Low-medium	Medium	Medium	Low
Email	Low-medium	High	Low	Low
VoIP	Low	Medium	High	High
E-commerce	Low	High	High	Low
SMS	Low	Medium-high	Low	Low
MMS	Low	Medium-high	Low	Low
Fax	Low	Medium	Low	Low
FTP	High	High	Low	Low

Table 5.2: Quantitative QoS Requirements for popular IP-based Applications

Application	Throughput		Delay	Jitter	Size (Typical)	Loss	Security	Supporting Protocols
	Fwd.	Return.						
Web-Browsing	64 Kb/s	16 Kb/s	2 s to 4 s/ page	N.A.	10 KB	0	Optional (HTTPS)	HTTP, HTTPS, TCP&IP
FTP	> 14 Kb/s		N.A.	N.A.	10 KB to 10 MB	0	Optional	FTP, TCP&IP
E-mail	> 20 Kb/s (POP/IMAP) ~1 Kb/s (SMTP)		2-4 s/email (POP/IMAP) < 5min/email (SMTP)	N.A.	10 KB (no attachm.)	0	Required	POP3 SMTP IMAP ESMTP TCP&IP
Messenger	As available		N.A.	N.A.	250B	0	Optional	MSNP8/9 /10 TCP&IP
E-Commerce	> 32 Kb/s		< 2 s to 4 s	N.A.	10 KB	0	Required	
VoIP	4 Kb/s to 64 Kb/s		< 400 ms	< 1 ms*	N.A.	< 3 % (PER)	Optional	H.323, SIP, RTP, UDP&IP
SMS	< 1 Kb/s		Few minutes	N.A.	1,5 Kbit	0		
MMS	1 kb/s		Few minutes	N.A.	100 Kbit	0		
Fax	1 kb/s		Few minutes	N.A.	10 KB	10^{-6} (BER)		
NOTE: For satellite links this requirement is too demanding. Realistic values are in the order of 40 ms to 50 ms.								

With respect to the required bit-rates it is observed that for the basic applications, which will be targeted by the medium-term architecture, a maximum bit-rate of 64 Kb/s is acceptable in the forward link. In the return link a maximum bit-rate requirement of 16 Kb/s seems to be acceptable for most types of targeted applications. Another observation is that the system should be able to support and allocate flexibly different channel bandwidths for different types of applications. This flexibility is important in making efficient use of the available system resources. On a resource management level, the system should be as flexible and intelligent as possible in order to optimise the use of resources under user and application priority constraints.

The 64 KB/s to 16 Kb/s reference bit rates (to be supportable by handheld mobile terminals) are obviously not sufficient for quick large file transfers or multimedia type of applications. Higher data rates should thus be supported, although these can be made accessible only to specific terminal classes with better characteristics and features (e.g. antenna diversity, higher gains and G/T, lower noise figures, etc.). A good reference, with respect to the supportable bit rates, is provided by the S-UMTS standard, which specified (useful) information bit rates of as low as 1,2 Kb/s up to 384 Kb/s, with many intermediate options.

5.2 Signal Processing and Diversity Techniques for Improving the Performance of Mobile Satellite Systems

Signal processing and diversity techniques are perhaps the driving factors which have allowed modern terrestrial radio systems to multiply their capacities without utilizing additional bandwidths. For instance the evolution of T-UMTS from Release 99, which achieves peak data rates of less than 0,5 Mbps (in a 5 MHz band), to Release 8 which achieves peak data rates in excess of 25 Mb/s (in a 5 MHz band), has been primarily driven by physical layer techniques such as MIMO. Advanced physical layer techniques have the potential to provide capacity improvements of similar order in satellite systems, though they have to be carefully adapted and optimised on the satellite channel conditions.

This clause provides a comprehensive review of classical and state-of-the-art signal processing and diversity techniques that could be considered in the system design within the medium-term architecture. More specifically, clause 5.3.1 reviews multi-signal detection techniques that can find application for performing multi-user detection (and thus suppressing intra-system interference) within CDMA based satellite systems, for detecting signals at the output of spatially-multiplexed (e.g. dual polarised) MIMO channels, and also for suppressing inter-spotbeam interference. Clause 5.3.2 reviews the classical diversity and diversity combining techniques, and also the latest MIMO techniques for achieving better system performance and coverage through the utilisation of multiple or dual-polarised antennas. Clause 5.3.3 reviews signal processing algorithms for applying adaptive beamforming, that could find use either in vehicular types of terminals (equipped with antenna arrays), or in the gateway for suppressing spatial interference occurring for example in the uplink from terrestrial systems that re-use the satellite band. Clause 5.3.4 reviews the most powerful FEC coding techniques, namely Turbo and LDPC coding. Finally clause 5.3.5 provides some recommendations on the possible application scenarios of the reviewed techniques.

5.2.1 Signal Detection Technique for Multi-User CDMA systems, Spatially/Polarized multiplexed MIMO and Inter-Spotbeam Interference Suppression

The higher spectral efficiencies that can be achieved by CDMA, compared to TDMA and FDMA [i.46], are not straight forwardly available within a practical radio system (such as S-UMTS). Multipath signal propagation and asynchronicity among user transmissions makes the design of orthogonal user codes a challenging task, a fact that accounts for the poor performance of simple rake receivers, especially when many users need to be accommodated within a limited bandwidth. Furthermore, the high sensitivity of the simple correlation receiver, due to the near-far effect, introduces further severe performance degradation, unless complex and bandwidth consuming power control is accommodated. A solution to the problems of inter-user interference and near-far effect in CDMA is offered by the joint processing of multiple users' signals; termed as Multi-User Detection (MUD) [i.47]. MUD techniques exploit precise knowledge of the structure of the interference component at the output of the conventional receiver and follow different strategies in order reject it.

MUD could find immediate application within the W-CDMA based S-UMTS standard. More importantly, in a satellite system MUD can be applied not only to mitigate intra-spot beam multi-access interference, but also to suppress inter-spot beam interference [i.91]. Eliminating inter-spot interference through signal processing is a very attractive proposition since the frequency re-use factor (and thus the capacity) of the system can be increased.

Similarly, achieving in practice the high data rates that are theoretically available in spatially multiplexed MIMO systems relies heavily on the design of the signal estimation algorithm at the receiver (MIMO channel equalizer). Recent results [i.40] and [i.41] indicate that the capacity of satellite can be significantly increased through the use of dual-polarized antennas; that create a virtual MIMO channel. The use of MIMO algorithms for the dual-polarized satellite channel has recently been investigated in [i.90].

The above techniques can play a significant role in enhancing the performance and capacity of future satellite systems, especially since they do not consume any additional power/bandwidth resources.

A thorough review of the different types of signal estimation algorithms that can be applied are provided in annex B.

Summary

Clause 5.2.1 can be summarized in the following main points:

- Multi-User detection, (dual-polarized) MIMO channel equalization and inter-spotbeam interference suppression can enhance the performance/capacity of satellite systems without requiring the use of additional power/bandwidth resources.
- From a mathematical modeling point of view, the three problems are very similar, and thus the same techniques can be applied for their solution.
- The complexity of the optimal signal detector/estimator grows exponentially with the problem's dimensionality and thus its application is limited to systems with small number of co-channel users/antennas/interfering spotbeams.
- Reduced complexity techniques that approach very closely the performance of the optimal detector are available.
- Reduced complexity Maximum-A-Posterior (MAP) detectors and Iterative (Turbo) estimation techniques have been shown to offer very significant performance/capacity improvements in coded systems, even when sub-optimal soft-input-soft-output signal detectors are used. Therefore these types of techniques represent the state-of-the-art with respect to the maximum performance gains that can be achieved through signal estimation techniques.

5.2.2 Diversity and MIMO Techniques

As discussed in clause 4.3, the performance of mobile satellite systems is limited mainly by large-scale fading effects (shadowing, blocking) and secondarily by multipath fast fading and atmospheric induced fading (assuming operation over L and S bands). Diversity techniques can provide QoS improvements (in terms of link availability) over fading environments; in the expense of increased user-equipment/satellite complexity and/or bandwidth resources. In principle both large and small scale fading can be mitigated through diversity techniques. However the first type, which also introduces the biggest limitations, presents a more challenging case to achieve in practice.

This clause reviews classical diversity (Time, Frequency, Space and Polarization) and diversity combining (switch, selection, equal gain and maximum ratio combining) techniques, and the more Space-Time Coding techniques for MIMO Systems.

The detailed review of the techniques is provided in annex C.

5.2.3 Adaptive Beamforming

Modern mobile satellite systems, such as Thuraya and Inmarsat-4 make use of multiple antenna feed elements in order to form up to several hundreds of service area spot-beam. The formation and positioning of spotbeams is controlled through digital onboard processing; by applying complex weighting and linear processing on the antenna elements. Despite the big capacity advantages offered by the multi-spotbeam technology, its full potential is limited by inter-spotbeam interference. Moreover, if a system operator would like to capitalize on recent regulatory developments that foresee the terrestrial reuse of satellite bands, interference issues between terrestrial and satellite users would need to be resolved.

In the return link, an efficient solution to the above problems can be achieved by performing adaptive beamforming at the gateway [i.91] and [i.171]. In particular the gateway can apply digital processing on the raw antenna element feed element signals in order to adaptively optimize the signal quality of each individual satellite user. On the other hand, in the forward link adaptive beamforming can be applied in vehicular mobile terminals which employ multiple antenna elements.

Adaptive beamforming is one of the key technologies that can allow reliable communications in an interference-limited environment. The term beamforming refers to shaping the antenna beam pattern in receive or transmit or both modes. Adaptive beamforming requires the communicating terminals to have more than one antenna. The desired beam pattern shape is obtained by optimizing the weight parameters for each array element accordingly. For a narrowband system single complex weights for each element are sufficient (spatial filtering). For a wideband system FIR filters need to be utilized in order to take into account the temporal filtering effect of the propagation environment. So in the wideband case channel equalization and beamforming are performed jointly (spatio-temporal filtering).

There are two main techniques for choosing the weight parameters in adaptive beamforming; *null steering* and *optimal combining*. Null steering directly exploits directional information about the desired user and all the interfering signals to be suppressed. This requires prior estimation of Direction-of-Arrival (DoA) of the signals before the weights of the beamformer are determined, which involves the solution of a linear system of equations [i.105]. The method ideally results in forming nulls in the antenna radiation pattern towards interferers and a beam towards the desired user. The performance of null-steering depends mainly on two factors:

- Assumption that the spatial structure of the impinging signal from each source is accurately described by only one steering vector, which means that there is insignificant azimuth spreading of the signals e.g. Line of Sight Scenario.
- Performance of the DoA method used and distortion by the antenna manifold.

An advantage of null-steering being a DoA based method is that the same beamformer weights can be used both for uplink and downlink both in TDD and FDD modes. This is because averaged DoA estimates remain approximately the same over large frequency offsets. Moreover, since DoA is usually a slowly changing parameter the update rate of the beamformer's weights does not need to be very fast.

In practice null-steering cannot achieve the optimal SNIR. Moreover there are difficulties in acquiring the directions of interfering sources. Optimal combining methods overcome these limitations [i.106]. In optimal combining, the beamformer weight vector is given as a solution to some optimization problem. The optimized quantity can be the MSE, SNIR, or interference power level - the MVDR criterion. Asymptotically all criteria lead to the same solution given by the Wiener-Hopf equation [i.107]. As opposed to null steering, optimal combining can maximize the output SNR even if the number of interfering sources exceeds N-1, where N is the number of antenna elements. Optimal Combining is also resilient to distortion caused by the antenna manifold as the modified optimization problem to be solved is not necessarily more difficult.

A more detailed review of optimal combining and DoA estimation algorithms for facilitating adaptive beamforming is given in annex D.

5.2.4 State-of-the-art Error Correction Codes

This clause reviews Turbo and LDPC codes, which are presently considered to be the best performing error correcting codes available. Over AWGN channels both algorithms approach closely the Shannon limit, while they offer practical encoding/decoding implementation complexities. Turbo codes have been adopted in T/S-UMTS, DVB-RCS, DVB-SH, IEEE 802.11 [i.269], IEEE 802.16 [i.333] standards, while LDPC codes are defined in the DVB-S2, IEEE 802.16 [i.333] (optional) and are also considered for the DVB-T2 standard. Generally LDPC codes achieve slightly better performance over AWGN channels, but Turbo codes perform better over multipath-fading channels. Of course, specific conclusions can only be reached by considering specific systems, code designs and propagation models.

The detailed review of Turbo and LDPC codes is given in annex E.

5.2.5 Time Interleavers

The existing DVB-SH and ETSI-SDR standards offer the use of time interleavers to facilitate adequate service continutity during short propagation fades. Specific details are given in the standards and their associated guidelines documents [i.1], [i.293], [i.327] and [i.328]. Time interleaving is viewed as a potential feature for inclusion in future architectures but care is required to handle the interactions between time slicing, FEC, block structures and the interleaver implementation. For more details see clause F.3.

5.2.6 Conclusions on Signal Processing and Diversity Techniques for Improving the Performance of Mobile Satellite Systems

This clause has reviewed advanced physical layer techniques that could be applied in order to enhance the capacity/availability of a system within the medium-term architecture.

Signal processing techniques, such as multi-user detection and inter-spotbeam interference suppression techniques could be applied relatively easily within CDMA and multi-beam systems for eliminating intra system interference.

The most serious limitation in land mobile satellite systems is signal shadowing and blocking, which practically translate into link unavailability for the affected terminals. Diversity techniques could in principle be applied in order to recover the system's performance, but in practice this is either very difficult or very costly to achieve. In particular, satellite diversity is a very expensive options, especially for GEO constellations, and time diversity is too bandwidth inefficient and not suitable for real time applications. For certain types of data applications long channel interleaving could be considered as a cost-effective solution, but again the maximum delay in a bidirectional service will be limited by TCP timers. Another practical and cost-effective solution would be to try to implement some type of adaptive time diversity, where the terminal and gateway detect when shadowing occurs and apply time diversity selectively only for the affected terminals.

Countering shadowing cost-effectively could also be achieved through Hybrid-ARQ techniques, since the large satellite propagation delays mean that retransmissions will be decorrelated not only with respect to the fast fading but also to the slow fading channel component. Upper-Layer FEC (alone or as part of H-ARQ mechanism) could also provide an efficient solution. These techniques are reviewed in clause 5.4.

Another use of signal processing techniques could be to allow the satellite bands to be re-used effectively by terrestrial networks; without the latter causing severe interference to the satellite system. This can be achieved through adaptive beam-forming techniques at the gateway (by post-processing jointly the spot-beam signals), in order to adaptively cancel interfering signals from terrestrial users.

5.3 Upper-Layer Error Control Techniques

5.3.1 Introduction

The diversity and advanced signal processing techniques discussed in clause 5.3.2, though they can provide significant improvements in the availability/capacity of mobile satellite systems, they cannot guarantee the 100 % reception reliability required by many types of bidirectional data applications. In fact, absolute reliability for each data packet can only be achieved through the use of upper-layer protocols (ARQ, HARQ) which facilitate, in some form, retransmissions for lost/erroneous data packets. On the other hand, recent advances in higher-layer Packet Level FEC (PL-FEC) and interleaving have introduced a new dimension in the design of error control mechanisms. For unicast applications, PL-FEC techniques provide additional error protection above the physical layer; increasing the link reliability especially over harsh mobile environments. By designing carefully the rate of the code, the system can achieve better efficiency and reduced delays as compared to purely ARQ based system. Also, by operating on large data blocks, PL-FEC techniques (possibly also combined with packet level interleaving) provide a form of time diversity against slow fading effects. PL-FEC is most useful for multicast applications as it allows the design of efficient reliable delivery protocols.

5.3.2 Upper-Layer Coding and Interleaving

As in classical error control coding, a PL-FEC encoder takes k source packets as input and generates n encoded packets with n > k. In the case of a systematic encoder, the n transmitted packets contain the original k packets and h newly generated parity packets. If the encoder is non-systematic, all n packets are encoded, namely the original k packets are no longer part of the transmitted set of packets. These n packets constitute a FEC block. The inverse of the code rate: $F = n/k$, is called the Stretch Factor (SF) and directly expresses the additional capacity requirements due to FEC. Hereafter, only systematic encoders are considered as is the case in standards such as MBMS, DVB-H, DVB-SH; systematic encoders are preferred as they allow direct access to the original information so UEs without a particular FEC decoder can also receive and process content.

A FEC decoder can recover the original k data packets as long as it receives enough packets out of the n transmitted packets. With some FEC codes, such as the Reed-Solomon (RS) codes, the decoder only requires any combination of k out of the n transmitted packets to recover the original k packets in a given FEC block. In contrast, other FEC codes, such as the Low Density Generator Matrix (LDGM) codes require more than k packets to recover the original k packets in a FEC block.

Detailed review of upper-layer FEC and interleaving techniques are given in annex F.

5.3.3 Feedback-based Schemes

5.3.3.1 ARQ

The most basic form of user feedback is ARQ. The three common types of ARQ are:

- Stop-and-Wait;
- Go-Back-N (GBN);
- Selective-Repeat (SR).

In stop-and-wait ARQ, one segment of data is transmitted and the sender stops to wait for an ACK from the receiver; if a receiver receives the segment correctly, it sends an ACK to the sender, which in turn sends another segment of data upon reception of the ACK for the previous segment. However, if the first segment is received in error, the receiver sends a NACK to the sender which then retransmits that particular segment before stopping to wait for an ACK. This stop-and-wait approach results in idle time, i.e. inefficiency, which is worse for environments with long propagation delays and high levels of data loss (such as mobile satellite).

GBN eliminates the idle time in stop-and-wait since the sender transmits data segments in a contiguous manner; the receiver still sends ACKs and NACKs accordingly. However, the sender reacts in different way to stop-and-wait upon the reception of a NACK; if the sender transmits data segments 1, 2, 3, 4, 5, 6, 7 and 8, and receives a NACK for data segment 4, it retransmits data segments 4 to 8 instead of data segment 4 only. The idea behind this behaviour is that a buffer is not required at the receiver to restore the order of the data segments since they all arrive in order.

Although better than stop-and-wait, GBN results in unnecessary retransmissions especially when the propagation delay is long and the data rate is high. SR can overcome this ineffectiveness since it improves on GBN by retransmitting a data segment if it receives a NACK for that specific data segment. Nonetheless, SR requires the use of a buffer at the receiver for restoring the order of the data segments, e.g. after transmitting data segments 1, 2, 3, 4, 5, 6, 7 and 8, a sender receives a NACK for data segment 4 which is then immediately retransmitted, so assuming correct reception for all the other data segments, the received sequence at the receiver is 1, 2, 3, 5, 6, 7, 8 and 4.

An example of a transport protocol which incorporates some form of ARQ is TCP. The form of ARQ in TCP works in a similar fashion to stop-and-wait with some exceptions:

- Often more than one data segment is sent at a time; there is a dynamic value for the number of data segments a sender can transmit at once and this dynamic value is determined by the congestion and flow control mechanisms of TCP.
- The receiver does not send NACKs; instead, the TCP sender has a retransmission timer to guard against lost data segments and/or ACKs. Upon the expiration of the retransmission timer for a particular data segment, the TCP sender retransmits it.

5.3.3.2 Hybrid ARQ

Hybrid ARQ (HARQ) improves the efficiency of standard ARQ through integration with physical or upper-layer FEC. The former variation has been adopted in UMTS Release 6.7 (HSDPA, HSUPA) and in WiMAX, while the latter variation has been proposed for improving the efficiency of reliable multicasting protocols ([i.138] to [i.140]). There are two main types of H-ARQ:

- Type I: The transmitter sends coded blocks/packets and the receiver requests their retransmission if an error is detected.
- Type II: The transmitter sends coded blocks/packets and the receiver requests additional parity check bits/packets if an error is detected (incremental redundancy).

Generally Type II is more capacity efficient, but an exact comparison would need to be carried out over specific system and channel assumptions. HARQ can be used in a stop-and-wait or a selective repeat mode. In HSDPA a stop-and-wait approach is followed but several HARQ process are initiated in parallel in order to achieve high throughputs.

HARQ can be effectively applied for enhancing the performance of HSDPA over a GEO satellite link. In fact the large propagation delay involved in the satellite channel results in achieving time diversity for the retransmissions; not only against the fast fading component but also against the slow fading one. The adaptation for satellite links can be achieved by increasing the parallel HARQ processes to match the propagation delay of the satellite link (although this requires large memory requirements in the user terminal).

In the context of multicasting applications, ARQ and HARQ schemes can result in feedback implosion; where the sender, network entities, and links on the path from receivers to the sender are overwhelmed with feedback information. If the number of receivers is X, and the uniform, independent packet loss rate is p, then the probability that a data packet needs to be retransmitted is $1-(1-p)^X$; evidently, this probability increases as X increases, thus making basic ARQ unsuitable for communication with large audiences. It is also difficult for the sender to keep track of the status of each packet at each receiver; this information is required so that the sender can release packets from its buffers if it has received X ACKs. In order to enhance the scalability and resilience of ARQ/HARQ against feedback implosion, ARQ or HARQ can be combined with one or more feedback suppression mechanisms, which can be either timer-based or structure-based.

5.3.4 Satellite Reliable Multicast Transport Protocols

5.3.4.1 Reliable Multicast over Unidirectional Satellite Link (RMUS)

RMUS [i.151] is asymmetric and does not employ any form of FEC. Data transmission is interspersed with the error recovery process which consists of two main polling stages: in the first, all the receivers are polled for reception reports and only the lost packets are retransmitted; the second stage entails polling for reception reports from receivers which reported missing packets during the first polling stage. If there are still receivers with missing packets, the polling process continues for a predefined number of times and a receiver which does not report during this period is ignored. User feedback is obtained via a low bit-rate return link, which may be a satellite uplink or dial-up connection.

5.3.4.2 Satellite Reliable Multicast Transport Protocol (SRMTP)

SRMTP [i.152] is window-based and the window size is optimised with respect to the delay-bandwidth product of the satellite channel to realise a high data rate. The network architecture under consideration has end-users, which receive multicast content directly from the satellite or via a local LAN connected to a receiving gateway. The return channels are via satellite. The recovery mechanism is based on ACKs and NACKs; FEC is not considered. It is shown that SRMTP is enhanced by satellite onboard processing and buffering, i.e. the satellite retransmits missing packets, to outperform the following:

- SRMTP over a bent pipe satellite.
- SRMTP with only onboard processing.
- Multicast File Transfer Protocol.

5.3.4.3 SATellite Reliable Multicast Transport Protocol (SAT-RMTP)

NOTE: Although SAT-RMTP and SRMTP have the same spelling, they refer to different protocols.

Unlike RMUS and SRMTP, SATellite Reliable Multicast Transport Protocol (SAT-RMTP) [i.153] adopts FEC and HARQ, thus making it more scalable. SAT-RMTP supports receivers with and without return links. The sender solicits feedback periodically throughout the transfer of a file. Receivers with return links employ random timers to avoid feedback implosion. A receiver without a return link can benefit from the response of the sender to feedback generated by receivers with return links; alternatively, a receiver with a return link but with a timer which is yet to expire can suppress a pending request if the provided feedback is sufficient.

5.3.5 Conclusions on Upper-Layer Error Control Techniques

This clause has reviewed feedback and feed forward upper layer techniques that guarantee or improve, respectively, the link availability for certain types of applications. The benefits of H-ARQ over S-UMTS have been reported within the EU FP6 MAESTRO research project. An issue in satellite systems is the large number of parallel ARQ processes that require a lot of memory in the user's terminal. H-ARQ techniques could be combined efficiently with upper-layer coding techniques; in an incremental redundancy mode.

Significant emphasis within this clause has also been given in reviewing the different types of available upper layer coding techniques. RS coding is suitable for the smaller packet sizes that would be encountered in the medium-term architecture scenario. Raptor codes could also be considered, particularly within H-ARQ, due to their flexible incremental redundancy feature. Both of these two coding techniques are specified within DVB mobile broadcasting standards (DVB-SH/H).

5.4 Existing satellite systems and standards

This clause considers existing technologies and their applicability to future mobile satellite architectures.

5.4.1 Fixed and Broadcast satellite systems and standards

5.4.1.1 DVB-SH, DVB-S/S2 and DVB-RCS

One approach for the definition of the communications protocols would be to take the recently defined DVB-SH standard as the basis for the forward-link and introduce a DVB-RCS based return-link access, in a similar fashion as DVB-S/S2 is complemented by DVB-RCS (of course a more optimized solution would be provided by a new "RCS" standard; optimized for below 3 GHz services and which is based on more spectrally efficient mobility access techniques (following developments in satellite and terrestrial UMTS - including Long Term Evolution multiple access methods), and which also specifies directly the use of advanced communication techniques). This approach would be more in line with the general motivation behind the medium-term architecture, i.e. the introduction of bidirectional data services over mobile broadcasting systems. It also has the potential advantage that DVB-SH receivers could breakthrough in the mass-market (for mobile broadcasting), and would thus be a more compatible/economical proposition to re-use DVB-SH for carrying unicast/multicast data as well (such approach could also allow a more gradual recommitment of resources; from broadcasting service to unicast/multicast services). DVB-RCS is already a commercialized technology for fixed satellite access (and thus not optimised for mobile-handheld services), but it is recently being investigated, within the related DVB group, for application in higher frequency (Ku/Ka band) mobile-vehicular environments. It turns out that mobility issues (mainly related to initial log-on, maintaining network timing synchronisation, and spot-beam handover) can relatively easily be accommodated by RCS (at least with GEO constellations). However, additional issues would arise in applying DVB-RCS in the spectrum-limited below -3 GHz bands, as DVB-RCS consumes substantial amounts of resources for signalling and control overheads. Moreover, RCS receivers would need to be adapted for LMS channel reception and data rates would have to be scaled down to a few Kbit/s. For interworking with DVB-SH, another issue is the delays introduced in the forward-link signalling due to the long time interleaving. DVB-SH standard today offers the flexibility to configure the interleaver length from 100 ms to 200 ms delay (class 1 receivers) to several seconds (class 2 receivers) Thus, Unicast applications would required the use of short physical layer interleaver while broadcast applications sharing the same multiplex would need to relay in the PL-FEC protection. An alternative approach would be the introduction of different simultaneous types of interleaver protection which is currently not allow by the DVB-SH standard.

5.4.1.2 SATMODE

A lower-rate and reduced overhead standard that is compatible with a DVB-based forward link is SATMODE [i.331]. However, specifying a contention based access method, and being designed for TV interactive applications, SATMODE would be too spectrally inefficient and would require mobility application studies; to confirm its suitability for a mobile service.

5.4.1.3 Conclusions on existing fixed and broadcast satellite systems and standards

In summary, the DVB-based approach is less compatible with 3G terrestrial systems and therefore can be considered as a more aggressive system development approach, where competitiveness rather than complementary towards terrestrial networks is followed (assuming a hybrid overall system architecture is to be deployed). In 2009 mobility features were added to the DVB-RCS standards.

An alternative approach would be to introduce non-DVB based communication protocols that would operate in isolation to DVB-SH (by re-committing a specific amount of bandwidth from broadcasting to unicast), though such an approach would lead to more complex multi-standard terminals.

5.4.2 Mobile satellite standards

5.4.2.1 Existing mobile satellite systems and standards

There are a range of Mobile Satellite Systems (MSS) that are expected to continue in operation over the medium term timescale. For example:

- Systems based on Geostationary satellites (GSO); for example:
    - The GMR-1 system, especially the latest Release 3 system (GMR-1 3G).
    - The Inmarsat system, especially the latest BGAN system.
    - The GMR-2 system.
- Systems based on Non-geostationary satellites (NGSO):
    - The Globalstar system.
    - The Iridium system.

Many of these systems are already being further developed and the second and third generations of these systems are evolving more capable systems that meet many of the requirements of the medium term architecture.

These MSS systems typically provide a satellite radio interface into a standard "mobile" core network (i.e. uses standard terrestrial mobile core network elements). This is an important advantage when compared to fixed and broadcast systems since the core network and the associated higher layer protocols (mobility, authentication, etc.) represent a large part of the overall system complexity.

5.4.2.2 New and emerging mobile satellite systems and standards

The WCDMA based S-UMTS standards [i.2] are another candidate for the medium term architecture. These standards are currently incomplete and have not been deployed. However, a WCDMA based S-UMTS systems would be more compatible with terrestrial systems; allowing convenient roaming options to be offered to the user.

Taking WCDMA based S-UMTS as a possible starting point, evolved specification could be derived; following for example the paradigm of LTE-UMTS and mobile WiMAX, which adopt more spectrally efficient access technologies and specify state-of-the-art physical and upper-layer features.

5.4.2.3 Conclusion on mobile satellite systems and standards

Mobile satellite systems have evolved significantly over the last decade and most of the current systems can support a range of terminals from small handheld terminals up to larger transportable terminals. The larger terminals can support data rates that are comparable to 2,5 G terrestrial systems, but in all cases these mobile terminals use a less directional antenna (to permit less precise pointing) and this limits the maximum data rates when compared to a fixed satellite terminal (which can have a highly directional and more accurately pointed antenna).

In general, any mobile satellite radio interface requires significant optimization to achieve optimum performance over the satellite link. Moreover, a different radio interface is needed for the different satellite orbits: i.e. a Geostationary (GSO) satellite and a Non-geostationary (NGSO).

Mobile satellite systems provide a complementary role to the terrestrial mobile systems, typically by providing increased geographic coverage. However, they can also supply another important complementary role in emergency situation (disaster relief, major emergence) by providing a telecommunications services that is completely independent of the terrestrial infrastructure.

Demand for mobile satellite services is expected to grow and the system requirements will need to evolve in line with the growing capabilities of the terrestrial mobile systems. Existing mobile satellite systems and standards are already evolving to use more advanced modulation and coding scheme to provide higher data rates, and further evolution can be expected.

In parallel, there is the additional option for one or more new mobile satellite technologies to emerge, based on a satellite optimization of the emerging terrestrial standards.

In all cases, a common element is likely to be a continued use of a standard terrestrial core network, with the satellite specific elements restricted to the satellite radio interface.

5.5 New existing mobile Terrestrial Standards

This clause reviews the latest terrestrial mobile standards, namely mobile WiMAX and 3GPP Long Term Evolution (LTE) of UMTS, which are expected to be the two main competitive technologies in the race towards 4G networks. Both WiMAX and LTE, adopt many of the state-of-the-art communications technologies reviewed in clauses 5.2 and 5.3, and can thus be used as strong references in defining enhanced SatCom standards.

5.5.1 IEEE Mobile WiMAX

The mobile WiMAX standard has been developed by the IEEE 802.16 [i.333] Working Group (on Wireless Access Standards), that is involved with preparation of specifications for broadband Wireless MANs. In 2004 the IEEE 802.16d [i.334] standard was released, which specifies fixed non-LOS wireless access (outdoor and indoor) in metropolitan areas. In 2005 this standard was extended (IEEE 802.16e [i.335]) for supporting mobile terminals, and this is the standard commonly referred as mobile WiMAX. The main innovation brought into IEEE 802.16e [i.335] is the introduction of Scalable OFDMA (SOFDMA) mode, which supports scalable channel bandwidths (from 1,25 MHz to 20 MHz), but is not backwards compatible with the OFDM mode in fixed WiMAX.

The IEEE 802.16 [i.333] standards provide a system definition up to the MAC layer. In order to address network architecture issues, the WiMAX Forum has been recently working on the specification of the network architecture. The first version of the Network Architecture specification was released in early 2007. Release 1.5 is expected to add support for telecom-grade mobile services, full IMS interworking, carrier-grade VoIP, and broadcasting applications like mobile TV.

Mobile WiMAX systems offer scalability in both radio access technology and network architecture, allowing flexibility in network deployment options and service offerings. Some of the key features supported by Mobile WiMAX are:

- **Advanced Physical Layer Techniques:** MIMO, sub-channelization schemes and ACM provide high spectral efficiencies (peak DL data rates up to 63 Mbps per sector and peak UL data rates up to 28 Mbps per sector in a 10 MHz channel).

- **Quality of Service (QoS):** It defines Service Flows which can map to DiffServ code points or Multi-Protocol Label Switching (MPLS) flow labels that enable end-to-end IP based QoS. Additionally, sub channelization and Media Access Protocol (MAP) based signalling schemes provide a flexible mechanism for optimal scheduling of space, frequency and time resources over the air interface on a frame-by-frame basis.

- **Scalability:** For operation in different channelizations from 1,25 MHz to 20 MHz to comply with varied variable bandwidth availability constraints.

- **Security:** Features provided are within the Extensible Authentication Protocol (EAP) class, Advanced Encryption Standard (AES) - CCM (Counter with Cipher-block chaining Message authentication code) based authenticated encryption, and CMAC (block Cipher-based Message Authentication Code) and HMAC (keyed Hash Message Authentication Code) based control message protection schemes. It also support a diverse set of user credentials exists including; SIM/USIM cards, Smart Cards, Digital Certificates, and Username/Password schemes based on the relevant EAP methods for the credential type.

- **Mobility:** supports optimized handover schemes with latencies less than 50 milliseconds to ensure real-time applications such as VoIP perform without service degradation. Flexible key management schemes assure that security is maintained during handover.

This review has been largely based on [i.154]. The reader is also directed to [i.155] which provides a comparative analysis between mobile WiMAX and 3GPP HSDPA/HSUPA. Also see annexes G and I.

5.5.2 3GPP evolutions including LTE (Long Term Evolution of UMTS)

Since the Release 99 of UMTS, which has been the basis for WCDMA based S-UMTS [i.2], the 3GPP has produced upgraded versions of the standard (Releases 5, 6 and 7) that reduce upper layer latencies, support new services (such as MBMS) and crucially introduce high throughput channels in the downlink and uplink. More specifically, in Release 5 a new downlink transport channel was introduced in 2002: High Speed Downlink Shared Channel (HSDPA) which provides peak data rates up to 14 Mb/s, and also reduces latencies. The High Speed Uplink Packet Access (HSUPA) transport channel was introduced in Release 6 (2004) in order to complement HSDPA in the uplink; by providing data rates up to 5,8 Mb/s. Moreover, in Release 6, MBMS and enhancements in IMS (such as Push to Talk over Cellular) were introduced, as well as support for integrated operation wit Wireless LAN networks. In mid 2007 Release 7 was made available which adds MIMO, and high order modulation (64-QAM in the DL and 16-QAM in the uplink), providing peak data rates of 42 Mb/s and 11,5 Mb/s in the DL and UL, respectively. Release 7 also specifies reduced latencies and improved QoS; including improvements in real-time applications like VoIP.

In summary the advanced features introduced by the later UMTS Releases for enhancing the system's cost-effectiveness and performance are:

- For Higher Data Rates:
  - Higher-order modulation formats (16-QAM, 64-QAM).
  - MIMO Techniques.
- For improved QoS and low latency:
  - Shorter Transmit Time Intervals (TTI), allowing lower round trip delays.
  - Dynamic Scheduling, allowing end-user traffic streams prioritized according to their service agreements.
- For higher capacity:
  - Shared channel transmission making efficient use of time, frequency, code and power resources.
  - Link adaptation for dynamic optimization of transmission parameters.
  - Channel-dependent scheduling assigning radio resources to users according to the instantaneous radio channel conditions.
  - HARQ for fast and efficient retransmission of corrupted data

This clause provides an overview of the latest 3GPP UMTS: Release 8, which is also referred as Long Term Evolution (LTE). In brief, LTE introduces OFDM/OFDMA in the downlink and SC-FDMA in the uplink. It also supports very high data rates; exceeding 100 Mb/s in the downlink and 50 Mb/s in the uplink. Moreover, the new set of standards supports bandwidth scalability, from 1,25 MHz up to 20 MHz, and also both FDD and TDD modes of operation.

Table 5.3: Performance requirements for the LTE of UMTS

Metric	Requirement
Peak data rate	DL: 100Mbps UL: 50Mbps (for 20MHz spectrum)
Mobility support	Up to 500kmph but optimized for low speeds from 0 to 15kmph
Control plane latency (Transition time to active state)	< 100ms (for idle to active)
User plane latency	< 5ms
Control plane capacity	> 200 users per cell (for 5MHz spectrum)
Coverage (Cell sizes)	5 – 100km with slight degradation after 30km
Spectrum flexibility	1.25, 2.5, 5, 10, 15, and 20MHz

See annex H and I for more details on this subject.

5.5.3 Conclusions on New existing mobile Terrestrial Standards

Mobile WiMAX and 3GPP LTE are the latest terrestrial mobile standards, which are expected to be competing for broadband data provision in metropolitan areas. Both standards are OFDM-based and adopt state-of-the-art communications techniques and evolved upper layer concepts (e.g. for efficient mobility management, vertical system handovers), which provide unprecedented spectral efficiencies, range of services and flexible system interoperability.

Any new mobile SatCom system based on adaptation of these emerging terrestrial standards would follow the paradigm of WiMAX and 3GPP LTE; in terms of the specification of state-of-the-art communication techniques, and also in terms of an all-IP network architecture, which permits efficient interoperability with terrestrial mobile systems.

6 Analysis of Long-Term MSS Architecture

6.1 4G concepts

During recent years there have been many attempts to define 4G (e.g. [i.281] to [i.284]), but despite the big efforts by industry and academia, a well established and widely accepted definition of 4G has not yet emerged. This is explained to a large extend by the complex interactions between the key involved players (i.e. terminal and infrastructure equipment manufacturers, academia, operators, service providers, regulatory bodies and governmental agencies), who do not always share the same interests, goals and plans. Thus the 4G arena appears to be fragmented, as this is evident by the several different visions and technological paths supported by different regional players around the world.

The first of the main competing visions for 4G is the so called "*vertical approach*", which foresees the linear extension of current 3G systems, aiming for higher data rates. This vision is limited to highlighting the high speed capabilities of future communication systems, and is still the prevalent approach to 4G in Asia, where notably Korea, Japan, China and India are the major players. In North America, emphasis on the high-data rate side of 4G has prevailed, though mainly through the development of wireless LANs. More recently Asia and America have concentrated on the development and enhancement wireless MANs.

The second "horizontal approach" or concurrent vision of 4G (often identified as the European vision of 4G) is based on the integrative role of 4G as a convergence platform of several networks, and includes the linear vision as one of its constituent component networks. The horizontal approach is more in line with the visions of the ITU. Indeed, the ITU-R Recommendation M.1645 [i.285], states that future wireless communications systems could be realized by functional fusion of existing, enhanced and newly developed elements of current 3G systems, nomadic wireless access systems and other wireless systems with high commonality and seamless inter-working. The ITU approach is flexible in allowing legacy systems (2G and 3G), the products of their evolutionary development, and new systems to coexist, each being a component part of a highly heterogeneous network: the 4G network. Backward compatibility and interoperability are thus made key characteristics of 4G.

On the European level, preliminary explorations tend to show that many new useful services could be developed under the assumption that a ubiquitous, high-speed wireless access is available. Another common belief among proponents of 4G architectures (and related technologies) is that future users will be attracted by rich-content based services that pervasively interact with the environment (e.g. see [i.210]). Thus, it appears that one of the main driving forces for 4G development is the growing demand for higher data throughput in virtually every possible scenario. Also, heterogeneity and convergence are two of the most commonly referred distinctive features of 4G, which apply to networks, terminals and services. 4G also appears to be open to the most recently developed technologies, and often departure from many conventional solutions used in previous generations is supported; implying a potentially revolutionary rather than evolutionary approach to IMT-2000 technologies. Multi-antenna techniques are (justifiably) identified as one of the key enabling technologies, pointing to a likely departure from the relatively simple single-antenna transceivers, to systems supporting several parallel receive and transmit branches. Network architectures are also expected to be highly diversified in 4G, with a more balanced participation of centralized and distributed network management approaches. Furthermore, interaction among wireless entities is expected to be considerably strengthened in a mutual effort to better use resources and improve performance, leading to cooperation. Cooperating wireless entities include not only wireless devices, but also layers (of the OSI stack), algorithms, networks, processors, etc.

Figure 6.1 shows the 4G architecture, from a network coverage perspective, as envisaged in the horizontal approach. Top on the network hierarchy, is the distribution layer that provides large geographical coverage with full mobility. In this level, links may convey chunks of composite information rather than signals from individual subscribers; for instance broadcast services such as DAB and DVB. Next in the hierarchy is the cellular layer, with typical macro-cells of up to a few tens of kilometres. This network also provides full coverage, full mobility but now connections are intended to cater to individual users directly. It is noted that the cellular layer encompasses both macro and micro cells. The metropolitan layer or network (e.g. IEEE 802.16 [i.333]), provides urban coverage with a range of a few kilometres at most, with moderate mobility and moderate data speed capabilities. In a further smaller scale and moving to the local-area layer, e.g. indoor networks or short-range communications, the network provides access in a pico-cell, typically not larger than a few hundred meters; in order to fulfil the high capacity needs of hotspots. Nomadic (local) mobility is supported as well as global roaming. 3G makes use of the cellular layer (typically micro-cells) in combination with hotspots (WLAN), through vertical handovers, to provide coverage in dedicated areas. Next in the wireless network hierarchy is the Personal Area Network (PAN), for supporting very-short range communication links (typically 10 m or less) in the immediate vicinity of the user. Within this layer we can also include Body Area Networks (BAN), and some other sub-meter wireless short-range access. Wireless Sensor Networks (WSN) are also one essential constituent part of 4G networks. WSNs provide solutions to the problem of efficiently monitoring, collecting and distributing information in a distributed network made of (typically) a large number of nodes.

It is widely accepted that the integration of the diverse types of wireless networks needs to be done at the IP networking layer, because of the cohesive role of IP which allows enabling wide seamless connectivity across heterogeneous networks. An all-IP network, embracing the access and core networks, is the most straightforward and effective way to integrate all the possible different networks constituting the 4G network. Horizontal and vertical handovers will assure seamless intra and inter-network connectivity, respectively.

As 4G is envisaged to be at the union of heterogeneous networks, it follows that it needs to support heterogeneous terminals. The current trend is to have either single-mode or multi-mode terminals. Even though both approaches could easily find considerable market share, the latter will inherently better match the capabilities of 4G networks, namely handling multimedia information of various types supporting advanced services. Multifunctional devices represent the convergence of several technologies. Multi-functionality implies several air interfaces on board (e.g. wide-area, local-area and very short range), audio, imaging, positioning and other features. This convergence will allow users to have simultaneous or independent access to different networks with a single terminal. 4G will ultimately facilitate and expedite the three-screen convergence, bringing together the TV, PC and mobile phone screens into a single portable device.

Heterogeneous networks and terminals need to be finally complemented by heterogeneous services. In other words, heterogeneous services imply a wide range of services able to operate across different networks and in various types of terminals. In addition, convergence is essential in this context as the concept of multi-access services is becoming a reality.

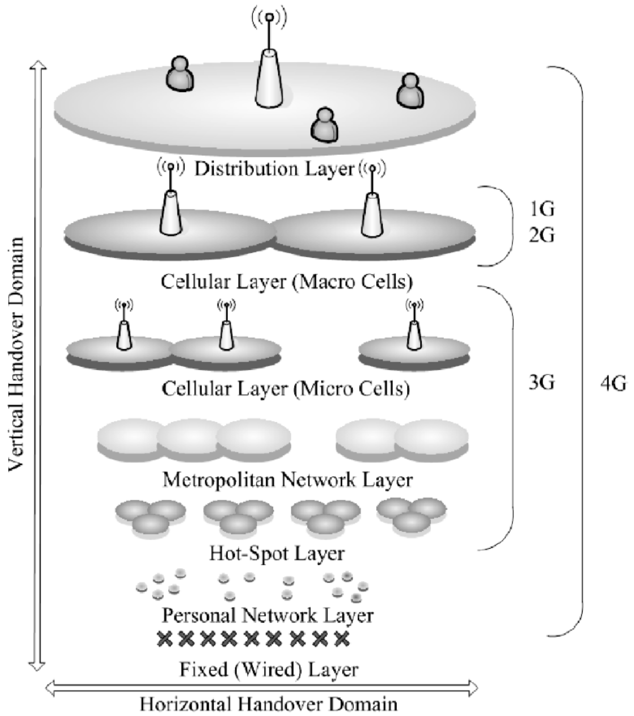


Figure 6.1: Network-Level Architecture in 4G

From a technology standpoint, 4G should embrace new techniques and technologies that overcome limitations and solve the problems of the previous generations. In particular, the difficulty for CDMA to achieve very high data rates in interference limited multi-user, multi-rate environments puts multi-carrier techniques in a favourable position in 4G. Also, another problem in current wireless systems is the difficulty of providing a full range of multi-rate services with different QoS requirements due to the constraints imposed on the core network by the air interface standard (it is not a fully integrated system). 4G needs to tackle also the lack of end-to-end seamless transport mechanism. Other important constraints of current mobile systems are the limited availability of spectrum and its particular allocation as well as the difficulty of roaming across distinct service environments in different frequency bands.

4G is also expected to favour short-range links, and air interfaces supporting local access. In addition to conventional narrow and wide band transmission techniques, Ultra Wide Band (UWB) techniques have lately received considerable attention, in particular as a non-intrusive, low power and low cost alternative to other short-range communications methods. In addition, optical wireless communications also provide a viable alternative for short-range links. Optical wireless systems can be used not only for point-to-point links, like those standardized through the Infrared Data Association (IrDA) [i.286], but also for full-mobility indoor applications based on either infrared or visible light [i.287] to [i.289]. Among the main advantages of optical wireless systems, are the virtually unlimited bandwidth, and inherent security, as the optical signal is confined within the operational scenario. Moreover, in optical systems no RF radiation is generated (consequently neither interference pollution nor possible health hazards are produced, thus they are well suited to sensitive environments).

Table 6.1 summarizes some of the main expected characteristics of "horizontally developed" 4G system architectures.

Table 6.1: Expected Characteristics of 4G Systems

Data transfer capability	100 Mbps (wide coverage)
	1 Gbps (local area)
	Design targets representing overall cell throughput.
Networking	All-IP network (access and core networks)
	Plug & Access network architecture
	An equal-opportunity network of networks
Connectivity	Ubiquitous
	Mobile
	Seamless
	Continuous
Network capacity	10-fold that of 3G.
Latency	Connection delay ¡ 500 ms
	Transmission delay ¡ 50 ms
Cost	Cost per bit: 1/10-1/100 that of 3G
	Infrastructure cost: 1/10 that of 3G
Connected entities	Person-to-person
	Person-to-machine
	Machine-to-machine
4G Keywords	Heterogeneity of networks, terminals and services
	Convergence of networks, terminals and services
	Harmonious wireless ecosystem
	Perceptible simplicity, hidden complexity
	Cooperation as one of its underlying principles.

The very ambitious requirements of 4G systems, in terms of performance, cost, coverage and network/terminals/services heterogeneity and convergence, have led to the necessity of considering, developing and adopting new technologies in all levels of the system design. Apart from the purely telecoms system and network design challenges, another important issue is the expected very high power consumption of 4G terminals. The higher power consumptions will come as a result of:

- The much higher data rates.
- The "always connected" requirement.
- The complex multi-standard devices.
- The shift towards higher frequency bands.
- Support for multimedia applications and powerful processors for multi-purpose computing tasks.
- Satellite reception capabilities for at least positioning services.

From the 4G terminal manufacturer perspective the power consumption problem is critical, not only technically but also taking into account the market expectations from a newly introduced technology. The long operational time capability of terminals is both satisfying and vital for users; it gives them a truly wireless experience. This feature has been put at the top of the wish list by consumers as shown recently in [i.290], and therefore it is to be taken seriously by the industry, and indirectly, by the research community.

One of the key strategic concepts that defines the focus of research into 4G enabling technologies is ***cooperation*** at all levels of the system and network architecture.

On the first level flexible spectrum cooperation between sub-systems and sub-networks operating within the transparent 4G entity, is widely considered as a necessary step in order to meet the very demanding data-rate, capacity and cost-per-bit requirements. Among other spectrum sharing models that have been proposed for different types of system scenarios, dynamic spectrum sharing between co-existing systems has recently gained a big research momentum through the concept of cognitive radios. Cognitive radios build on the flexibility of software defined radios, adding the ability of mobile nodes to analyze the spectrum environment and take autonomous, local and dynamic spectrum access decisions. Thus, in principle cognitive radios allow highly flexible dynamic spectrum access to be realized, which in turn offers a fundamental solution to the spectrum scarcity problem; presently limiting the achievable system capacities and costs of services.

On a second level, the fact that 4G is a platform embracing different networks makes the ideal setting for exploring network-level cooperation. A few years ago wireless LANs and decentralised ad-hoc networking ideas, were perceived as threatening competition by cellular operators. However this pessimistic view is changing as operators get more and more accustomed to the 4G vision of converged-transparent networks, and the big new market opportunities this new model will bring. Currently, convergence is becoming a reality with the availability of multi-network devices, which although does not imply cooperation, it is an important step towards it. The next step is expected to be the introduction of complementary between networks, mainly for supporting vertical handovers. However network cooperation can be taken much further than that, as it is advocated by the World Wireless Research Forum [i.210]. As the constituent 4G elements range from distribution (e.g. broadcast) networks down to personal networks, the possibilities for inter network cooperation are in principle numerous. Cooperation strategies between more than one network can be devised around a single user, assuming that his terminal is equipped with multiple air interfaces, and cooperation can in principle take place at any of the OSI layers. Examples of network-level cooperation are:

- ***Air-interface diversity*****:** Data rate or QoS can be improved by combining multiple signals jointly provided by the cellular and ad-hoc networks. Multiple description coding could be used for this purpose, for instance for video streaming, as suggested in [i.291]. Extending the centralized network with a locally distributed one, formed by an ad-hoc group surrounding the target user, will help to convey more efficiently the high-data rate information to the destination.

- ***Security Support to Ad-Hoc Networks*****:** Peer-to-peer communications over a short-range links could be seen sometimes as risky, from the point of view of interacting with an untrusted (or unknown) counterpart. Thus, through cellular-controlled short-range communication the base station could take the role of verifying, authenticating and making secure a given transaction. If service is requested over a short-range link to a machine the infrastructure network could intervene providing initial secure configurations for the transaction (including distribution of keys) and billing services.

- ***Local retransmission*****:** The interaction between centralized and decentralized networks can be exploited to improve, among others, spectral and power efficiency. Typically, centralized approaches (e.g. cellular networks) consume spectrum and require more power, while decentralized approaches (e.g. WLAN) operate in unlicensed frequency bands and require lower power levels. Cooperation between these two networks will aim to use as much as possible short-range links, bringing advantages to users (in the ad-hoc networks) as well as to the operators.

- ***Synchronization*****:** For some purposes local synchronization may be required, leading to a common reference time among a number of nodes. This common timing could be defined at different layers, e.g. physical and application layers. In the former, a distributed process may need a precise common temporal reference, which may not be straightforwardly distributed by a central entity. By combining master-slave and mutual synchronization approaches provided by the cellular and ad-hoc networks respectively, local synchronization can be obtained. At the application layer, some services shared by a group may need to have a common timing reference, e.g. aligning video and audio signals on the group users, as suggested in [i.292].

- ***Routing Support for Ad-hoc Networks*****:** A fundamental difficulty in deploying large scale decentralised ad-hoc networks is the complexity, stability and overheads required for maintaining the routing mechanism. This limitation can be overcome when an overlay infrastructure network is supporting the routing functions for the ad-hoc network. Specific protocol designs have been recently demonstrated the advantages brought by this level of cooperation.

Cooperation can also be applied at the physical layer through fixed and mobile relaying techniques, which bring higher link capacities, fading diversity benefits, and coverage extensions for infrastructure networks. Different types of relaying protocols (such as amplify and forward, decode and forward, compress and forward, etc.) have recently received high research interest and practical physical layer architectures have been proposed for practical system implementations. An extension of relay and ad-hoc techniques are the virtual antenna arrays, which are formed by neighbouring terminals; in order to apply distributed space-time coding and beam-forming techniques.

6.2 Dynamic Spectrum Sharing and Cognitive Radio

6.2.1 Introduction

Dynamic spectrum sharing has recently been identified by telecoms regulatory bodies as one of the timeliest technological objectives, that needs to be pursued and realised to the maximum possible degree in future radio communications systems. The idea foresees the alteration of the current regulatory regime, in which pieces of bandwidth are allocated and operated exclusively by a single network operator on a national, regional or global basis. The need for changes in the allocation, management and usage policies of the spectrum was made clear by recent spectrum usage studies, which conclude that lack of available spectrum in many bands is primarily a result of low (on average across time and location) spectrum occupancy, and not due to the lack of physical spectrum. The high relevance of evolving radio regulations and associated technologies towards more open and dynamic spectrum usage models, is made clear by the major findings and recommendations in the study by the FCC's Spectrum Policy Task Force, published in 2002 [i.157]:

- Advances in technology create the potential for systems to use spectrum more intensively and to be much more tolerant of interference than in the past.
- In many bands, spectrum access is a more significant problem than physical scarcity of spectrum, in large part due to legacy command-and-control regulation that limits the ability of potential spectrum users to obtain such access.
- To increase opportunities for technologically innovative and economically efficient spectrum use, spectrum policy is likely to evolve towards more flexible and market-oriented regulatory models.
- Such models are likely to be based on clear definitions of the rights and responsibilities of both licensed and unlicensed spectrum users, particularly with respect to interference and interference protection.
- No single regulatory model should be applied to all spectrum: the Commission should pursue a balanced spectrum policy that includes both the granting of exclusive spectrum usage rights through market-based mechanisms and creating open access to spectrum "commons," with command-and-control regulation used in limited circumstances.
- The Commission should seek to implement these policies in both newly allocated bands and in spectrum that is already occupied, but in the latter case, appropriate transitional mechanisms should be employed to avoid degradation of existing services and uses.

The FCC recommendations report [i.157] highlights different types of evolved spectrum access models that differ with respect to the access rights of different systems and in the degree of supported dynamic access; and are therefore more or less suitable for different bands. A well presented summary of the different types of dynamic access models is found in [i.158]:

Dynamic Exclusive Use Model: This model maintains the basic structure of the current spectrum regulation policy: Spectrum bands are licensed to services for exclusive use. The main idea is to introduce flexibility to improve spectrum efficiency. Two approaches have been proposed under this model:

- ***Spectrum property rights [*i.159*]* and [**i.160**]:** This approach allows licensees to sell and trade spectrum and to freely choose technology. Economy and market will thus play a more important role in driving toward the most profitable use of this limited resource. Note that even though licensees have the right to lease or share the spectrum for profit, such sharing is not mandated by the regulation policy.
- ***Dynamic spectrum allocation [*i.161*]*:** This approach was brought forth by the European DRiVE project. It aims to improve spectrum efficiency through dynamic spectrum assignment by exploiting the spatial and temporal traffic statistics of different services. In other words, in a given region and at a given time, spectrum is allocated to services for exclusive use. This allocation, however, varies at a much faster scale than the current policy.

Based on an exclusive-use model, these approaches cannot eliminate white space in spectrum resulting from the bursty nature of wireless traffic.

Open Sharing Model: Also referred to as spectrum commons [i.162] and [i.163] this model employs open sharing among peer users as the basis for managing a spectral region. Advocates of this model draw support from the success of wireless services operating in the unlicensed industrial, scientific, and medical (ISM) radio band (e.g. Wi-Fi). Centralized [i.164] and [i.165] and distributed [i.166] to [i.168] spectrum sharing strategies have been initially investigated to address technological challenges under this spectrum management model.

Hierarchical Access Model: This model adopts a hierarchical access structure with primary and secondary users. The basic idea is to open licensed spectrum to secondary users while limiting the interference perceived by primary users (licensees). Two approaches to spectrum sharing between primary and secondary users have been considered:

- ***Spectrum underlay*:** imposes severe constraints on the transmission power of secondary users so that they operate below the noise floor of primary users. By spreading transmitted signals over a wide frequency band (UWB), secondary users can potentially achieve short-range high data rate with extremely low transmission power. Based on a worst-case assumption that primary users transmit all the time, this approach does not rely on detection and exploitation of spectrum white space.

- ***Spectrum overlay*:** This approach was first envisioned by Mitola [i.169] under the term spectrum pooling and then investigated by the DARPA Next Generation (XG) program under the term opportunistic spectrum access. Differing from spectrum underlay, this approach does not necessarily impose severe restrictions on the transmission power of secondary users, but rather on when and where they may transmit. It directly targets at spatial and temporal spectrum "white spaces" by allowing secondary users to identify and exploit local and instantaneous spectrum availability in a non-intrusive manner.

The different dynamic access models that have been proposed so far are summarized in figure 6.2.

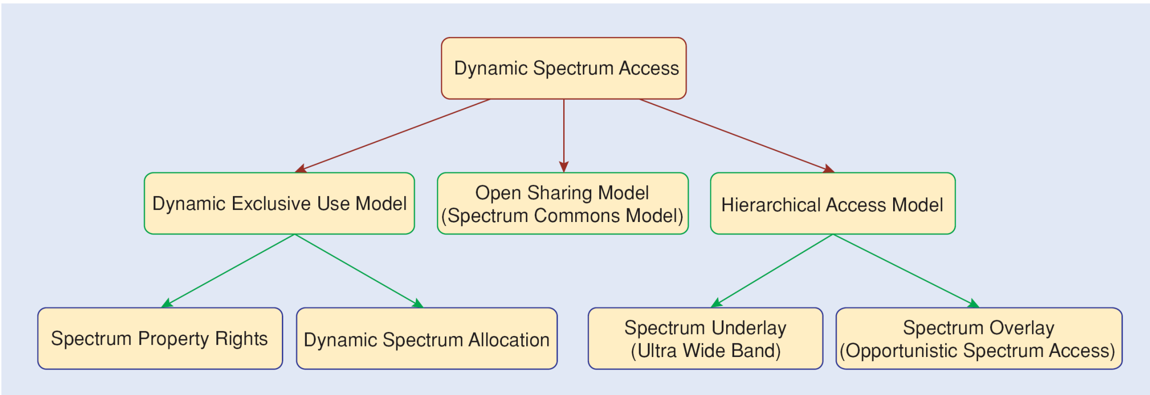


Figure 6.2: A taxonomy of dynamic spectrum access

6.2.2 Existing Examples of Dynamic Spectrum Access

Most of the above dynamic access models have already been followed in different bands, and have evidently stimulated development of new commercial systems and more effective utilization of the spectrum resource. Examples of dynamic access model adoptions in different bands are reviewed briefly in this clause.

a) ISM band (2,4 GHz): Open Sharing Model

Perhaps the most successful example has been the adoption of the open sharing model in the ISM band, which has allowed the development of the Wi-Fi and Bluetooth standards. Systems operating under the open sharing model need to follow some strategy for handling inter and intra system interference and facilitating coexistence of different devices (from the same or from different systems). Interference management approaches can be based on:

1) A cross-layer approach: e.g. though spectrum sensing in order to detect occupied parts of the spectrum and choose accordingly active OFDM sub-carriers.

2) Interference-robust modulation techniques: e.g. through the use of spread-spectrum, or frequency-hopping techniques for achieving interference diversity.

3) Advanced signal processing techniques: e.g. adaptive beamforming, multi-user detection, precoding, dirty-paper coding.

b) VHF-UFH bands (54 MHz to 862MHz): Spectrum Overlay Model

Another example of dynamic spectrum access model adoption is the one promoted by FCC, for the dynamic sharing of terrestrial TV bands (VHF-UHF) with wireless wide-region area networks, that can be operating freely but on a secondary basis. This type of dynamic sharing falls in the category of Spectrum Overlay (within the Hierarchical Access Model), and requires the secondary network to actively scan the spectrum in order to avoid causing interference to the primary TV network.

A Wireless Regional Area Network (WRAN) system that will be utilizing "white spaces" (i.e. idle frequencies) in the terrestrial TV radio spectrum is currently being standardised by the by the IEEE 802.22 [i.346] work group. Different approaches for the implementation of the interference avoidance mechanism are being considered. A centralised approach requires the user terminals (Access Points) to be equipped with a GPS receiver in order report their position to centralised servers (e.g. managed by FCC in USA) in order to determine the availability of free TV channels in the specific area. A more decentralised approach is to allow local spectrum sensing by the access points in order to determine independently which of the bands are locally available. This second approach falls within the cognitive radio and cognitive network concepts proposed in [i.169]. A combination of the two approaches is also considered as a more robust option.

c) 2 GHz, L, and1.6 GHz to 2,4 GHz bands: Dynamic Spectrum Allocation

In 2003 the FCC approved a proposal filed by several mobile satellite services operators (including ICO and Mobile Satellite Ventures (MSV)) to allow the terrestrial re-use (over the US landmass) of MSS bands in order to support reliable high-rate services in urban areas, through the use of Ancillary Terrestrial Components (ATCs). The integrated satellite-terrestrial architecture envisaged by MSV, aims at a transparent network coverage accessed through a common cellular-style handheld terminal, as illustrated in figure 6.3.

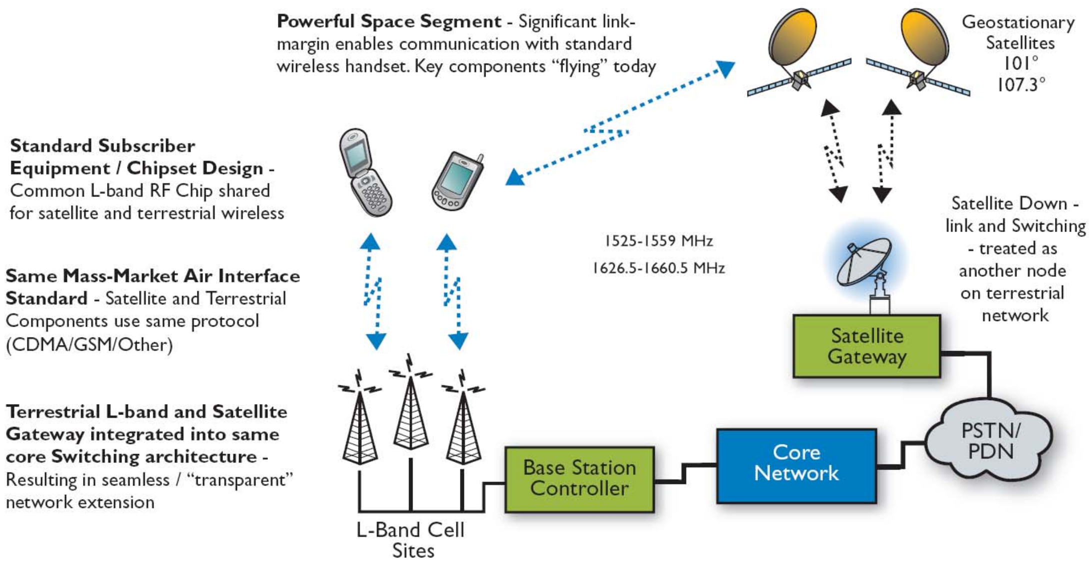


Figure 6.3: Hybrid Terrestrial/Satellite Wireless Network

The FCC has established limits on a US-wide Ancillary Terrestrial Network (ATN) deployment such that the uplink interference potential to a co-channel satellite of another MSS operator would not exceed a predetermined limit. In predicting uplink interference, the FCC model took into account several interference mitigation factors, such as outdoor blockage, power control, vocoder factor, voice activity, and polarization isolation. Further interference studies have been carried out by MSV [i.171] in order to quantify the inter-system interference levels generated to other MSS systems by the ATN. Within the hybrid network, interference between the satellite and terrestrial component is avoided by allocating frequencies to ATCs which are not used within the wider satellite spot beam in which they are situated. Still, intra-system interference can occur in the return-link and this needs to be mitigated at the gateway through joint adaptive beam-forming and multi-user detection, as it is discussed in [i.172].

d) Ultra-Wideband (3,1 GHz to 10,6 GHz): Spectrum Underlay

Ultra-Wideband (UWB) radio systems operate over very large bandwidths but with very low power spectral densities, i.e. according to the spectrum underlay model. They have recently emerged as a competitive option for providing short range (PAN) data communications and other short range radar and positioning applications. ITU-R and FCC classify as UWB systems which operate over bandwidths which exceed the lesser of 500 MHz or 20 % of the centre frequency. In 2002 FCC authorized the unlicensed use of UWB over the 3,1 GHz to 10,6 GHz frequency range, but restricted the emitted power spectral density to the limit of unintentional emissions (-41,3 dBm/MHz). Following recommendations in a report on UWB by ITU-R in 2005, other nations are also acting on the regulation of UWB.

6.2.3 Cognitive Radios for Realising Opportunistic Spectrum Access

NOTE: See also annex J.

Much of the research effort in dynamic spectrum access is focusing on the Opportunistic Spectrum Access (OSA) (or Spectrum Overlay) model. This is firstly because the OSA model provides the highest flexibility in terms of allowing secondary systems to exploit "spectral holes", and secondly because hierarchical models are most compatible with current spectrum management policies and legacy wireless systems. Furthermore, the practical realization of efficient and at the same non-intrusive OSA presents big technical challenges at various levels in the system's hardware and software (i.e. DSP and protocols) design.

The Cognitive Radio (CR) paradigm proposed by Mitola [i.169], which relies on the ability of individual network nodes (terminals, base stations, etc.) to analyze the radio environment and adapt their radio profile and take autonomous (or more centralized) spectrum access decisions, is widely considered as one of the most suitable models for developing OSA systems. The cognitive radio model builds on software-defined radios, which in their general form are multiband radios that support multiple air interfaces and protocols and is reconfigurable through software run on DSP or general-purpose microprocessors [i.173] The addition of cognition is intended to allow software radios to be aware of different aspects of the surrounding radio environment and be capable of performing autonomous reconfigurations through learning. OSA can be viewed as one of the potentially many applications of cognitive radio.

According to [i.158], the application of CR for OSA (under the assumption of no cooperation between the primary and secondary systems) consists of three main functions:

a) Spectrum Opportunity Identification

This is the most fundamental function that needs to be carried out by the pair of secondary transmitter and receiver in order to identify available channels for exploitation. A spectrum opportunity requires that no primary receivers are within the "harmful vicinity" of the secondary transmitter, and conversely that the secondary receiver is not in the "harmful vicinity" of primary transmitters. The former task is challenging to carry out directly, and usually the problem is transformed to that of detecting primary transmitters [i.158]. The term "harmful vicinity" implies that some levels of interference might be allowed to occur, subject to possible regulatory rules. The requirement for joint detection of spectrum opportunities by the secondary transmitter and receiver means that there are not only signal processing functionalities that need to be implemented, but networking ones as well. A related MAC layer protocol design has been proposed in [i.177].

Spectrum opportunity identification can serve not only in order to make immediate decisions on OSA, but also in order to acquire statistical information on spectrum occupancy so that more rewarding sensing decisions can be made in the future [i.158]. More specifically, assuming that the primary traffic statistics are modelled as a Markov process, then the opportunity detection process allows building information on the parameters of the underlying statistical process (e.g. state transition probabilities). This statistical knowledge can be used as prior information in order to enhance the reliability of future sensing measurements. The design of optimal sensing strategies has been formulated and addressed within the framework of Partially Observable Markov Decision Processes (POMDP) in [i.177] and [i.178].

Spectrum opportunity identification is of course an imperfect process, since it depends on the estimation performance of signal processing algorithms that carry out spectral estimation. The performance of detection algorithms is characterized by the probability of false alarm versus probability of missed detection curve. The imperfectness of the signal processing algorithms needs to be managed by the MAC layer on the second: *Spectrum Opportunity Exploitation* step.

b) Spectrum Opportunity Exploitation

Once spectrum opportunities are detected secondary users need to decide whether and how to exploit them. Specific issues include whether to transmit given that opportunity detectors will make mistakes, what modulation and transmission power to use, and how to share opportunities among secondary users to achieve a network-level objective.

Secondary users need an access strategy to determine whether to transmit over a particular channel based on the detection outcome. If the spectrum detector was perfect, the design of the access strategy would have been straightforward. In the presence of detection errors, the access strategy is complicated by the need to decide when and to what degree to trust the detector. The trade-off is between minimizing overlooked spectrum opportunities and avoiding collisions with primary users. The optimal access strategy should take into account the operating characteristics of the spectrum detector. Intuitively, when the miss detection probability of the detector is large (i.e. a busy channel is often detected as idle); the access policy should be conservative to avoid excessive collisions. On the other hand, when the detector has a high false alarm probability, the access policy should be aggressive to reduce overlooked spectrum opportunities. A framework for optimal access has been established in [i.179].

With respect to modulation, OFDM is an attractive option for the secondary systems since it allows utilizing non-contiguous frequency bands. Nevertheless there are issues that need to be addressed such as matching the sub-carrier spacing and symbol time interval to the spectral and temporal durations of spectrum opportunities [i.180]. Also, cross-channel spectrum leakage caused by signal truncation in the time domain and nonlinearity of the transmitter's power amplifier needs to be controlled to ensure non-intrusive communication.

c) Compliance to Regulatory Policy

Exploitation of spectrum opportunities by secondary systems should also conform to general and more specific regulatory policies. These can for example dictate the maximum probability of collision due to opportunity identification errors. This can range from very restrictive policy, where systems should maintain absolute orthogonally at all times, to free access by secondary systems (e.g. in cases of national emergency).

Policies might have a limited validity depending upon several factors such as the local time, the country where the radio is residing, thus CRs might have to use the policies in an adaptive manner. Therefore, a well-defined policy framework is needed. Such a framework implies language constructs for specifying a policy, a machine-understandable representation of these policies and a reasoning instance. Policy confirmation validation that is responsible for downloading, updating and validating policies is also needed. The syntactical correctness of the policy that has been downloaded should be verified and after the validation process, the policies are converted into a machine-understandable language such as the Web Ontology Language (OWL) to enable the computations. The policy based learning process in CR is shown in figure 6.4.

The DARPA XG policy description language has typical three main elements: first, a selector description, which is used to filter policies to a specific environment; secondly, the opportunity description that specifies under which conditions the spectrum is considered unused; and thirdly, a constraint description which specifies the behaviour of the CR when using a spectrum opportunity [i.217]. The Universal Modelling Language structure of policies in DARPA XG policy descriptions language is shown in figure 6.5.

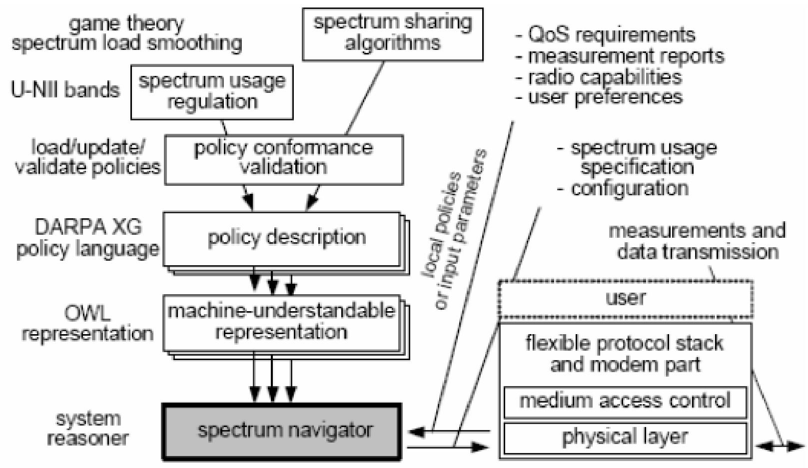


figure 6.4: Policy based learning in a CR

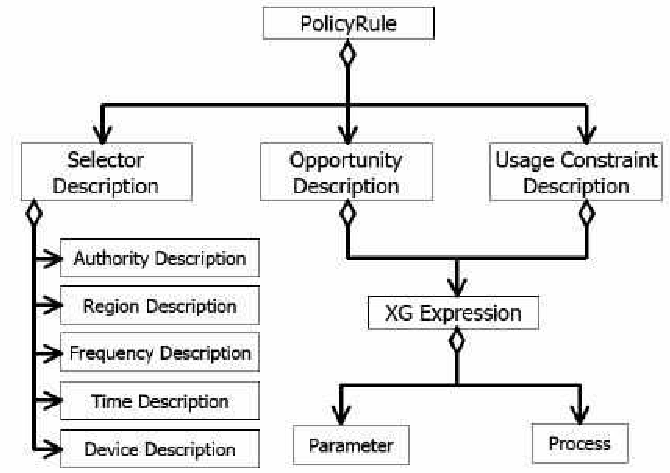


Figure 6.5: UML structure of a policy in DARPA XG language

6.2.4 Interference Mitigation Techniques

Interference mitigation techniques can play an important role in facilitation dynamic spectral co-existence of different systems. They can be applied complementarily to CR (for realizing the spectrum overlay model), in order to increase the rate of spectrum opportunities for secondary systems (particularly in the heavily used cellular bands), and also to suppress interference caused by opportunity detection mistakes. Interference mitigation can also be applied for facilitating system coexistence under the open sharing model, where all systems have equal access rights and are not protected by intra-system interference regulations.

Among other factors, the range of applicable interference mitigation techniques within different sharing scenarios depends on the degree of cooperation between systems (i.e. with respect to sharing system information, having access to each other transmitted data and possibly also performing cooperative processing tasks). In practical spectrum sharing scenarios, the most reasonable assumption is that co-existing systems do not cooperate in any degree, and therefore interference mitigation techniques should be designed accordingly.

However, in the domain of satellite systems some practical scenarios of inter-system cooperation can be envisaged, particularly with respect to sharing the satellite spectrum with terrestrial networks. One possible "co-operative" spectrum sharing scenario is between a hybrid S-DMB system and a terrestrial Wireless Metropolitan Area Network (e.g. WiMAX). In principle, the WMAN network can have prior (non-causal) access to the broadcasting transmissions and thus use some Dirty Paper Coding (DPC) [i.195] technique in order to allow interference-free reception. The rate-losses incurred to the broadcasting system due to the caused interference, can also be moderated, as the WMAN transmitter can commit some of its power resources in order to assist the broadcasting transmission. Recent information theoretic results (see [i.191]) show that such types of co-operative precoding techniques provide higher total capacities to the two co-existing systems, relative to conventional TDM sharing model. However in order to exploit the high capacity potentials of such type of co-operative precoding techniques, practical encoding and decoding algorithms should be devised and also feedback channels need to be used in order to provide the transmitter with channel state information. Thus, though promising, this type of advanced interference suppression co-operation techniques are still in early stages and their benefits have to be demonstrated under practical system assumptions.

Different types of available interference suppression techniques are reviewed in **annex K**, namely:

- Multi-User Detection for CDMA Co-Existing Systems.
- Linear Precoding in MIMO Systems.
- Dirty-Paper Coding for Cooperative Systems.

6.2.5 Conclusions on Dynamic Spectrum Sharing and Cognitive Radio

This clause has firstly reviewed the different types of spectrum sharing models that are currently being considered for different types of systems and associated bands. The models differ mainly with respect to the dynamicity of the frequency allocations, and on the access priorities of the co-existing systems. Different models are thus most suitable for certain types of system scenarios as this is evident by the so far examples of practical spectrum sharing systems. However, the OSA (or Spectrum Overlay) model is currently receiving a lot of interest from the research community, as it allows high flexibility (to secondary systems) in exploiting spectrum holes and allows compatibility with legacy systems. Key enabling technology for realising the OSA model is the cognitive radio paradigm, proposed by Mitola, which essentially adds intelligence to reconfigurable radio devices, in order to allow them to take decisions about the status of channels (free, occupied, or partially occupied) and exploiting autonomously detected opportunities.

The use of the cognitive radio concept for realising the OSA model appears to be a very promising proposition for terrestrial networks, as their coverage is highly dependent on location. Indeed the spectrum overlay model has been adopted by the IEEE 802.22 [i.346] standard, which specifies a wireless regional areas network operating opportunistically in the UHF-VHF (TV) bands. Different methods of avoiding harmful inference to TV receivers are being considered in the IEEE 802.22 standard [i.346], ranging from centralised control, to position based and spectrum scanning approaches. With national and regional regulatory bodies demonstrating high interest towards the OSA model, it is expected that the cognitive-radio enabled OSA model will be adopted in other terrestrial system scenarios in the near future. What is left to be proved is whether the fully decentralised decision making proposed by cognitive radio can be sufficiently robust in practice, or whether OSA will need to be supported by centralised entities or through cooperation between the primary and secondary systems.

With respect to satellite systems, the only significant development toward spectrum sharing has been the recent FCC regulations, which essentially adopt the Dynamic Spectrum Sharing model for allowing sharing the MSS bands with terrestrial networks. This allows MSS frequencies to be reused by terrestrial systems, in a non-overlapping coverage basis; according to the frequency re-use pattern of the satellite spot-beams. This development is very significant for satellite operators since it allows them to develop optimised hybrid system architectures that provide competitive services in urban regions, and coverage continuity in rural regions.

Subject to achieving the necessary amendments in regulatory policies and rules, sharing of the satellite bands with terrestrial networks can in principle be taken much further, particularly within the context of 4G networks where network level cooperation (between heterogeneous systems) is expected to be an inherent feature. This high-level cooperation can allow flexible and robust OSA of the MSS bands by terrestrial networks. A more decentralised cognitive-enabled approach can also be envisaged, where local area or other types of terrestrial networks detect and exploit idle spectrum and time slots in the satellite downlink transmission. The local nature of such terrestrial networks and the low densities of satellite terminals, combined with the cognitive radio procedures, would provide safeguards that no interference is caused to satellite terminals. Even more open to sharing is the satellite uplink, since the low powered (and possibly restricted to indoors) terrestrial terminals would be unlikely to cause any damaging interference to the satellite uplink; though this needs to be confirmed per scenario, through specific link-budget calculations. Such type of spectrum sharing would be in line with the spectrum underlay model, where secondary transmitters are allowed to radiate freely subject to power spectral density constraints. Similarly the feeder link frequencies could be utilised by terrestrial systems.

Obviously spectrum sharing between satellite systems would be difficult to achieve in a non-centrally coordinated manner. However one could envisage cooperation between systems at the resource management and MAC layer, in order to share their resources for the benefit of both.

Another important enabling factor for realising spectrum sharing between systems, are interference mitigation techniques, which can facilitate open sharing co-existence or safe-guard against unintended harmful interference due to the imperfections of dynamic sharing techniques. Similar to the dynamic MAC decisions taken by system operating within the OSA model, interference mitigations techniques can be designed subject to cooperation or non-cooperation assumptions by the co-existing systems. In the non-cooperation case, blind multi-user detection and MIMO precoding techniques have been proposed in the scientific literature. Under cross-system cooperation assumptions improved MIMO precoding can be derived, and also Dirty Paper Coding techniques can be considered. The latter approach has the advantage that it does not require multiple transmit antennas, but further research is required in developing practical coding schemes.

6.3 Radio Relays and Co-Operative Transmission Techniques

6.3.1 Introduction

The ambitious spectral efficiency and peak rates of 4G systems cannot be delivered cost-effectively by the classical cellular architectures (i.e. through "brute-force" increase of base-station densities). According to [i.210], co-operative transmission techniques (such as multi-hop/co-operative relaying and virtual MIMO transmissions), offer one of the most promising architectural upgrade in wireless systems that has been proposed for long time.

The specific benefits brought by such cooperative transmission techniques include:

- Reduction of propagation losses between source and destination, resulting in larger link data rates and potentially solving the problem of achieving high rates in large cells.
- In co-operative transmission schemes (i.e. not simple multi-hop relaying), capacity gains are also achieved by the more raw power contributed by the relay and the space diversity provided by the highly uncorrelated links. This is a key advantage of co-operative techniques, relative to classical multi-antenna schemes, which are generally unable to provide space diversity against slow-fading effects.
- When relaying through fixed infrastructure, the challenges of designing robust routing protocols are avoided. Even when the relays are mobile, the routing can be assisted by the overlay fixed/cellular network and thus the routing challenge is not as significant as in ad-hoc networks.
- Relays are low-cost and low-transmit power network elements, compared to base stations.
- The relay to receiver links could use a different system/band (e.g. Wi-Fi integrated in cellular), yielding significant gains from load balancing through the relays [i.219].

The advantages brought in by relay-enabled architectures are equally and perhaps even more significant in satellite systems, where in most scenarios even light shadowing reduces dramatically the link availability. In fact fixed relaying is already implemented in hybrid architectures such as the S-DMB system in order to facilitate urban coverage. In principle, the introduction of mobile relaying/co-operation capabilities in mobile terminals can yield further benefits in the availability of land-mobile satellite systems, especially in rural regions where fixed relaying infrastructure would not be economical to deploy. Even in terrestrially covered areas, co-operation between mobile nodes can ease the requirements for dense repeater networks and/or can in principle enhance further the coverage and data rates of the system. Nevertheless mobile relaying presents more design challenges, as this will be discussed later on in this clause, and also it cannot guarantee any predictable network coverage improvements. The latter is particularly true for general land mobile satellite system scenarios, where the geographical densities of users are low. Nevertheless the concept can prove useful for particular system scenarios (e.g. disaster relief).

Cooperative techniques can be applied in many different types of systems and application scenarios. However the motivation of using cooperation can be different and also the design of the co-operative system and related protocols will need to be adapted according to the applicable communication scenario (e.g. system bandwidth, required range, etc.). Possible systems and application scenarios for which cooperation techniques could be relevant (as an enabling/ enhancing technology) include:

- **Cellular systems:** Cellular systems are generally characterized by the requirements of long range links and wide coverage. In this case cooperative transmission would probably be implemented as a piggy back short range cooperative links, in order to enhance the common throughput.

- **Private Mobile Radio (PMR):** In this case it is assumed that some terminals can have similar range to the cellular case, while others are only within a short range from each other, so as to operate in high stress safety/distress situations. In this scenario, the motivation behind cooperative transmission would be the improvement of the link reliability, whereas increasing the system throughput would be of a lesser concern. An example of a PMR system is the Terrestrial Trunked RAdio (TETRA) system.

- **Wireless Local Area Network (WLAN):** The inherent assumption for cooperative transmission in this scenario is that the terminals handle data and require internet access. In this case, both individual throughputs as well as overall system throughput are of interest.

- **Personal Area Network (PAN):** This scenario is a heterogeneous setup of varying application possibilities with varying data rate requirements. The main motivators for cooperation would be flexibility and low power consumption.

- **Sensor networks:** Similar to PANs, the main motivation for cooperative transmission is the fact that sensor nodes are power limited. Cooperation is expected to contribute to power savings, as well as range extension over the monitored area.

6.3.2 Cooperative Techniques in the IEEE 802.16 [i.333] Standard (a.k.a Wimax)

Despite the fact cooperative techniques (in general - including mesh networking) have only recently received considerable research attention, and many theoretical (e.g. capacity limits) and practical problems remain open, The IEEE 802 [i.269] working groups have already started working on incorporating cooperative techniques into current standards. See table 6.2.

Cooperative techniques appear at several levels of the network:

- Cooperative transmission among mobile stations (in centralized or non-centralized networks).

- Cooperation among networks (e.g. for traffic load balancing, handover, spectrum sharing).

- Cooperation among mobiles and networks in unlicensed operation.

- Cooperation between licensed and unlicensed spectrum users.

The specific cooperative transmission techniques that will be reviewed in this clause (i.e. relaying and Virtual Antenna Areas) are being considered for enhancing the 802.16e (Mobile WiMAX) standard [i.335].

A study group was formed in July 2005 for developing methods and investigation modifications in the PHY and MAC of IEEE 802.16e [i.335], in order to support multi-hop relay and cooperative techniques. The study group was named the Mobile Multi-hop Relay Study Group (MMR-SG), and defined its goals as coverage extension and throughput enhancement. Achieving these goals requires modification of the frame structure and the addition of new protocols for relay operation, while keeping the backward compatibility for the point-to-multipoint mode in IEEE 802.16e [i.335]. As the mesh type of operation is already incorporated in the IEEE 802.16-2004 [i.280], it was not be considered in this study group. A key requirement in upgrading the IEEE 802.16e [i.335] standard for supporting multi-hop/relay operation, is to efficiently provide a multi-hop or relay path to a mobile station or to a base station with a small number of hops. The operating scenarios under consideration in the mobile multi-hop relay study group are summarized in table 6.3. It is observed that the mobile client relay will not be considered by this Study Group due the complexity, battery life of the client relay, and security.

Table 6.2: IEEE 802 standardization activities that address cooperative techniques

IEEE group	Scope	Operation scenario	Type of cooperation
802.15, TG 3, 4, 5	High rate Wireless Personal Area Network (WPAN)	Mesh networking	Cooperative retransmission
802.11s	Local Area Network (LAN) MAC enhancement for reliable and easily scalable network	Mesh networking	Peer-to-peer cooperation
802.16, 2004	Metropolitan Area Network (MAN) MAC enhancement for reliable and easily scalable network	Mesh networking	Peer-to-peer cooperation
802.16, MMR-SG	Coverage extension, Through-put enhancement, Spectral efficiency improvement in MAN	Relay	Multihop relay, cooperative transmission
802.22	Wireless Regional Area Network (WRAN)	Fixed centralized point-to-multipoint for unlicensed operation in TV bands	Cognitive radios

Table 6.3: Topologies and operation scenarios considered in IEEE 802.16 [i.333] MMR-SG

Topology	Scenario	
	Infrastructure	Client
Mesh operation	No	No
Fixed	Yes	Yes
Nomadic	Yes	Yes
Mobile	Yes	No

The IEEE 802.16j [i.347] Project Authorization Request has been approved by the IEEE-SA Standards Board on April 2006. IEEE 802.16j [i.347] is aiming to enhance IEEE 802.16e [i.335] standards to gain coverage extension, throughput enhancement, by providing specifications for mobile multi-hop relay features, functions and interoperable relay stations. IEEE 802.16j [i.347] specifies OFDMA PHY and MAC layer enhancement to IEEE 802.16 [i.333] for licensed bands to enable the operation of relay stations. The relay station can be fixed, nomadic or mobile, which pretends to be a base station for MT and to be a MT for base station. The relay links are assumed to operate at 2,4 GHz and/or 5,8 GHz unlicensed bands. As it is shown further on in this clause, there are many cooperative transmission techniques proposed in the literature for coverage extension and throughput enhancement. The 802.16 mobile multi-hop relay study group is the first attempt to introduce them into any standard.

Research on cooperative transmission techniques is focusing on two main levels. Firstly a theoretical performance analysis into the performances and capacity limits of different cooperation scenarios (and associated co-operative coding approaches), and secondly on the development of practical co-operative techniques in practical system scenarios (e.g. assuming single relay and CDMA signals). The following clauses provide a review on the research progress in each of the above themes.

6.3.3 Conclusions on Radio Relays and Co-Operative Transmission Techniques

This clause has reviewed cooperative transmission through the use of relaying and virtual antenna arrays. In their simplest form, multi-hop relaying techniques facilitate coverage extensions to infrastructure and power-limited ad-hoc networks. More sophisticated cooperative coding techniques (AF, DF, CF, etc.) have been shown to provide higher link capacities, even in AWGN channel conditions, though their main benefit in practical systems is the fading diversity advantage. Similarly virtual antenna arrays allow applying distributed space-time coding across neighbouring terminals in order to exploit optimally the diversity of the virtual MIMO channel.

Relaying and cooperative transmission techniques are currently being studied within the IEEE 802.16j [i.347] project. The project's objective is to define enhancements in the physical and MAC layer of the IEEE 802.16 [i.333] (for licensed bands) to enable operation of relay stations.

According to the IEEE 802.16 [i.333] group:

"*the purpose of this amendment is to enhance coverage, throughput and system capacity of 802.16 networks by specifying 802.16 multi-hop relay capabilities and functionalities of interoperable relay stations and base stations. The multi-hop relay is a promising solution to expand coverage and to enhance throughput and system capacity for IEEE 802.16 [i.333] systems. It is expected that the complexity of relay stations will be considerably less than the complexity of legacy IEEE 802.16 [*i.333*] base stations. The gains in coverage and throughput can be leveraged to reduce total deployment cost for a given system performance requirement and thereby improve the economic viability of IEEE 802.16 [i.333] systems. Relay functionality enables rapid deployment and reduces the cost of system operation. These advantages will expand the market opportunity for broadband wireless access. This project aims to enable exploitation of such advantages by adding appropriate relay functionality to IEEE 802.16[i.333] through the proposed amendment. Stakeholders include manufacturers and operators of IEEE 802.16 [i.333] networks*".

The above statement, which is based on preliminary investigations by the IEEE 802.16 [i.333] relay group, leaves no doubts about the significance of relaying techniques in enhancing the coverage, capacity and cost-effectiveness of radio network.

Fixed relaying is already a defining element of the S-DMB architecture, as it allows penetrations of satellite signals into urban regions. Cooperative transmission techniques can also be applied between neighbouring mobile terminals (which form a terrestrial ad-hoc network) in order to improve the availability of the satellite link to the group. The large distance spreads between satellite terminals accessing commercial services is, however, likely to be an important factor which will limit the impact of the technique in practice. On the other hand non-commercial usage scenarios, especially those related to emergency services, are very likely to benefit in terms of throughputs and link-availability enhancements. The satellite can also serve in coordinating the operation (e.g. routing) and administration of the terrestrial ad-hoc network, in order to achieve performance optimizations, security, and signalling overheads reductions.

Some of the specific cooperation techniques that have been proposed in the literature have been reviewed in this clause. Obviously specific studies are required to assess the performance benefits and optimization of these techniques over satellite channels. The study results and selection of techniques made within the IEEE 802.16j [i.347] project will also set an important reference for developing solutions in satellite systems.

6.4 Mobile Ad-hoc Networks

NOTE: See also annex M.

6.4.1 Introduction

Very closely related to cooperative and relaying techniques is the concept of ad-hoc networking, since it can facilitate the required spontaneous network-level connections between nodes. However, ad-hoc networks have a more general purpose of interconnecting communications devices that are not necessarily coordinated by a central network entity, in order to support local communications, and local service provision. The concept, though not very new, it has recently received significant attention since it is expected to play a key role in 4G networks (see [i.210]). In particular ad-hoc networks are envisaged to be a component of the "multi-network" architectures envisaged in 4G, as it can provide cost-effective support of multi-hop communications (and thus provide coverage extensions to infrastructure networks). In these system configurations, the overlay (infrastructure) networks can assume organisation roles for supporting the ad-hoc network (mainly supporting routing functionalities).

To that end ad-hoc networking between mobile terminals can be an important element of future satellite architectures, where neighbouring mobile terminals establish an ad-hoc network not only in order to inter-communicate and offer services between themselves, but also in order to extend the satellite coverage/availability through cooperative transmission/reception techniques. On the other hand the satellite can help the ground network of maintaining low protocol complexities and low signalling overheads by providing routing support functionalities.

Ad-hoc networks have the distinguishing capability of being self-configuring, in the sense that there is no (central) management system with configuration responsibilities. Some, if not all, nodes are capable of assuming router functionality when needed. This enables terminals to communicate with each other when they are out of (radio) range, provided they can reach each other via intermediate hosts acting as routers that relay the packets from source to destination. The structure of the network can change constantly because of the movement of the nodes. Networks can significantly vary in size (from a few devices comprising a PAN, to hundreds of sensors comprising a wireless sensor network). All network functions and protocols are distributed and executed by all network participants.

It can be expected that in the near future there will be a proliferation of wireless devices (laptops, PDAs, camcorders, mobile phones, MP3 players, game stations, sensors, etc.), and with various characteristics like throughput, transmission power, energy resources, size or cost. However the common features of all ad-hoc network devices are limited energy resources and capability to communicate using one or more wireless technologies. Bluetooth, WLAN 802.11 and UWB are the most frequently considered technologies for use in various ad hoc network scenarios. The following are some of the possible ad-hoc networking scenarios [i.252] to [i.256]:

- **Personal use:** non-commercial transfer of data between devices or persons; communication in areas without adequate wireless coverage or short range peer-to-peer communications in an ad hoc group, in which it does not make (economic) sense to use an operator network (e.g. group of hikers wishing to communicate).
- **Commercial use:** setting up communication in exhibitions, conferences or sales presentations.
- **Sensor networks:** communication between, or with, intelligent sensors.
- **Search and rescue operations:** communication in areas without adequate wireless coverage, or when the existing communication infrastructure is non-operational due to a natural disaster or a global war.
- **Vehicle communication networks:** crash avoidance warning system, safety distance for cruise control for cars, trains, airports, etc.

Ad-hoc networks can be viewed as stand-alone groups of mobile terminals, but they may also be connected to a pre-existing network infrastructure and use it to access hosts which are not part of the ad-hoc network. The multi-hop communication capability can be used to extend the coverage of existing wireless access technologies. Another interesting aspect of ad-hoc networks is their self-configurability and neighbour discovery capability, which imply that these networks will be a key element for enhancing the interoperability among different wireless technologies.

The main features of ad-hoc networks can be summarized as follows:

- **Dynamic network topology:** due to the node mobility and radio propagation, network topology is constantly changing. This requires specific network protocol functions for topology construction and maintenance.
- **Distributed nature:** this is an inherent characteristic of ad-hoc networks. As it is not necessary to have a permanent central authority, all networking functions have to be distributed across participating nodes.
- **Multi-hop communications:** due to the limited range of wireless interfaces, usually it is not possible to establish direct communication links with all nodes. As there is no infrastructure to support establishment of multi-hop routes, the nodes themselves have to run routing algorithms to establish routes in the network, and to forward packets destined for other nodes.
- **Limited energy resources:** as ad-hoc network nodes will usually be battery driven, optimization of energy consumption across all protocol stack layers is extremely important.
- **Limited bandwidth:** wireless technologies that are envisaged to be suitable for ad-hoc networks provide throughputs of a few hundred kilobits per second to a few megabits per second, which is enough for many applications. However, the wireless environment is a harsh one and can cause significant error rates, which are aggregated along the multi-hop links.

Presently, there are indications that ad-hoc networking is finding its place, and has also good possibilities to be adopted for commercial purposes, perhaps not as an alternative, but as an extension to existing networks. There is on-going interest to apply ad-hoc networking principles towards a range of possibilities such as (community) mesh networking, range-extension of cellular and mesh networks, and small-scale special purpose ad hoc-networks such as Personal Area Networks for games and entertainment. This is in part reflecting the enhanced technological capabilities, but also the fact that real applications cases.

6.4.2 Ad-hoc Networking Capability in IEEE 802 Standards

The IEEE 802 [i.269] standards working groups are perusing the introduction of state of the art cooperative and ad-hoc networking techniques in Wireless PANs, LANs, and MANs (IEEE 802.15 [i.279], IEEE 802.11 [i.269], and IEEE 802.16 [i.333]).

6.4.2.1 IEEE 802.11s

An evolution of the IEEE 802.11 [i.269] standard using mesh networking, named 802.11s [i.341], is currently being investigated in the 802.11 working group. The objective is to upgrade the 802.11 MAC layer for supporting self configuring and multi-hop topologies. It may support broadcast, multicast and unicast traffic in the network. There are a few network element functionalities defined in the task group, such as mesh point, mesh access point, and mesh portal. The mesh point is the basic element, as it collects information about the neighboring mesh points, communicating with them and forwarding the traffic. The mesh access point is a mesh point that has the capability to function as the 802.11 access point. The mesh portal is a mesh point, which connects the mesh network and a non-802.11 network, as it is shown in figure 6.6.

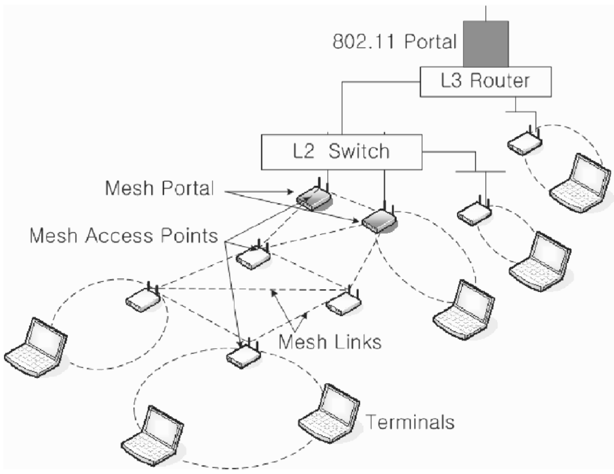


Figure 6.6: The IEEE 802.11s mesh architecture

The IEEE 802.11s task group received around 15 initial proposals but only two of those were considered for further study. These are the "SEE mesh" and the "Wi-Mesh Alliance" proposals. The SEE mesh proposal introduced the concept of mesh portal for interoperability in mesh networks and to accommodate other 802.11 WLAN (old or new) services in the network. The Wi-Mesh alliance claims to be equipment vendor independent and operable in indoor and outdoor situations.

The usage models for 802.11s [i.341] are categorized into four main items depending on the deployment, propagation characteristics and required service. The basic residential model contains a small number of nodes and its main characteristic is to provide a low-cost, high performance and easily deployable mesh network to remove the radio frequency dead-spots. Other usage models include the office, campus/public access network and public safety networks. The office and campus/public access models contain a relatively large number of nodes and a wider coverage area. The public safety model is to form a relatively smaller easily deployable network during emergency situations.

6.4.2.2 IEEE 802.15 [i.279]

The IEEE 802.15 [i.279] Standards defines the physical and MAC layers for short-range communications Wireless PAN using the ultra wideband (UWB). Data rates from 250 kbps (802.15.4) to 55 Mbps (802.15.3), with communication distances from 1 metre to 75 metres, are intended. The IEEE 802.15.5 [i.343] standard is the mesh extension to 802.15 [i.340].

In comparison with the mesh operation in 802.11s [i.341], the 802.15 [i.340] differs in the way terminals act as nodes in the mesh network. In 802.11, which is an infrastructure mesh, only Access Points are nodes of the mesh network, whereas in 802.15 [i.340], which is a client mesh, user terminals are the nodes of the mesh network. As a result, the mesh control layer also addresses network performance and control, in addition to coverage and range extension. This feature requires collective behaviours to be implemented. Thus cooperation is required at the network level. In particular in large mesh networks, local routing decisions result in sub-optimal global routing, leading to unacceptable QoS performance. In order to guarantee QoS to critical applications, local network information is shared globally.

The challenges of propagating network information to every node lie in the overhead required for transmitting that information, and the delay between the time the information is sent by a node and received by all other nodes, which could render the information obsolete due to the time-varying nature of the Mesh WPAN. Nodes in the mesh network should therefore cooperate to propagate control and data streams of other nodes, hopefully resulting in a benefit for every single node in the network. Moreover, cooperative re-transmission mechanisms using nodes as relays, built on ARQ protocols and cooperative coding, also offer further enhancement to the physical and MAC layers.

An important characteristic of WPAN is the low transmission powers due to energy-limited battery-operated devices. Another distinguishing feature of Wireless PAN Networks is proactive power management. It is well known that relaying, multi-hop and cooperative transmission techniques can help save energy. MAC protocols can also be designed to allow nodes to participate in cooperative routing for power savings, and to go into energy-saving modes as often as possible.

Cooperation is also often required for the coexistence or sharing of resources by collocated networks. In addition to contention-based access to the channel for delay-insensitive applications (with a Carrier Sense Multiple Access (CSMA) approach for collision avoidance), delay sensitive applications rely on beacons to ensure isochronous transmissions in IEEE 802.15.3. In the scenario of simultaneous operating mobile piconets, collisions of such beacons would prevent the successful transmission of delay-sensitive data. Cooperation between the piconets is thus necessary to avoid this undesirable situation. The beacon mode of operation specifies a super-frame structure with a sub-frame for the transmission of beacons, and a PAN coordinator to address coexistence. However, the beacon mode of operation is currently not allowed in the mesh mode. To conclude on PAN, due to the short communication ranges, mesh architecture is natural, but it requires advanced cooperative techniques in order to be scalable and reliable. The power-limited nature of the devices is also addressed by cooperative transmission and routing techniques.

6.4.2.3 IEEE 802.16 [i.333]

The IEEE 802.16, [i.280] is an OFDM, OFDMA and single carrier based fixed wireless LAN/MAN standard in 10 GHz to 66 GHz bands. It improved and consolidated the previous standards such as 802.16, 2001 [i.333], 802.16a [i.344], 2003, and 802.16c, 2002 [i.345]. The MAC layer supports the point-to-multipoint and optional mesh network topology [i.280]. The optional mesh mode operation was initially defined in the 802.16a-2003 standard with basic signalling, message formats, etc. Subsequently, the mesh mode specifications were integrated and improved in the IEEE 802.16 [i.280], 2004 revised Standard.

Unlike the point-to-multipoint mode, there are no clearly separate downlink and uplink sub-frames in the Mesh mode. Each terminal communicates with a number of neighbouring stations instead of communicating with a base station. There are a few terminals, which function as gateway to the backhaul network and provide some of the base station functions.

In the IEEE 802.16-2004 standard [i.280], centralized scheduling, distributed scheduling, and a combination of both scheduling schemes are used. If centralized scheduling is employed, the mesh base station nodes functions are similar to the base station in the point-to-multipoint mode. The mesh base station provides the control and scheduling decisions. When distributed scheduling is employed, all terminals, including the mesh base station, transmit their data after coordinating with the two-hop neighbourhood and broadcast their scheduling information, such as available resources, requests and grants [i.280] It is assumed that no interference occurs between nodes that are two hops away. Thus, the mesh with two-hop neighbourhood suffers from the hidden terminal problem.

The inter node interference is one of the major factors affecting the network capacity and the scalability in mesh networks. If the inter node interference is taken into account in the radio resource allocation, better spectral efficiency may be obtained. In centralized scheduling, resources are allocated in a more centralized manner. The mesh base station gathers requests for resources in uplink and downlink from the terminals within a range of a few hops. It makes the decision and transmits the scheduling message which is not the actual schedule to the terminals. The terminals use a predetermined method to calculate the actual scheduling information depending on the system parameters [i.280]. The mesh network with centralized scheduling has limited scalability. It can only support around 100 subscribers due to the structure of centralized scheduling messages.

6.4.3 Conclusions on Mobile Ad-hoc Networks

Ad-hoc networking is another key technology that is expected to enable 4G architectures achieving their performance and network/service flexibility objectives. Key benefits are establishment of infrastructure-less local communications and local service provision and support of multi-hop and cooperative transmission techniques for enhancing link performances and overall network capacities. IEEE wireless LAN, MAN, and PAN standards have or presently are being enhanced for supporting mesh topologies and ad-hoc networking. Bluetooth and UWB are also expected to be popular platforms for implementing ad-hoc PANs.

Within the framework of 4G cooperative networks, ad-hoc networks will be operating in cooperation with overlaying infrastructure networks. This cooperation is important for the ad-hoc networks in order to access non-local services and also to receive administrative (e.g. billing), security and network management support.

Although terrestrial infrastructure systems are considered to the main "partners" to ad-hoc networks, clearly this role can also be played by satellite systems particularly for specific types of services where terrestrial infrastructure networks are not available. At this point in time this type of cooperation between satellites and ad-hoc networks has been explored primarily for military applications. However additional usage scenarios can be commercially relevant, such as in disaster relief and sensor networks. In any case, in the context of 4G networks, future SatCom standards should follow terrestrial ones in providing ad-hoc networking support capabilities, as this will strengthen the position of satellites in becoming an important component of 4G systems.

7 Candidate System Architecture for Beyond 3G or 4G Satellite Component

In the present clause, several candidate system architectures are identified which can implement the advanced techniques described in clauses 5 and 6. They can be used as basis for the definition of medium or long term systems.

7.1 Overview

Beyond 3G systems are mobile service systems that include the new capabilities of IMT that go beyond those of IMT-2000. Such systems provide access to a wide range of telecommunication services including advanced mobile service, supported by mobile and fixed networks, which are increasing packet-based. Beyond 3G systems support low to high mobility applications and a wide range of data in accordance with user and service demands in multiple user environments. These systems also have capabilities for high quality multimedia applications in a wide range of services and platforms anywhere, providing a significant improvement in performance and quality of service.

In order to provide the seamless service over a global coverage, the satellite component of beyond 3G or 4G systems will always be considered because the terrestrial component only will not be possible to be deployed all over the world.

The following service scenarios can be considered:

- Two ways communication using multi-spot coverage with frequency re-use.
- Broadcasting using linguistic beams with national coverage.
- Bidirectional data services over mobile broadcasting systems.

The following clause describes a few examples of candidate system architectures to illustrate some of these options in more detail. It highlights their main characteristics.

7.2 Examples of candidate system architecture

7.2.1 Application examples

This example considers a satellite component that has a high degree of commonality with the terrestrial mobile radio interface for the scenario of "two ways communications using multi-spot coverage with frequency re-use".

The two ways communication scenario can be regarded as coverage extension and service continuity of the terrestrial part. One of the major key satellite advantages is its possibility to provide wide geographic area coverage. In particular, it can offer services in regions without terrestrial coverage. The areas not adequately covered by the terrestrial part include physical isolated regions; other gaps of terrestrial network and "emergency areas" in the event that the terrestrial system collapses due to disaster. In the scenario, handover technique with terrestrial part would be most importantly considered. For the cost-effective handover, future satellite radio interfaces should be compatible and have a high degree of common functionality with an envisaged terrestrial radio system like LTE (or LTE-Advanced) and WiMAX (or WiMAX evolution). It would also be possible to reuse terrestrial part technology to minimize user terminal chipset and network equipment for low cost and fast development.

In addition, the considered satellite radio interface can provide efficient interactive multimedia broadcasting services as well since the envisaged terrestrial mobile radio interfaces can handle services for broadcast as well as a bi-directional communications in a cellular system. Indeed, the satellite component has an advantage over terrestrial component for delivery of same content to spread over a wide geographic area.

7.2.2 Possible services

The development of mobile telecommunication networks and the need for higher data rate allow the development of new user services and concurrently create the need to stay connected everywhere in the best possible condition. The services required for the satellite component of such a system would be similar to terrestrial services assuming that new applications and new users services may appear, these future main telecommunication services may include:

- Data services with elastic traffic and variable data rate for numerical object transfer.
- Streaming services with guaranteed (though possibly variable) data rate.
- VoIP services.

Table 7.1 provides the list of a wide range of telecommunication service offering that will increase as new applications are developed based on the advancement of technology.

Table 7.1: Possible services and service examples

User experience class	Service class	Example services
Interactive	Interactive high delay	E-Education (e.g. data search); Consultation (e.g. data search); Internet browsing; Mobile commerce; Location-based services; ITS-enabled services.
	Interactive low delay	Emergency calling; e-mail (IMAP server access); Remote collaboration (e.g. desktop sharing);, Public alerting (e.g. with feedback); Messaging (instant messaging); Mobile broadcasting/multicasting (mobile interactive personalized TV); Interactive gaming.
Streaming	Streaming live	Emergency calling; Public alerting, e-Education (e.g. remote lecture); Consultation (e.g. remote monitoring); Machine-to-machine (e.g. observation); Mobile broadcasting/multicasting; Multimedia.
	Streaming non-live	Mobile broadcasting/multicasting; e-Education (e.g. education movies); Multimedia; Mobile commerce; Remote collaboration.
Background	Background	Messaging; Video messaging; Public alerting; e-mail (transfer RX/TX, e.g. POP); Machine-to-machine; File transfer/download; e-Education (file download/upload); Consultation (file download/upload); Internet browsing; Location-based service.

To make provision for some of these services that would be provided in broadcast and/or multicast mode, specific satellite resource may be dedicated to broadcast and/or multicast. Their role may be similar to the E-MBMS cells in the LTE cellular system.

The system network layer has to support these different types of data traffic with different requirements. Moreover, the satellite system constraints will have impact on the latency, jitter and average and peak throughput and corresponding QoS services. The parameters described in clause 5.1 for the medium term architecture may differ for the long term architecture. And support of VoIP at the same level of QoS as on the terrestrial networks may be a rough constraint on the satellite component of the system but not noticeably different than for other satellite air interfaces.

7.2.3 System requirements

7.2.3.1 Integrated satellite and terrestrial system case

Satellite allows global coverage of a large geographical area and is particularly appropriate in rural areas where the deployment of high speed terrestrial mobile telecommunication networks is likely to occur later than urban areas (if at all). In these areas, the terrestrial network could be replaced by the satellite component of a hybrid integrated network.

An important requirement for the integration with the terrestrial cellular systems is to consider a multi beam coverage which will enable integration in a cellular mobile system. Users will be connected to the network through the terrestrial cell or through the satellite beam according to its geographical position or the quality of the transmission link and propagation conditions:

- Different options are available:
  - The use of different frequency between the terrestrial and satellite cells with frequency re-use.
  - Single Frequency Network (SFN) reception between satellite and terrestrial cell on the geographical coverage of the beam.

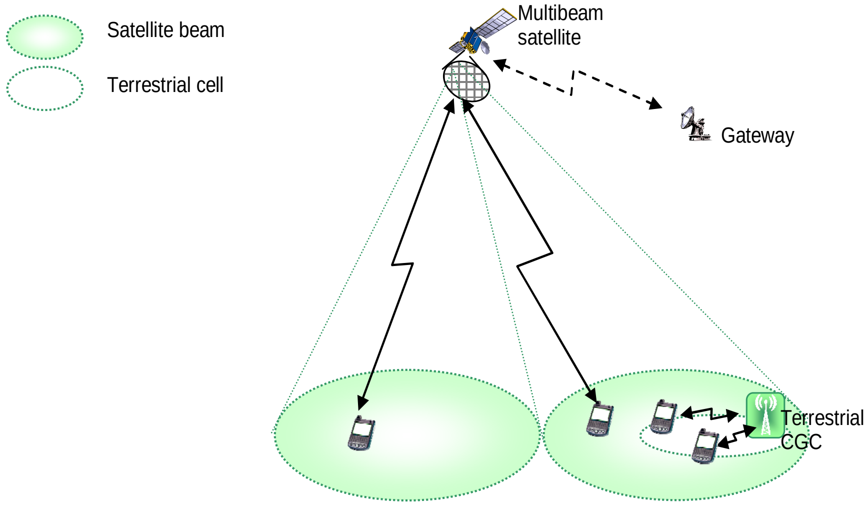


Figure 7.1: Integrated terrestrial and satellite system architecture

The system uses terrestrial complementary ground components (CGCs) which interact directly to the terrestrial core network. The CGC is also called by CGC type 1.

7.2.3.2 Hybrid satellite and terrestrial system case

A typical large-scale hybrid system comprises multi-spot beam satellites sharing resource between services and broadcast services. For broadcast services support, this satellite component can be associated with a nationwide or regional ensemble of CGC wherein both satellite and CGC segment communicate with user equipment using a common set of MSS frequencies to handle broadcasting scenario. For bidirectional services support, the terrestrial component using a separated set of frequencies provides ubiquitous coverage of broadband two ways communications to end users with generalized mobility requirements in complement to the bidirectional services provided by the satellite component in rural zone.

The satellite component of beyond 3G system should have sufficient power and receiver sensitivity to establish communications with user terminals that are similar to terrestrial terminals. Large satellite antennas, providing high gain reconfigurable spot beams are among the key attributes of hybrid systems. These features enable increased spectrum reuse as well as communications via typical low-cost handheld terminals. User terminals select either the terrestrial or satellite network based on the received signal level and network availability to keep certain service quality over a wide and continuous service area. It would be preferable that the hybrid system should have roaming capability with unique user/subscriber identifier across both terrestrial and satellite system. Vertical handover between satellite and terrestrial component in a hybrid system should be carried out within the extent that execution of handover does not significant decrease the system capacity or increase system complexity.

The concept for this hybrid system is shown in figure 7.2. The difference between figures 7.2(a) and 7.2(b) is that some processing is made in the satellite in figure 7.2(b) i.e. figure 7.2(b) scenario includes an onboard processing satellite.

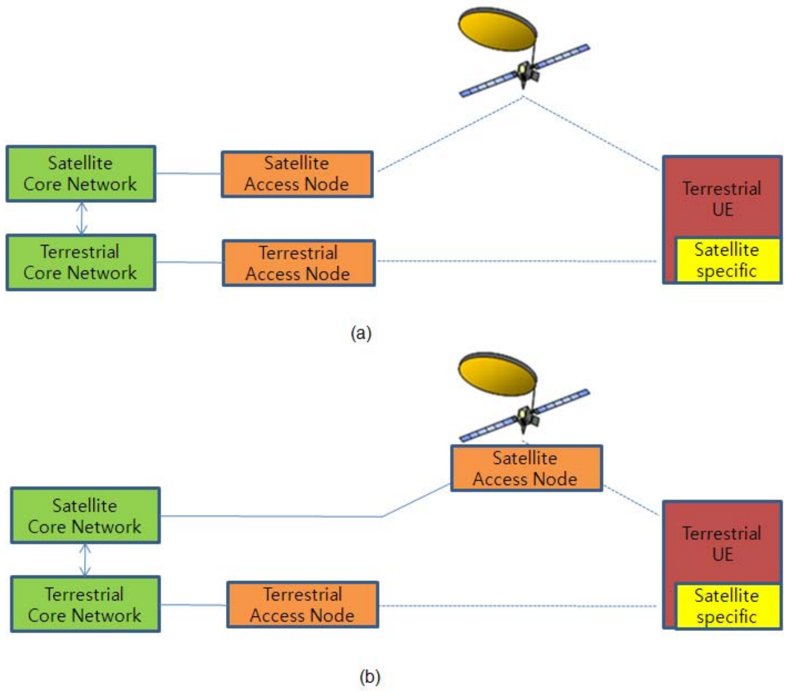


Figure 7.2: Hybrid terrestrial and satellite system architecture

Figure 7.2 shows a separate core network for the satellite and terrestrial component, but this is primarily a logical separation and they could be merged into a common core network. In both cases, they are expected to use the same core network technologies.

7.2.4 Specific features

7.2.4.1 Integrated satellite and terrestrial system case

This clause describes the different options possible for the integrated system features. The complete system architecture may be a combination of all the features or the selection of the most relevant options.

- Combining architecture: in figure 7.3, the terminals which receive terrestrial data link and satellite data link simultaneously can make the combining of the signals from data source and terrestrial source in case of MFN reception.

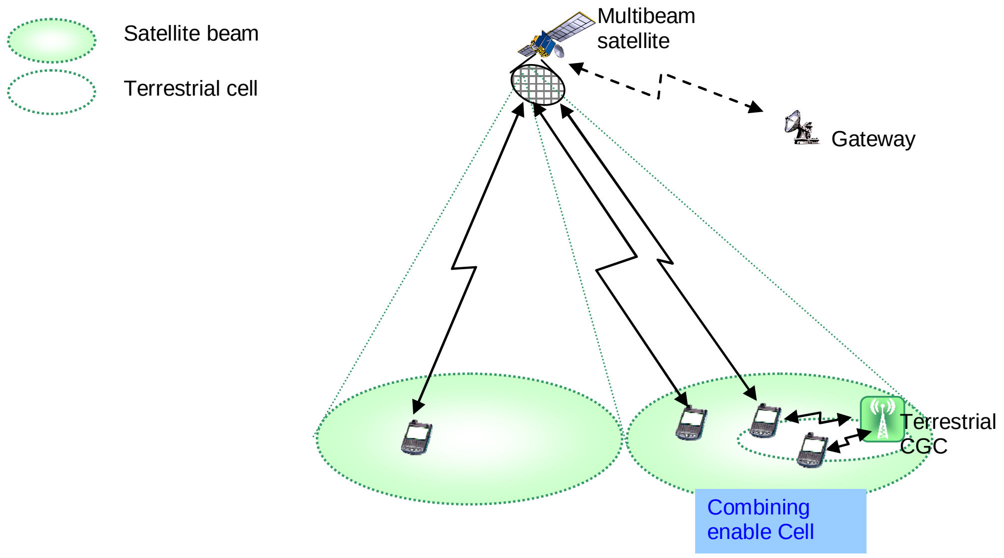


Figure 7.3: Terrestrial and Satellite combining architecture

- SFN architecture can be considered in the case of use the same interface air between the terrestrial and the satellite components of the system and this will enable the use of the same frequency between terrestrial and satellite component of the system. The terrestrial component of the system is called by terrestrial CGC or CGC (type1).

- CGC (repeater): the addition of CGCs used as repeaters can improve the coverage. This type of CGC (repeater) could also be use for the indoor coverage. The system uses CGC which received its signal directly from the satellite. The CGC is called by CGC type 2.

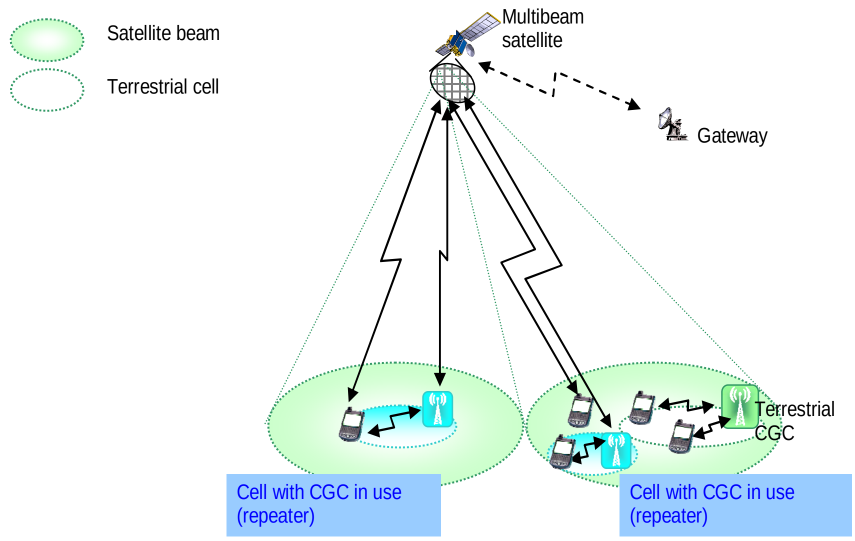


Figure 7.4: System architecture with the use of complementary ground component (type2) used as a repeater

- Trunking: figure 7.5 is showing system architecture with trunking capability added for the delivery of data in isolated zone with no terrestrial network. The reception in this cell is not hybrid as only the satellite is received. In this case, the data is delivered to a CGC type 2. CGC can be only a repeater or transcoding CGC. The direct link between the terminal and the satellite is not precluded but it will add some constraints on the system.

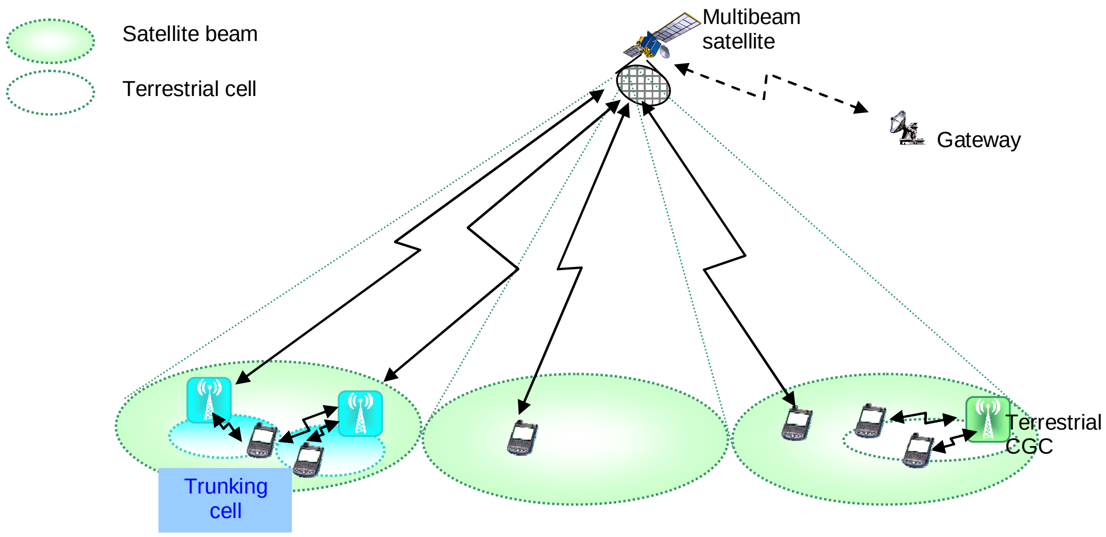


Figure 7.5: System architecture with trunking capability

- Ad-hoc networking: another solution in case of isolated users with no terrestrial coverage, another option is the use of Ad-hoc network: cooperative communication between the users as it is shown in figure 7.6.

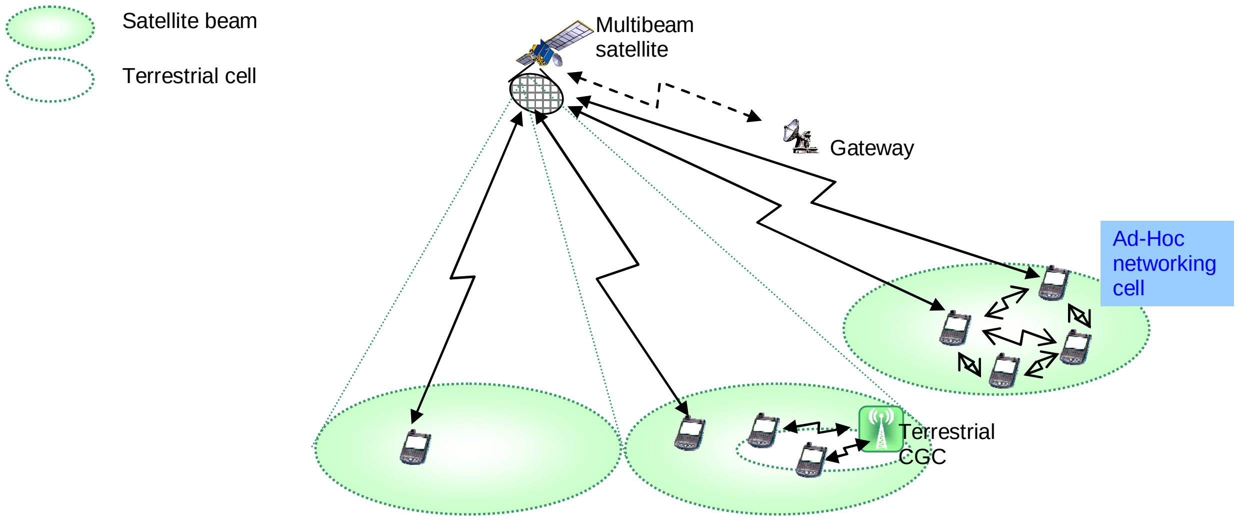


Figure 7.6: System architecture with Ad-hoc networking capability

These co-operative transmission techniques are described in details in clause 6.4 and in annex M.

7.2.4.2 Hybrid satellite and terrestrial system case

Figure 7.7 describes overall system features for the considered hybrid system concept. The following factor can be considered.

- Satellite: It will provide services and applications similar to those of terrestrial systems outside terrestrial and CGC coverage under the inherent constraints imposed by power limitation and long round trip delay.

- CGC: In order to provide mobile broadcasting/multicasting services, they can be deployed in areas where satellite reception is difficult, especially in urban areas. They may be collocated with terrestrial cell sites or standalone. Several kinds of them can be considered such as simple amplifying and forwarding CGC like simple repeaters, a demodulation and forwarding CGC for high modulation and a decoding and forwarding CGC for better traffic quality. These CGCs are type 2 CGC.

- Terrestrial component: Satellite component can provide voice and data communication service in regions outside terrestrial coverage. The areas not adequately covered by terrestrial component include physically isolated regions, gap of terrestrial component and areas where terrestrial component permanently, or temporarily, collapses due to disaster.

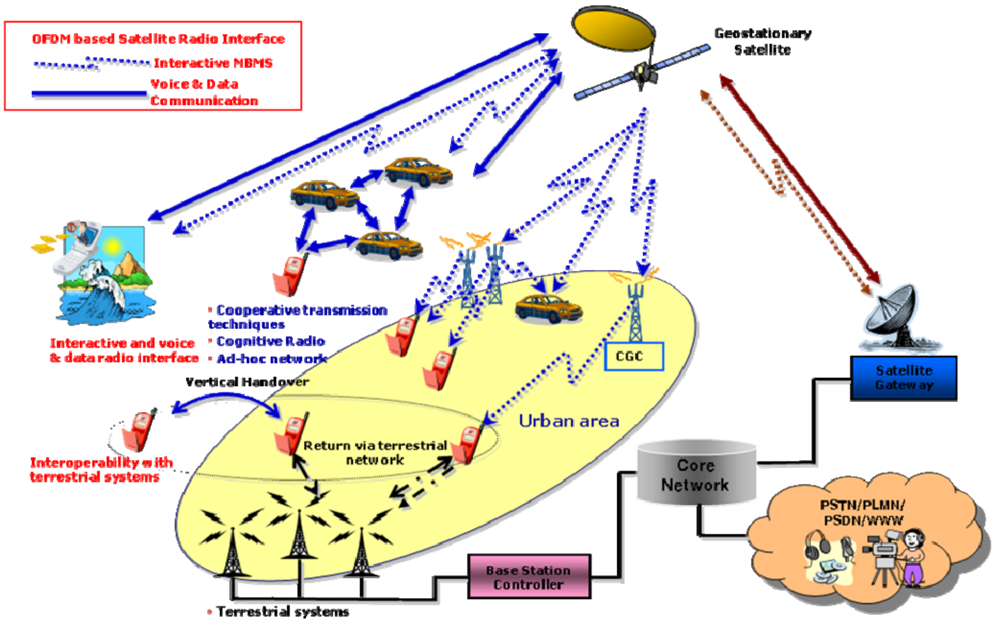


Figure 7.7: Hybrid satellite and terrestrial system architecture

In figure 7.7, different modes of communication modes are described:

- Reception of broadcast information from the satellite.
- Reception of broadcast information from CGC type 2:
  - possible combining with the signal transmitted by the satellite.
- Bi-directional transmission between the user terminal and the satellite:
  - possible communications with other terrestrial terminal using cooperative network.
- Bi-directional transmission between the user terminal and the terrestrial network:
  - Use of terrestrial network to establish a return link for broadcast services.
  - Enable interoperability and roaming with other terrestrial systems for bi-directional links.

Each of this mode, the selection of the same or different air interface for each communication link may have great impact on the terminal architecture and complexity.

7.2.5 Possible technical issues

7.2.5.1 Integrated satellite and terrestrial system case

7.2.5.1.1 Mobility between terrestrial and satellite coverage

The terrestrial and satellite component of the integrated system case may have different coverage. An important technical issue is to enable the user terminal mobility between the terrestrial and the satellite component. The two components may use the same network management system and may share the same core network. Consequently specific procedures for the handover can be defined. However, the difference of transmission conditions and the difference between the propagation delays for the two components will impact the system design. In particular for a seamless handover, some innovative technologies are required to switch between the satellite and the terrestrial context.

7.2.5.1.2 Mobility between terrestrial system and MSS integrated satellite and terrestrial case

Another important technical issue will be to enable the user terminal mobility between a terrestrial system and the integrated terrestrial and satellite case. Indeed, according to the description of the system provided in clause 7.1.3.1, the CGC type 1 interacts directly with the terrestrial core network. Some mechanisms and technologies have to be defined to enable a seamless handover between these two systems and further studies are required on the subject of vertical handover between a MSS integrated system and a terrestrial system.

7.2.5.1.3 Terminal architectures

The definition of the satellite radio interface for the system is an important topic and it will affect the different mode of communication for the terminal (Satellite, Terrestrial or hybrid communication mode, etc.).

Various options can be considered for the satellite radio interface as noted in clause 5.

A satellite-specific radio interface is likely to offer improved performance and increased satellite traffic (i.e. traffic capacity for a given satellite power) relative to an interface based on an adapted terrestrial radio interface. However, this approach will require a dual-mode terminal and this may result in higher terminal development costs and production costs.

By contrast, a satellite radio interface based on an adaptation of a terrestrial radio interface could help to reduce both the development cost and the production costs of the terminal, particularly if the satellite mode and the terrestrial mode use a similar radio interface.

In both cases, the design of multimode terminals will benefit from the development of innovative technologies as the Software Definable Radio.

7.2.5.1.4 Application of long term techniques

The application of advanced technologies can also be used to improve the functionalities of the system (cooperative communications), to improve the performance and the spectrum efficiency (use of combining or MIMO techniques, use of method for interference cancellation, optimization of the spectrum sharing, etc.).

7.2.5.2 Hybrid satellite and terrestrial system case

In hybrid satellite and terrestrial system, two components can be operated independently of each other and can have separate network management systems. Therefore, in order to provide seamless services, vertical handover between satellite and terrestrial components should be considered as on of the most important techniques. For the cost-effective vertical handover, future satellite radio interfaces of IMT-Advanced may be compatible and have a high degree of commonality with a terrestrial interface of IMT-Advanced. It may bring possibility to reuse terrestrial part technology to minimize user terminal chipset and network equipment for low cost and fast development. Of course, it can also have several different features from terrestrial radio interface to reflect the satellite-specific characteristics such as long round trip delay and power limitation.

In addition, some interesting "beyond 3G enabling techniques" can be considered in enhancing the cost-effectiveness and competiveness of the satellite component as follows.

- Horizontal integration of services and networks on personal mobile devices (SDR).
- Optimized communication techniques (satellite MIMO, MUD, Turbo detection, HARQ, ACM, Pre-equalization, IPv6).
- Introduction of new concepts and techniques for increased coverage, data speeds and spectral efficiencies, such as ad-hoc networking, cooperative multi-point transmission (CoMP) and relaying, cognitive radio techniques for dynamic spectral sharing, improved inter-beam interference management.

7.2.6 Example of proposed beyond 3G system

An example of a beyond 3G satellite services envisaged for the Republic of Korea is presented in annex N.

8 Spectrum Requirements

The ITU has undertaken studies into the estimated spectrum requirements for MSS and presented the results in REPORT ITU-R Recommendation M.2077 [i.329].

This ITU Report provides traffic forecasts including MultiMedia (MM) distribution services, and estimates of spectrum requirements for the satellite component of IMT-2000 and systems beyond IMT-2000 for the period 2010 to 2020. It builds upon previous material, such as ITU-R Report ITU-R Recommendation M.2023 [i.336], as well as more recent subscriber forecasts and traffic models. Application of the latest spectrum calculation methodology in compliance with ITU-R Recommendation M.1391 [i.337] is employed to determine the spectrum requirements.

Table 8.1: Current MSS spectrum allocations in the 1 GHz to 5 GHz band

Uplink band (MHz) (see note)	Downlink band (MHz) (see note)	Bandwidth (MHz)
1 626,5 to 1 645,5, 1 646,5 to 1 660,5	1 525 to1 544, 1 545 to 1 559	2 × 33
1 610 to 1 626,5	2 483,5 to 2 500	2 × 16,5
1 668 to 1 675	1 518 to 1 525	2 × 7
1 980 to 2 010	2 170 to 2 200	2 × 30
2 670 to 2 690	2 500 to 2 520	2 × 20
2 655 to 2 670	2 520 to 2 535	2 × 15
Total allocated spectrum		**2 × 121,5**
NOTE: Some of these frequency bands are not necessarily available to Mobile Satellite Services in all countries.		

ITU-R Assumptions and conclusion

According to the ITU-R [i.329] the traffic forecasts and anticipated spectrum requirements for the satellite component of IMT-2000 and beyond IMT-2000 are presented for the period 2010 to 2020 for a pessimistic and an optimistic scenario are shown in the following table. The main reason for the imbalance between Earth-to-space and space-to-Earth directions are distribution applications and asymmetric multimedia services showing higher spectrum requirements on space-to-Earth links.

The ITU-R goes on to state that the anticipated low traffic scenario is based on 9 % MSS subscriber growth per year starting with the known number of MSS subscribers of 1,4 million at the beginning of 2006. Incentives for new investments resulting in continuing introduction of high multi-beam systems with more than 200 beams will therefore be low, thus not leading to a significant reduction of percentage of traffic in hot spots. Further assumptions are asymmetric MM traffic loss in hot spots around 50 %, no growth of asymmetric MM services and 50 % of asymmetric return MM traffic relative to the forward direction. Distribution with 17 data channels to optimize use of the assumed 30 MHz available bandwidth by 2010, and 26 channels beyond 2015.

The anticipated high traffic scenario is based on a 14 % MSS subscriber growth per year. Strong incentives for investments in new MSS technologies will result in introduction of high multi-beam systems with larger reflectors and around 600 beams which will be introduced gradually until 2020. Further assumptions are asymmetric MM traffic loss in hot spots around 25 %, annual growth rate of 5 % for asymmetric MM services and 71 % for asymmetric return MM traffic. Distribution with 26 data channels up to 2015 and 35 data channels thereafter.

Distribution applications should consider maximizing the number of channels made available to the subscriber as a function of totally available bandwidth, channel bandwidth, channel quality, number of spot beams and satellite systems.

Unlike non-multimedia applications, the traffic from multimedia satellite service applications will continue to grow rapidly.

The dominant contribution to the spectrum requirements is from the multimedia distribution and asymmetric services.

Support of a specific grade of mobility at a given bandwidth is only possible up to a certain operating frequency due to fast fading phenomena, and therefore, for high mobility, suitable operating frequencies would be up to 6 GHz.

Table 8.2: Required spectrum for the low and high traffic scenarios

Spectrum required (MHz)	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Low traffic scenario											
Spectrum in Earth-to space direction	55	61	62	68	71	74	83	89	93	98	105
Spectrum space-to-Earth excluding distribution	70	76	79	85	92	98	106	116	122	130	140
Total spectrum without distribution	125	137	141	153	163	172	189	205	215	228	245
Spectrum for MM distribution services	30	30	30	30	30	30	60	60	60	90	90
Grand total spectrum including distribution	155	167	171	183	193	202	249	265	275	318	335
High traffic scenario											
Spectrum in Earth-to-space direction	74	83	89	98	108	123	134	146	157	165	176
Spectrum space-to-Earth excluding distribution	89	98	107	119	131	149	163	178	193	206	223
Total spectrum without distribution	163	181	196	217	239	272	297	324	350	371	399
Spectrum for MM distribution services	30	30	30	30	30	60	90	90	90	120	120
Grand total spectrum including distribution	193	211	226	247	269	332	387	414	440	491	519

To investigate the sensitivity of the spectrum requirements, various parameters have been modified relative to the nominal scenario. This analysis revealed that the annual growth rate of MSS subscribers and the percentage of global traffic in the worst-case cluster are driving factors for the required spectrum. To some extent, there is a compensating factor between the increasing number of subscribers and percentage of traffic in hot spots as a strongly increasing MSS market will be a driving incentive to develop new higher spot-beam MSS satellites. Of significant influence is also the annual growth rate for asymmetric MM services. Less significant are variations to traffic loss in hot spots due to sharing with terrestrial services. The return to forward ratio of asymmetric MM services has a rather minor influence.

The following table shows the required spectrum taking into account existing allocations. Considering that out of $2 \times 121,5$ MHz of currently allocated MSS spectrum in the range 1 GHz to 5 GHz, only around 2×86 MHz are globally available, additional global MSS allocations are needed commencing with around 14 MHz by the year 2010 and increasing up to 114 MHz by the year 2020 for a low traffic scenario, including 30 MHz and 90 MHz for distribution applications by 2010 and 2020, respectively. A high traffic scenario would require around 33 MHz by 2010, increasing to 257 MHz by 2020. These estimates include 30 MHz and 120 MHz for multimedia distribution by 2010 and 2020, respectively.

Table 8.3: Required new global MSS spectrum allocations in the 1 GHz to 6 GHz range

Required spectrum (MHz)	Low traffic scenario		High traffic scenario	
	2010	2020	2010	2020
Spectrum in Earth-to-space direction	55	105	74	176
Spectrum in space-to-Earth direction excluding distribution	70	140	89	223
Spectrum for multimedia distribution in space-to-Earth direction	30	90	30	120
Total required spectrum	155	335	193	519
Required new allocations in Earth-to-space direction		19		90
Required new allocations in space-to-Earth direction excluding distribution		54	3	137
Required new allocations in space-to-Earth direction including distribution	14	144	33	257

Conclusions on spectrum requirements

Based upon the extensive work conducted in the ITU-R there is a clear requirement for additional spectrum to meet the emerging MSS traffic forecasts.

In the 1 GHZ to 6 GHz range, for the year 2020, the shortfall in spectrum requirements is between 19 MHz and 90 MHz in the uplink direction and between 144 MHz and 257 MHz in the downlink direction, including the distribution applications.

There is an emerging interest in exploring spectrum above 6 GHz (both Ku and Ka-band) for MSS type services, in particular to vehicle mounted terminals. Spectrum requirements in these bands have not been addressed in the ITU-R Report and represents additional as yet un-quantified requirements.

9 Conclusions

Two possible future SatCom system architectures have been identified and analysed from a technological perspective in this technical report:

1) A "medium-term" architecture based upon the evolution of existing and emerging radio interfaces for mobile satellite services (e.g. GMR, B-GAN, S-UMTS Family SL, DVB-RCS, Satmode, DVB-SH and ETSI SDR). An alternative would be adaptation of emerging terrestrial mobile radio interfaces for satellite services. In this regard, this document focuses on the WiMAX and LTE standards. Consideration of such an adapted radio interface has been incorporated in the "Medium-Term Architecture". This architecture may include a complementary broadcast component.

2) A "long-term" 4G multi-network architecture where the potential roles of satellites in beyond 3G converged networks and in the context of related-future communication technologies on a bottom-up basis.

9.1 Medium-Term Architecture

Regarding the medium-term architecture, the analysis in the present document has focused on the satellite aspects of the architecture. The following specific aspects were analysed and related inferences were drawn:

- Identification of key interactive services to be supported and the related QoS requirements.
- Applicable propagation radio-channel models, for analyzing the system's link performances and optimizing performance-enhancing communication techniques.
- Signal processing and diversity transmission techniques.
- Upper-Layer Error Control Techniques.
- Three main paths for developing a medium term architecture were identified:
  - An evolution of mobile satellite technology. For example, a satellite optimized evolved standard which will adopt some of the latest communication techniques reviewed in the present document, and also adapt techniques specified within the next generation mobile terrestrial standards.
  - An evolution of fixed satellite or broadcast satellite technology. For example, evolution of a DVB-based return-link standard, optimized for operation below 3GHz, which will be complementing the DVB-SH standard. Such approach would require a significant standardisation development effort, and would also be a more competitive approach towards new generation mobile terrestrial standards.
  - An adaptation of an emerging terrestrial interface such as LTE or WiMAX.

9.2 Long-Term Architecture

The analysis of the long-term architecture was started by reviewing the prevailing visions of 4G as these are identified in industrial telecommunications fora. The "vertical" approach (which is supported mainly in Asia) puts the increase of data rates as the main target to be achieved by future systems. On the other hand the "horizontal" approach (supported mainly in Europe) foresees the convergence of heterogeneous networks, which will lead to higher efficiencies and new services (that will be accessible on integrated multi-network terminals).

As far as the analysis of the long-term architecture is concerned, the horizontal vision of beyond 3G, which is also foreseen within ITU recommendation documents, is the one that has been thought to be the most likely one to be realized. In this context, clear objective of SatCom systems are to identify and claim possible roles within the union of heterogeneous networks that will provide beyond 3G services.

Clause 6 of the present document has been devoted in analyzing some of the latest communication technologies that are widely considered as key enablers for achieving the very ambitious performance and spectral efficiency targets of beyond 3G sub-systems. The following main conclusions have been drawn regarding the relevance of these technologies to SatCom systems:

- Spectrum Sharing Models and Related Technologies.
- Cooperative Transmission and Relay Techniques.
- Mobile Ad-hoc Networks.

No specific service and performance requirements are outlined for the Long-Term Architecture as this was considered inappropriate at this time.

10 Recommendations

As a result of the analysis conducted for the present document the following future work is recommended:

- Focus on Medium term architectures including the development of a first set of requirements.
- Define more explicitly performance requirements taking into account 3GPP TS 22.105 [i.330].
- Define more explicitly anticipated services and applications taking into account 3GPP TS 22.105 [i.330].
- Examine appropriate radio interfaces for the different architectures.
- Address the need for future research activity that will enable suitable channel modelling above 3 GHz, including satellite MIMO.
- Investigate emerging network technology elements such as the NGN and IMS.

Annex A: Detailed Review of Land Mobile Satellite Channel Models

A.1 Empirical Models

Empirical models are based on experimental data and provide accurate characterization of the particular environment type where measurements have been carried out. However they are difficult to generalize to other environment types.

The Modified Exponential Decay (MED) model has been built upon a large database of measurements with different vegetation types and humidity conditions, and allows calculating the mean path loss due to vegetation as a function of the "vegetation path length" D_v and frequency f :

$$L_v = a_v(f)D_v \tag{A-1}$$

where the form of $a_v(f)$ is defined in the ranges $0 \le D_v \le 400m$ $200 \le f \le 9500MHz$ by ITU [i.9]:

$$a_v(f) = 0.2 f^{0.3} D_v^{-0.4} \tag{A-2}$$

Thus for example at 2 450 MHz and for $D_v = 1$ the mean path loss attenuation is 2 dB.

The Empirical Roadside Shadowing (ERS) model [i.5] to [i.8] permits the prediction of the roadside tree attenuation as a function of elevation angle (θ), frequency (f) and percentage of time (P). This model is recommended by ITU for rural environments [i.9]. The original model is valid for $f = 1.5GHz$, $1\% \le P \le 20\%$, $20^o \le \theta \le 60^o$, and formulates the road-side attenuation L as:

$$L(P,\theta) = -A\ln(P) + B \text{, where} \tag{A-3}$$

$$A = 3{,}44 + 0{,}0975\theta - 0{,}002\theta^2$$

$$B = -0{,}443\theta + 34{,}76$$

The model has been extended in the percentage range $20\% < P \le 80\%$ by applying the following scaling on (A-3) [i.6].

$$L(P,\theta) = \frac{L(20\%,\theta)}{\ln(4)}\ln(80/P) \tag{A-4}$$

Also the frequency range can be translated in the $0{,}8GHz \le f \le 20GHz$ range as [i.5]

$$L(f_2) = L(f_1)e^{1.5\left(\frac{1}{\sqrt{f1}} - \frac{1}{\sqrt{f2}}\right)} \tag{A-5}$$

Figure A.1 provides the excess attenuation distributions predicted by the ERS model for different elevation angles at 2,45 GHz. It is observed that the attenuation caused by road-side obstructions is quite severe for significant percentages of time.

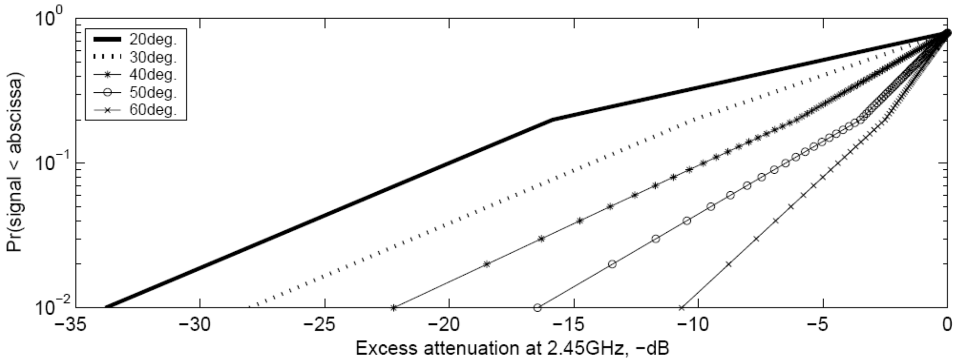


Figure A.1: Excess attenuation distributions predicted by the ERS model for different elevation angles at 2,45 GHz

Despite its accuracy, the ERS model does not allow to generate a time-series of the modelled fading process, which would be useful in simulating the link-level performance of the system. Also lacking a higher order spatio-temporal statistical characterization of the process does not allow designing anti-fading mechanisms, such as space/time diversity, time interleaving, hybrid ARQ and higher layer coding.

A.2 Statistical Models

Statistical models are based on the use of known parameterized statistical distributions and probabilistic state transition models (Markov). Different distributions, state transition models and related parameters are used to model different types of environments. Statistical models are used extensively for simulating the link-level performance of the system, mainly because it is easy to synthesize a (baseband) discrete time-series.

Statistical LMS models can be classified with respect to the number of "high-level" states they use in order to characterize different types of propagation conditions (within different types of environments and terrains). Two and three state models are well established but higher-order state models are being proposed, mainly in order capture in higher precision different types of shadowing effects. Within each state, the fading process is assumed to be stationary-ergodic, and its first and second order statistics are typically modelled through specific parameterized distributions and analytically derived (based on specific system assumptions) auto-correlation functions, respectively.

A.2.1 Single-State Narrowband (stationary) Models

Rice Model

In LOS conditions the Rice distribution is used for modelling the fading amplitude statistics. As a reminder, a Rician process is obtained by adding a complex LOS component $m(t) = m_1(t) + jm_2(t) = \rho_m \exp(j(2\pi f_\rho + \theta_m))$, where f_ρ is the Doppler frequency, to a zero mean complex Gaussian noise process $k(t) = k_1(t) + jk_2(t)$ with variance of σ_0^2 per dimension, and taking the absolute of the sum. The pdf of the received signal amplitude is given by:

$$p_X(x) = \frac{x}{\sigma_0^2} \exp(-\frac{x^2 + \rho^2}{2\sigma_0^2}) I_0(\frac{x\rho}{\sigma_0^2}) \qquad \text{(A-6)}$$

where $I_0(.)$ denotes the zeroth order modified Bessel function of the first kind. The K-factor is defined as: $\rho_m^2 / 2\sigma_0^2$, i.e. the ratio of coherent (LOS) to incoherent (multipath) power, and is dependent on the morphology of the scattering environment, the frequency and the antenna patterns. For a low-gain terminal operating in L/S bands in rural types of areas, the K-factor takes values in the range of 7 dB to 12 dB.

Rayleigh Model

As the K-factor is reduced to zero, the Rice distribution reduces to the Rayleigh distribution (obtained by setting $\rho = 0$ in (A-6)) which is typically assumed for modelling the short-term (by short term it is meant within a time interval which is significantly shorter than the coherence time of the large-scale fading process) statistics of the non-LOS narrowband channel factor.

Shadowing Model

For modelling the slower varying large-scale fading statistics, a log-normal distribution is typically assumed. This can be justified if the signal attenuation contributions along the shadowing path act independently, since in this case the total attenuation is the multiplication of signal attenuation power ratios. Expressed in dB, this is the sum of attenuation contributions in dB. If these contributions are taken as random variables, then the central limit theorem holds and hence the shadowing follows a log-normal (Gaussian) distribution [i.10].

In order to model the combined statistics of small and large scale fading "mixture" models have been proposed (that adopt the log-normal modelling of the mean signal attenuation), as these are described below.

Rice-Lognormal model

This model, developed by Corazza and Vatalaro in, is based on the assumption that both the direct and the multipath components are affected by shadowing. The pdf of a log-normal random variable S is given us:

$$p_S(s) = \int_0^\infty \frac{1}{s\sqrt{2\pi}h\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln s - \mu}{h\sigma}\right)^2\right] \quad \text{(A-7)}$$

where $h = \ln(10)/20$ and μ and $(h\sigma)^2$ are the mean and variance of the associate normal distributed variable. In [i.12] it is shown that the pdf of the received signal is given us:

$$p_R(r) = \int_0^\infty \frac{1}{S} p_X(\frac{r}{S}) p_S(s) dS \quad \text{(A-8)}$$

where $p_X(\frac{r}{S}) = \frac{r}{\sigma_R^2 S^2} \exp\left[-\frac{1}{2}\left(\frac{r^2}{\sigma_R^2 S^2} + 2K\right)\right] I_0\left(\frac{r\sqrt{2K}}{S\sigma_R}\right)$, $\sigma_R^2 = 1/2(K+1)$ and K is the K-factor.

Suzuki model

If $K = 0$ in (A-8), then the Rician distribution reduces to Rayleigh and the overall density given by (A-8) reduces to Suzuki. The Suzuki pdf applies when it is assumed that the direct LOS component is completely blocked by obstacles. In order to use this type of statistical modelling requires the ability to extract the parameters μ , σ and K based on measurement data for different types of environments.

Loo Model

The Loo process [i.12] is the sum of a log-normal random phasor and a Rayleigh phasor:

$$r = \rho e^{j\phi_0} + w e^{j\phi} \quad \text{(A-9)}$$

where ρ and w are log-normally and Rayleigh distributed, respectively. The phase random variables are uniformly distributed over the $[0, 2\pi]$ range. The conditional pdf $p(r/\rho)$ is Rician as given by (A-6). Combining this conditional density with the log-normal density, the density of the Loo envelope is given as:

$$p_R(r) = \frac{r}{\sigma_R^2 S\sqrt{2\pi}} \int_0^\infty \frac{1}{\rho} \exp\left[-\frac{(\ln\rho - \mu)^2}{2S} - \frac{r^2 + \rho^2}{2\sigma_R^2}\right] I_0\left(\frac{r\rho}{\sigma_R^2}\right) \quad \text{(A-10)}$$

Hwang Model

The Hwang model [i.13] also considers lognormal process affecting both the LOS and the multipath component but with total decorrelation between them. This model has been shown to include the Rice, Loo and Corazza models as special cases.

A.2.2 Second order statistics of single-state models

Additionally to deriving the distribution and first order statistics, accurate characterization and modelling of the fading process, also requires determination of its second order statistics, which provide a measure of the rate at which the signal level changes with time. They are important in the choice of bit-rate, frame length and the design of interleavers, channel estimators, channel coding (to name some examples). Common methods to describe the rate of change of the signal level are:

- The auto-correlation function (with respect to temporal delay) or Doppler spectrum (which is the frequency domain equivalent).
- Levels crossings per second and fade durations in seconds, as illustrated in figure A.2.

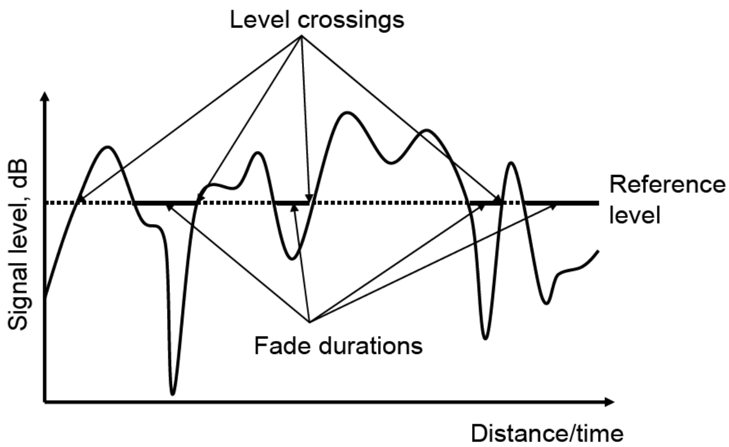


Figure A.2: Level crossing and fade duration

Work in [i.19] used measured data to describe the CDF of fade and non-fade durations, measured in metres. Whilst in fade conditions, with a threshold ranging from 2 dB to 8 dB, the best-fit curve was log-normal. The probability of a fade duration lasting more than x metres is given by:

$$P(d_F > x) = \frac{1}{2}\left[1 - erf\left(\frac{\ln x - \ln a}{\sigma_F \sqrt{2}}\right)\right] \quad \text{(A-11)}$$

where a and σ_F were computed from the measurement data. In non-fade conditions, the CDF curve followed a power law:

$$P(d_F > x) = bc^{-c} \quad \text{(A-12)}$$

where b and c are derived from the measurement data.

A study [i.20], based on measurements with elevations ranging from 60^o to 80^o, found the shadow fading to be log-normally distributed, and suggested an auto-correlation formulation:

$$C_s(\tau) = \sigma_S^2 e^{-v|\tau| / X_C} \quad \text{(A-13)}$$

where σ_S is the standard deviation of the log-normal shadow fading, v is the velocity and x_C is the effective correlation distance, defined as the distance when the correlation falls to e^{-1}, and is found to be 9,17 m in wooded areas and 16,2 m in suburban areas.

An alternative autocorrelation of the log-normal process is derived in [i.21], based on the assumption that the Doppler spectrum of the associated Gaussian process is itself a Gaussian function.

The second order statistics of the small-scale fading component depend mainly on the distribution of the angle of arrival of the multipath components, relative to the direction of motion, and also on the pattern of the receiving antenna. For terrestrial mobile systems, equations for the Doppler spectrum have been derived under various antenna and propagation assumptions [i.14]. The directional antenna results, presented in [i.14], were extended in [i.15] for an arbitrary antenna direction relative to the direction of motion, although some corrections to these results were made in [i.16]. A common assumption made in [i.14], [i.15] and [i.16] is that multipath signal power arrives to the mobile unit only along the azimuth angular dimension.

A 3D extension of the results given in [i.14] and [i.17] was made in [i.18], where specific assumptions about the distribution of the received power along the elevation dimension were made.

A.2.3 Multi-State Narrowband Models

The above single-state models are based on stationary stochastic processes with fixed parameters. In order to model very large areas with diverse types of environments multi-state non-stationary Markov models can be used [i.31]. Popular multi-state models include the Lutz two-state model [i.23] to [i.25] and the Fontan three-state model [i.27] to [i.30]. In [i.14] it is concluded, based on experimental results that for most channels a four-state mode is sufficient.

Lutz two-state Markov model

Lutz et al [i.23] to [i.26] conducted a Europe wide measurement campaign from the MAREC B2 satellite at an elevation of 13^0 to 43^0. They derived a two state model from their data. When in a "good" state, a state with no shadowing, the envelope of the fading was modelled as Ricean with a different Rice K factor for each environment and vehicle antenna. When in a shadowed "bad" state, the envelope was modelled as a Rayleigh process with log-normal mean distribution. They proposed the following model for simulation, as shown in figure A.3.

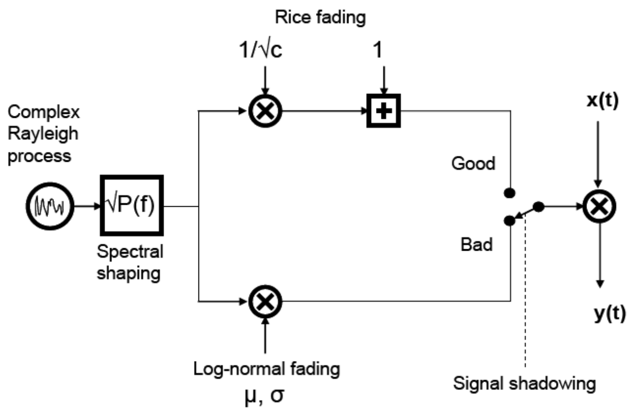


Figure A.3: Lutz two-state simulation model

The time shares of the two states are defined by the state transition probability matrix which is derived empirically from measurement data. The parameters c (Rice-factor), μ and σ (mean and standard deviation of the log-normal process) are derived from the measured data and presented for various elevations, environments and vehicle antenna.

Fontan three-state Markov model

The Fontan model provides a refinement of the Lutz model by discriminating between "moderate" and "deep" shadowing. For the derivation of the model, measured data were low pass filtered into "very slow" fading, which was used for calculating the Markov state and transition matrices, W and P. Log-normal fading was used to model the "slow" fading within each state, and the small scale fading in each state was dependent on the large scale fading level using the Loo approach.

Many measurement campaign datasets were used to derive the parameters for their three state model. 3 × 1 state and 3 × 3 transition Markov matrices were derived for each measurement campaign, and therefore given as a function of elevation and environment. The Loo parameters were also calculated separately for each state. Attempts were also made to derive a wideband model in addition to the narrowband model, but were not fully validated due to lack of wideband measurement data.

The Fontan model has been adopted by the TM-SSP group for evaluating the performance of DVB-SH in different types of environments. Tables A.1 and A.2 provide the model parameters for different types of environments: open, suburban, intermediate tree-shadowing and heavy tree-shadowing.

Table A.1: Fontan model parameters per state based on measurements parameters for 40° elevation as presented in [i.30]

Environment	State 1: LOS			State 2: Shadowing			State 3: Deep shadow		
	α (dB)	ψ (dB)	MP (dB)	α (dB)	ψ (dB)	MP (dB)	α (dB)	ψ (dB)	MP (dB)
Open (*)	0.1	0.37	-22.0	-1.0	0.5	-22.0	-2.25	0.13	-21.2
Suburban	-1.0	0.5	-13.0	-3.7	0.98	-12.2	-15.0	5.9	-13.0
Intermediate Tree-Shadow	-0.4	1.5	-13.2	-8.2	3.9	-12.7	-17.0	3.14	-10.0
Heavy Tree-Shadow (2)	-	-	-	-10.1	2.25	-10.0	-19.0	4.0	-10.0

Table A.2: Fontan model parameters per state based on measurements parameters for 40° elevation as presented in [i.30]

Environment	[P]			[W]	d_{corr} (m)	L_{frame} (m)	L_{trans} (m)
Open (2)	0.9530	0.0431	0.0039	0.5	2.5	8.9	12.4
	0.0515	0.9347	0.0138	0.375		7.5	
	0.0334	0.0238	0.9428	0.125		4.0 (1)	
Suburban	0.8177	0.1715	0.0108	0.4545	1.7	5.2	2.2
	0.1544	0.7997	0.0459	0.4545		3.7	
	0.1400	0.1433	0.7167	0.091		3.0 (1)	
Intermediate Tree-Shadow	0.7193	0.1865	0.0942	0.3929	1.5	6.3	2.6
	0.1848	0.7269	0.0883	0.3571		6.3	
	0.1771	0.0971	0.7258	0.25		4.5	
Heavy Tree-Shadow (2)	0.7792	0.0452	0.1756	0	1.7	-	3.5
	0	0.9259	0.0741	0.5		4.8	
	0	0.0741	0.9259	0.5		4.5	

Note 1 : These values have been extrapolated since they are not given in.[Fontan2]
Note 2 : Not simulated, for information only
The parameters values are :

- α Average value of the attenuation on the LOS link for a state
- ψ Standard deviation of the attenuation on the LOS link for a state
- [P] Probability of occurrence of a transition (3x3 matrix)
- [W] Total probability of having a given state
- dcorr Correlation distance of the channel
- L_{Frame} Minimum state frame length as defined in [Fontan2]
- L_{Trans} Transition region length as defined in [Fontan]

A.2.4 Wideband Models

Most research on LMS channel modelling has created narrowband models as the transmitted bandwidths are usually small compared with the channel coherence bandwidths. At higher elevations associated more often with LMS systems, the coherence bandwidths are usually wider than low elevation terrestrial systems. However, in systems with high-rate channelization (e.g. WCDMA based S-UMTS), wideband channel modelling becomes increasingly important, to model intersymbol interference for example. Mobile-satellite wideband channel modelling usually takes the form of a time-variant impulse response, which is often modelled as a delay-line with each tap defined by different weights and distributions. Empirical wideband modelling, based on extensive measurement campaigns, has been presented in [i.32] to [i.34].

Table A.3 to table A.5 provide the tap-delay line wideband model parameters that have been adopted for evaluating the performance of WCDMA based S-UMTS [i.2]. It is noted that the actual number of channel taps will depend on the system's data-rate. For example assuming a rate of 5 Ms/s the symbol-chip duration will be 200 ns and thus only the sub-urban and urban models would be wideband; the rural model would collapse to a single tap model.

Table A.3: Rural wideband satellite channel model parameters used for the evaluation of WCDMA based S-UMTS

Tap number	Relative tap delay value (ns)	Tap amplitude distribution	Parameter of amplitude distribution (dB)	Average amplitude with respect to free space propagation	Rice factor (dB)	Doppler spectrum
1	0	LOS: Rice NLOS: Rayleigh	10 log c 10 log P_m	0,0 -7,3	10 -	Rice Classic
2	100	Rayleigh	10 log P_m	-23,6	-	Classic
3	180	Rayleigh	10 log P_m	-28,1	-	Classic

Table A.4: Sub-urban wideband satellite channel model parameters used for the evaluation of WCDMA based S-UMTS

Tap number	Relative tap delay value (ns)	Tap amplitude distribution	Parameter of amplitude distribution (dB)	Average amplitude with respect to free space propagation	Rice factor (dB)	Doppler spectrum
1	0	LOS: Rice NLOS: Rayleigh	10 log c 10 log P_m	0,0 -9,5	7 -	Rice Classic
2	100	Rayleigh	10 log P_m	-24,1	-	Classic
3	250	Rayleigh	10 log P_m	-25,1	-	Classic

Table A.5: Urban wideband satellite channel model parameters used for the evaluation of S-UMTS

Tap number	Relative tap delay value (ns)	Tap amplitude distribution	Parameter of amplitude distribution (dB)	Average amplitude with respect to free space propagation	Rice factor (dB)	Doppler spectrum
1	0	LOS: Rice NLOS: Rayleigh	10 log c 10 log P_m	0,0 -12,1	3 -	Rice Classic
2	60	Rayleigh	10 log P_m	-17,0	-	Classic
3	100	Rayleigh	10 log P_m	-18,3	-	Classic
4	130	Rayleigh	10 log P_m	-19,1	-	Classic
5	250	Rayleigh	10 log P_m	-22,1	-	Classic

A.2.4.1 Hybrid Satellite-Terrestrial Channel Models

In the hybrid satellite/terrestrial scenario foreseen in the 4G study, the services delivered via the satellite are characterized by a reduced bandwidth with respect to what foreseen in the 3GPP LTE context; this characteristic makes the hybrid propagation channel similar to what considered in the DVB-SH framework [i.293]. Accordingly, the definition of the tapped delay line model for the hybrid SFN network under investigation can be in general done in two different ways:

- For a given site for which a 3D representation is available, deterministic tools can be used based on ray tracing or ray launching as done in the framework of the MAESTRO project. This approach yields to a site-specific model.

- For a generic urban or suburban environment, another possibility is to define a macro-cellular geometric configuration for a given cell radius (e.g. 2 km has been selected in the framework of the SATIN project), and then, for a given position of the user terminal, the SFN channel PDP is obtained as a combination of the elementary PDP coming from each terrestrial repeater. A single elementary PDP is selected and repeated for each repeater. The DVB-SH guidelines propose for example to adopt the ITU Vehicular a PDP which is used in the context of UMTS performance assessment, or the GSM-TU6 which has been selected for DVB-H assessment.

In this analysis, the wideband channel models proposed in the framework of the MAESTRO project are selected [i.43], [i.44] and [i.45], which have been generated through ray tracing deterministic tools and geographical 3D databases of the cities of Munich and Milan for the urban cases and Tuningen (Germany) for the rural case. These models are derived as the combination of a satellite tapped delay line plus two terrestrial tapped delay line models, the first one of which corresponds to the predominant terrestrial repeater, while the second represents all other residual repeaters that significantly contribute to the received signal. The models support a dynamic range of 25 dB as the difference between the strongest and the weakest path. The tap power is expressed in absolute terms (i.e. in dBm) to ease possible models rescaling (satellite and repeaters EIRP are indeed attached to the model itself).

These wideband models are representative of a wide class of user environments, considering both worst and typical operation conditions, considering a GEO satellite and carrier frequency 2 197,5 MHz, see table A.6.

Table A.6: Propagation scenarios for wideband channel models

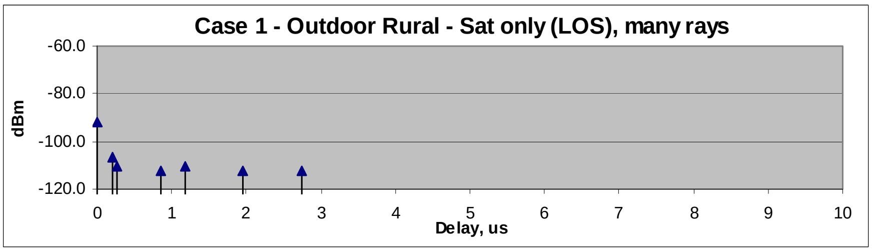
Case Number	Environment	Scenario description
Case 1	Outdoor rural	Satellite LoS with many rays
Case 2	Outdoor urban	Satellite LoS with few rays
Case 3	Outdoor urban	Satellite NLoS with many rays
Case 4	Outdoor urban	Satellite + 3 Repeaters (without processing delay) - street canyon
Case 5	Outdoor urban	Satellite + 3 Repeaters (without processing delay) - open area
Case 6	Outdoor urban	Satellite + 3 Repeaters (with processing delay) - large delay
Case 7	Indoor urban	Satellite NLoS only
Case 8	Indoor urban	Satellite + 3 Repeaters (without processing delay)
Case 9	Outdoor urban	Satellite + 3 Repeaters (without processing delay) - very large delays

The corresponding PDPs are listed in the following tables.

Case 1 and case 2 scenarios are reported in tables A.7 and A.8, respectively. These cases are representative of outdoor scenarios with satellite only reception in LoS without the presence of IMRs. Accordingly, the PDPs present a predominant path and a scattered component characterized by an maximum dispersion in the order of 2,7 µs in case 1 (which is the worst case) and 0,13 µs in case 2.

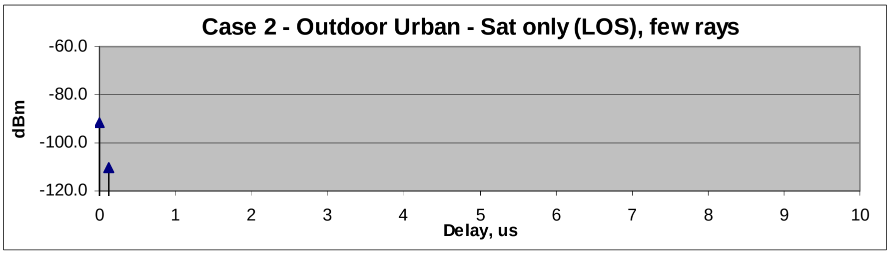
The corresponding delay spread results to be T_S(case 1) = 334 ns and T_S(case 2) = 16 ns. These values of the maximum delay are compatible with the frequency non-selectivity requirement over the OFDM intercarrier spacing for both LTE and WiMAX.

Table A.7: Wideband propagation channel, Case 1



Case 1: Satellite LOS with many rays		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-91,9	10
195,3	-106,3	-inf
260,4	-110,1	-inf
846,3	-112,5	-inf
1 171,9	-110,2	-inf
1 953,1	-112,5	-inf
2 734,3	-112,5	-inf

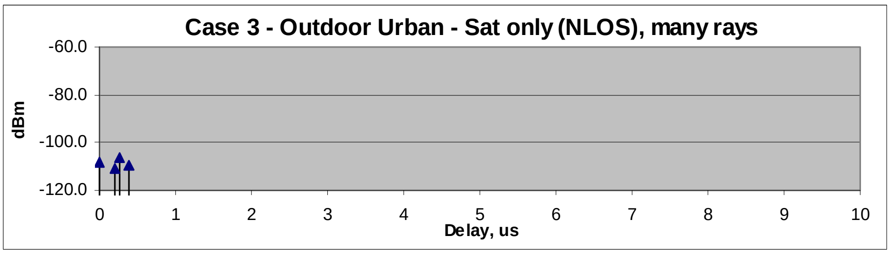
Table A.8: Wideband propagation channel, Case 2



Case 2: Satellite LOS with few rays		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-91,8	7
130,2	-110,1	-inf

Case 3 is representative of outdoor reception without repeaters, in Non-LoS conditions. The corresponding PDP, shown in table A.9, presents a direct component Rayleigh distributed as the scattered paths, and with power comparable to the following taps. The corresponding delay spread results to be T_s(case 3) = 138 ns. This value is compatible with the frequency non-selectivity requirement over the OFDM intercarrier spacing for both LTE and WiMAX.

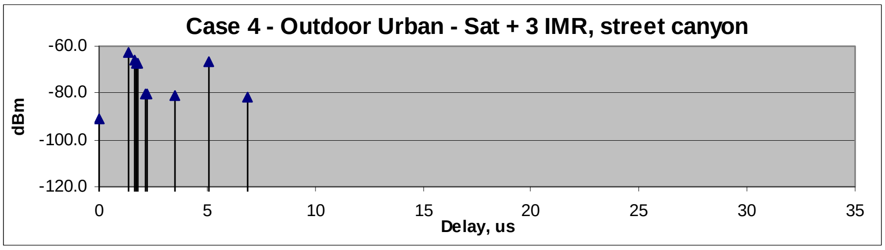
Table A.9: Wideband propagation channel, Case 3



Case 3: Satellite NLOS with many rays		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-108,5	-inf
195,3	-110,9	-inf
260,4	-106,6	-inf
390,6	-109,3	-inf

Case 4 considers a hybrid scenario with the satellite in LoS plus 3 terrestrial repeaters without processing delay, i.e. which are able to perform frequency conversion only. The first repeater (with 4 associated taps) is considered to illuminate the street canyon, yielding a wave guide effects that dominates the reception. Overall a total delay in the order of 7µs is achieved, which can be considered typical for a network of terrestrial repeaters in cities. The corresponding delay spread results to be T_s (case 4) = 1,305 µs. This value of the delay spread meets the frequency non-selectivity requirement over Δf for LTE, while is on the boundary for WiMAX with large Δf schemes.

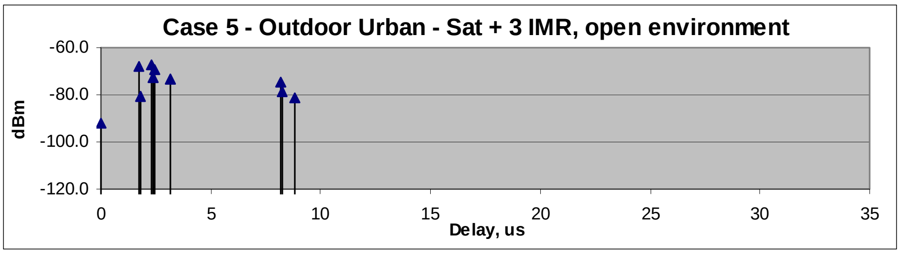
Table A.10: Wideband propagation channel, Case 4



Case 4: Satellite+3 Repeaters (without processing delay) - street canyon		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-90,9	7
1 367,2	-62,3	-inf
1 627,6	-65,7	-inf
1 692,7	-66,9	-inf
1 822,9	-67,0	-inf
2 148,4	-80,6	-inf
2 213,5	-80,4	-inf
3 515,6	-81,1	-inf
5 078,0	-66,5	-inf
6 835,8	-81,5	-inf

Case 5 PDP is reported in table A.11. It considers again the reception from the satellite plus 3 transparent repeaters, but an open area is addressed, so that a number of repeaters can be equally well received. The contribution from all 3 repeaters is in this case comparable in power, with relative delay of 2,3 µs, 3,2 µs, and 8,8 µs. The corresponding delay spread results to be T_s(case 5) = 1,774 µs. Again, this value meets the frequency non-selectivity requirement over Δf for LTE, while is on the boundary for WiMAX with large Δf schemes.

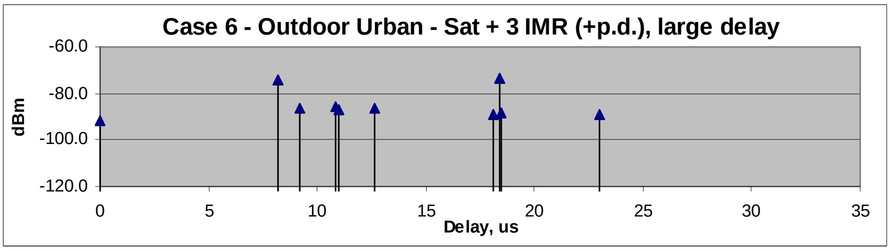
Table A.11: Wideband propagation channel, Case 5



Case 5: Satellite+3 Repeaters (without processing delay) - open area		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-91,8	7
1 692,7	-67,8	-inf
1 757,8	-80,7	-inf
2 278,6	-67,5	-inf
2 343,7	-72,8	-inf
2 408,8	-69,6	-inf
3 190,0	-73,1	-inf
8 203,0	-74,8	-inf
8 268,1	-78,4	-inf
8 788,9	-81,6	-inf

Case 6 scenario considers reception from satellite in LoS plus 3 repeaters with processing delay (equal to 8 µs which is typical for commercial on-channel repeaters). The corresponding PDP is reported in table A.12. The corresponding delay spread results to be T_s(case 6) = 5,098 µs. This value meets the frequency non-selectivity requirement over Δf for LTE, while is on the boundary for WiMAX with large Δf schemes. Further, the delay spread results to be larger than the guard period for LTE with normal cyclic prefix and for a series of WiMAX configurations.

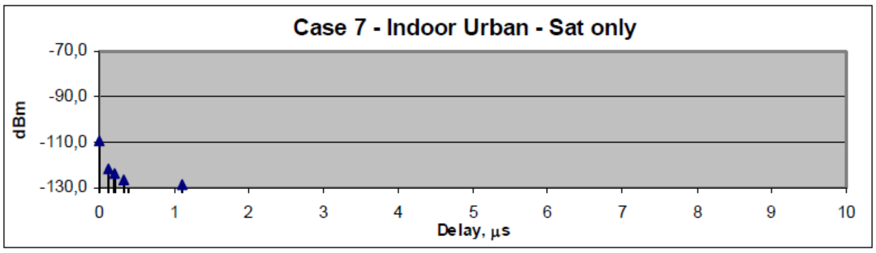
Table A.12: Wideband propagation channel, Case 6



Case 6: Satellite+3 Repeaters (with processing delay 8µs) - large delay		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-91,7	7
8203,0	-74,4	-inf
9179,5	-86,3	-inf
10872,2	-85,4	-inf
11002,4	-86,8	-inf
12630,0	-86,4	-inf
18098,6	-89,2	-inf
18424,1	-73,6	-inf
18498,2	-88,6	-inf
22981,3	-89,3	-inf

Case 7 scenario considers reception Indoor Urban - NLoS satellite only.

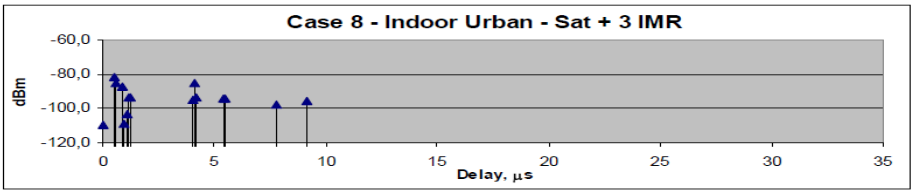
Table A.13: Wideband propagation channel, Case 7



Case 7: Indoor Urban - NLoS Satellite only		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-109,5	-inf
130,2	-122,0	-inf
195,3	-124,1	-inf
325,5	-126,6	-inf
390,6	-130,8	-inf
1106,8	-128,6	-inf

Case-8 scenario considers reception for Indoor Urban -Satellite plus 3 repeaters without processing delay.

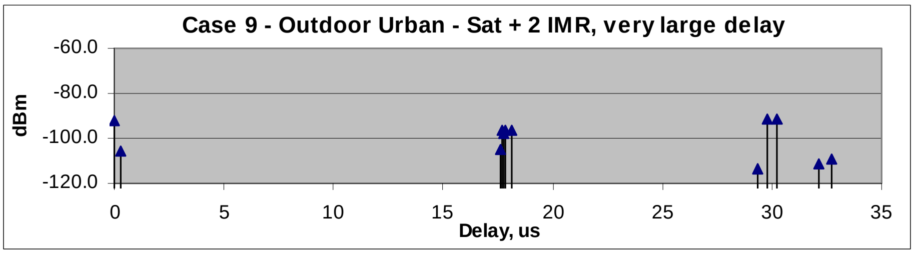
Table A.14: Wideband propagation channel, Case 8



Case 8: Indoor Urban -Satellite + 3 IMRs (without processing delay)		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-109,5	-inf
520,8	-81,5	-inf
585,9	-85,0	-inf
846,3	-87,0	-inf
911,4	-108,9	-inf
1106,8	-103,3	-inf
1171,9	-93,2	-inf
1237,0	-93,5	-inf
4036,4	-94,9	-inf
4101,5	-85,2	-inf
4166,6	-93,2	-inf
5403,5	-94,5	-inf
5468,6	-94,2	-inf
7812,3	-97,4	-inf
9114,4	-95,3	-inf

Case-9 shows the case of an outdoor scenario with contribution from 3 transparent repeaters, when the user is at large distance from the repeaters, so that the levels of the signals from the repeaters is comparable with that from the LoS satellite reception. The overall delay is in this case in the order of 33 µs and thus this is the worst case PDP, as illustrated in table A.15. The corresponding delay spread results to be T_s(case 9) = 11,823 µs, which is critical for the frequency selectivity over Δf for LTE at 15 kHz and for WiMAX with carrier spacing above 10,94 kHz, and which is larger than the guard interval period for LTE wi.mal cyclic prefix and for a series of WiMAX configurations.

Table A.15: Wideband propagation channel, Case 9



Case 9: Satellite + 2 IMRs (without processing delay) - very large delay		
Delay [ns]	Power [dBm]	Rice Factor [dB]
0	-91,9	7
325,5	-106,0	7
17642,9	-104,7	-inf
17708,0	-96,2	-inf
17773,1	-98,2	-inf
17838,2	-96,6	-inf
18098,6	-96,8	-inf
29361,4	-113,3	-inf
29817,1	-91,4	-inf
30207,7	-91,1	-inf
32160,8	-111,2	-inf
32746,7	-109,5	-inf

A.3 Physical and Physical-Statistical Models

Physical models rely on a deterministic modelling of propagation phenomena (reflection, diffraction, refraction), and also of the considered environment. These models have been efficiently used for planning purposes in terrestrial radio-communication or broadcast networks. Due to the global area coverage of satellites, deterministic modelling is not often used in the mobile satellite scenario, as the range of environments to cover are vast. However, if a particular application is required, for example satellite to indoor propagation, then 2D, 2.5D or 3D ray tracing can be used to estimate the wideband shadowing and small scale fading. Some details of applying ray tracing to the mobile satellite environment is given in [i.35].

Whilst physical or deterministic models can provide accurate and detailed electrical channel characteristics, they are impractical for a mega-cell environment. On the other hand purely empirical channel models, based on measurement data, can provide excellent statistical accuracy. However their application is limited to similar environments and frequencies. A compromise between these approaches, known as physical-statistical modelling, was developed in [i.36]. They combine the statistical accuracy, ease-of-use and low computational requirements of empirical models, yet with the physical insights of deterministic models. The approach uses geometrical optics and the geometrical theory of diffraction, known as ray-tracing, on statistically accurate environment parameters. Different environments can be categorized by a set of these environment statistics - for example the sporadic nature of buildings and their height distributions, and vegetation statistics can be inputted into ray tracing simulations. Physical-statistical channel modelling also lends itself well to obtaining Markov state and transition matrices in multiple-state models. Some examples of physical-statistical modelling are given in [i.36] to [i.39].

A.3.1 MIMO (multi-satellite and dual-polarized)

A physical-statistical LMS-MIMO channel model has been presented in [i.40] which permits simulation of the multiple satellite and/or dual polarization channels under different environments and satellite elevations. Based on this model the capacity gains of the LMS MIMO channel have been predicted in [i.41].

A second empirical-statistical LMS-MIMO model, formed from the measurement campaign data, has been reported in [i.42]. The model uses Markov chains for the very slow fading effects, log-normal distributions for the large scale fading and conditioned Ricean distributions for the small scale fading. A narrowband and a wideband model were presented. The model is particularly useful for the understanding of the dual polarized LMS-MIMO channel. The capacity of the dual polarized LMS-MIMO channel has been predicted in [i.43] based on the experimental data presented in [i.42].

Annex B: Detailed Review of Multi-Signal Detection Techniques

B.1 DS-CDMA up-link model and the formulation of the classical Multi-User Detection Problem

U active users are assumed in the system, where each desires to transmit a bit sequence $\mathbf{b}_u, 1 \le u \le U$ of length *N*. $\mathbf{b}_u, 1 \le u \le U$ is mapped on the BPSK constellation to produce symbol sequences $\mathbf{a}_u, 1 \le u \le U$. Subsequently, these are modulated by a high rate spreading code $\mathbf{s}_u, 1 \le u \le U$ of length *L* chips, which is unique for each user. Furthermore, the codes are chosen and normalized so that they satisfy the condition $\mathbf{s}_i \mathbf{s}_j^H = \begin{Bmatrix} 1, i = j \\ \rho_{ij} < 1, i \ne j \end{Bmatrix}$.

The spread symbols are transmitted through multipath channels, which in the discrete time domain can be modelled as Finite Impulse Response (FIR) filters of memory order *M*:

$$\mathbf{h}_u = \sum_{m=1}^{M+1} h_u^m \delta[(k-m)T_c] \tag{B-1}$$

where *k* is a discrete unit delay variable, h_u^m are the filter's tap weights, T_c represents the chip period $\delta[k] = \begin{Bmatrix} 1, k = 0 \\ 0, k \ne 0 \end{Bmatrix}$. The relative delays between user transmissions are also incorporated in the channel responses. The receiver observes a superposition of the filtered signals in AWGN:

$$\mathbf{r} = \sum_{u=1}^{U}\sum_{n=1}^{N} A_u a_u[n] \cdot (\mathbf{s}_u * h_u) + \mathbf{n} = \sum_{u=1}^{U}\sum_{n=1}^{N} A_u a_u[n] \cdot \mathbf{g}_u + \mathbf{n} \tag{B-2}$$

$(*)$ denotes discrete convolution between two sequences, A_u is the received amplitude of user *u*, $\mathbf{g}_u$ can be thought of as the effective user code, whose duration however spreads beyond the duration of an information symbol T_S resulting in ISI, and $\mathbf{n}$ is a sampled realisation of a zero mean white Gaussian process with variance σ^2 and is of length (*LN*+*M*).

The sufficient statistic for estimating the user information is extracted by matched-filtering $\mathbf{r}$ with the effective user codes [i.47]. At a system level this is equivalent to passing $\mathbf{r}$ through a bank of *U* parallel matched-.filters with responses $\mathbf{g}_u^H$. It is useful if (B.2) is re-written in a compact matrix form:

$$\mathbf{r} = \sum_{u=1}^{U} \mathbf{G}_u \mathbf{A}_u \mathbf{a}_u + \mathbf{n} = \mathbf{GAa} + \mathbf{n} \tag{B-3}$$

where $\mathbf{G}_u$ is a $[(LN+M)\times N]$ matrix, each column of which contains a downward shift of $\mathbf{g}_u$ (the detailed description of $\mathbf{G}_u$ can be found in i.49), $\mathbf{A}_u = diag(A_u, A_u, .., A_u)$ and its size is $N \times N$, $\mathbf{A} = diag(\mathbf{A}_1, \mathbf{A}_2, .., \mathbf{A}_U)$ (i.e. a diagonal matrix of size $(UN \times UN)$ with the first N diagonal elements being A_1 then next N, A_2 and so on) and $\mathbf{a} = [\mathbf{a}_1^T, \mathbf{a}_2^T, .., \mathbf{a}_U^T]^T$. At the matched filter bank's output:

$$\mathbf{z} = \mathbf{G}^H\mathbf{GAa} + \mathbf{G}^H\mathbf{n} = \mathbf{RAa} + \mathbf{v} \tag{B-4}$$

In (B-4), $\mathbf{R}$ is $(UN \times UN)$ block filtering matrix, which is symmetric semi-definite positive, and $\mathbf{v} \sim N(0, \sigma^2 \mathbf{R})$ i.e. coloured Gaussian noise. $\mathbf{z}$ is the sufficient statistic for estimating $\mathbf{a}$, usually with assuming knowledge of $\mathbf{R}$ and $\mathbf{A}$.

B.1.1 Equivalence with the Spatially Multiplexed MIMO Equalization and Inter-Spotbeam Interference Suppression Problems

It can be easily shown that the MUD and the spatial-multiplexing MIMO equalization and inter-spot beam interference suppression problems are mathematically equivalent. Thus a signal estimation technique developed for one system can in principle be readily applied to the other. In fact all three systems are described by the general linear model:

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \boldsymbol{\varepsilon} \qquad \text{(B-5)}$$

where $\mathbf{F}$ is a $[m \times n]$ linear mixing matrix (which could have some special-filtering structure), $\boldsymbol{\varepsilon} \sim N(0, [\mathbf{C}_{\boldsymbol{\varepsilon}} \mathbf{C}_{\boldsymbol{\varepsilon}}^H])$ and $\mathbf{x}$ is a random information vector whose elements are independently and equiprobably drawn from a finite signal alphabet Ω.

B.2 Review of MUD Methods and Algorithms

Optimal Signal Detection

The optimal joint signal detection algorithm, which was presented in i.49, solves the Maximum Likelihood (ML) problem within the set of feasible solutions: Ω^n. The constrained ML problem is formulated as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \Omega^n} p(\mathbf{y} / \mathbf{x}) = \arg\min_{\mathbf{x} \in \Omega^n} \left| \mathbf{C}_{\boldsymbol{\varepsilon}}^{-1} (\mathbf{y} - \mathbf{F}\mathbf{x}) \right|^2 \qquad \text{(B-6)}$$

where the second equality follows because $\boldsymbol{\varepsilon}$ is Gaussian. In general the constrained ML problem is known to be NP-hard, meaning that it requires an exhaustive search through the solution set, which makes the complexity of the optimal detector grow exponentially with the problem's dimensionality: $O(\omega^n)$, where ω is the cardinality of the set Ω. Similarly calculating the marginal posterior signal probabilities for each transmit symbol ($p(x_i \ / \mathbf{y} \ , \mathbf{F}), 1 \le i \le n$) as soft reliability information to be forwarded to the channel decoder, also requires exhaustive enumeration over all signal combinations.

For the MUD problem some special structures of the code correlation matrix have been identified which allow optimal detection in polynomial complexity [i.50] and [i.51]. In practice however the randomness of the radio environment tends to destroy any imposed structure on $\mathbf{F}$. The impractically high complexity of the optimum detector has been the motivation for the development of sub-optimal techniques that can approach the optimum performance in realistic complexity.

Relaxed ML detection

A basic approach for developing reduced complexity sub-optimal techniques is to relax the hard constraint in the ML problem, in a way which makes the problem tractable by polynomial complexity algorithms. One such relaxation is achieved by the Semi Definite (SD) constraint (see [i.52]), according to which the ML problem is solved under the constraint $\mathbf{x}\mathbf{x}^T \succ 0$ i.e. $\mathbf{x}\mathbf{x}^T$ is semi-definite positive. Semi-definite programming can be applied to the MUD problem has been studied in [i.53], where it was shown that very good performance is achieved in polynomial complexity: $O(n^{3.5})$.

A different kind of relaxation is to assume that the solution lies within a closed convex set such as a hyper-cube or a hyper-sphere. The formal proof that the Convex Constraint (CC)-ML problem can be solved by a polynomial time algorithm can be found in [i.54]. An early proposed polynomial complexity algorithm for solving CC quadratic problems is the Gradient Projection Method (GPM) [i.55]. In [i.56] a polynomial-time algorithm, which was developed in [i.58], has been applied in MUD for solving the ML problem under the hyper-cube constraint. In [i.56] and [i.57] it is shown that previously proposed Parallel Interference Canceller (PIC) algorithms are special cases of the general algorithm given in [i.58]. In [i.56] it is also commented that there is a strong relationship between the hyper-sphere constraint ML detector and the linear Minimum Mean Squared Error (MMSE) detector (see references within [i.56]) and it s demonstrated through simulations that the two have almost identical performance. In general the CC-ML detectors perform worse than the SD programming algorithms but they are computationally simpler.

Sphere DEcoding (SDE) has been proposed in [i.59] and [i.60] for MUD and MIMO detection respectively. In SDE an exhaustive search is performed within a subset Λ of Ω^n, which when linearly transformed by $\mathbf{F}$, the resulting set is enclosed within a hyper-sphere defined in R^n with radius $\sqrt{C}$ and cantered at $\mathbf{y}$ (the method assumes that the observation noise is white and noise whitening needs to precede the application of the algorithm if noise is coloured). This guarantees that only the lattice points within the squared distance C from $\mathbf{y}$ are considered for metric minimization. In this way the ML problem is solved under the constraint relaxation that $\mathbf{x} \in \Lambda$. The magnitude of C controls the performance and complexity of the algorithm; as $C \to \infty$ the optimum solution is guaranteed through a complete search in Ω^n. On the other hand as $C \to 0$ the search space is minimized but no solution will be obtained (as Λ will be empty). In the algorithmic implementation of the SDE proposed in [i.59] and [i.60], n = 64 is a practical limit since in the worst case the complexity order is $O(n^6)$. The variable computational complexity (depending on the realization of the system matrix and the noise power) is an important drawback of the SDE, which limits its suitability in many practical applications.

Linear Detection

Linear signal detection methods include the Least Squares (LS) and MMSE detectors:

$$\hat{\mathbf{x}}_{LS} = \Psi_\Omega\{(\mathbf{F}^H[\mathbf{C}_\varepsilon\mathbf{C}_\varepsilon^H]^{-1}\mathbf{F})^{-1}\mathbf{F}^H[\mathbf{C}_\varepsilon\mathbf{C}_\varepsilon^H]^{-1}\mathbf{y}\} \quad \text{(B-7)}$$

$$\hat{\mathbf{x}}_{MMSE} = \Psi_\Omega\{(\mathbf{F}^H[\mathbf{C}_\varepsilon\mathbf{C}_\varepsilon^H]^{-1}\mathbf{F}+\sigma^2\mathbf{I})^{-1}\mathbf{F}^H[\mathbf{C}_\varepsilon\mathbf{C}_\varepsilon^H]^{-1}\mathbf{y}\} \quad \text{(B-8)}$$

Where Ψ_Ω signifies the nearest point projection operator on the signal alphabet Ω. (B.7) is more widely known as the Zero-Forcing (ZF) detector, as inter-user and/or inter-antenna interference is forced to zero but at the expense of amplifying the noise in cases where $\mathbf{F}$ is badly conditioned. (B.8) offers a fine balance between residual interference in the estimate and noise variance and generally provides a more reliable estimate than the ZF detector. Both approaches relax the constraint that the solution belongs in a discrete finite set and they solve the problem within n-dimensional complex vector space. In fact, in the studied problems the LS solution coincides with the ML estimate by relaxing the hard solution constraint to the continuous vector space.

In general both linear detectors offer moderate performance while the need for a matrix inversion makes their complexity of the order of $O(n^3)$. However, iterative methods for optimizing convex quadratic problems [i.61] can offer reduced complexity implementations, especially in cases where $\mathbf{F}$ has many zero entries (see [i.62] and [i.63] for applications in MUD). For the MUD problem it is important to note that both linear methods are optimum in terms of near-far resistance [i.64]. Moreover, a blind implementation of the MMSE detector which does not require knowledge of the code sequences and user amplitudes has been proposed in [i.65].

Decision driven detectors

Iterative decision-driven MUD techniques were early low-complexity solutions for suppressing Multi-User Interference in DS-CDMA systems. Examples of such non-linear detectors are the Successive Interference Canceller (SIC) [i.66], the Parallel Interference Canceller (PIC) [i.67] and [i.68], the Decision Feedback (DF) detector [i.69] and [i.70] and the Group Decision Feedback (GDF) [i.71] detector. The SIC and the PIC in general provide inferior performance compared to linear MUD methods (Decorrelator, MMSE) but their complexity depends only quadratic ally on the problem's dimensionality, unlike the linear detectors where the dependency is cubic. The DF and the GDF detectors, on the other hand, offer performance improvements relative to linear detectors while their complexity, though increased, remains of the same order.

The idea in the Interference Cancellation (IC) approaches is to reuse estimated symbols, at the output of the user-code matched filter(s), in order to reconstruct and cancel interference from the channel observation. In the Successive Interference Canceller (SIC), the interference cancellation process proceeds sequentially, i.e. user-by-user. Performance improvements are achieved by ordering users from the strongest to the weakest in terms of received power [i.66]. On the other hand, the PIC reconstructs and cancels in parallel interference caused by all interfering users so that the user of interest is detected more reliably. Performance enhancement in the PIC can be achieved by repeating the whole procedure several times before a data decision is made. However, in the case where hard data decisions are made, this multi-stage approach cannot guarantee performance improvements as successive stages enhance further the increased interference caused by wrong data decisions. This problem is avoided by the partial interference cancellation approach proposed in [i.72], according to which a smaller fraction of the reconstructed interference is cancelled in initial stages and this fraction increases in further stages. Several types of PIC have been proposed with respect to the function which is applied on the soft estimates prior to IC; in [i.73] this function is linear so the soft estimate is used directly for IC, while in [i.74] a hard-limiting function is applied to the soft-estimate. The unit-clipper function proposed in [i.74] and [i.75] has been proven to optimize the ML problem under the hyper-cube constraint relaxation ([i.56] and [i.57]).

In the DF detector only one user's data is decided at the output of a linear detector (ZF or MMSE). The hard decision is used to cancel the interference caused by this user and linear detection is applied again for the remaining undecided users. In the iterative detection process, users are optimally ordered [i.76]. In GDF detection the same principle is applied but in each iteration a group of users is decided. Optimizing the group sizes for the GDF detector has been proposed while optimal group ordering has been derived in [i.76].

Though decision-driven iterative detectors offer an attractive performance trade-off, in most realistic scenarios they fail to approach closely the optimal performance offered by the constrained ML detector.

Heuristic search algorithms

Heuristic search algorithms [i.77] perform a limited search in Ω^n; seeking the global optimum of the ML cost function. Such algorithms, though they cannot guarantee convergence to the optimal solution, can in practice provide good performance. Examples of heuristic algorithms which have been applied to signal detection include the Local Search Algorithm (LSA) [i.78] and [i.79], the Genetic Algorithm (GA) [i.80] and [i.81], the Tabu search algorithm [i.82] and the Boltzmann machine [i.83]. The basic idea behind these heuristic algorithms is to penalize the cost function associated with search directions which have a small probability. Another common characteristic, which aims to avoid convergence to a local minimum, is the introduction of a random rule for updating the solution. For example, in the Tabu search this random rule is realized by forcing flips in some signal values if no improved solution can be found by the normal procedure of the algorithm. In the GA on the other hand, this random rule is realized through the random process by which the "solution pool" is updated in each iteration; i.e. through random crossovers and mutations of the surviving solutions. In terms of complexity, heuristic algorithms suffer from slow convergence as they typically require hundreds of iterations to converge. Moreover their complexity is further burdened by the requirement of cooperation with some other suboptimal detector to provide a good starting solution, if good performance is to be achieved.

Probabilistic Data Association

Probabilistic Data Association (PDA), initially proposed for target tracking [i.84], is an iterative technique for updating the signal probabilities. The main idea in PDA is to make the probability updates by approximating the interfering signals plus noise (which have a multi-modal Gaussian distribution) by a single Gaussian distribution with matched mean and covariance. This is an unjustifiable assumption, but simulation results have shown that the PDA detector provides near optimal MUD performance (with Gold spreading codes) in polynomial complexity $O(n^3)$ [i.85].

Turbo Detection Techniques

In coded systems substantial performance improvements can be achieved through iterative turbo-detection techniques, in which soft signal information produced by the channel decoder are used to update prior signal probabilities; for improving the performance of the MUD detector/MIMO equalizer, or alternatively for constructing soft signal estimates for cancelling inter-antenna interference. Iterative detection should be expected to find application in future systems since it provides very significant performance/system capacity improvements, while many efficient algorithmic implementations can be found in the published literature.

The best performing iterative detector makes use of an exact a-posterior probability detector, but its complexity is exponentially dependent on the problem's dimensionality. Approximate a-posterior detectors, such as the one proposed in [i.86], provide near-optimal performance in a computationally efficient manner. Single-user performance in convolutionally coded system over AWGN channel is also reported in [i.87] where a soft-interference canceller is complemented by a linear MMSE filter. The PDA algorithm can also be applied for computationally efficient and high performing iterative detection [i.88].

An iterative detector in LDPC coded systems is reported in [i.89].

Annex C: Detailed Review of Diversity Techniques and MIMO

C.1 Types of Diversity

C.1.1 Frequency Diversity

The "classical" frequency diversity concept consists of transmitting the same signal over multiple carriers that are separated by at least the coherence bandwidth of the fading channel. Frequency diversity offers improved link-performances over frequency selective channels, and is thus more suitable for very high-rate systems. This classical type of frequency diversity is obviously not a cost-effective option, since it involves sacrificing large amounts of frequency resources for solving a type of problem (frequency selective fading), which is not so critical in mobile satellite systems.

In high rate satellite systems more cost-effective approaches for providing frequency diversity gains, are to make use of coded and interleaved OFDM or fast frequency hopping. These type of solutions, though they do not offer any increase in the average received power (as compared to convention frequency diversity), are effective in exploiting the diversity in frequency selective channels without sacrificing any spectrum.

Frequency diversity is currently specified in mobile broadcasting systems, such XM Radio and is also implicitly available in the OFDM mode of the DVB-SH system (though frequency domain coding and interleaving).

C.1.2 Time Diversity

Time diversity involves transmitting the same signal on more than one time slots that are separated in time by at least the coherence time of the fading process. Time diversity can be designed to offer both macro and micro diversity benefits; by setting the retransmission delay to be larger than the correlation time of the slow-fading effects.

As with frequency diversity, the classical version of time diversity results in big system capacity sacrifices. Bit interleaving combined with error control coding offer time diversity in a more cost-effective way; provided the interleaver's depth is larger than the channel coherence time. Very long time interleaving (i.e. in the order of several seconds) can in principle be introduced in order to provide resistance against large scale fading. This approach has been adopted in the DVB-SH standard where the time interleaver can be flexibly specified up to tens of seconds [i.1]. Simulation results in annex A of [i.1] show that the time interleaver plays a key role in providing sufficient QoS for vehicular reception in rural and sub-urban types of environments (link-level simulations in the TM-SSP group have used the three-state Fontan model for modelling the LMS channel (see annex A.2.3))

Though time interleaving is a suitable option for broadcasting/multicasting, its application in bi-directional data systems is limited by the delay requirements of interactive applications. Even for relatively delay-insensitive applications such as web-browsing and on-line messaging, the depth of the time-interleaver would have to be limited to a couple of seconds; in order also to avoid TCP timeouts. However, some types of applications would be suitable to undergo very long time interleaving, namely e-mails, FTP, SMS, MMS. Though not compatible with any existing standards, one could envisage a cross-layer or a multi-standard approach where the network and link-layer would forward these types of applications to a "long-interleaved" physical layer chain. Such approach would enhance the availability of certain types of applications in a resource-economical manner.

C.1.3 Space diversity

Space transmit/receive diversity involves the use of more than one antenna on both or either communicating ends. The use of multiple receiving antennas increases the average SNR and thus provides a performance benefit even in LOS conditions. On the other hand space-transmit diversity (which is enabled through space-time coding) only provides increased robustness against fading effects, but offers little improvements in LOS conditions. Diversity against small-scale fading is relatively easy to achieve since only half-wavelength antenna separations are acceptable for achieving sufficient channel decorrelation. On the other hand diversity against shadowing effects requires the antennas need to be separated by large distances. In satellite systems macro-diversity can be practically achieved through the use of multiple satellites (although this is an expensive solution - particularly with GEO constellations), or in special types of vehicular platforms such as trains and buses.

The use of space receive diversity (on the mobile terminal) has been investigated by the TM-SSP group, and it has been concluded that it provides significant improvements in the DVB-SH service availability.

On the other hand, multi-satellite diversity has been studied thoroughly and simulations were carried out in order to determine the satellite visibility and multiple satellite diversity probabilities, for various constellation orbit architectures and various mobile terrestrial environments. The target design aim has been to ensure at least one satellite is visible to the mobile for the highest amount of time. The geographical mobile-population density has also been accounted for; in that constellation orbits are often optimized around mid-latitudes. Lastly, diversity combining has been assumed when more than one satellite is visible to a mobile, signals can be combined in a diversity arrangement.

For the case of two satellites, Figure C-1 shows the correlation coefficient between satellite visibilities as a function of azimuth separation at various elevations, street widths ($W = 12m$ shown here), mobile position from buildings ($d[m]$]) and building height ($H_b = 15m$ shown here). The optimum azimuth separation is 90^o , where a negative cross correlation coefficient occurs (meaning when one satellite is blocked there is a higher probability that the other satellite is visible). This is also intuitively correct, as when a mobile is moving along a street, one satellite is more likely to be visible if the other is blocked when separated in azimuth by 90^o . When satellites are close together, as they would be in a satellite cluster, the shadowing will be highly correlated. Similarly if satellites are positioned at either side of the mobile at 180^o azimuth, blockage is also likely to be highly correlated.

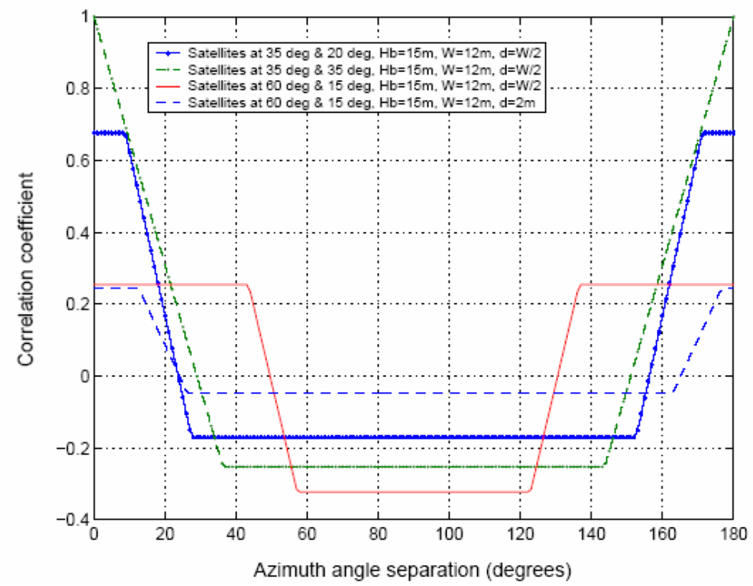


Figure C.1: Optimum position of two satellites for maximizing satellite visibility

C.1.4 Polarization diversity

Polarization diversity involves transmission of the same signal over a dual polarized antenna (e.g. LHCP/RHCP). The capacity improvements achievable over dual-polarized land mobile satellite channel have been recently characterized through experimental results [i.42]. In shadowed conditions the dual polarized channel behaves much like a 2×2 MIMO system and thus MIMO detection should be employed in order to detect reliably the transmit signals [i.90].

C.2 Receive Diversity Combining Techniques

C.2.1 Switch diversity/Selection Diversity

The simplest receive diversity schemes switch (and stay) and selection diversity. In both schemes the only one receiving antenna is active at any time. In switch and stay a switching threshold is used, which is often a loss in signal level on the antenna being used. In selection diversity the signal strengths of the two antennas are continuously monitored and the strongest one is selected. The switching may be performed at RF to avoid the need for a down converter for each antenna.

Switch diversity can also be applied effectively in transmit diversity mode for countering large scale fading effects. For example this can be achieved in an open loop fashion by estimating the received signal strength in the forward link; in order to detect shadowing/blockage. Of course this approach assumes large physical separations of the antenna, and is thus practical for trains/buses.

C.2.2 Maximal Ratio and Equal Gain Combining

In this method the combiner makes use of all received signals instead of only one branch. Every signal is multiplied with a weighting factor before the combining process. Equal gain is similar to maximal ratio but all the weights have the same magnitude and only the phases of the signals are manipulated in order to ensure the constructive addition of the signals at the receiver. Both Maximal Ratio and Equal Gain Combining rely on a-prior knowledge of the channel and the combining needs to be done after the demodulation process.

Given the advances in RF and digital technologies, maximum ratio combining offers the most cost-effective solution, particularly in satellite systems where it brings a minimum of 3 dB benefit in LOS conditions.

C.2.3 Space-Time Coding

Space Time (ST) codes (see [i.93]) introduce redundancy (in a systematic manner) in systems with multiple transmit antennas, with the aim to make use of the diversity available due to the (uncorrelated) radio sub-channels, and possibly to provide a coding gain. This is a different type of utilization of the transmit antennas in a MIMO system, relative to the spatial multiplexing architecture, which aims to make the communication link more robust against the fading effects of the radio channel. Although utilizing space transmit diversity is inefficient spectrally and computationally compared to utilizing space receive diversity (see [i.94]), in systems with an unbalanced number of antennas (between the receiver and transmitter), it can be the best (or the only) option. In [i.93] and [i.171] a summary is provided regarding the categorization of transmit diversity techniques; a) schemes with feedback, b) schemes with training information but no feedback, and c) blind schemes (see references within [i.93]). STC belongs to the second category.

There are two main families of ST codes, both designed with the purpose of enabling space diversity; ST Block Codes (STBC) [i.95] and [i.96] and ST Trellis Codes (STTC) [i.97]. STBC constitute the construction of an orthogonal ST matrix of information symbols (i.e. each row of this matrix corresponds to a different transmit antenna and each column to a different symbol period). For a system with 2 transmit antennas, the size of the ST matrix is 2x2; allowing full transmit diversity without a sacrifice in the transmission rate (Alamouti scheme). This is generally possible also for transmitters with more antennas provided that only real signalling constellations are used. For systems with more than 2 transmit antennas, that make use of complex signalling, the transmission rate of STBC cannot exceed 0,75 symbols per channel [i.98]. Despite this inefficiency, STBC enjoy very low decoding complexity; similar to that of the Maximum Ratio Combiner (MRC) in systems which use space receive diversity. On the other hand, STTC utilize a convolutional encoder and thus require a trellis type decoder. The high decoding complexity is the major drawback of STTC, which nonetheless can allow small rate sacrifices and also offer coding gains as well as exploiting optimally the available diversity. STTC also suffer from the fact that they lack closed form construction, while the code design difficulty (based on the criteria proposed in [i.98]) grows exponentially with the number of transmit antennas and the transmission rate. Other recently proposed ST codes [i.99] and [i.100] are designed with the aim of achieving full diversity with no sacrifice in transmission rate. However, they suffer from high computational complexity and as for optimal decoding they require exhaustive search detection.

STC schemes have been proposed with the basic assumption that the MIMO channel undergoes flat fading. While in OFDM systems the standard "narrowband" ST codes can be applied directly per sub-carrier, in wideband single-carrier systems an adaptation of the code design is required. An example of such adaptation for STTC is found in [i.101]. A block orthogonal design for adapting STBC to ISI channel conditions has been proposed in [i.101]. This scheme can achieve maximum diversity $N_t(M+1)$ [i.102]. This STC architecture can be viewed as a direct extension of the Alamouti scheme ([i.95]), designed for frequency flat channel conditions. The two schemes however differ distinctively in two ways; the orthogonal transmit matrix in the wideband scheme is constructed by information symbol vectors (rather than discrete information symbols) and on the receiver side it requires separate equalization of the decoupled-maximally ratio combined received sequences. A different approach for extending the application of STBC in wideband systems can be found in [i.103], where the original (symbol based) Alamouti scheme is utilized and widely linear equalization [i.104] is utilized for jointly equalizing and decoding.

As far as mobile satellite systems are concerned, where multiple antennas can be fit on the mobile terminal, transmit diversity cannot be expected to provide the same as significant benefits as receive diversity (mainly because it does not increase the average SNR). However if the terminal is equipped with two antennas then it would make sense to apply a simple STC on the return link, in order to provide better resistance against Rice fading, and possible assist the interleaver and channel decoder to improve the availability over lightly shadowed environments.

Annex D: Review of Optimal Combining and Direction of Arrival Algorithms for Beamforming

D.1 Optimal Combining Algorithms

There are numerous algorithms for optimizing the beam former weights with respect to some chosen optimization criterion. In this clause some algorithms for optimizing the SINR will be reviewed. The optimization problem has the form:

$$\mathbf{w}_{MSINR} = \max_{\mathbf{w}} \frac{\sigma_{SS}^2}{\sigma_{I+N}^2} = \max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_{SS} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{I+N} \mathbf{w}} \tag{D-1}$$

where $\mathbf{R}_{SS}$ is the spatial correlation matrix of the desired signal only and $\mathbf{R}_{I+N}$ is the spatial correlation matrix of interference plus noise only. Equation (D.1) can be reduced to i.108:

$$\mathbf{w}_{MSINR} = \max_{\mathbf{w}} \frac{\sigma_{SS}^2 \left|\mathbf{w}^H \mathbf{a}_{SS}\right|^2}{\mathbf{w}^H \mathbf{R}_{I+N} \mathbf{w}} \tag{D-2}$$

σ_{SS}^2 is the signal power and $\mathbf{a}_{SS}$ is the spatial signature of the user of interest. The latter can be obtained from the DoA estimation step, which should precede the beamforming step. Other ways for obtaining the spatial signature of the desired user is to make use of training sequences or (semi-)blind source separation. The optimization problem can equivalently be posed as:

$$\mathbf{w}_{MSINR} = \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{I+N} \mathbf{w} \text{ subject to } \mathbf{a}_{SS}^H \mathbf{w} = 1 \tag{D-3}$$

The solution is given as:

$$\mathbf{w}_{MSINR} = \gamma \mathbf{R}_{I+N}^{-1} \mathbf{a}_{SS} \tag{D-4}$$

where γ is an arbitrary scalar constant which does not affect the output SINR. It can be observed that determining the optimal weights requires knowledge of $\mathbf{R}_{I+N}$, which is not available, and also accurate knowledge of $\mathbf{a}_{SS}$. The first approach to get a solution for (D.4) is to approximate $\mathbf{R}_{I+N}$ by the sample covariance matrix:

$$\hat{\mathbf{R}}_{XX} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}(k) \mathbf{x}^H(k) \tag{D-5}$$

where $\mathbf{x}$ is the observed signal and K is the number of snapshots available. The approach is known as Sample Matrix Inversion (SMI). It can be shown that the SMI solution for the maxSINR problem is the same as the optimum solution for the MMSE problem [i.109]:

$$\mathbf{w}_{MSINR} = \mathbf{w}_{MMSE} = \hat{\mathbf{R}}_{XX}^{-1} \mathbf{a}_{SS} \tag{D-6}$$

The performance of the algorithm is severely degraded when the signal component is in the data snapshots, in which case $\hat{\mathbf{R}}_{XX}$ is a bad approximation for the actual $\mathbf{R}_{I+N}$. Moreover, the algorithm does not provide sufficient robustness against mismatch between the available estimate for $\mathbf{a}_{SS}$ and its actual value. This sensitivity of SMI to spatial signature mismatch is alleviated to some extend by the Loaded SMI algorithm [i.110] which "loads" uniformly the diagonal of $\hat{\mathbf{R}}_{XX}$ by some small constant α prior to inversion. The method attempts to provide a regularization effect to the ill-posed inverse problem given by (D.6). Although performance improvements can be achieved in a seemingly simple manner, optimal choice for α represents a serious problem in practical applications.

Another approach for robust adaptive beamforming in the presence of spatial signature mismatch is the eigenespace-based beam former proposed in [i.111] and [i.112]. In this approach instead of the estimated spatial signature $\hat{\mathbf{a}}_{SS}$ being used directly, the projection of this vector on to the signal plus interference subspace is used instead:

$$\mathbf{w}_{eig} = \hat{\mathbf{R}}_{XX}^{-1}\mathbf{P}_E\mathbf{a}_{SS} \tag{D-7}$$

where $\mathbf{P}_E = \mathbf{E}\mathbf{E}^H$ with $\mathbf{E}$ containing the eigenvectors that span the signal plus interference subspace. The method requires a subspace decomposition on $\hat{\mathbf{R}}_{XX}$, which involves with identifying which of its eigenvectors span the signal plus interference subspace and which the noise subspace. The method is fundamentally akin to Loaded SMI as in both cases a filtering action is imposed on the spectrum of $\hat{\mathbf{R}}_{XX}$. In the eigenspace beam former, however, this is done in some optimal way as opposed to LSMI where there is no clear criterion for choosing α. The eigenspace method can only perform well in high SNR conditions as in low SNR there is a high probability of subspace swaps [i.113]. Additionally, the eigenspace beam former is efficient only if the dimensionality of the signal plus interference subspace is small and known exactly. These limitations are crucial for radio systems as scattering makes the signal plus interference subspace of uncertain (and perhaps high as well) dimensionality. The high SNR requirement is also an important limitation to be considered for any system under consideration.

A different robust beamforming algorithm has been proposed in [i.113] and [i.114]. This technique is based on the optimization of the worst case performance using Second Order Cone (SOC) programming. The idea there is to bound the spatial signature mismatch $\mathbf{\Delta}$ by some known constant $\varepsilon > 0$ ($|\mathbf{\Delta}| \leq \varepsilon$) and modify the constraints of the optimization problem described by (clause D.2) so that optimization is performed over all possible mismatched realizations of the spatial signature $\mathbf{a}_{SS} + \mathbf{\Delta}$. This implies that $\mathbf{a}_{SS} + \mathbf{\Delta}$ belongs to a continuous set that satisfies $|\mathbf{\Delta}| \leq \varepsilon$. The direct problem does not have a straight forward solution but a reformulation is given in [i.113] and [i.114] which permits the modified problem to belong to the class of convex SOC programs. This allows the problem to be efficiently solved using modern convex optimization tools. In particular the complexity of the algorithm is comparable to the conventional adaptive beam formers (SMI, LSMI) [i.115]. The performance of the algorithm has been investigated in [i.115] for a Time Division Synchronous CDMA system where it is concluded that the SOC beam former is the best among the various beamforming techniques. More specifically, it is concluded that it is nearly optimal in terms of robustness against spatial signature mismatches, applicability to arrays with small number of elements and computational complexity.

D.2 Direction of Arrival (DoA) Estimation

DoA is a necessary step (prior to beamforming) for many systems either employing null-steering or optimal combining. The performance of the beam former in most cases is directly depended on the quality of the DoA estimate. In this clause a brief reference to the various DoA estimation methods is made. These methods can be classified to spectral and parametric based techniques [i.116]. In the former a spectrum-like function is computed, i.e. the DoA spectrum. The locations of the highest (separated) peaks are the desired DoA estimates. Typical methods in this group include (in order of performance): classical Fourier, Capon [i.117] and MUSIC. While these methods are computationally attractive, they do not always provide the required accuracy, especially in the mobile radio environment with coherent multi-path sources. If high accuracy is required, the alternatives are the parametric methods. The increased robustness and accuracy is obtained at the expense of complexity since multidimensional optimization is required. If the multidimensional search is performed iteratively, an initial guess can be provided by the spectral methods. Typical methods in this group include: Deterministic and stochastic Maximum Likelihood [i.118] and [i.116], ESPRIT [i.119] and Weighted Subspace Fitting [i.120].

The requirement for DoA can be relaxed as recently proposed in [i.121] by using Blind Source Separation (BSS) techniques. Motivation for applying BSS arises especially from the potential performance gains that can be achieved when uncalibrated arrays are available. In [i.122] a semi-BSS method is proposed for a DS-CDMA under narrowband intentional jamming, which is based on Independent Component Analysis (ICA). The technique provides seems to provide good jammer mitigation both in the cases when the jammer's carrier frequency is locked and unlocked to the desired signal's carrier frequency.

Annex E: Detailed Review of State-of-the-Art Error Correcting Codes

E.1 Turbo Codes

Turbo Codes (TC), were introduced in [i.123], where a systematic encoder consisting of two parallel concatenated recursive convolutional encoders separated by a bit interleaver, was proposed. Optimal decoding of this code is practically impossible, as the number of states in the trellis grow exponentially not only with the convolutional encoder's memory, but also with the interleaver's length. However, the big innovation in the TC proposal was the possibility of suboptimal decoding with tractable complexity. The parallel concatenation of two BCJR algorithms [i.124] based on Soft Input Soft Output (SISO) decoders interconnected through an interleaver and a de-interleaver allowed performance as close as 0,7 dB (at BER $=10^{-5}$) to the Shannon performance bound, after a small number of iterations. The very high coding gains offered by the turbo encoder can be credited to the combination of its various features. Indeed, although in conventional convolutional coding recursive encoders offer no benefits compared to non-recursive schemes, in TC recursive encoders have a great influence on the error probability as they introduce an interleaver gain [i.125]. The serial concatenation (separated by interleaving) of recursive convolutional encoders was proposed in [i.126] and has been established as the serial counterpart of the original parallel turbo encoder. Once again, crucial for the performance of the code is the recursive nature of the constituent encoders.

Figure E.1 illustrates the structure of "classical" Rate 1/3 parallel concatenated turbo encoder. Higher rate codes can be constructed from this basic 1/3 code by puncturing the output bits according to some pattern. Puncturing essentially consists of periodically deleting some code bits from the encoded sequence. The puncturing matrix for producing a 1/2 rate code is:

$$\mathbf{P}_{1/2} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{(E-1)}$$

where the zeros in each row indicate the deleted bits for each pair of code bits on the corresponding branch of the encoder.

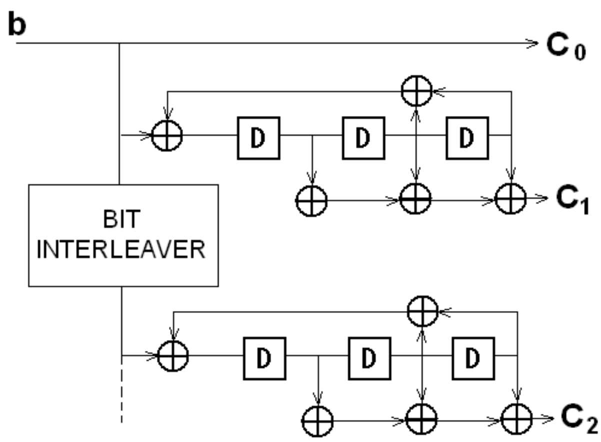


Figure E.1: Rate 1/3 Parallel Concatenated Systematic Turbo Code

The generic structure of the Turbo decoder is illustrated in figure E-2. In the initial iteration Soft-Input-Soft-Output (SISO) decoder 1 accepts the channel observations of the code bit sequences that correspond to the systematic and first code branches. In this initial decoding step, no useful prior information is available, so these are set to: $\Pr[b_i = 1] = \Pr[b_i = 0] = 0.5$ for all data bits. The Turbo decoder produces a-posterior probabilities for all data bits: $\Pr[b_i = 1/\mathbf{y}_0, \mathbf{y}_1]$ and $\Pr[b_i = 0/\mathbf{y}_0, \mathbf{y}_1]$. The posterior probability can be decomposed into three parts:

$$\Pr[b_i = 1/\mathbf{y}_0, \mathbf{y}_1] = p_1 p_2 p_3 \tag{E-2}$$

where:

- p_1 is the prior input probability $\Pr[b_i = 1]$ (or $\Pr[b_i = 0]$).
- p_2 is posterior information associated to the systematic branch $\mathbf{y}_0$, which is not related to the code's trellis, and it is also directly available to the second SISO decoder (by interleaving $\mathbf{y}_0$).
- p_3 is posterior information associated to the 1st coded branch $\mathbf{y}_1$. This is the only new information yielded by the code's constraints and is not available to the second decoder (through the interleaved $\mathbf{y}_0$ and $\mathbf{y}_2$). This is known as extrinsic information.

p_3 is the soft information output by the first decoder. This is interleaved (in order to match the input to the second encoder) and used to update the prior probabilities $\Pr[b_i = 1]$, $\Pr[b_i = 0]$ of all bits, that will be assumed by the second decoder. This Turbo process continues by feeding the extrinsic information produced by one decoder as input information for the other one. In the final iteration the posterior probabilities are used to decide on the value of each bit.

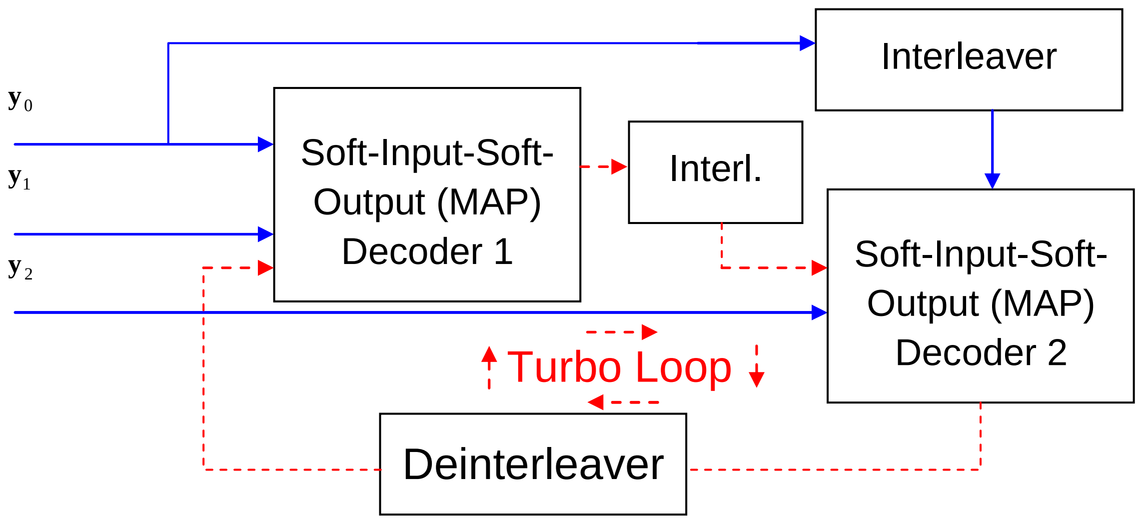


Figure E.2: Turbo Decoder

The above description of the turbo decoder assumes that the constituent SISO decoders are based on the Maximum A-Posterior (MAP) algorithm [i.124]. Several variations of the MAP algorithm have been proposed in order to reduce the numerical complexity of the decoder. Some examples are the Log-MAP and Max-Log-MAP [i.127], and the Soft Output Viterbi Algorithm [i.128]. Efficient analogue implementations of turbo decoders have also been receiving attention recently (see for example [i.129]).

E.2 LDPC Codes

Low Density Parity Check (LDPC) were originally proposed in [i.130] and rediscovered many years later in [i.131] and [i.132], where they were shown to achieve near Shannon-limit performance with a practical decoding algorithm. LDPC codes are a class of linear blocked codes in which some specific structure is imposed on their parity check matrix. The main feature of LDPC codes is the very low density of 1's in the parity check matrix **H**. A second structural feature of **H** is that any two of its columns have no more than one non-zero entry in common. As an example, the parity check matrix given in equation (clause E.3) does not classify as LDPC, since the 1st and 2nd columns have more than one non-zero entries in common and also it has a high density of 1's.

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad \text{(E-3)}$$

An LDPC code is further classified as regular if all its columns have exactly γ 1's and all its rows have exactly ρ 1's. In this case the LDPC code is referred to as a (γ, ρ) regular LDPC code.

Similar to the trellis representation of convolutional codes, LDPC codes can be represented graphically by a "Tanner" Graph. This consists of two node sets: variable nodes set: whose n elements correspond to the columns of **H**, and check nodes set, whose *n-k* elements correspond to the rows of **H**. Connections between the two node sets are determined by the coordinates of 1's on **H**. The Tanner graph for the parity check-matrix given in equation (E.3) (though it is not of an LDPC code), is shown in figure E.3, where the variable nodes are circular and the parity nodes square. The Tanner graph gives a graphical representation of the parity check equations a valid codeword needs to satisfy.

Some formal definitions:

- Defining $I = \{i_1, .., i_n\}$ as the variable node set and $J = \{j_{1,} .., j_{n-k}\}$ as the check node set, then the "neighbourhood" $N_v(i_k)$ of some variable node $i_k \in I$ is the set of check nodes connected to i_k and similarly the neighbourhood $N_c(j_k)$ of some check node $j_k \in J$ is the set of variable nodes connects to j_k.
- The number of elements (γ_{i_k}) in $N_v(i_k)$ is referred to as the degree of the variable node i_k and the number of elements (ρ_{j_k}) in $N_c(j_k)$ is referred to as the degree of the check node j_k.

In regular LDPC codes γ_{i_k} are equal for all i_k and similarly all ρ_{j_k} are equal for all j_k.

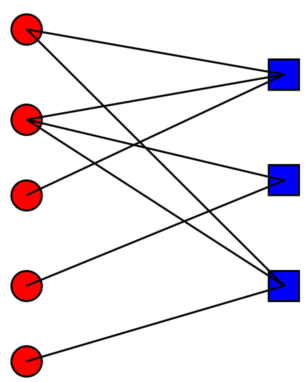


Figure E.3: Graph of the code described by equation (E.3)

E.2.1 Optimal Maximum Likelihood Decoding

Assuming an AWGN channel, the sampled noisy channel observation: $\mathbf{y} = \mathbf{c} + \mathbf{n}$ is available at the receiver, where $\mathbf{c}$ is the transmit codeword that consists of n bits, and $\mathbf{n}$ is a realization of n i.i.d. Gaussian random variables with zero means and variances σ^2. The Maximum Likelihood (ML) transmit codeword is determined by solving the optimization problem:

$$\hat{\mathbf{c}}_{ML} \arg\max_{\mathbf{c}\in\mathbf{C}} p(\mathbf{y}/\mathbf{c}) \qquad \text{(E-4)}$$

where $p(\mathbf{y}/\mathbf{c})$ is the likelihood function. In words, the ML codeword is the one, among all valid codewords that maximizes the (Gaussian) likelihood function. However this requires an exhaustive search throughout the set of valid codewords $\mathbf{C}$ (2^n words!), which is generally prohibitive in terms of search time and memory requirements.

E.2.2 Bit-Flipping Algorithms

Bit-Flipping (BF) [i.130] is the simplest practical (sub-optimal) approach for decoding LDPC codes, however its performance is typically far from optimal. The BF algorithm is classified as a "hard" decoding approach since it operates on hard-decisions on the noisy channel observations. The algorithm consists of the following main steps:

- Make hard decisions on the noisy channel observations: $\hat{c}_i = 0,\ \text{if}\ y_i < 0{,}5$ and $\hat{c}_i = 1,\ \text{if}\ y_i > 0{,}5$.
- Check which parity equations are satisfied and which not.
- Classify as unreliable those bits which are involved in most failed parity check equations, and flip them.
- Continue this process until all parity checks are satisfied, or for some pre-set maximum number of iterations.

The very small computational requirements of the BF algorithms have motivated research on modified routines which provide improved performance. Various improved performance variations of the BF algorithm have been reported in [i.133] and [i.134].

E.2.3 The Belief Propagation Algorithm

The BP algorithm [i.130] is the standard decoding algorithm used in practice, since it typically achieves near-optimal performance and involves manageable decoding complexity. The BP algorithms is an iterative technique for computing posterior probabilities on graph based models (like the Tanner Graph). It has since been rediscovered in other disciplines, and is also known as Sum-Product Algorithm (SPA) or Message Passing Algorithm (MPA). Each decoding iteration consists of two "passes".

In the first pass, messages in the form of conditional probabilities are passed from each variable node to its neighbouring check nodes, as it is shown in figure E.4. The message is in the form of the conditional probability that the variable-bit $c_i = 0/1$ given the input messages (passed in the previous iteration) from all other neighbouring check nodes, except the one the message is sent to. The variable node performs processing on the input messages in order to generate the outgoing message. Each variable node produces such messages for all of its neighbouring check nodes.

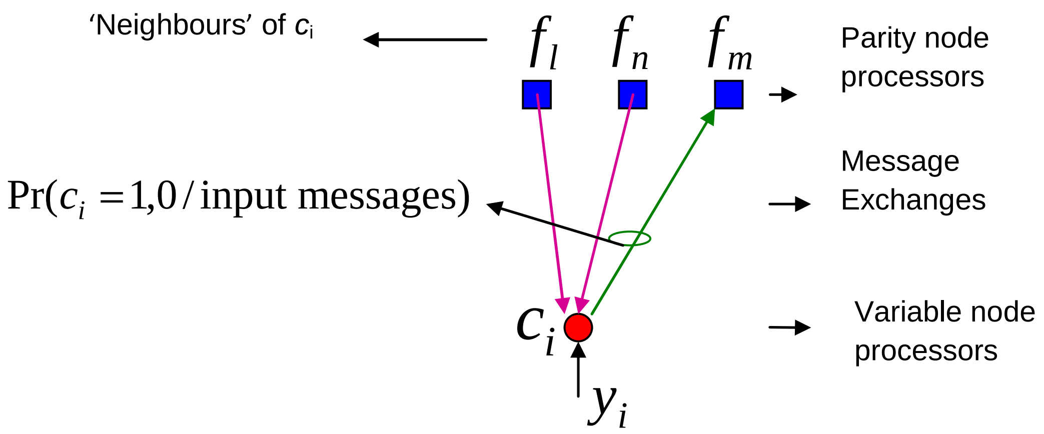


Figure E.4: First Pass in the BP Algorithm

In the second pass each check node sends a message to each of its neighbouring variable nodes, as it is shown in figure E.5. The message is in the form of the conditional probability that the parity check equation f_i is satisfied given all the input messages from the first pass. In this pass the parity checks need to carry out some processing in order to produces messages for all of their neighbouring variable nodes.

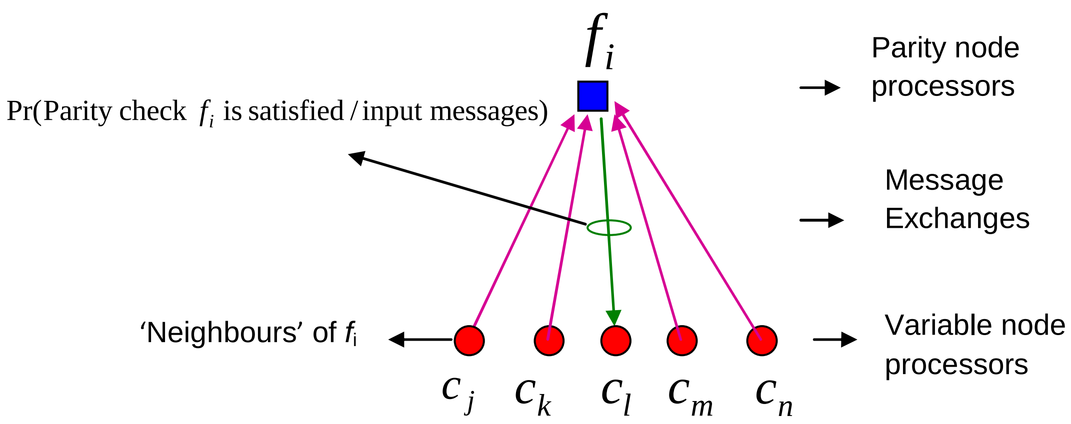


Figure E.5: Second Pass in the BP Algorithm

The iterative procedure continues until all parity checks are satisfied, or for some maximum number or iterations. Typically about 100 iterations are required for the algorithm to converge. The algorithm finally produces the posterior probabilities: $\Pr(c_i = 1/\mathbf{y})$ and $\Pr(c_i = 0/\mathbf{y})$, based on which decisions are made for the value of each code bit.

It has been shown that the BP algorithm is optimal provided that the Tanner graph of the code has no cycles. A cycle is a path on the Tanner graph which ends at the starting node. An example of a cycle is shown in figure E.6. In practice LDPC codes have cycles and thus the BP algorithm is sub-optimal. However provided that these cycles are not short then in practice the algorithm achieves near-optimal performance. A measure of the cyclic-structure of the code, is the length (in number of connections: edges) of the shortest cycle, which is termed as the "girth" of the code. Good LDPC codes are characterized by a large girth.

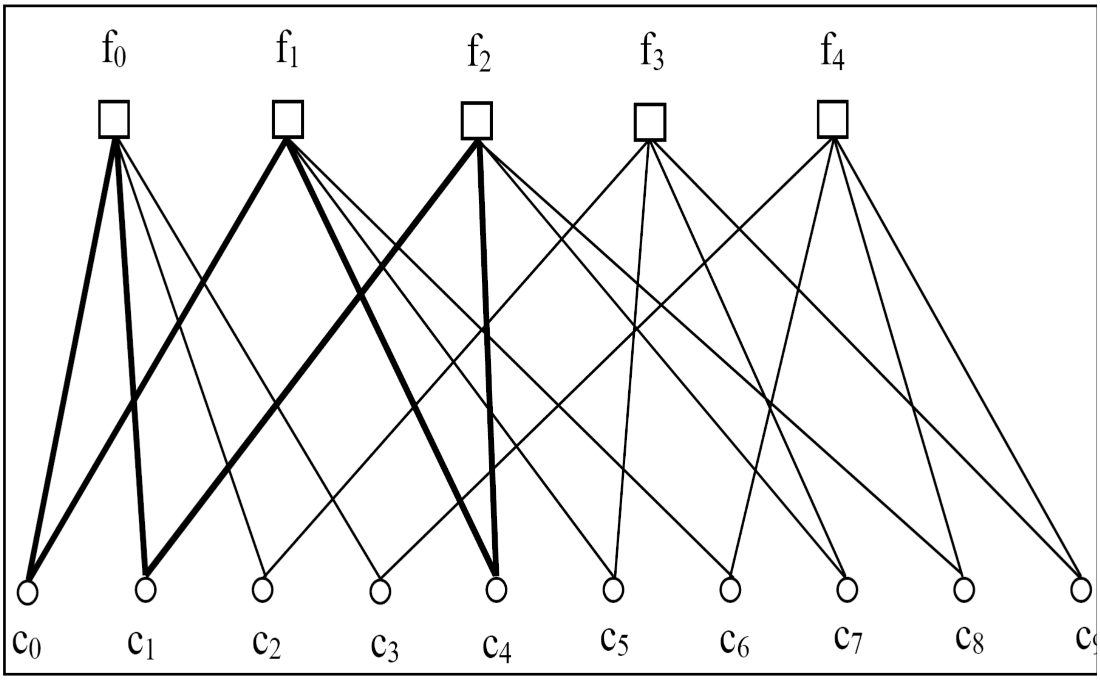


Figure E.6: A Cycle on the Tanner Graph of an LDPC code

The Min-Sum algorithm [i.135] is a reduced complexity variation of the BP algorithm, which typically introduces 0,5 dB to1 dB performance degradation. Analogue implementations of LDPC decoding algorithms have also been gaining momentum recently [i.136].

Annex F: Review of Upper-Layer FEC Codes and Upper-Layer Interleaving

F.1 Characteristics of Small and Large FEC Codes

FEC codes can be categorised into two classes: small and large. Small codes, e.g. RS, are better suited to small FEC blocks, since the computational complexity of their encoding/decoding processes becomes prohibitive for large FEC blocks. On the contrary, large codes, e.g. LDGM, require simpler encoding/decoding operations. As a result, they have higher codec throughputs, and can encode whole files into one or very few large FEC blocks when compared to small codes. Encoding a whole file within very few large FEC blocks is beneficial because for a given overall FEC redundancy level, the error correction capability of a code increases with the block size. Moreover, simpler FEC decoder operations, consequently higher decoder throughputs, are particularly attractive for energy-constrained handheld devices.

The ratio of the decoding throughput of the decoder of a particular flavour of LDGM codes called LDGM Triangle over the RS decoder (speed-up factor) depends on several factors including the FEC block parameters {k, n} used for RS. For example, with F = 1,5 for a file with 20,000 1 024-byte packets and k = 51 (or n = 77) for RS, the speed-up factor is about 1,8, whereas for n = 255 (the maximum number of packets in an RS FEC block with a Galois Field (GF) containing 8-bit elements) the speed-up factor is about 8,3.

Nevertheless, the actual advantage of large codes with respect to the achievable codec throughput has to be assessed taking into account the transmission rates supported by the different systems. For low transmission rates, e.g. the upper limit of 384 kb/s for MBMS and S-DMB, the bottleneck in data transport is the transmission capacity rather than the codec speed. On the other hand, the specific advantage of large codes becomes more relevant for low-end handheld devices with limited processing resources.

Small codes have the advantage of having no block reception overhead whereas large codes do. In general, the percentage reception overhead increases with higher SF values and smaller file sizes. The lower bound of reception overhead for LDGM Triangle is a little over 5 %. Nonetheless, even though RS codes do not have a reception overhead with respect to a single FEC block, they introduce some overhead indirectly; when they are used for protecting large files. Since the computationally intensive GF arithmetic of RS encoding and decoding procedures necessitates splitting up large files into small and more easily manageable FEC blocks, the decoder may not be able to recover a particular file even if the total number of received packets is greater than the original packets. This is better shown in figure F.1, where full file recovery is not possible because fewer than k packets have been received for the third FEC block. This problem is referred to as the coupon collector problem [i.142]. To distinguish the two aforementioned types of reception overheads, the one pertaining to a FEC block is called block reception overhead (ro), and the one pertaining to a file global reception overhead.

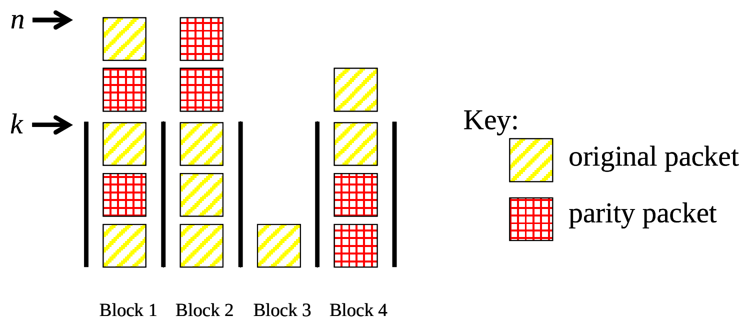


Figure F.1: Demonstration of global reception overhead for RS codes with *k* = 3, *n* = 5

F.2 Common FEC Codes

In the following, common FEC Codes, namely RS, LDGM, Low Density Parity Check (LDPC), and Raptor, are reviewed in more detail. These codes that have been considered for reliable multicast transport in different standards including the IETF RMT WG, MBMS, DVB-H, DVB-SH, and S-DMB.

F.2.1 Reed Solomon (RS)

During encoding, packets are arranged in an array and the RS code is applied column-wise on corresponding RS symbols from each original packet to form parity packets as illustrated in figure F.2. The $n \times k$ matrix **G**, called a generator matrix, is central to both encoding and decoding of RS codes. To create a FEC block, original data symbols represented by a $k \times 1$ input vector **x** are multiplied by the generator matrix according to GF arithmetic rules [i.137], the output is an $n \times 1$ vector **y**. For systematic codes, the first k of the n output symbols are the original data symbols. A key fact to note is that all the original data symbols are involved in the generation of each parity symbol. A RS code with symbol size m bits has the parameters ($n = 2^m - 1$, k). In order to reduce complexity in handheld devices, it is generally agreed that this symbol size should be 8 bits, resulting in GF(28) i.e. a GF with 256 elements; if the multiplication of a particular file size by F is greater than *n*, then the application of FEC to this particular file results in more than one FEC block.

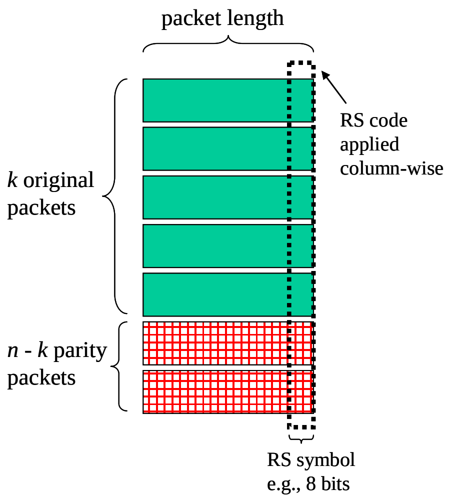


Figure F.2: Concept of symbols and packets in a FEC block

When a decoder receives packets from **y**, it checks if all the original packets have arrived correctly. If they have, it does no work and passes these packets to a higher layer; otherwise, if some original packets are missing, the decoder initiates the recovery process; if it has received at least *k* packets from **y**. The missing original data packets can be recovered by solving the linear system as in equation (F-1), where **y**' is a subset of *k* components of **y** available at the receiver. **G**' is the subset of rows from **G** corresponding to subset **y**'. The amount of decoding work or complexity increases with increasing numbers of lost data packets [i.142].

$$\mathbf{y}' = \mathbf{G}'\mathbf{x} \rightarrow \mathbf{x} = \mathbf{G}'^{-1}\mathbf{y}' \qquad \text{(F-1)}$$

The type of RS codes described so far as are called one-Dimensional (1D) RS. The performance of 1D RS can be improved for large files by using RS product codes or two-Dimensional (2D) RS. To ease complexity, it could be advantageous to have one FEC code implementation in a terminal, thus one RS mother code could be fed appropriate parameters to act as 1D RS or 2D RS, instead of having to switch from RS for small files and to LDGM for large files for example.

With 2D RS, packets are assembled in a 2D matrix to produce row-wise and column-wise parity packets as shown in figure F.3. Each row or column is essentially a single block of 1D RS. In the decoding stage of 2D RS, missing data packets from a given row can be recovered if that particular row receives at least *k* packets or if their respective columns receive at least *k* packets; however, even if both conditions are false, the respective columns may eventually have *k* packets through the decoding of other rows. Decoding of a 2D block starts with rows/columns and then columns/rows before the process iterates; the process terminates if one of the following conditions is satisfied:

- All the original packets have been recovered.
- An iteration does not yield any more recovered packets.

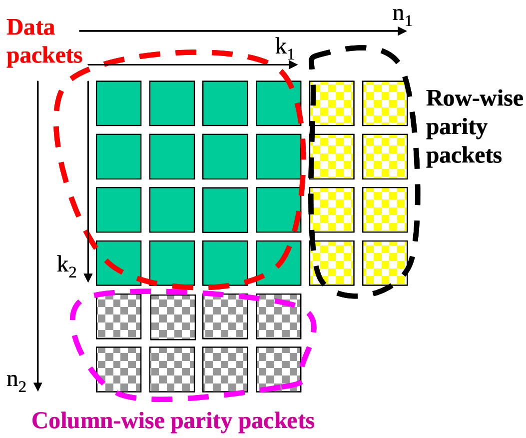


Figure F.3: An example of a 2D RS block

There is also another option for improving 1D RS which could be explored: zigzag RS. This alternative could be similar to the zigzag codes analysed in [i.143]. The motivation behind zigzag RS is to enhance 1D RS by ensuring that adjacent FEC blocks carry information about each other, thus mitigating the coupon collector problem.

F.2.2 Low Density Generator Matrix (LDGM)

LDGM codes are open source FEC codes developed by INRIA. There are three types of LDGM codes: LDGM or LDGM Standard, LDGM Staircase, and LDGM Triangle. These codes differ in the properties of their respective parity check matrices. However, once the parity matrix is created, the encoding and decoding processes for these three codes are the same.

Unlike, RS, the generation of each parity packet does not directly involve all the original packets. This is the root cause of the block reception overhead in large codes. In addition, the generation of a parity packet requires no higher order GF arithmetic, as only the simple XOR-ing of the packets identified by a parity check matrix (see figure F.4), thus allowing higher codec throughputs compared to RS. Each row of the parity check matrix is called a *check node*, which is simply a constraint defining the relationship between specific packets identified "1"s, e.g. in figure F.4, at *c1*, $A \oplus C \oplus P1 = 0$ or in other words, parity packet *P1* is created by XOR-ing packets *A* and *C*.

$$
\begin{array}{cc}
\begin{array}{cccc:ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{P1} & \mathrm{P2} & \mathrm{P3} \end{array} & \\
\left(\begin{array}{cccc:ccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) & \begin{array}{c} \mathrm{c1} \\ \mathrm{c2} \\ \mathrm{c3} \end{array}
\end{array}
$$

Figure F.4: An example of a parity check matrix for LDGM Standard

Decoding takes places iteratively as follows:

1) When a packet arrives at the decoder, it is replaced in all the check nodes.

2) Additional recovered data and parity packets are recovered, as a result of the packet received in step 1 or step 3 (go to step 4 if there are none); a packet is recovered if it is the only one still missing from a particular check node.

3) Replace each recovered packets in all the check nodes (go to step 2).

4) Go to step 1 if there are original packets which are yet to be recovered and if new packets are still arriving at the decoder, otherwise stop decoding.

Table F.1 demonstrates this algorithm by showing the packet(s) recovered upon the reception of each packet at a decoder using the parity check matrix in figure F.4.

Table F.1: An example of iterative decoding with the use of the parity check matrix in figure F.4

Packets received	*A*	*C*	*P2*	*D*
Packets recovered	*A*	*C*, *P1*	*P2*, *B*	*D*, *P3*

In S*tandard LDGM*, all the parity packets are linked to exactly one check node, thus yielding an identity matrix in the second part of the parity check matrix.

In *LDGM Staircase*, each parity packet is linked to one or more check nodes in such a way that a the identity matrix in the second half of the parity check matrix form LDGM is now replaced by a staircase matrix as depicted in figure F.5. Since parity packets are now protected, LDGM significantly reduces the block reception overhead compared to LDGM Standard, but at the expense of codec throughput. For example, packet *P2* can be recovered through packets *A*, *B*, and *P1*; this allows *P2* to be used in check node 3 if required. On the contrary, with LDGM, an erased parity packet cannot be recovered unless all original packets in the associated check node are known, but in that case the parity packet is not required.

$$
\begin{array}{c}
\begin{array}{ccccc:ccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{P1} & \mathrm{P2} & \mathrm{P3} & \mathrm{P4} & \mathrm{P5} \end{array} \\
\left(\begin{array}{cccc:ccccc}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1
\end{array}\right)
\begin{array}{c} \mathrm{c1} \\ \mathrm{c2} \\ \mathrm{c3} \\ \mathrm{c4} \\ \mathrm{c5} \end{array}
\end{array}
$$

Figure F.5: An example of a parity check matrix for LDGM Staircase

The *LDGM Triangle* parity check matrix is formed by the addition of "1"s to the empty triangle beneath the staircase diagonal in LDGM Staircase (see figure F.5). This variation leads to a performance increase, i.e. lower block reception overhead, compared to LDGM Staircase for $F < 2.5$; on the other hand, the encoding is slightly slower since there are more "1"s per row, which lead to more XOR operations.

F.2.3 Raptor Codes

Raptor codes [i.144] and [i.145 were designed by researchers from Digital Fountain, Inc., by building on their earlier work on Luby Transform (LT) [i.146] and Tornado codes [i.141]. Although Raptor codes are protected by patents, they have been selected for use in both MBMS and DVB-H, and have been under consideration in DVB-SH. Apart form having a noticeably low block reception overhead in the order of 1 % to 2 %, Raptor codes are *expandable*, meaning that an encoder can generate as many parity packets as possible on demand. Flexibility is thus higher, since *n* does not have to be fixed beforehand, which is the case with RS, LDGM, and LDPC codes. For example, if a multicast session involved the transmission of on demand parity packets (parity packets which were not transmitted in the first round), the use of LDGM and RS codes would be inefficient compared to Raptor; LDGM and RS codes would require transmission of earlier packets upon exhaustion of the on demand parity packets. A Raptor encoder, instead, would simply generate new parity packets, thus avoiding repetition or inefficiency. For this reason, Raptor codes are also referred to as rate less codes, i.e. the code rate approaches zero as $n >> k$.

The encoding process in Raptor codes occurs in two stages:

1) The original k packets are encoded into intermediate packets using a block code which guarantees that first k packets in the output of stage two are source packets.

2) The output from stage one is passed to an LT encoder which works as follows:

   a) A degree *d* for an output packet is chosen based on a given distribution (robust soliton distribution in the case of the standard LT code [i.146]).

   b) A particular output packet is created by XOR-ing *d* distinct packets which are selected uniformly at random from the intermediate packets.

As is the case with other systematic codes, the decoder does no work if all the original data packets arrive intact. However, if some data packets are missing, the decoder attempts to generate sufficient intermediate packets from the available packets so that it can then generate the missing data packets.

F.2.4 Other FEC Codes

- ***Simple FEC codes [i.147]:*** entail simple XOR operations. A group of *k* data packets are XOR-ed together to generate a single parity packet; this allows a receiver to recover the original *k* data packets as long as it receives any *k* of the $k + 1$ transmitted packets. Packets can even be grouped in a 2D matrix with each row and column having a single parity packet. However, as admitted by the authors in [i.147], these simple FEC codes are for very low loss conditions which are quite rare in satellite environments.

- ***Online codes [i.148]:*** are rate less codes as is the case with LT and Raptor codes. The encoding and decoding processes of online and LT codes have some similarities. The designers of online codes even claim that their codes are more computationally efficient than LT codes. However, a study comparing the two codes does not exist. It is worth pointing out that LT codes appear prominently in various standardisation activities as a component of Raptor codes, and are protected by patents. On the other hand online codes only appear in a few publications.

- ***Convolutional codes [i.149]:*** are not rate less. A convolutional encoder acts on a continuous stream of packets and its output depends on the current input and one or more of previous inputs. A convolutional code is described by the parameters (*n*, *k*, and *m*). For every *k* information packets, *n* output packets, called a code group, are generated. The parameter *m* denotes the constraint length, which shows how many previous code groups affect one redundant packet. Reference [i.149] presents a study comparing convolutional and RS codes; convolutional codes are shown to outperform RS codes in terms of bandwidth efficiency. Nevertheless, the work only considers low packet loss rates in the region of 1 % which are more relevant for the Internet than a satellite environment. Studies on how convolutional codes compare with large PL-FEC codes such as LDGM are also yet missing.

- ***LDPC Copper codes:*** can be considered a flavour of LDGM codes since their parity check matrix can also be used to directly provide information for the generation of parity packets. NEC, Inc., proposed LDPC Copper codes for protecting MBMS content, but Raptor codes were selected ahead of these codes and RS due to their superior performance when considering the aggregate effect of both block reception overhead and decoding complexity. Although, NEC, Inc., showed that LDPC Copper codes outperform RS codes under specific settings, their comparison with LDGM and Raptor is yet to be made.

It is noted that comparisons of FEC codes typically consider uniform packet losses or burst losses where the burst length consists of a few packets. However, in LMS channel models, which are of interest in this study, error burst lengths of many packets can occur and this may affect the balance of power between the FEC codes.

F.3 Interleaving

The motivation for interleaving information at packet-level is similar as in the physical layer. Bursts of erasures may occur as a result of large shadowing/blocking events or due to upper layer packet losses. Similar to FEC, interleaving is also highly scalable since no user feedback is required. Interleaving can only be effective when combined with FEC (except in audio/video streaming applications where interleaving improves the performance of error concealment techniques, and hence the perceived media quality). Nonetheless, interleaving is applied carefully in streaming applications because it adds latency. This latency is of an order that does not pose major concerns for file download services.

In general, there are two types of interleavers at the physical layer: *block* and *convolutional* [i.150]. These are also applicable at packet-level. In addition, *random* interleaving is also recommended for LDGM at packet-level. There is no easy answer with regards to the best interleaver and parameters: this depends on the packet loss characteristics, which in turn depend on the specific system and transmission environment. The actual characterisation of data loss at packet level has to take into account the impact of the physical and link layers of the radio interface. The three aforementioned interleaving techniques are now described in turn.

F.3.1 Block Interleaving

Block interleaving is more relevant to RS since this code is likely to have several FEC blocks. A critical parameter is the interleaving depth or degree, which is defined as the number of blocks spanned by packets from a certain FEC block when spread; its value is two or greater. In general, the performance improves with increasing interleaving depth. If a file has four FEC blocks and all are involved in interleaving, there are two possibilities:

- interleaving depth = 2: send the first packet from block one, then the first from block two, followed by the second packet from block one, and so on; do the same to blocks three and four;
- interleaving depth = 4 (or maximum depth): send the packets in this order - first packet from block one, first packet from block two, first packet from block three, first packet from block four, second packet from block one, second packet from block two, and so on.

The example in figure F.6 shows the benefits of interleaving: each FEC block can only recover from a maximum of two erasures, but interleaving spreads the three erasures into one for FEC block *A*, and two for FEC block *B*, thus enabling full recovery.

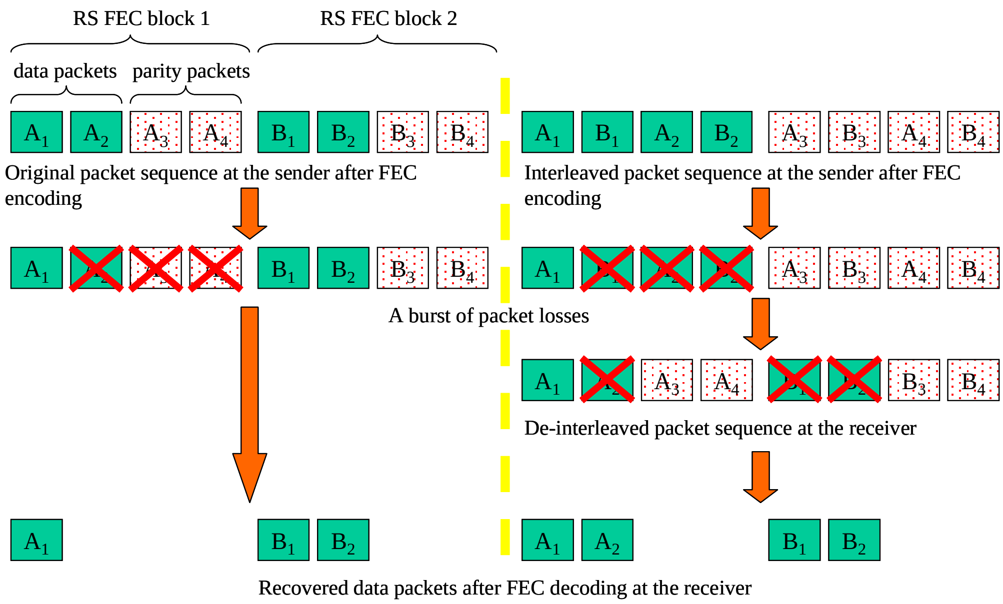


Figure F.6: An illustration of the benefits of block interleaving in the presence of a burst of erasures

F.3.2 Convolutional Interleaving

Convolutional interleaving is applicable to both small and large codes. A convolutional interleaver takes a packet stream at its input and ensures that there are a constant number of packets between packets, which were adjacent in the original stream. For example, an original packet sequence 0, 1, 2, 3, 4, 5, 6, 7, 8 could be **0**, -3, -6, -9, 4, **1**, -2, -5, 8, 5, **2**, -1, etc., after interleaving which results in once adjacent packets being separated by four packets.

F.3.3 Random Interleaving

Random interleaving takes an original sequence of packets and randomises it before transmission. This can work for one FEC block, usually the case with large codes, or many FEC blocks, which is the norm for RS. It is possible to combine block interleaving and random interleaving for RS. On this occasion, it is possible to have an interleaving depth of one so that the packet sequence is randomised per FEC block. An interleaving depth of two results in the packet sequence being randomised per two FEC blocks as illustrated in figure F.7.

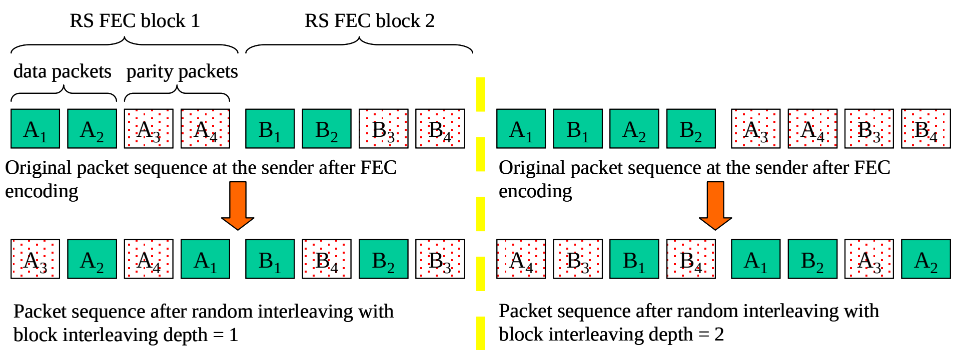


Figure F.7: An example of integrated random and block interleaving with block interleaving depths one and two

Annex G: Review of Mobile WiMAX

G.1 Physical Layer Aspects

Multiple access in Mobile WiMAX is OFDMA based. The OFDMA symbol structure consists of three types of sub-carriers:

a) Data sub-carriers for data transmission.

b) Pilot sub-carriers for estimation and synchronization purposes.

c) Null sub-carriers for no transmission; used for guard bands and DC carriers.

Active (data and pilot) sub-carriers are grouped into subsets of sub-carriers (sub channels). The WiMAX OFDMA physical layer supports sub-channelization in both DL and UL. The minimum frequency-time resource unit of sub-channelization is one slot, which is equal to 48 data tones (sub-carriers).

There are two types of sub-carrier permutations for sub-channelization; diversity and contiguous. The diversity permutation draws sub-carriers pseudo-randomly to form a sub-channel. It provides frequency diversity and inter-cell interference averaging. The diversity permutations include DL Fully Used Sub-Carrier (FUSC), DL Partially Used Sub-Carrier (PUSC) and UL PUSC and additional optional permutations. With DL PUSC, for each pair of OFDM symbols, the available or usable sub-carriers are grouped into clusters containing 14 contiguous sub-carriers per symbol period, with pilot and data allocations in each cluster in the even and odd symbols.

Interleaving is used to form groups of clusters such that each group is made up of clusters that are distributed throughout the sub-carrier space. A sub-channel in a group contains 2 clusters and is made up of 48 data sub-carriers and 8 pilot subcarriers. The data sub-carriers in each group are further permutated to generate sub channels within the group. The data sub-carriers in the cluster are distributed to multiple sub-channels. Analogous to the cluster structure for DL, a tile structure is defined for the UL PUSC.

The available sub-carrier space is split into tiles and 6 tiles, chosen from across the entire spectrum by means of a re-arranging/permutation scheme, are grouped together to form a slot. The slot comprises 48 data sub-carriers and 24 pilot sub-carriers in 3 OFDM symbols. The contiguous permutation groups a block of contiguous sub-carriers to form a sub channel. The contiguous permutations include DL ACM and UL ACM, and have the same structure. A bin consists of 9 contiguous sub-carriers in a symbol, with 8 assigned for data and one assigned for a pilot. A slot in ACM is defined as a collection of bins of the type ($N \times M = 6$), where N is the number of contiguous bins and M is the number of contiguous symbols. Thus the allowed combinations are [(6 bins, 1 symbol), (3 bins, 2 symbols), (2 bins, 3 symbols), (1 bin, 6 symbols)]. ACM permutation enables multi-user diversity by choosing the sub-channel with the best frequency response.

In general, diversity sub-carrier permutations perform well in mobile applications while contiguous sub-carrier permutations are well suited for fixed, portable, or low mobility environments. These options enable the system designer to trade-off mobility for throughput.

G.1.1 Scalable OFDMA

The OFDMA mode specified in IEEE 802.16e-2005 [i.335] is based on the concept of *Scalable OFDMA* (S-OFDMA). S-OFDMA supports a wide range of bandwidths to flexibly address the need for various spectrum allocation and usage model requirements. The scalability is supported by adjusting the FFT size while fixing the sub-carrier frequency spacing at 10,94 kHz. Since the resource unit sub-carrier bandwidth and symbol duration is fixed, the impact to higher layers is minimal when scaling the bandwidth.

The S-OFDMA parameters are provided in table G.1. The system bandwidths for two of the initial planned profiles developed by the WiMAX Forum Technical Working Group for Release-1 are 5 and 10 MHz (highlighted in table G.1).

Table G.1: OFDMA Scalability Parameters

Parameters	Values			
System Channel Bandwidth (MHz)	1.25	5	10	20
Sampling Frequency (F_p in MHz)	1.4	5.6	11.2	22.4
FFT Size (N_{FFT})	128	512	1024	2048
Number of Sub-Channels	2	8	16	32
Sub-Carrier Frequency Spacing	10.94 kHz			
Useful Symbol Time ($T_b = 1/f$)	91.4 microseconds			
Guard Time ($T_g = T_b/8$)	11.4 microseconds			
OFDMA Symbol Duration ($T_s = T_b + T_g$)	102.9 microseconds			
Number of OFDMA Symbols (5 ms Frame)	48			

G.1.2 TDD Frame Structure

IEEE 802.16e [i.335] supports TDD and Full and Half-Duplex FDD operation; however the initial release of Mobile WiMAX certification profiles only included TDD. With ongoing releases, FDD profiles will be considered by the WiMAX Forum to address specific market opportunities where local spectrum regulatory requirements either prohibit TDD or are more suitable for FDD deployments. To counter interference issues, TDD requires system-wide synchronization; nevertheless, however it provides some important advantages relative to FDD:

- Enables adjustment of the downlink/uplink ratio to efficiently support asymmetric downlink/uplink traffic, while with FDD, downlink and uplink always have fixed and generally, equal DL and UL bandwidths.
- Provides channel reciprocity for better support of link adaptation, MIMO and other closed loop advanced antenna technologies.
- Only requires a single channel for both downlink and uplink providing greater flexibility for adaptation to varied global spectrum allocations.
- Transceiver designs for TDD implementations are less complex and therefore less expensive.

Figure G.1 illustrates the OFDM frame structure for the TDD implementation. Each frame is divided into DL and UL sub-frames separated by Transmit/Receive and Receive/Transmit Transition Gaps (TTG and RTG, respectively) to prevent DL and UL transmission collisions. In a frame, the following control information is used to ensure optimal system operation:

- **Preamble:** used for synchronization, is the first OFDM symbol of the frame.
- **Frame Control Header (FCH):** The FCH follows the preamble. It provides the frame configuration information such as MAP message length and coding scheme and usable sub-channels.
- **DL-MAP and UL-MAP:** The DL-MAP and UL-MAP provide sub-channel allocation and other control information for the DL and UL sub-frames respectively.
- **UL Ranging:** The UL ranging sub-channel is allocated for Mobile Stations (MS) to perform closed-loop time, frequency, and power adjustment as well as bandwidth requests.
- **UL CQICH:** The UL CQICH channel is allocated for the MS to feedback channel-state information.
- **UL ACK:** The UL ACK is allocated for the MS to feedback DL HARQ acknowledge.

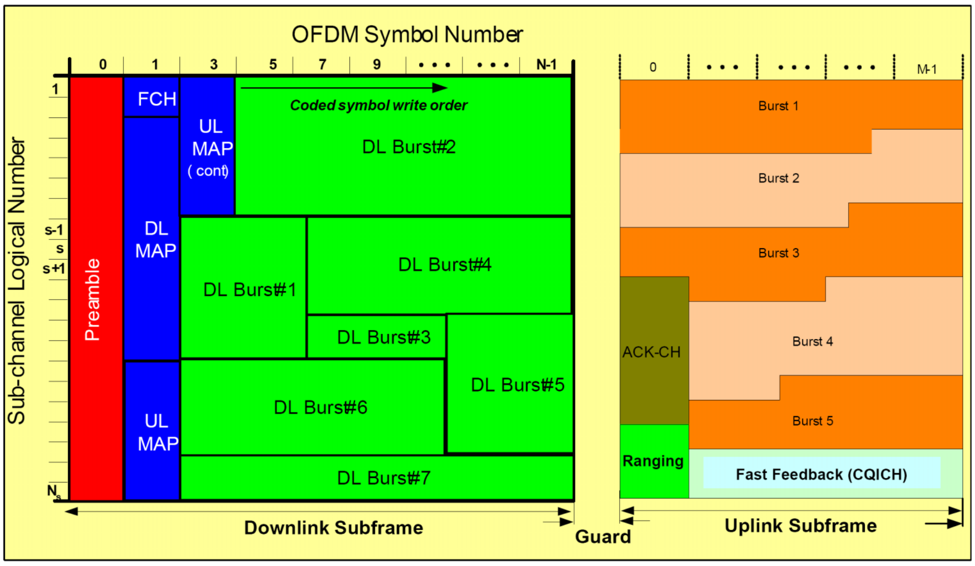


Figure G.1: WiMAX OFDM Frame Structure

G.1.3 Advanced Physical Layer Features

The WiMAX standards specify a number of advanced physical layer techniques, such as ACM, HARQ and Fast Channel Feedback (CQICH) that enhance system capacity and coverage in mobility environments. QPSK, 16-QAM are mandatory for both UL and DL, whereas 64-QAM is only mandatory for the DL and optional for UL. Both Convolutional Codes (CC) and Convolutional Turbo Codes (CTC) with variable code rate and repetition coding are supported. Block Turbo Code and Low Density Parity Check Code (LDPC) are supported as optional features.

Table G.2 summarizes the coding and modulation schemes supported in the mobile WiMAX (optional features shown in Italics).

Table G.2: Supported coding and modulation formats

		DL	UL
Modulation		QPSK, 16QAM, 64QAM	QPSK,16QAM, *64QAM*
Code Rate	CC	1/2, 2/3, 3/4, 5/6	1/2, 2/3, *5/6*
	CTC	1/2, 2/3, 3/4, 5/6	1/2, 2/3, *5/6*
	Repetition	x2, x4, x6	x2, x4, x6

The combinations of various modulations and code rates provide a fine resolution of data rates as shown in table G.3 which shows the data rates for 5 MHz and 10 MHz channels with Partially Used Sub-Carrier (PUSC) sub-channels. The frame duration is 5 milliseconds. Each frame has 48 OFDM symbols, with 44 OFDM symbols available for data transmission. The highlighted values indicate data rates for optional 64-QAM in the UL.

The base station scheduler determines the appropriate data rate (or burst profile) for each burst allocation based on parameters such as the buffer size, and the channel propagation conditions at the receiver. A Channel Quality Indicator (CQI) channel is utilized to provide channel-state information from the user terminals to the base station scheduler. Relevant channel-state information can be fed back by the CQICH including: Physical CINR, effective CINR, MIMO mode selection and frequency selective sub-channel selection. With TDD implementations, link adaptation can also take advantage of channel reciprocity to provide a more accurate measure of the channel condition.

HARQ is enabled using N channel "Stop and Wait" protocol which provides fast response to packet errors and improves cell edge coverage. Chase Combining and optionally, Incremental Redundancy are supported to further improve the reliability of the retransmission. A dedicated ACK channel is also provided in the uplink for HARQ ACK/NACK signalling. Multi-channel HARQ operation is supported. Multi-channel stop-and-wait ARQ with a small number of channels is an efficient, simple protocol that minimizes the memory required for HARQ and stalling. The standard also provides signalling to allow fully asynchronous operation. The asynchronous operation allows variable delay between retransmissions which gives more flexibility to the scheduler at the cost of additional overhead for each retransmission allocation. HARQ combined together with CQICH and AMC provides robust link adaptation in mobile environments at vehicular speeds in excess of 120 km/hr.

Table G.3: Mobile WiMAX data rates with PUSC sub-channel

Parameter	Downlink	Uplink	Downlink	Uplink
System Bandwidth	5 MHz		10 MHz	
FFT Size	512		1024	
Null Sub-Carriers	92	104	184	184
Pilot Sub-Carriers	60	136	120	280
Data Sub-Carriers	360	272	720	560
Sub-Channels	15	17	30	35
Symbol Period, T_S	102.9 microseconds			
Frame Duration	5 milliseconds			
OFDM Symbols/Frame	48			
Data OFDM Symbols	44			

Mod.	Code Rate	5 MHz Channel Downlink Rate, Mbps	5 MHz Channel Uplink Rate, Mbps	10 MHz Channel Downlink Rate, Mbps	10 MHz Channel Uplink Rate, Mbps
QPSK	1/2 CTC, 6x	0.53	0.38	1.06	0.78
	1/2 CTC, 4x	0.79	0.57	1.58	1.18
	1/2 CTC, 2x	1.58	1.14	3.17	2.35
	1/2 CTC, 1x	3.17	2.28	6.34	4.70
	3/4 CTC	4.75	3.43	9.50	7.06
16QAM	1/2 CTC	6.34	4.57	12.67	9.41
	3/4 CTC	9.50	6.85	19.01	14.11
64QAM	1/2 CTC	9.50	*6.85*	19.01	*14.11*
	2/3 CTC	12.67	*9.14*	25.34	*18.82*
	3/4 CTC	14.26	*10.28*	28.51	*21.17*
	5/6 CTC	15.84	*11.42*	31.68	*23.52*

G.1.4 MIMO Techniques

Mobile WiMAX supports a wide range of MIMO techniques, namely:

- Beam-Forming (BF).
- Space-Time Coding (STC).
- Spatial Multiplexing (SM).
- Collaborative SM between two users in the uplink

Table G.4 summarises the different MIMO configurations in the DL and UL.

Table G.4: MIMO configuration options in mobile WiMAX

Link	Beamforming	Space Time Coding	Spatial Multiplexing
DL	Nt≥2, Nr≥1	Nt=2, Nr≥1 Matrix A	Nt=2, Nr≥2 Matrix B, vertical encoding
UL	Nt ≥1, Nr≥2	N/A	Nt=1, Nr≥2 Two-user collaborative SM

The standard also supports adaptive switching between these options, based on the instantaneous channel conditions, in order to maximize the benefit of the MIMO techniques. For instance, SM improves peak throughput. However, when channel conditions are poor, the Packet Error Rate (PER) can be high and thus the coverage area where target PER is met may be limited. STC on the other hand provides large coverage regardless of the channel condition but does not improve the peak data rate. Mobile WiMAX supports *Adaptive MIMO Switching* (AMS) between multiple MIMO modes to maximize spectral efficiency with no reduction in coverage area.

Table G.5 provides a summary of the theoretical peak data rates for various DL/UL ratios assuming a 10 MHz channel bandwidth, 5 ms frame duration with 44 OFDM data symbols (out of 48 total OFDM symbols) and PUSC sub-channelization. With 2 × 2 MIMO, the DL user and sector peak data rate are theoretically doubled. The maximum DL peak data rate is 63,36 Mbps when all the data symbols are dedicated to DL. With UL collaborative SM, the UL sector peak data rate is doubled while the user peak data rate is unchanged. The UL user peak data rate and sector peak data rate are 14,11 Mbps and 28,22 Mbps respectively when all the data symbols are dedicated to UL. By applying different DL/UL ratio, the bandwidth can by adjusted between DL and UL to accommodate different traffic patterns. It should be noted that the extreme cases such as all DL and all UL partition are rarely used. The WiMAX profile supports DL/UL ratios ranging from 3:1 to 1:1 to accommodate different traffic profiles. The resulting peak data rates that will typically be encountered are in between the two extreme cases.

Table G.5: Achievable data rates with different MIMO configurations

DL/UL Ratio			1:0	3:1	2:1	3:2	1:1	0:1
User Peak Rate (Mbps)	SIMO (1x2)	DL	31.68	23.04	20.16	18.72	15.84	0
		UL	0	4.03	5.04	6.05	7.06	14.11
	MIMO (2x2)	DL	63.36	46.08	40.32	37.44	31.68	0
		UL	0	4.03	5.04	6.05	7.06	14.11
Sector Peak Rate (Mbps)	SIMO (1x2)	DL	31.68	23.04	20.16	18.72	15.84	0
		UL	0	4.03	5.04	6.05	7.06	14.11
	MIMO (2x2)	DL	63.36	46.08	40.32	37.44	31.68	0
		UL	0	8.06	10.08	12.10	14.12	28.22

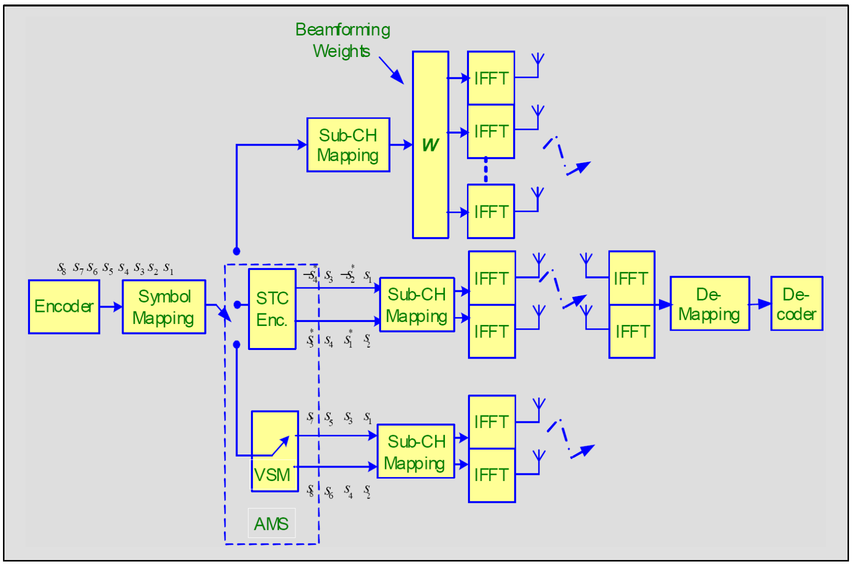


Figure G.2: Adaptive MIMO switching

G.2 MAC Layer

The MAC layer of 802.16 is based on the DOCSIS standard and can support bursty data traffic with high peak rate demand while simultaneously supporting streaming video and latency-sensitive voice traffic over the same channel. The resource allocated to one terminal by the MAC scheduler can vary from a single time slot to the entire frame, thus providing a very large dynamic range of throughput to a specific user terminal at any given time. Furthermore, since the resource allocation information is conveyed in the MAP messages at the beginning of each frame, the scheduler can effectively change the resource allocation on a frame-by-frame basis to adapt to the bursty nature of the traffic.

G.2.1 Quality of Service (QoS) Support

Mobile WiMAX can meet QoS requirements for a wide range of data services and applications due to the high achievable throughputs, asymmetric downlink/uplink capability, fine resource granularity and a flexible resource allocation mechanism. QoS is provided via service flows as illustrated in figure G.3. This is a unidirectional flow of packets that is provided with a particular set of QoS parameters. Before providing a certain type of data service, the base station and user-terminal first establish a unidirectional logical link between the peer MACs called a connection. The outbound MAC then associates packets traversing the MAC interface into a service flow to be delivered over the connection. The QoS parameters associated with the service flow define the transmission ordering and scheduling on the air interface. The connection-oriented QoS therefore, can provide accurate control over the air interface. Since the air interface is usually the bottleneck, the connection-oriented QoS can effectively enable the end-to-end QoS control. The service flow parameters can be dynamically managed through MAC messages to accommodate the dynamic service demand. The service flow based QoS mechanism applies to both DL and UL to provide improved QoS in both directions. Table G.6 summarises the data services and applications (with different QoS requirements) supported by Mobile WiMAX.

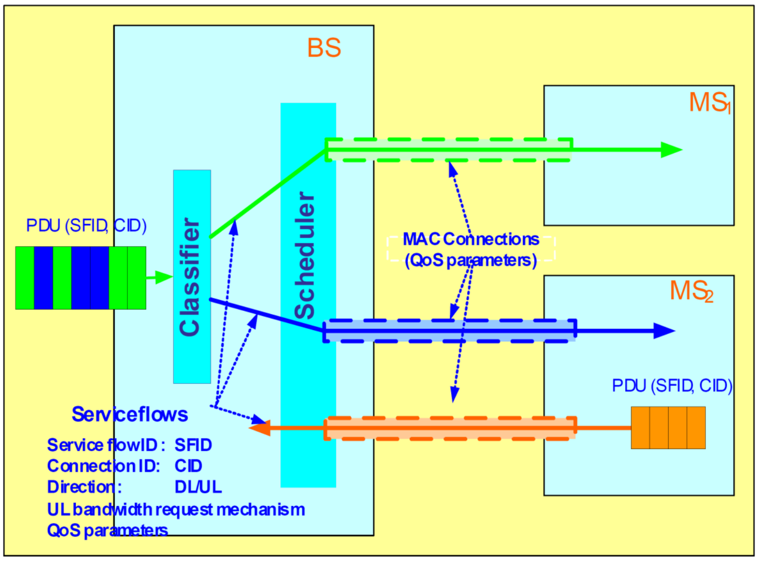


Figure G.3: QoS support

Table G.6: Mobile WiMAX applications and QoS

QoS Category	Applications	QoS Specifications
UGS Unsolicited Grant Service	VoIP	• Maximum Sustained Rate • Maximum Latency Tolerance • Jitter Tolerance
rtPS Real-Time Polling Service	Streaming Audio or Video	• Minimum Reserved Rate • Maximum Sustained Rate • Maximum Latency Tolerance • Traffic Priority
ErtPS Extended Real-Time Polling Service	Voice with Activity Detection (VoIP)	• Minimum Reserved Rate • Maximum Sustained Rate • Maximum Latency Tolerance • Jitter Tolerance • Traffic Priority
nrtPS Non-Real-Time Polling Service	File Transfer Protocol (FTP)	• Minimum Reserved Rate • Maximum Sustained Rate • Traffic Priority
BE Best-Effort Service	Data Transfer, Web Browsing, etc.	• Maximum Sustained Rate • Traffic Priority

G.2.2 MAC Scheduling Service

The MAC scheduling service, which is designed to deliver efficiently broadband data services including voice, data, and video over time varying broadband wireless channel, has the following features:

- **Fast Data Scheduler:** The MAC scheduler efficiently allocates available resources in response to bursty data traffic and time-varying channel conditions. The scheduler is located at each base station to enable rapid response to traffic requirements and channel conditions. The data packets are associated to service flows with well defined QoS parameters in the MAC layer so that the scheduler can correctly determine the packet transmission ordering over the air interface. The CQICH channel provides fast channel information feedback to enable the scheduler to choose the appropriate coding and modulation for each allocation. The adaptive modulation/coding combined with HARQ provide robust transmission over the time varying channel.

- **Scheduling for both DL and UL:** The scheduling service is provided for both DL and UL traffic. In order for the MAC scheduler to make an efficient resource allocation and provide the desired QoS in the UL, the UL feeds back accurate and timely information as to the traffic conditions and QoS requirements. Multiple uplink bandwidth request mechanisms, such as bandwidth request through ranging channel, piggyback request and polling are designed to support UL bandwidth requests. The UL service flow defines the feedback mechanism for each uplink connection to ensure predictable UL scheduler behaviour. Furthermore, with orthogonal UL sub-channels, there is no intra-cell interference. UL scheduling can allocate resource more efficiently and better enforce QoS.

- **Dynamic Resource Allocation:** The MAC supports frequency-time resource allocation in both DL and UL on a per-frame basis. The resource allocation is delivered in MAP messages at the beginning of each frame. Therefore, the resource allocation can be changed frame-by-frame in response to traffic and channel conditions. Additionally, the amount of resource in each allocation can range from one slot to the entire frame. The fast and fine granular resource allocation allows superior QoS for data traffic.

- **QoS Oriented:** The MAC scheduler handles data transport on a connection-by-connection basis. Each connection is associated with a single data service with a set of QoS parameters that quantify the aspects of its behaviour. With the ability to dynamically allocate resources in both DL and UL, the scheduler can provide superior QoS for both DL and UL traffic. Particularly with uplink scheduling - the uplink resource is more efficiently allocated, performance is more predictable, and QoS is better enforced.

Frequency Selective Scheduling: The scheduler can operate on different types of sub-channels. For frequency-diverse sub-channels such as PUSC permutation, where sub-carriers in the sub-channels are pseudo-randomly distributed across the bandwidth, sub-channels are of similar quality. Frequency-diversity scheduling can support a QoS with fine granularity and flexible time-frequency resource scheduling. With contiguous permutation such as AMC permutation, the sub-channels may experience different attenuation. The frequency-selective scheduling can allocate mobile users to their corresponding strongest sub-channels. The frequency-selective scheduling can enhance system capacity with a moderate increase in CQI overhead in the UL.

G.3 Mobility Management

G.3.1 Power Saving Features

The standard supports Sleep Mode and Idle Mode to enable power-efficient MS operation. Sleep Mode is a state in which the MS conducts pre-negotiated periods of absence from the Serving Base Station air interface. These periods are characterized by the unavailability of the MS, as observed from the Serving Base Station, to DL or UL traffic. Sleep Mode is intended to minimize MS power usage and minimize the usage of the Serving Base Station air interface resources. The Sleep Mode also provides flexibility for the MS to scan other base stations to collect information to assist handoff.

Idle Mode provides a mechanism for the MS to become periodically available for DL broadcast traffic messaging without registration at a specific base station as the MS traverses an air link environment populated by multiple base stations. Idle Mode benefits the MS by removing the requirement for handoff and other normal operations and benefits the network and base station by eliminating air interface and network handoff traffic from essentially inactive MSs while still providing a simple and timely method (paging) for alerting the MS about pending DL traffic. It also supports seamless handoff to enable the MS to switch from one base station to another at vehicular speeds without interrupting the connection.

G.3.2 Handoff

There are three handoff methods supported within the 802.16e standard - Hard HandOff (HHO), Fast Base Station Switching (FBSS) and Macro Diversity Handover (MDHO). HHO is mandatory while FBSS and MDHO are two optional modes. The WiMAX Forum has developed several techniques for optimizing hard handoff within the framework of the 802.16e standard. These improvements have been developed with the goal of keeping Layer 2 handoff delays to less than 50 milliseconds.

When FBSS is supported, the MS and BS maintain a list of BSs that are involved in FBSS with the MS. This set is called an Active Set. In FBSS, the MS continuously monitors the base stations in the Active Set. Among the BSs in the Active Set, an Anchor BS is defined. When operating in FBSS, the MS only communicates with the Anchor BS for uplink and downlink messages including management and traffic connections. Transition from one Anchor BS to another (i.e. BS switching) is performed without invocation of explicit HO signalling messages. Anchor update procedures are enabled by communicating signal strength of the serving BS via the CQI channel. A FBSS handover begins with a decision by an MS to receive or transmit data from the Anchor BS that may change within the active set. The MS scans the neighbour BSs and selects those that are suitable to be included in the active set. The MS reports the selected BSs and the active set update procedure are performed by the BS and MS. The MS continuously monitors the signal strength of the BSs that are in the active set and selects one BS from the set to be the Anchor BS. The MS reports the selected Anchor BS on CQICH or MS initiated HO request message. An important requirement of FBSS is that the data is simultaneously transmitted to all members of an active set of BSs that are able to serve the MS.

For MSs and BSs that support MDHO, the MS and BS maintain an active set of BSs that are involved in MDHO with the MS. Among the BSs in the active set, an Anchor BS is defined. The regular mode of operation refers to a particular case of MDHO with the active set consisting of a single BS. When operating in MDHO, the MS communicates with all BSs in the active set of uplink and downlink unicast messages and traffic. A MDHO begins when a MS decides to transmit or receive unicast messages and traffic from multiple BSs in the same time interval. For downlink MDHO, two or more BSs provide synchronized transmission of MS downlink data such that diversity combining is performed at the MS. For uplink MDHO, the transmission from a MS is received by multiple BSs where selection diversity of the information received is performed.

G.4 Security

Mobile WiMAX supports best in class security features by adopting the best technologies available today. Support exists for mutual device/user authentication, flexible key management protocol, strong traffic encryption, control and management plane message protection and security protocol optimizations for fast handovers. The usage aspects of the security features are:

- **Key Management Protocol:** Privacy and Key Management Protocol Version 2 (PKMv2) is the basis of Mobile WiMAX security as defined in 802.16e. This protocol manages the MAC security using PKM-REQ/RSP (Public Key Management Request/Response messages) PKM EAP authentication, Traffic Encryption Control, Handover Key Exchange and Multicast/Broadcast security messages all are based on this protocol.

- **Device/User Authentication:** Mobile WiMAX supports Device and User Authentication using IETF EAP protocol by providing support for credentials that are SIM-based, USIM-based or Digital Certificate or **Usernames/Password-based:** Corresponding EAP-SIM, EAP-AKA, EAP-TLS or EAP-MSCHAPv2 authentication methods are supported through the EAP protocol. Key deriving methods are the only EAP methods supported.

- **Traffic Encryption:** AES-CCM is the cipher used for protecting all the user data over the Mobile WiMAX MAC interface. The keys used for driving the cipher are generated from the EAP authentication. A Traffic Encryption State machine that has a periodic key (TEK) refresh mechanism enables sustained transition of keys to further improve protection.

- **Control Message Protection:** Control data is protected using AES based CMAC, or MD5-based HMAC schemes.

- **Fast Handover Support:** A 3-way Handshake scheme is supported by Mobile WiMAX to optimize the re-authentication mechanisms for supporting fast handovers. This mechanism is also useful to prevent any man-in-the-middle-attacks.

G.5 Multicast and Broadcast Service (MBS)

Multicast and Broadcast Service (MBS) supported by Mobile WiMAX combines the best features of DVB-H, MediaFLO and 3GPP E-UTRA and satisfies the following requirements:

- High data rate and coverage using a Single Frequency Network (SFN).
- Flexible allocation of radio resources.
- Low MS power consumption.
- Support of data-casting in addition to audio and video streams.
- Low channel switching time.

The Mobile WiMAX Release-1 profile defines a toolbox for initial MBS service delivery. The MBS service can be supported by either constructing a separate MBS zone in the DL frame along with unicast service (embedded MBS) or the whole frame can be dedicated to MBS (DL only) for standalone broadcast service.

Figure G.4 shows the DL/UL zone construction when a mix of unicast and broadcast service are supported. The MBS zone supports multi-BS MBS mode using Single Frequency Network (SFN) operation and flexible duration of MBS zones permits scalable assignment of radio resources to MBS traffic. It may be noted that multiple MBS zones are also feasible. There is one MBS zone MAP IE descriptor per MBS zone. The MS accesses the DL MAP to initially identify MBS zones and locations of the associated MBS MAPs in each zone. The MS can then subsequently read the MBS MAPs without reference to DL MAP unless synchronization to MBS MAP is lost. The MBS MAP IE specifies MBS zone PHY configuration and defines the location of each MBS zone via the OFDMA Symbol Offset parameter. The MBS MAP is located at the 1st sub-channel of the 1st OFDM symbol of the associated MBS zone. The multi-BS MBS does not require the MS be registered to any base station. MBS can be accessed when MS in Idle mode to allow low MS power consumption. The flexibility of Mobile WiMAX to support integrated MBS and unicast services enables a broader range of applications.

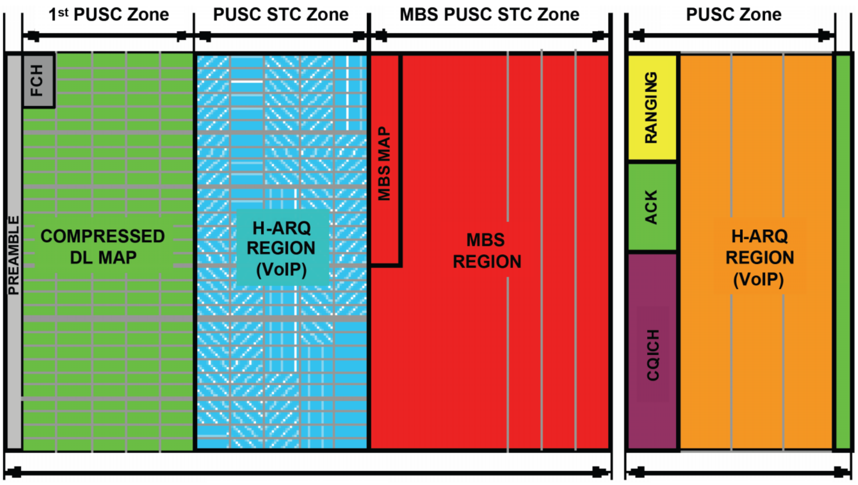


Figure G.4: Embedded MBS support - MBS zones

G.6 End-to-End WiMAX Architecture

The IEEE only defined the Physical (PHY) and Media Access Control (MAC) layers in IEEE 802.16 [i.333]. This approach has worked well for technologies such as Ethernet and Wi-Fi, which rely on other bodies such as the Internet Engineering Task Force (IETF) to set the standards for higher layer protocols such as TCP/IP, SIP, VoIP and IPSec. In the mobile wireless world, standards bodies such as 3GPP and 3GPP2 set standards over a wide range of interfaces and protocols because they require not only airlink interoperability, but also inter-vendor inter-network interoperability for roaming, multi-vendor access networks, and inter-company billing. Vendors and operators have recognized this issue, and have formed additional working groups to develop standard network reference models for open inter-network interfaces. Two of these are the WiMAX Forum's Network Working Group, which is focused on creating higher-level networking specifications for fixed, nomadic, portable and mobile WiMAX systems beyond what is defined in the IEEE 802.16 [i.333] standard, and Service Provider Working Group which helps write requirements and prioritizes them to help drive the work of the Network WG.

The Mobile WiMAX End-to-End Network Architecture is based on an All-IP platform, all packet technology with no legacy circuit telephony. It offers the advantage of reduced total cost of ownership during the lifecycle of a WiMAX network deployment. The use of All-IP means that a common network core can be used, without the need to maintain both packet and circuit core networks, with all the overhead that goes with it. A further benefit of All-IP is that it places the network on the performance growth curve of general purpose processors and computing devices, often termed "Moore's Law". Computer processing advances occur much faster than advances in telecommunications equipment because general purpose hardware is not limited to telecommunications equipment cycles, which tend to be long and cumbersome. The end result is a network that continually performs at ever higher capital and operational efficiency, and takes advantage of 3rd party developments from the Internet community. This results in lower cost, high scalability, and rapid deployment since the networking functionality is all primarily software-based services.

In order to deploy successful and operational commercial systems, there is need for support beyond 802.16 (PHY/MAC) air interface specifications; especially the need to support a core set of networking functions as part of the overall End-to-End WiMAX system architecture.

Some general aspects have guided the development of Mobile WiMAX Network Architecture and include the following:

a) Provision of logical separation between such procedures and IP addressing, routing and connectivity management procedures and protocols to enable use of the access architecture primitives in standalone and interworking deployment scenarios.

b) Support for sharing of Access Service Network (ASN) of a Network Access Provider (NAP) among multiple Network Service Provider (NSP).

c) Support of a single NSP providing service over multiple ASN(s) - managed by one or more NAPs.

d) Support for the discovery and selection of accessible NSPs by an MS or Subscriber Station (SS).

e) Support of NAPs that employ one or more ASN topologies.

f) Support of access to incumbent operator services through internetworking functions as needed.

g) Specification of open and well-defined reference points between various groups of network functional entities (within an ASN, between ASNs, between an ASN and a CSN, and between CSNs), and in particular between an MS, ASN and CSN to enable multi-vendor interoperability.

h) Support for evolution paths between the various usage models subject to reasonable technical assumptions and constraints.

i) Enabling different vendor implementations based on different combinations of functional entities on physical network entities, as long as these implementations comply with the normative protocols and procedures across applicable reference points, as defined in the network specifications.

j) Support for the most trivial scenario of a single operator deploying an ASN together with a limited set of CSN functions, so that the operator can offer basic Internet access service without consideration for roaming or interworking.

G.6.1 Support for Services and Applications

The end-to-end architecture includes the support for:

a) Voice, multimedia services and other mandated regulatory services such as emergency services and lawful interception.

b) Access to a variety of independent Application Service Provider (ASP) networks in an agnostic manner.

c) Mobile telephony communications using VoIP.

d) Support interfacing with various interworking and media gateways permitting delivery of incumbent/legacy services translated over IP (for example, SMS over IP, MMS, WAP) to WiMAX access networks.

e) Support delivery of IP Broadcast and Multicast services over WiMAX access networks.

G.6.2 Interworking and Roaming

Interworking and Roaming is another key strength of the End-to-End Network Architecture with support for a number of deployment scenarios. In particular, there will be support of:

a) Loosely-coupled interworking with existing wireless networks such as 3GPP and 3GPP2 or existing wireline networks such as DSL and MSO, with the interworking interface(s) based on a standard IETF suite of protocols.

b) Global roaming across WiMAX operator networks, including support for credential reuse, consistent use of AAA for accounting and billing, and consolidated/common billing and settlement.

c) A variety of user authentication credential formats such as username/password, digital certificates, Subscriber Identify Module (SIM), Universal SIM (USIM), and Removable User Identify Module (RUIM).

WiMAX Forum industry participants have identified a WiMAX Network Reference Model (NRM) that is a logical representation of the network architecture. The NRM identifies functional entities and reference points over which interoperability is achieved between functional entities. The architecture has been developed with the objective of providing unified support of functionality needed in a range of network deployment models and usage scenarios (ranging from fixed - nomadic - portable - simple mobility - to fully mobile subscribers).

Figure G.5 illustrates the NRM, consisting of the following logical entities: Mobile Station (MS), Access Service Network (ASN), and Connectivity Service Network (CSN) and clearly identified reference points for interconnection of the logical entities. The figure depicts the key normative reference points R1-R5. Each of the entities, MS, ASN and CSN represent a grouping of functional entities. Each of these functions may be realized in a single physical device or may be distributed over multiple physical devices. The grouping and distribution of functions into physical devices within a functional entity (such as ASN) is an implementation choice; a manufacturer may choose any physical implementation of functions, either individually or in combination, as long as the implementation meets the functional and interoperability requirements.

The intent of the NRM is to allow multiple implementation options for a given functional entity, and yet achieve interoperability among different realizations of functional entities. Interoperability is based on the definition of communication protocols and data plane treatment between functional entities to achieve an overall end-to-end function, for example, security or mobility management. Thus, the functional entities on either side of a reference point represent a collection of control and bearer plane end-points.

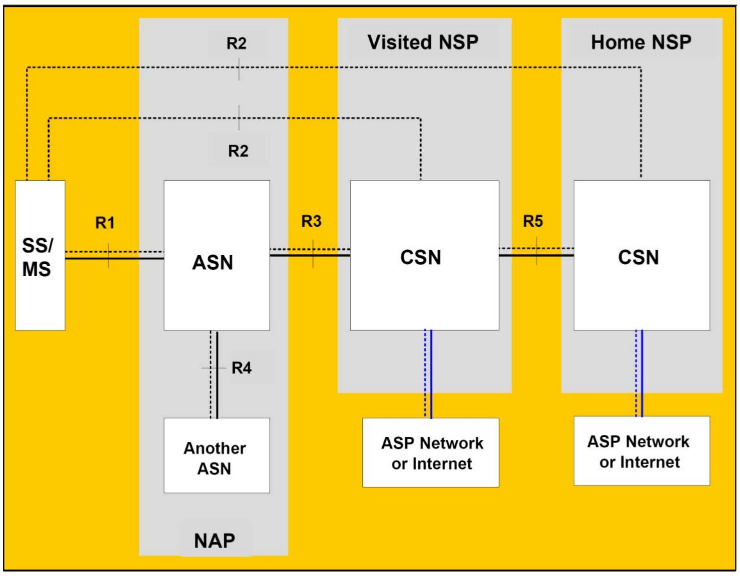


Figure G.5: WiMAX Network Reference Model

The ASN defines a logical boundary and represents a convenient way to describe aggregation of functional entities and corresponding message flows associated with the access services. The ASN represents a boundary for functional interoperability with WiMAX clients, WiMAX connectivity service functions and aggregation of functions embodied by different vendors. Mapping of functional entities to logical entities within ASNs as depicted in the NRM may be performed in different ways.

Connectivity Service Network (CSN) is defined as a set of network functions that provide IP connectivity services to the WiMAX subscriber(s). A CSN may comprise network elements such as routers, AAA proxy/servers, user databases and Interworking gateway devices. A CSN may be deployed as part of a Greenfield WiMAX Network Service Provider (NSP) or as part of an incumbent WiMAX NSP.

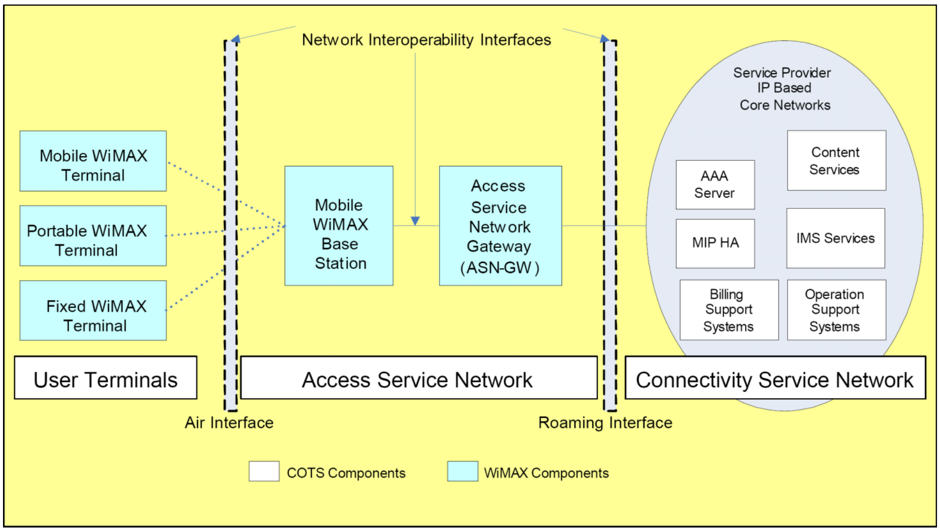


Figure G.6: Overview of the Network IP-based architecture

G.6.3 Network-Level Mobility Handover

The end-to-end WiMAX Network Architecture has extensive capability to support mobility and handovers. In particular it is intended to provide:

- Vertical or inter-technology handovers - e.g. to Wi-Fi, 3GPP, 3GPP2, DSL, or MSO - when such capability is enabled in multi-mode MS.
- Support IPv4 or IPv6 based mobility management. Within this framework, and as applicable, the architecture accommodates MS with multiple IP addresses and simultaneous IPv4 and IPv6 connections.
- Support roaming between NSPs.
- Utilize mechanisms to support seamless handovers at up to vehicular speeds - satisfying well defined (within WiMAX Forum) bounds of service disruption.

Some additional capabilities for mobility support include:

- Dynamic and static home address configurations.
- Dynamic assignment of the Home Agent in the service provider network as a form of route optimization, as well as in the home IP network as a form of load balancing.
- Dynamic assignment of the Home Agent based on policies.

Annex H: Review of UMTS Long Term Evolution (LTE)

H.1 LTE System Architecture

Figure H.1 illustrates the high-level system LTE system architecture. The functional elements and their interface points are described below.

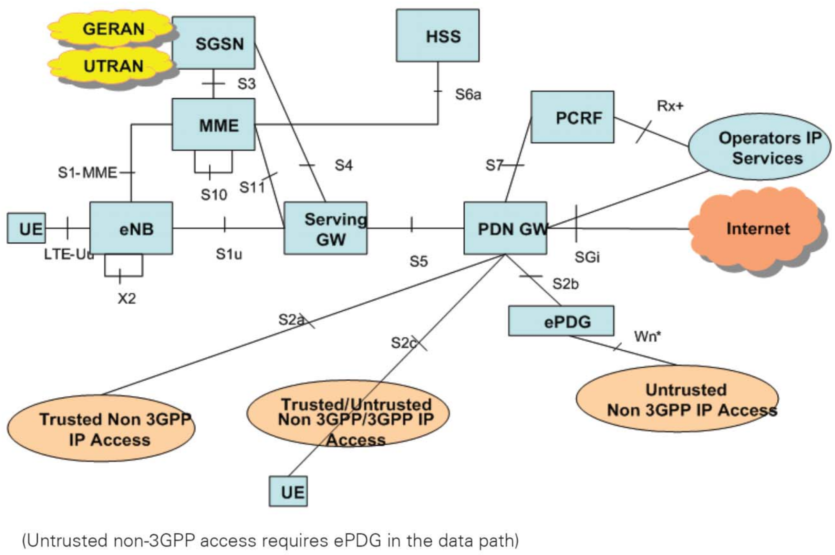


Figure H.1: High-level architecture of 3GPP LTE

Functional Elements of the LTE System Architecture:

- **Evolved Radio Access Network (RAN):** The evolved RAN for LTE consists of a single node, i.e. the eNodeB (eNB) that interfaces with the UE. The eNB hosts the PHYsical (PHY), Medium Access Control (MAC), Radio Link Control (RLC), and Packet Data Control Protocol (PDCP) layers that include the functionality of user-plane header-compression and encryption. It also offers Radio Resource Control (RRC) functionality corresponding to the control plane. It performs many functions including radio resource management, admission control, scheduling, enforcement of negotiated UL QoS, cell information broadcast, ciphering/deciphering of user and control plane data, and compression/decompression of DL/UL user plane packet headers.

- **Serving GateWay (SGW):** The SGW routes and forwards user data packets, while also acting as the mobility anchor for the user plane during inter-eNB handovers and as the anchor for mobility between LTE and other 3GPP technologies (terminating S4 interface and relaying the traffic between 2G/3G systems and PDN GW). For idle state UEs, the SGW terminates the DL data path and triggers paging when DL data arrives for the UE. It manages and stores UE contexts, e.g. parameters of the IP bearer service, network internal routing information. It also performs replication of the user traffic in case of lawful interception.

- **Mobility Management Entity (MME):** The MME is the key control-node for the LTE access-network. It is responsible for idle mode UE tracking and paging procedure including retransmissions. It is involved in the bearer activation/deactivation process and is also responsible for choosing the SGW for a UE at the initial attach and at time of intra-LTE handover involving Core Network (CN) node relocation. It is responsible for authenticating the user (by interacting with the HSS). The Non-Access Stratum (NAS) signalling terminates at the MME and it is also responsible for generation and allocation of temporary identities to UEs. It checks the authorization of the UE to camp on the service provider's Public Land Mobile Network (PLMN) and enforces UE roaming restrictions. The MME is the termination point in the network for ciphering/integrity protection for NAS signalling and handles the security key management. Lawful interception of signalling is also supported by the MME. The MME also provides the control plane function for mobility between LTE and 2G/3G access networks with the S3 interface terminating at the MME from the SGSN. The MME also terminates the S6a interface towards the home HSS for roaming UEs.

- **Packet Data Network Gateway (PDN GW):** The PDN GW provides connectivity to the UE to external packet data networks by being the point of exit and entry of traffic for the UE. A UE may have simultaneous connectivity with more than one PDN GW for accessing multiple PDNs. The PDN GW performs policy enforcement, packet filtering for each user, charging support, lawful Interception and packet screening. Another key role of the PDN GW is to act as the anchor for mobility between 3GPP and non-3GPP technologies such as WiMAX and 3GPP2 (CDMA 1X and EvDO).

Interfaces Points in the LTE Architecture:

- **S1-MME** Reference point for the control plane protocol between EUTRAN and MME. The protocol over this reference point is eRANAP and it uses Stream Control Transmission Protocol (SCTP) as the transport protocol.

- **S1-U** Reference point between EUTRAN and SGW for the per-bearer user plane tunneling and inter-eNB path switching during handover. The transport protocol over this interface is GPRS Tunnelling Protocol-User plane (GTP-U).

- **S2a** It provides the user plane with related control and mobility support between trusted non-3GPP IP access and the Gateway. S2a is based on Proxy Mobile IP. To enable access via trusted non-3GPP IP accesses that do not support PMIP, S2a also supports Client Mobile IPv4 FA mode.

- **S2b** It provides the user plane with related control and mobility support between evolved Packet Data Gateway (ePDG) and the PDN GW. It is based on Proxy Mobile IP.

- **S2c** It provides the user plane with related control and mobility support between UE and the PDN GW. This reference point is implemented over trusted and/or untrusted non-3GPP Access and/or 3GPP access. This protocol is based on Client Mobile IP co-located mode.

- **S3** It is the interface between SGSN and MME and it enables user and bearer information exchange for inter 3GPP access network mobility in idle and/or active state. It is based on Gn reference point as defined between SGSNs.

- **S4** It provides the user plane with related control and mobility support between SGSN and the SGW and is based on Gn reference point as defined between SGSN and GGSN.

- **S5** It provides user plane tunnelling and tunnel management between SGW and PDN GW. It is used for SGW relocation due to UE mobility and if the SGW needs to connect to a non-collocated PDN GW for the required PDN connectivity. Two variants of this interface are being standardized depending on the protocol used, namely, GTP and the IETF based Proxy Mobile IP solution.

- **S6a** It enables transfer of subscription and authentication data for authenticating/authorizing user access to the evolved system (AAA interface) between MME and HSS.

- **S7** It provides transfer of (QoS) policy and charging rules from Policy and Charging Rules Function (PCRF) to Policy and Charging Enforcement Function (PCEF) in the PDN GW. This interface is based on the Gx interface.

- **S10** Reference point between MMEs for MME relocation and MME to MME information transfer.

- **S11** Reference point between MME and SGW.

- **SGi** It is the reference point between the PDN GW and the packet data network. Packet data network may be an operator-external public or private packet data network or an intra-operator packet data network, e.g. for provision of IMS services. This reference point corresponds to Gi for 2G/3G accesses.

- **Rx+** The Rx reference point resides between the Application Function and the PCRF in the TS 123 203 [i.338].

Wn* This is the reference point between the Untrusted Non-3GPP IP Access and the ePDG. Traffic on this interface for a UE initiated tunnel has to be forced towards ePDG.

H.2 Protocol Architecture

This clause describes the functions of the different protocol layers and their location in the LTE architecture. Figures H.2 and H.3 show the control plane and the user plane protocol stacks, respectively.

In the control-plane, the NAS protocol, which runs between the MME and the UE, is used for control-purposes such as network attach, authentication, setting up of bearers, and mobility management. All NAS messages are ciphered and integrity protected by the MME and UE. The RRC layer in the eNB makes handover decisions based on neighbour cell measurements sent by the UE, pages for the UEs over the air, broadcasts system information, controls UE measurement reporting such as the periodicity of Channel Quality Information (CQI) reports and allocates cell-level temporary identifiers to active UEs. It also executes transfer of UE context from the source eNB to the target eNB during handover, and does integrity protection of RRC messages. The RRC layer is responsible for the setting up and maintenance of radio bearers.

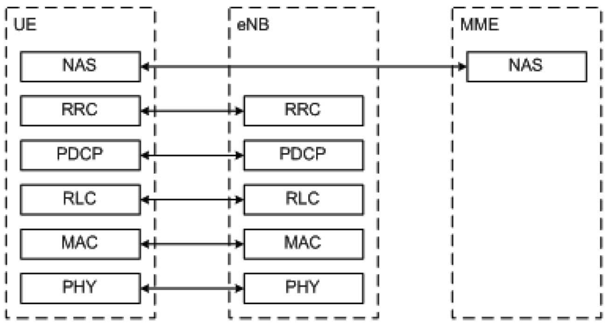


Figure H.2: Control plane protocol stack

In the user-plane, the PDCP layer is responsible for compressing/decompressing the headers of user plane IP packets using Robust Header Compression (ROHC) to enable efficient use of air interface bandwidth. This layer also performs ciphering of both user plane and control plane data. Because the NAS messages are carried in RRC, they are effectively double ciphered and integrity protected, once at the MME and again at the eNB.

The RLC layer is used to format and transport traffic between the UE and the eNB. RLC provides three different reliability modes for data transport- Acknowledged Mode (AM), Unacknowledged Mode (UM), or Transparent Mode (TM). The UM mode is suitable for transport of Real Time (RT) services because such services are delay sensitive and cannot wait for retransmissions. The AM mode, on the other hand, is appropriate for non-RT (NRT) services such as file downloads. The TM mode is used when the PDU sizes are known a priori such as for broadcasting system information. The RLC layer also provides in-sequence delivery of Service Data Units (SDUs) to the upper layers and eliminates duplicate SDUs from being delivered to the upper layers. It may also segment the SDUs depending on the radio conditions.

Furthermore, there are two levels of re-transmissions for providing reliability, namely, the HARQ at the MAC layer and outer ARQ at the RLC layer. The outer ARQ is required to handle residual errors that are not corrected by HARQ that is kept simple by the use of a single bit error-feedback mechanism. An N-process stop-and-wait HARQ is employed that has asynchronous re-transmissions in the DL and synchronous re-transmissions in the UL. Synchronous HARQ means that the re-transmissions of HARQ blocks occur at pre-defined periodic intervals. Hence, no explicit signalling is required to indicate to the receiver the retransmission schedule. Asynchronous HARQ offers the flexibility of scheduling re-transmissions based on air interface conditions.

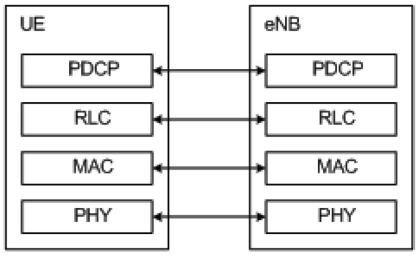


Figure H.3: User plane protocol stack

Figures H.4 and H-5 show the structure of layer 2 for DL and UL, respectively. The PDCP, RLC and MAC layers together constitute layer 2.

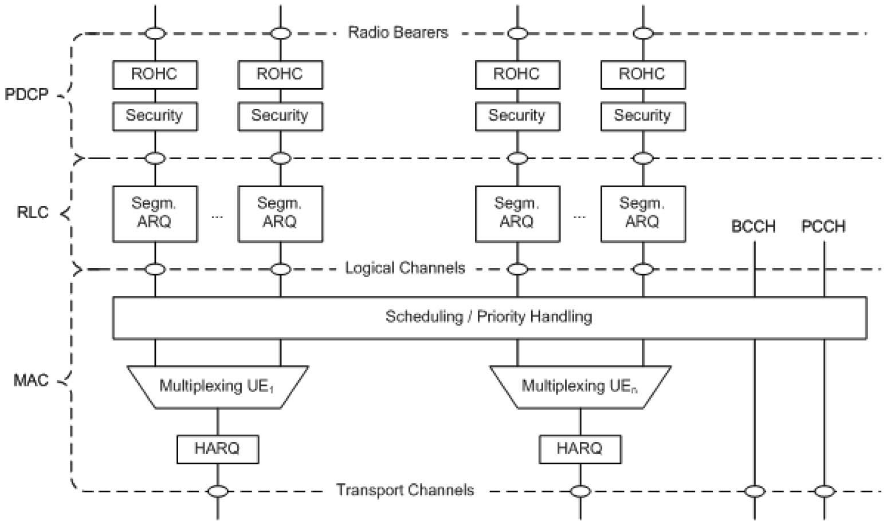


Figure H.4: Layer 2 structure for the DL

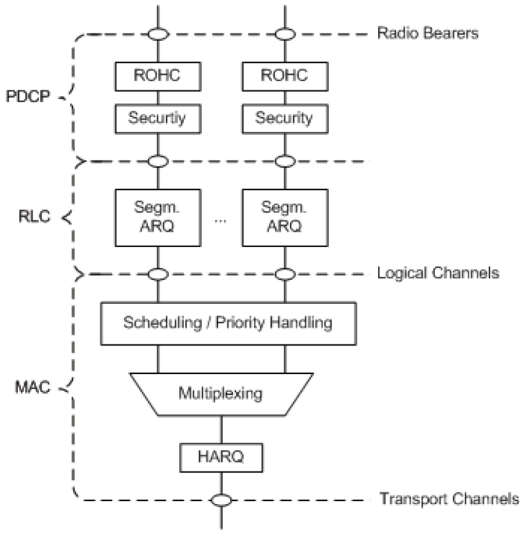


Figure H.5: Layer 2 structure for the UL

In LTE, there is significant effort to simplify the number and mappings of logical and transport channels. The different logical and transport channels in LTE are illustrated in figures H.6 and H.7 respectively.

The transport channels are distinguished by the characteristics (e.g. adaptive modulation and coding) with which the data are transmitted over the radio interface. The MAC layer performs the mapping between the logical channels and transport channels, schedules the different UEs and their services in both UL and DL depending on their relative priorities, and selects the most appropriate transport format. The logical channels are characterized by the information carried by them.

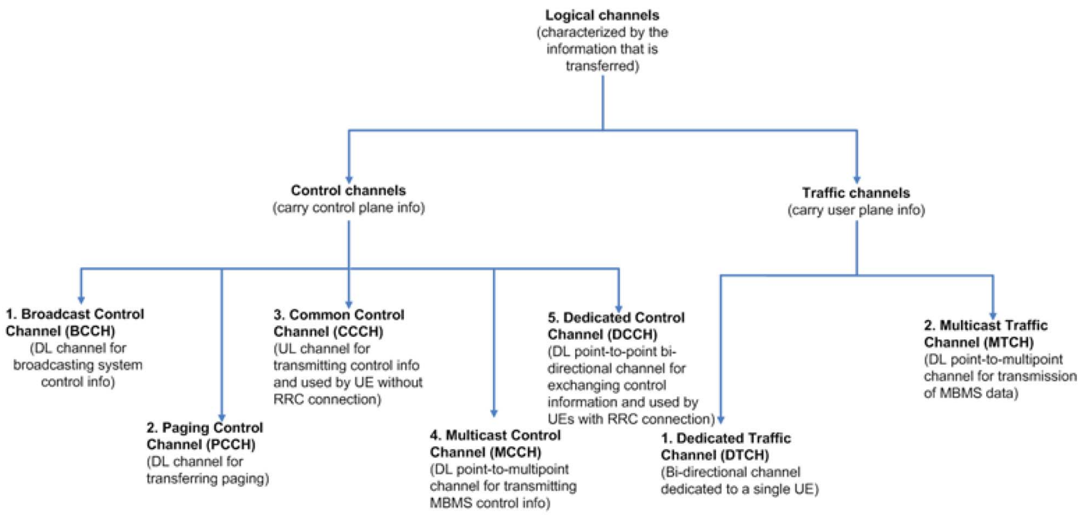


Figure H.6: Logical Channels in LTE

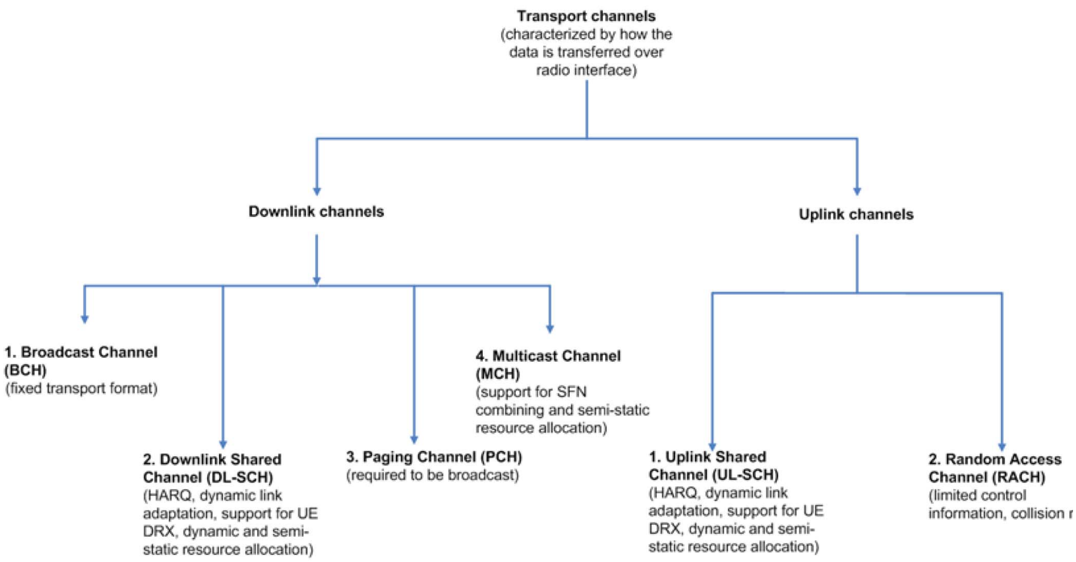


Figure H.7: Transport Channels in LTE

The mapping of the logical channels to the transport channels is shown in figure H.8.

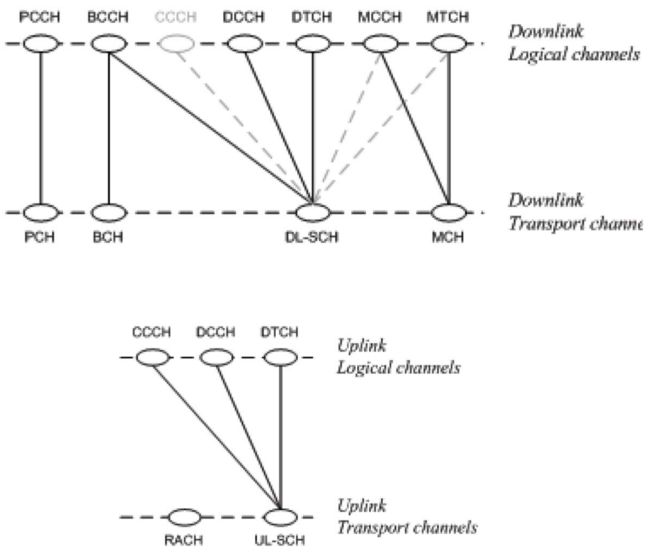


Figure H.8: Logical to Transport channel mappings

The physical layer at the eNB is responsible for protecting data against channel errors using Adaptive Modulation and Coding (AMC) schemes based on channel conditions. It also maintains frequency and time synchronization and performs RF processing including modulation and demodulation. In addition, it processes measurement reports from the UE such as CQI and provides indications to the upper layers.

The minimum unit of scheduling is a time-frequency block corresponding to one sub-frame (1 ms) and 12 sub-carriers. The scheduling is not done at a sub-carrier granularity in order to limit the control signalling. QPSK, 16-QAM and 64-QAM will be the DL and UL modulation schemes in E-UTRA. For UL, 64-QAM is optional at the UE.

Multiple antennas at the UE are supported with the 2 receive and 1 transmit antenna configuration being mandatory. MIMO is also supported at the eNB with two transmit antennas being the baseline configuration.

OFDMA with a sub-carrier spacing of 15 kHz and Single Carrier Frequency Division Multiple Access (SC-FDMA) have been chosen as the transmission schemes for the DL and UL, respectively. Each radio frame is 10 ms long containing 10 sub-frames with each sub-frame capable of carrying 14 OFDM symbols.

H.3 Mobility Management

Mobility management can be classified based on the radio technologies of the source and the target cells, and the mobility-state of the UE. From a mobility perspective, the UE can be in one of three states, LTE_DETACHED, LTE_IDLE, and LTE_ACTIVE as shown in figure H.9.

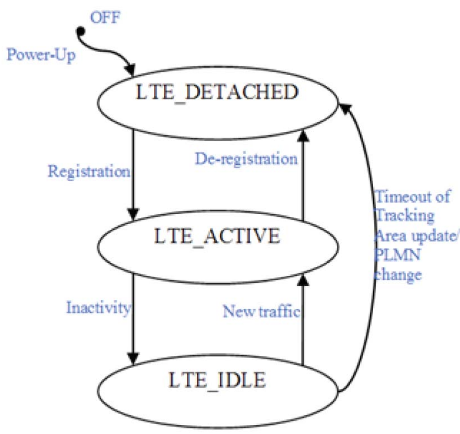


Figure H.9: Mobility states of the UE in LTE

LTE_DETACHED state is typically a transitory state in which the UE is powered-on but is in the process of searching and registering with the network. In the LTE_ACTIVE state, the UE is registered with the network and has an RRC connection with the eNB. In LTE_ACTIVE state, the network knows the cell to which the UE belongs and can transmit/receive data from the UE. The LTE_IDLE state is a power-conservation state for the UE, where typically the UE is not transmitting or receiving packets. In LTE_IDLE state, no context about the UE is stored in the eNB. In this state, the location of the UE is only known at the MME and only at the granularity of a Tracking Area (TA) that consists of multiple eNBs. The MME knows the TA in which the UE last registered and paging is necessary to locate the UE to a cell.

In idle mode, the UE is in power-conservation mode and does not inform the network of each cell change. The network knows the location of the UE to the granularity of a few cells, called the Tracking Area (TA). When there is a UE-terminated call, the UE is paged in its last reported TA. Extensive discussions occurred in 3GPP on the preferred tracking area mechanism. Static non-overlapping tracking areas were used in earlier technologies, such as, GSM. However, there are newer techniques that avoid ping-pong effects, distribute the TA update load more evenly across cells and reduce the aggregate TA update load. Some of the candidate mechanisms that were discussed include overlapping TAs, multiple TAs and distance-based TA schemes. It has been agreed in 3GPP that a UE can be assigned multiple TAs that are assumed to be non-overlapping. It has also been agreed that TAs for LTE and for pre-LTE RATs will be separate i.e. an eNB and a UMTS Node-B will belong to separate TAs to simplify the network's handling of mobility of the UE when UE crosses 3GPP RAT boundaries.

In LTE_ACTIVE, when a UE moves between two LTE cells, "backward" handover or predictive handover is carried out. In this type of handover, the source cell, based on measurement reports from the UE, determines the target cell and queries the target cell if it has enough resources to accommodate the UE. The target cell also prepares radio resources before the source cell commands the UE to handover to the target cell.

In LTE, data buffering in the DL occurs at the eNB because the RLC protocol terminates at the eNB. Therefore, mechanisms to avoid data loss during inter-eNB handovers is all the more necessary when compared to the UMTS architecture where data buffering occurs at the centralized Radio Network Controller (RNC) and inter-RNC handovers are less frequent. Two mechanisms were proposed to minimize data loss during handover: Buffer forwarding and bi-casting. In buffer forwarding, once the handover decision is taken, the source eNB forwards buffered data for the UE to the target eNB. In bicasting, the SGW bi-casts/multi-casts packets to a set of eNBs (including the serving eNB), which are candidates for being the next serving eNB. The bicasting solution requires significantly higher backhaul bandwidth, and may still not be able to avoid data loss altogether. Moreover, the determination of when to start bi-casting is an important issue to address in the bi-casting solution. If bi-casting starts too early, there will be a significant increase in the backhaul bandwidth requirement. If bi-casting starts too late, it will result in packet loss. Therefore, the decision in 3GPP is that buffer forwarding would be the mechanism to avoid packet loss for intra-LTE handovers. The source eNB may decide whether or not to forward traffic depending on the type of traffic, e.g. perform data forwarding for NRT traffic and no data forwarding for RT traffic.

H.4 Evolved MBMS

There will be support for MBMS right from the first version of LTE specifications. However, specifications for E-MBMS are in early stages. Two important scenarios have been identified for E-MBMS: One is single-cell broadcast, and the second is MBMS Single Frequency Network (MBSFN). MBSFN is a new feature that is being introduced in the LTE specification. MBSFN is envisaged for delivering services such as Mobile TV using the LTE infrastructure, and is expected to be a competitor to DVB-H-based TV broadcast. In MBSFN, the transmission happens from a time-synchronized set of eNBs using the same resource block. This enables over-the-air combining, thus improving the Signal-to-Interference plus Noise-Ratio (SINR) significantly compared to non-SFN operation. The Cyclic Prefix (CP) used for MBSFN is slightly longer, and this enables the UE to combine transmissions from different eNBs, thus somewhat negating some of the advantages of SFN operation. There will be six symbols in a slot of 0,5 ms for MBSFN operation versus seven symbols in a slot of 0,5 ms for non-SFN operation.

The overall user-plane architecture for MBSFN operation is shown in figure H.10. 3GPP has defined a SYNC protocol between the E-MBMS gateway and the eNBs to ensure that the same content is sent over-the-air from all the eNBs. As shown in figure H.10, eBM-SC is the source of the MBMS traffic, and the E-MBMS gateway is responsible for distributing the traffic to the different eNBs of the MBSFN area. IP multicast may be used for distributing the traffic from the E-MBMS gateway to the different eNBs. 3GPP has defined a control plane entity, known as the MBMS Coordination Entity (MCE) that ensures that the same resource block is allocated for a given service across all the eNBs of a given MBSFN area. It is the task of the MCE to ensure that the RLC/MAC layers at the eNBs are appropriately configured for MBSFN operation. 3GPP has currently assumed that header compression for MBMS services will be performed by the E-MBMS gateway.

Both single-cell MBMS and MBSFN will typically use point-to-multipoint mode of transmission. Therefore, UE feedback, such as, ACK/NACK and CQI cannot be used as one could for the point-to-point case. However, aggregate statistical CQI and ACK/NACK information can still be used for link adaptation and retransmissions.

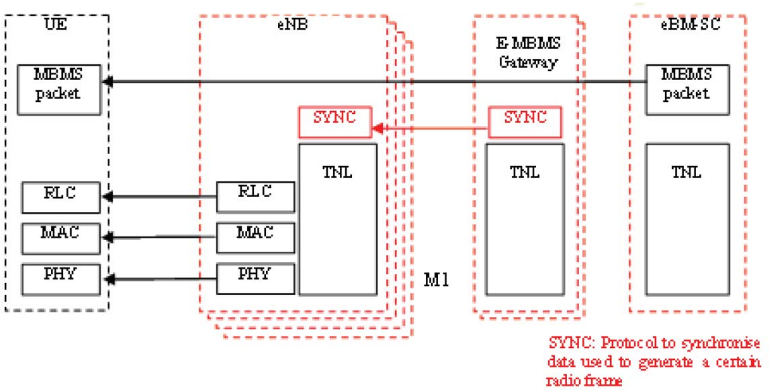


Figure H.10: The overall U-plane architecture of the MBMS content synchronisation

H.5 Physical Layer of LTE UMTS

LTE adopts OFDMA in the DL and SC-FDMA in the UL, where the selection of the latter is mainly intended to avoid the high PAPR in OFDMA signals, which result in reduced power efficiencies.

Both the DL and UL share the same frame structure. The standard specifies two types of frame structures, where the first type applies both to FDD and TDD modes, while the second type is only compatible with TDD mode. Figure H.11 illustrates the first type of frame structure. It is observed that the LTE frames have 10 ms in duration, and they are divided into 10 sub frames, each sub frame being 1 ms long. Each sub frame is further divided into two slots, each of 0,5 ms duration. Slots consist of either 6 or 7 ODFM/SC-FDMA symbols, depending on whether the normal or extended cyclic prefix is employed.

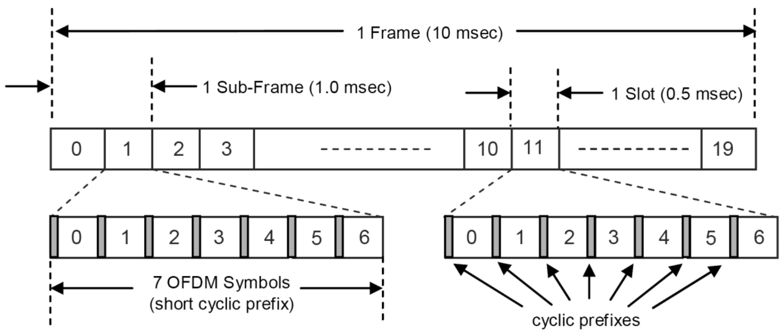


Figure H.11: First type of frame structure
(compatible both with FDD and TDD modes)

H.5.1 Downlink

The basic subcarrier spacing in OFDMA is 15 kHz, with a reduced subcarrier spacing of 7,5 kHz available for some multimedia broadcasting SFN scenarios. Table H.1 summarizes the OFDM modulation parameters.

Table H.1: OFDM Parameters (DL)

Transmission BW		1.25 MHz	2.5 MHz	5 MHz	10 MHz	15 MHz	20 MHz
Sub-frame duration		0.5 ms					
Sub-carrier spacing		15 kHz					
Sampling frequency		192 MHz (1/2 x 3.84 MHz)	3.84 MHz	7.68 MHz (2 x 3.84 MHz)	15.36 MHz (4 x 3.84 MHz)	23.04 MHz (6 x 3.84 MHz)	30.72 MHz (8 x 3.84 MHz)
FFT size		128	256	512	1024	1536	2048
OFDM sym per slot (short/long CP)		7/6					
CP length (usec/ samples)	Short	(4.69/9) x 6, (5.21/10) x 1	(4.69/18) x 6, (5.21/20) x 1	(4.69/36) x 6, (5.21/40) x 1	(4.69/72) x 6, (5.21/80) x 1	(4.69/108) x 6, (5.21/120) x 1	(4.69/144) x 6, (5.21/160) x 1
	Long	(16.67/32)	(16.67/64)	(16.67/128)	(16.67/256)	(16.67/384)	(16.67/512)

Depending on the channel delay spread, either short or long CP is used. When short CP is used, the first OFDM symbol in a slot has slightly longer CP than the remaining six symbols, as shown in table H.2. This is done to preserve slot timing (0,5 ms). It is noted that the CP duration is described in absolute terms (e.g. 16,67 μsec for long CP) and in terms of standard time units, Ts. Ts is used throughout the LTE specification documents. It is defined as Ts = 1 / (15 000 × 2 048) seconds, which corresponds to the 30,72 MHz sample clock for the 2 048 point FFT used with the 20 MHz system bandwidth.

Table H.2: Long and short CP durations

Configuration		Cyclic Prefix Length	
		T_s	μsec
Normal CP	Δf = 15 kHz	160 for l = 0 144 for l = 1, 2…5	5.21 for l = 0 4.69 for l = 1, 2…5
Extended CP	Δf = 15 kHz	512	16.67
	Δf = 15 kHz	1024	33.33

In OFDMA, users are allocated a specific number of subcarriers for a predetermined amount of time. These in LTE are referred to as Physical Resource Blocks (PRBs). PRBs thus have both a time and frequency dimension. Allocation of PRBs is handled by a scheduling function at the 3GPP base station (eNodeB). A PRB is defined as consisting of 12 consecutive subcarriers for one slot (0,5 ms) in duration. A PRB is the smallest element of resource allocation assigned by the base station scheduler.

The transmitted downlink signal consists of N_{BW} subcarriers for duration of N_{symb} OFDM symbols. It can be represented by a resource grid as depicted in figure H.12. Each box within the grid represents a single subcarrier for one symbol period and is referred to as a resource element. In MIMO applications, there is a resource grid for each transmitting antenna.

In contrast to packet-oriented networks, LTE does not employ a PHY preamble to facilitate carrier offset estimate, channel estimation, timing synchronization, etc. Instead, special reference signals are embedded in the PRBs as shown in figure H.13. Reference signals are transmitted during the first and fifth OFDM symbols of each slot. Reference symbols are transmitted every 6^{th} subcarrier. Further, reference symbols are staggered in both time and frequency. The channel response on subcarriers bearing the reference symbols can be computed directly. Interpolation is used to estimate the channel response on the remaining subcarriers.

Synchronization signals use the same type of pseudo-random orthogonal sequences as reference signals. These are classified as primary and secondary synchronization signals, depending how they are used by UE during the cell search procedure. Both primary and secondary synchronization signals are transmitted on the 72 subcarriers centred around the DC subcarrier during the 0th and 10th slots of a frame (recall there are 20 slots within each frame).

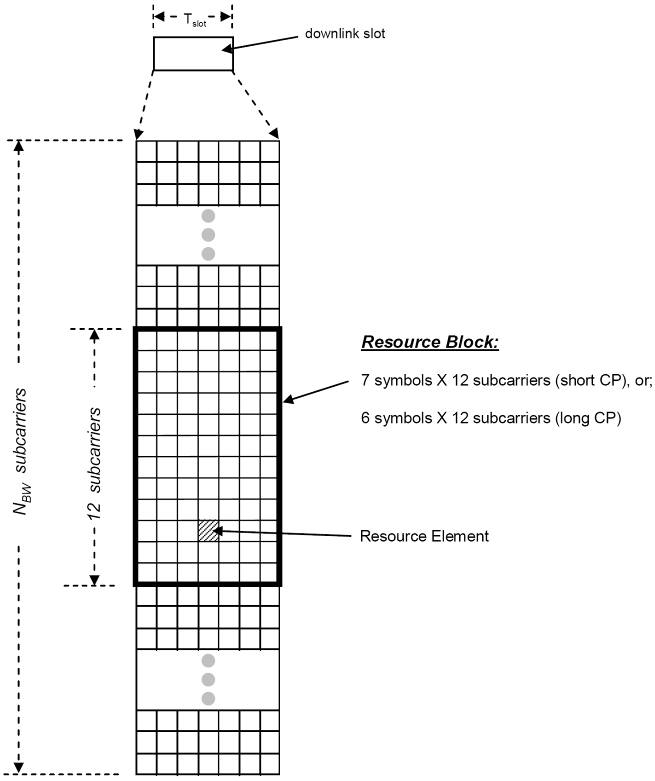


Figure H.12: DL resource grid

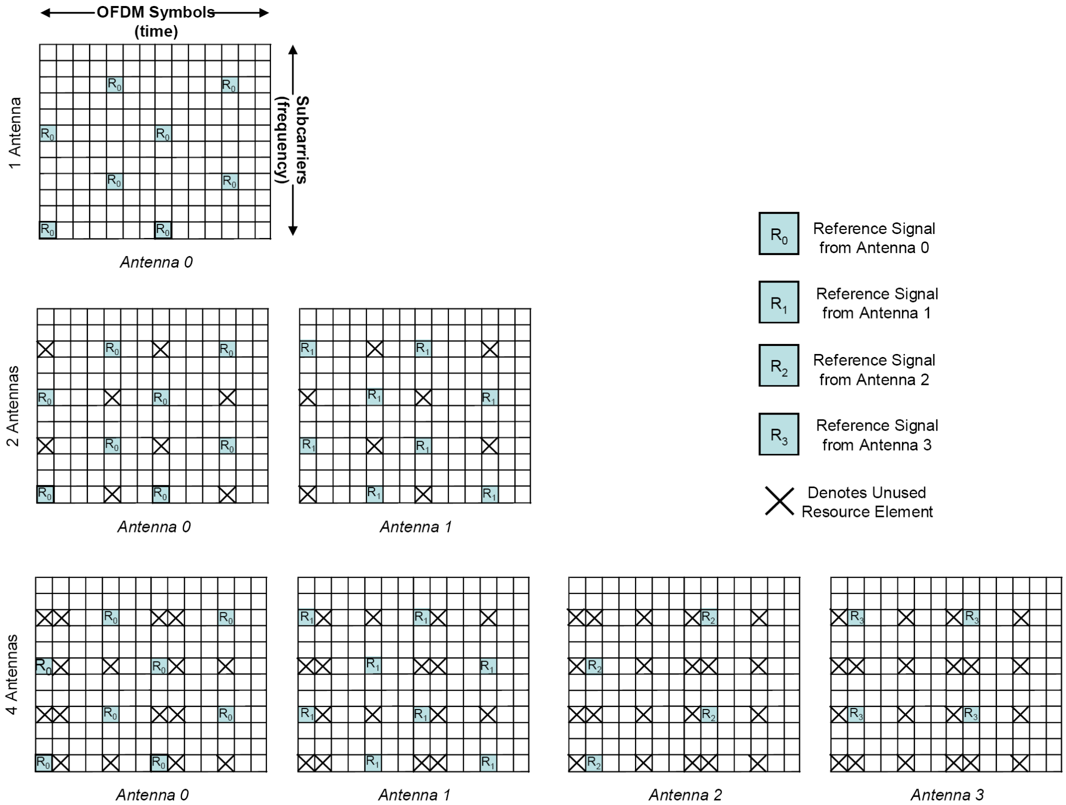


Figure H.13: LTE Reference signals (distributed among resource elements)

The following downlink physical channels are defined in LTE DL:

- **Physical Downlink Shared Channel (PDSCH):** The PDSCH is utilized basically for data and multimedia transport. It therefore is designed for very high data rates. Modulation options therefore include QPSK, 16-QAM and 64-QAM. Spatial multiplexing is also used in the PDSCH. In fact, spatial multiplexing is exclusive to the PDSCH.

- **Physical Broadcast Channel (PBCH).**

- **Physical Multicast Channel (PMCH).**

- **Physical Control Format Indicator Channel (PCFICH).**

- **Physical Downlink Control Channel (PDCCH):** The PDCCH conveys UE-specific control information. Robustness rather than maximum data rate is therefore the main consideration. QPSK is the only available modulation format. The PDCCH is mapped onto resource elements in up to the first three OFDM symbols in the first slot of a sub frame.

- Physical Hybrid ARQ Indicator Channel (PHICH).

The baseband signal representing a downlink physical channel is defined in terms of the following steps:

- scrambling of coded bits in each of the code words to be transmitted on a physical channel;

- modulation of scrambled bits to generate complex-valued modulation symbols;

- mapping of the complex-valued modulation symbols onto one or several transmission layers;

- precoding of the complex-valued modulation symbols on each layer for transmission on the antenna ports;

- mapping of complex-valued modulation symbols for each antenna port to resource elements;

- generation of complex-valued time-domain OFDM signal for each antenna port.

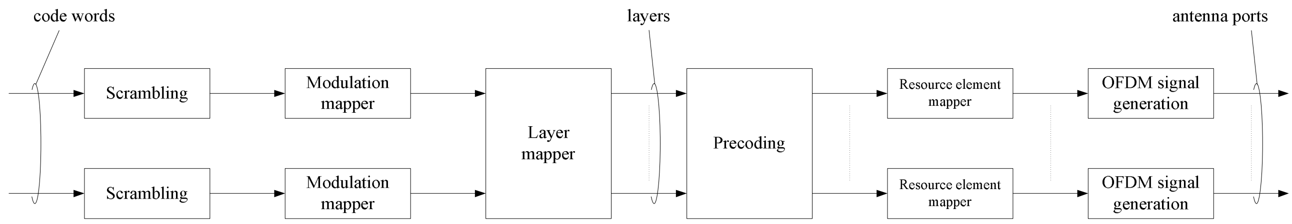


Figure H.14: Overview of physical channel processing

Different coding schemes are supported for the DL physical channels, including QPSK, 16-QAM and 64-QAM. The exact processing specifications for each channel are detailed in 3GPP TS 36.211 V8.0.0 [i.294].

Layer mapping and pre-coding are related to MIMO applications. A layer corresponds to a spatial multiplexing channel. MIMO systems are defined in terms of $N_{transmitters} \times N_{receivers}$. For LTE, defined configurations are 1 × 1, 2 × 2, 3 × 2 and 4 × 2. Note that while there are as many as four transmitting antennas, there are only a maximum of two receivers and thus a maximum of only two spatial multiplexing data streams.

For a 1 × 1 or a 2 × 2 system, there is a simple 1:1 relationship between layers and transmitting antenna ports. However, for a 3 × 2 and 4 × 2 system, there are still only two spatial multiplexing channels. Therefore, there is redundancy on one or both data streams. Layer mapping specifies exactly how the extra transmitter antennas are employed.

Precoding is also used in conjunction with spatial multiplexing. Recall that MIMO exploits multipath to resolve independent spatial data streams. In other words, MIMO systems require a certain degree of multipath for reliable operation. In a noise-limited environment with low multipath distortion, MIMO systems can actually become impaired.

H.5.2 Uplink

SC-FDMA is also known as "DFT-Spread-OFDM", since blocks of modulation symbols are processed through a Discrete Fourier Transform (DFT) before they are mapped to orthogonal sub-carriers. Thus, in effect the DFT spreads the energy of individual modulation symbols to more than a single sub-carrier. Figure H.15 provides a block diagram of the main processing functions involved in SC-FDMA. It is observed that the technique is very much similar to OFDM, except for the DFT and IDFT blocks at the transmitter and receiver respectively. In a sense the DFT, depending on its size relative to IFFT, tends to cancel the orthogonal sub-carrier mappings. The end effect is that the transmit Peak to Average Power Ratio (PAPR) is reduced relative to OFDM, and thus the technique is less susceptible to the non-linear response of the amplifier.

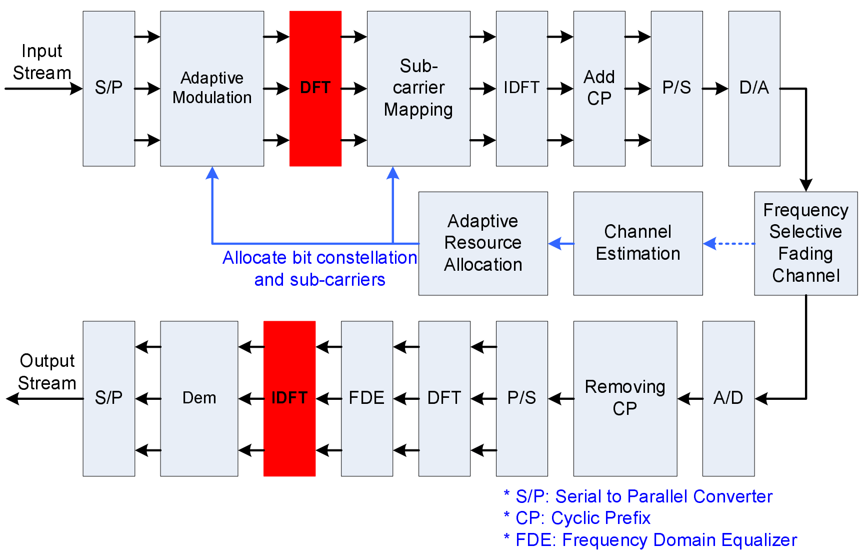


Figure H.15: Block Diagram of Signal Processing functions in SC-FDMA

In FDD the uplink uses the same subcarrier spacing of 15 kHz and PRB width (12 subcarriers). Uplink PRBs are assigned to UE by the base station scheduler via the downlink CCPCH. Uplink PRBs consist of a group of 12 contiguous subcarriers for the duration of one slot time.

Defined UL physical channels are:

- **Physical Uplink Shared Channel (PUSCH):** Resources for the PUSCH are allocated on a sub-frame basis by the UL scheduler. Subcarriers are allocated in multiples of 12 (PRBs) and may be hopped from sub-frame to sub-frame. The PUSCH may employ QPSK, 16-QAM or 64-QAM modulation.
- **Physical Uplink Control Channel (PUCCH):** As the name implies, the PUCCH carries uplink control information. It is never transmitted simultaneously with PUSCH data. PUCCH conveys control information including channel quality indication (CQI), ACK/NACK, HARQ and uplink scheduling requests. The PUCCH transmission is frequency hopped at the slot boundary (as shown in figure H.16) for added reliability.
- **Physical Random Access Channel (PRACH).**

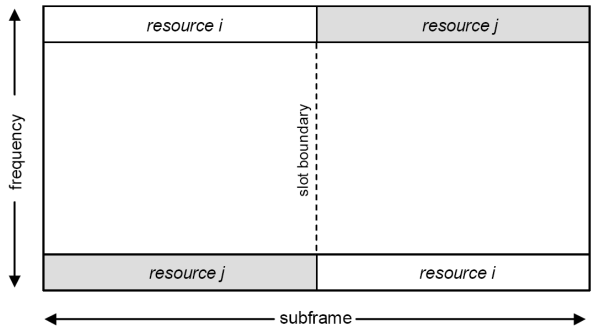


Figure H.16: PUCCH - hopped at slot boundary

The following UL physical signals are specified in LTE:

- Uplink Reference Signal: There are two variants of the UL reference signal. The demodulation signal facilitates coherent demodulation. It is transmitted in the fourth SC-FDMA symbol of the slot and is the same size as the assigned resource. There is also a sounding reference signal used to facilitate frequency dependent scheduling. Both variants of the UL reference signal are based on Zadhoff-Chu sequences.
- Random Access Preamble: The random access procedure involves the PHY and higher layers. At the PHY layer, the cell search procedure is initiated by transmission of the random access preamble by the UE. If successful, a random access response is received from the base station. The random access preamble format is shown in figure H.17. It consists of a cyclic prefix, a preamble and a guard time during which there is no signal transmitted.

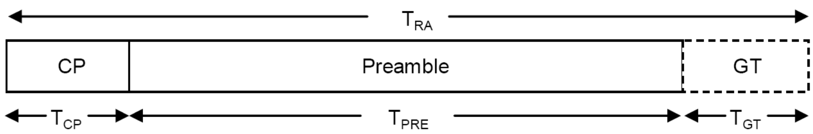


Figure H.17: Random access preamble format

For the generic frame structure, the timing parameters are:

- T_{RA}: 30720 T_S
- T_{GT}: 3152 T_S
- T_{PRE}: 24576 T_S

where T_S = period of a 30,72 MHz clock.

Random access preambles are derived from Zadoff-Chu sequences. They are transmitted on blocks of 72 contiguous subcarriers allocated for random access by the base station. In FDD applications, there are 64 possible preamble sequences per cell.

The exact frequency used for transmission of the random access preamble is selected from available random access channels by higher layers in the UE. Other information provided to the PHY by higher layers includes:

- Available random access channels.
- Preamble format (which preamble sequences).
- Initial transmission power.
- Power ramp step size.
- Maximum number of retries.

Annex I: Detailed Analysis of LTE and WiMAX air interfaces over satellite links

I.1 Application Scenarios

As a result of a market and business analysis, two application scenarios are considered in this annex:

- broadcasting using linguistic beams with national coverage;
- two-way communications using multi-spot coverage with frequency reuse.

Starting from the standard 3GPP LTE and WiMAX specifications, this chapter performs a selection of physical and access layer configurations to be considered for ad-hoc and end-to-end analyses and simulations for each of the two scenarios.

In both cases, the driving principle is to identify and finalise numerology such that a fair comparison between the two considered standards (3GPP LTE and WiMAX) is performed.

I.1.1 Considered OFDM Numerology

Tables I.1 and I.2 report the chosen OFDM numerology. FDD duplexing is considered for both standards.

The LTE frame is subdivided into 10 sub frames, each of them lasting 1 ms. WiMAX frame, which is not strictly specified by the standardization documents, has been designed so as to fit an integer number of clusters (1,4 tones for 2 OFDM symbols, used in the forward link) and of tiles (4 tones for 3 OFDM symbols, used in the reverse link). Thus, 12 OFDM symbols are considered in the WiMAX frame, plus the preamble. With this choice the number of useful OFDM symbols per allocation unit is the same for the two standards. Also, with the selected frame design, TTI is comparable for the two systems, namely 1 ms for LTE and 1,49 ms for WiMAX.

Table I.1: LTE OFDM Numerology

Selected LTE OFDM Numerology							
Channel Bandwidth	**FFT size**	**Subcarrier spacing**	**Number of OFDM Symbols per TTI**	**OFDM Symbol duration**	**Sampling time**	**TTI**	**CP Length**
5 MHz	2 048	15 kHz	12	83,33 µs	32,55 ns	1 ms	16,67 µs [1/4]
1,25 MHz	2 048	15 kHz	12	83,33 µs	32,55 ns	1 ms	16,67 µs [1/4]

Table I.2: WiMAX OFDM Numerology

Selected WiMAX OFDM Numerology							
Channel Bandwidth	**FFT size**	**Subcarrier spacing**	**Number of OFDM Symbols per TTI**	**OFDM Symbol duration**	**Sampling time**	**TTI**	**CP Length**
5 MHz	512	10,94 kHz	13 [Pre + 12]	114,29 µs	179 ns	1,49 ms	22,85 µs [1/4]
1,25 MHz	128	10,94 kHz	13 [Pre + 12]	114,29 µs	714 ns	1,49 ms	22,85 µs [1/4]

I.1.2 Reference signals patterns

The basic reference signals (pilots) patterns are reported in the following figures, for the considered Downlink and Uplink configurations. The density of these patterns is used in the following to determine the number of available allocation units per frame.

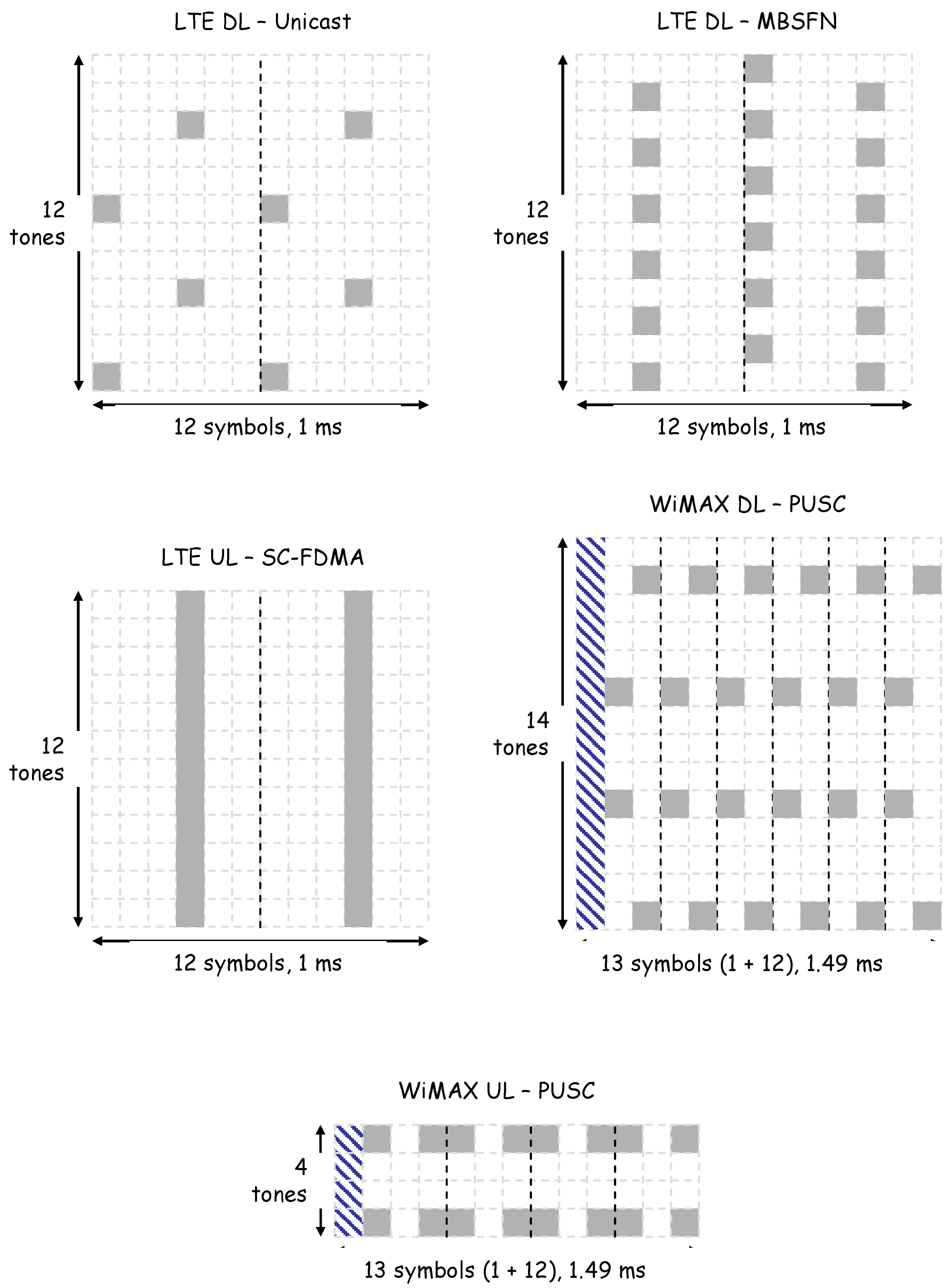


Figure I.1

I.1.3 Broadcasting scenario - Physical Layer Configuration

In the following, taking into consideration the numerology outlined in tables I.1 and I.2, a selection of physical layer configurations is carried out. A sub-set of these configurations is taken to determine the simulation scenarios for end-to-end simulations.

Table I.3: LTE Physical Layer Configuration for Broadcasting

LTE					
Number of jointly coded channels / number of channel groups	**Information bits per packet**	**Allocated data carriers per sub-frame (MBSFN RS) [RBs x OFDM symbols]**	**Mod**	**Actual Code rate**	**Bit Rate**
1/8	312	378 [3 x 12]	QPSK	0,41	312 kb/s
8/1	2 496	3 150 [25 x 12]	QPSK	0,40	2,50 Mb/s
1/12	312	252 [2 x 12]	QPSK	0,62	312 kb/s
12/1	3 776	3 150 [25 x 12]	QPSK	0,60	3,78 Mb/s
1/16	312	198 [3 x 6, even slot]	QPSK	0,78	312 kb/s
16/1	4 992	3 150 [25 x 12]	QPSK	0,80	4,99 Mb/s
1/25	312	126 [1 x 12]	16-QAM	0,62	312 kb/s
24/1	7552 [3 776 + 3 776]	3150 [25 x 12]	16-QAM	0,.60	7,49 Mb/s

Table I.4: WiMAX Physical Layer Configuration for Broadcasting

WiMAX					
Number of jointly coded channels / channel groups	**Information bits per packet**	**Allocated data carriers per frame, PUSC [sub channels]**	**Mod**	**Actual Code rate**	**Bit Rate**
1/9	480	480 [10]	QPSK	0,50	322 kb/s
2/4	960	960 [20]	QPSK	0,50	646 kb/s
1/12	448	336 [7]	QPSK	0,67	301 kb/s
3/4	1 344	1 008 [21]	QPSK	0,67	905 kb/s
1/15	480	288 [6]	QPSK	0,83	322 kb/s
4/3	1 920	1 152 [24]	QPSK	0,83	1 292 kb/s
1/22	512	192 [4]	16-QAM	0,67	345 kb/s
3/7	1 536	576 [12]	16-QAM	0,67	1 031 kb/s

I.1.3.1 End-to-end simulation cases

As it can be seen, three different modulation/coding pairs have been selected for LTE and for WiMAX. These have been extracted from the broadcasting physical layer configurations, and correspond to the configurations where several physical layer services are jointly coded at the PHY level. This results in a coding gain due to the larger encoded packet size. Nonetheless, extended simulation scenarios have been considered in order to assess the performance loss when each channel is separately encoded. In this document, a relevant subset of the produced results is reported.

Table I.5: PHY-FL Broadcasting Scenarios

PHY-FL Broadcasting Scenario						
	BC_LTE			BC_WiMAX		
	_QPSK_25	_QPSK_45	_16QAM_35	_QPSK_12	_QPSK_56	_16QAM_23
PL/OFDM Numerology						
Channel Bandwidth	5 MHz			5 MHz		
FFT Size	2048			512		
Subcarrier Spacing	15 kHz			10.94 kHz		
Active Subcarriers	300			420		
Number of data OFDM symbols per TTI	12			12		
TTI Duration	1 ms			1.49 ms		
OFDM Symbol Duration	83.33 us			114.29 us		
Sampling Time	32.55 ns			179 ns		
CP Length	16.67 us [1/4]			22.85 us [1/4]		
Pilot Pattern	MBSFN			DL-PUSC		
Data Source						
Information Bits	2496	4992	7552	960	1920	1536
Turbo Coding						
Rate	2/5	4/5	3/5	1/2	5/6	2/3
Fractional Predistortion						
LUT [LUT Size]	1024			1024		
On-Board HPA (TWTA)						
IBO	4 dB			4 dB		
Channel Model						
Model name	MAESTRO_CASE_5 --> Outdoor urban with repeaters			MAESTRO_CASE_5 --> Outdoor urban with repeaters		
Number of antennas (Tx, Rx)	(1,1)			(1,1)		
Terminal speed	30 kph			30 kph		
Normalized Doppler spread (OFDM sample rate)	3.63E-06			2.00E-05		
Channel Impairments						
Phase Noise Mask	DVB-SH mask			DVB-SH mask		
Normalized Frequency Offset	PRE/POST-FFT (LUT with residual estimation errors)			PRE/POST-FFT (LUT with residual estimation errors)		
Normalized Time Shift	uniform]-0.5:0.5] wrt OFDM sampling time			uniform]-0.5:0.5] wrt OFDM sampling time		
AWGN Noise						
Eb/No	Step 1.0 [dB] - Adjustable Range			Step 1.0 [dB] - Adjustable Range		
Estimation Algorithms						
Frequency	PRE/POST-FFT (LUT with residual estimation errors)			PRE/POST-FFT (LUT with residual estimation errors)		
OFDM symbol synch	Assumed to be recovered by PRE/POST-FFT acquisition and tracking			Assumed to be recovered by PRE/POST-FFT acquisition and tracking		
Channel	POST-FFT			POST-FFT		
SNIR	Ideal			Ideal		
Performance Measurements						
Target PER	**1.00E-03**			**1.00E-03**		
Estimation Performance						
Channel Estimator	Mean and variance of estimation error			Mean and variance of estimation error		
Reference PHY-FL Broadcasting Scenario						
_**IC** Ideal Channel	No HPA, No Fading/Shadowing, No Channel Impairments. As a reference for all considered code rate/modulation pairs					
_**IE** Ideal Estimation	All channel impairments in. Perfect Channel State Information					
_COMPLETE	Including whannel estimation, phase noise, freq offset, timing offset					
Extended PHY-FL Broadcasting Scenario						
_**KPH** Different speeds	**3 (M5), 150 (M2)**	**3 (M5), 150 (M2)**	**3 (M5), 150 (M2)**	**3 (M5), 150 (M2)**	**3 (M5), 150 (M2)**	**3 (M5), 150 (M2)**
_INTERTTI_3KPH	**40, 80, 160 TTIs**					
_INTERTTI_30KPH	**4, 8, 16 TTIs**					
_PAPR	**ACE**		**ACE**	**ACE**		**ACE**

OFDM numerology considers 5 MHz channelization, reference signal pattern foreseen for SFN networks for LTE, cyclic prefix of length 1/4, and frequency selective propagation channel, considering the presence of complementary ground components repeating the signal coming from the satellite. Further, different terminal speeds are considered, using as a baseline 30 km/h and also considering 3 km/h and 150 km/h. For the 150 km/h case the MAESTRO case 2 channel model has been considered, modelling propagation in open area.

The satellite input back-off has been set to 4 dB, resulting as a trade-off for the joint use of QPSK and 16-QAM modulations within the same sub frame. Predistortion techniques have been assumed at the transmitter, linearizing the satellite HPA behaviour. Further, a simulation scenario considering PAPR reduction techniques has been foreseen in order to assess the achievable performance improvement.

At the receiver, impairments such as phase noise (DVB-SH mask), timing error, and frequency offset are considered and the impact on performance is assessed in complete simulations.

The simulation chains considered in for this scenario are reported in clause I.1.5.1.

I.1.4 Two-Way communications - Physical Layer Configuration

In the two-way communications scenario channel bandwidth is 1,25 MHz, compared to the 5 MHz allocated in the forward link of the interactive broadcasting scenario. Thus the tables reported in the following clauses refer to the 1,25 MHz numerology. The peak data rates that have been targeted are for nomadic terminals.

I.1.4.1 Forward Link

Table I.6: LTE Physical Layer Configuration for Two-Way communications forward link

Number of users per 1,25 MHz channel	Information bits per packet	Allocated data carriers per sub-frame (Standard RS) [RBs x OFDM symbols]	Modulation	Actual Code rate	Bit Rate
Up to 3	272	272 [2 x 12]	QPSK	0,50	272 kb/s
Up to 4	304	204 [3 x 6]	QPSK	0,75	304 kb/s
Up to 6	272	136 [1 x 12]	16-QAM	0,50	272 kb/s

Table I.7: WiMAX Physical Layer Configuration for Interactive Broadcasting forward link

Number of users per 1,25 MHz channel (18 sub channels per frame)	Information bits per packet	Allocated data carriers per frame (PUSC) [sub channels]	Modulation	Actual Code rate	Bit Rate
Up to 2	384	384 [8]	QPSK	0,50	258 kb/s
Up to 3	432	288 [6]	QPSK	0,75	291 kb/s
Up to 4	384	192 [4]	16-QAM	0,50	258 kb/s

Analogously with what done for the broadcasting scenario, PHY configurations have been identified for the two-way communications scenario. They are reported in table I.8.

The differences with broadcasting configurations are:

- Channel bandwidth → 1,25 MHz instead of 5 MHz.
- Lower data rates → from 258 to 304 kb/s → lower protection from coding.
- Reference symbol pattern for LTE → Unicast instead of MBSFN.
- Propagation channel → Frequency flat.
- Possible use of Transmit Diversity and Spatial Multiplexing MIMO configurations.

Table I.8: PHY-FL Two-Way Scenarios

PHY-FL Two-Way Scenarios						
	2FL_LTE			2FL_WiMAX		
	_QPSK_12	_QPSK_34	_16QAM_12	_QPSK_12	_QPSK_34	_16QAM_12
PL/OFDM Numerology						
Channel Bandwidth	1.25 MHz			1.25 MHz		
FFT Size	2048			128		
Subcarrier Spacing	15 kHz			10.94 kHz		
Active Subcarriers	72			84		
Number of data OFDM symbols per TTI	12			12		
TTI Duration	1 ms			1.49 ms		
OFDM Symbol Duration	83.33 us			114.29 us		
Sampling Time	32.55 ns			714 ns		
CP Length	16.67 us [1/4]			22.85 us [1/4]		
Pilot Pattern	DL-Unicast			DL-PUSC		
Data Source						
Information Bits	272	304	272	384	432	384
Turbo Coding						
Rate	1/2	3/4	1/2	1/2	3/4	1/2
Fractional Predistortion						
LUT [LUT Size]	1024			1024		
On-Board HPA (TWTA)						
IBO	4 dB			4 dB		
Channel Model						
Model name	MAESTRO_CASE_2 --> Rural satellite			MAESTRO_CASE_2 --> Rural satellite		
Number of antennas (Tx, Rx)	(1,1)			(1,1)		
Terminal speed	**30 kph**			**30 kph**		
Normalized Doppler spread (OFDM sample rate)	3.63E-06			7.93E-05		
Channel Impairments						
Phase Noise Mask	DVB-SH mask			DVB-SH mask		
Normalized Frequency Offset	PRE/POST-FFT (LUT with residual estimation errors)			PRE/POST-FFT (LUT with residual estimation errors)		
Normalized Time Shift RMS	uniform]-0.5:0.5] wrt OFDM sampling time			uniform]-0.5:0.5] wrt OFDM sampling time		
AWGN Noise						
Eb/No	Step 1.0 [dB] - Adjustable Range			Step 1.0 [dB] - Adjustable Range		
Estimation						
Frequency	PRE/POST-FFT (LUT with residual estimation errors)			PRE/POST-FFT (LUT with residual estimation errors)		
Time	Assumed to be recovered by PRE/POST-FFT acquisition and tracking			Assumed to be recovered by PRE/POST-FFT acquisition and tracking		
Channel	POST-FFT with Interpolation			POST-FFT with Interpolation		
SNIR	Ideal			Ideal		
Performance Measurements						
Target PER	**1.00E-03**			**1.00E-03**		
Estimation Performance						
Channel Estimator	Mean and variance of estimation error			Mean and variance of estimation error		
Reference PHY-FL Two-Way Scenarios						
_IC Ideal Channel	No HPA, No Fading/Shadowing, No Channel Impairments. As a reference for all considered code rate/modulation pairs					
_IE Ideal Estimation	All channel impairments in. Perfect Channel State Information					
Extended PHY-FL Two-Way Scenarios						
_MIMO_TD Transmit Diversity	**2x2 Tx Div**		**2x2 Tx Div**			
_MIMO_SM Spatial Multiplexing	**2x2 MIMO**		**2x2 MIMO**			

I.1.4.2 Reverse Link

Table I.9: LTE Physical Layer Configuration for Two-Way communications reverse link

Number of users per 1,25 MHz channel	Information bits per packet	Allocated data carriers per sub-frame [Standard RS]	Modulation	Actual Code rate	Bit Rate
Up to 2	264	360 [3 x 12]	QPSK	0,36	264 kb/s
Up to 3	264	240 [2 x 12]	QPSK	0,55	264 kb/s
Up to 9	128	80 [8 subcarriers x 12] *Not standard compliant	QPSK	0,80	128 kb/s

Table I.10: WiMAX Physical Layer Configuration for Two-Way communications reverse link

Number of users per 1,25 MHz channel (16 sub channels per frame)	Information bits per packet	Allocated data carriers per frame (PUSC) [sub channels]	Modulation	Actual Code rate	Bit Rate
Up to 2	384	384 [8]	QPSK	0,50	258 kb/s
Up to 2	432	288 [6]	QPSK	0,75	291 kb/s
NOTE: Minimum packet size for WiMAX limiting minimum bit rate.					

Finally, PHY configurations foreseen for the RL of the two-way scenario are reported in table I.11.

Taking as a reference the FL PHY configurations, the differences are:

- Different HPA model → SSPA instead of TWTA.
- Channel model → Rice K = 7 dB.
- For LTE → SC-FDMA waveform → robustness to frequency errors and to NL distortion.
- Lower IBO → 1 dB for LTE, 2 dB for WiMAX.

Table I.11: PHY-RL Two-Way Scenarios

PHY-RL Two-Way Scenarios					
	2RL_LTE			2RL_WiMAX	
	_QPSK_13	_QPSK_12	_QPSK_45	_QPSK_12	_QPSK_34
PL/OFDM Numerology					
Channel Bandwidth	1.25 MHz			1.25 MHz	
FFT Size	2048			128	
Subcarrier Spacing	15 kHz			10.94 kHz	
Active Subcarriers	72			96	
Number of data OFDM symbols per TTI	12			12	
TTI Duration	1 ms			1.49 ms	
OFDM Symbol Duration	83.33 us			114.29 us	
Sampling Time	32.55 ns			714 ns	
CP Length	16.67 us [1/4]			22.85 us [1/4]	
Pilot Pattern	UL-SC-FDMA			UL-PUSC	
Data Source					
Information Bits	264	264	128	384	432
Turbo Coding					
Rate	1/3	1/2	4/5	1/2	3/4
Fractional Predistortion					
LUT [LUT Size]	1024			1024	
User terminal HPA (SSPA)					
IBO	1 dB			2 dB	
Channel Model					
Model name	Rice, K = 7 dB			Rice, K = 7 dB	
Number of antennas (Tx, Rx)	(1,1)			(1,1)	
Terminal speed	30 kph			30 kph	
Normalized Doppler spread (OFDM sample rate)	3.63E-06			7.93E-05	
Channel Impairments					
Phase Noise Mask	3GPP mask (TBC)				
Normalized Frequency Offset	PRE/POST-FFT (LUT with residual estimation errors)				
Normalized Time Shift RMS	uniform]-0.5:0.5] wrt OFDM sampling time				
AWGN Noise					
Eb/No	Step 1.0 [dB] - Adjustable Range			Step 1.0 [dB] - Adjustable Range	
Estimation					
Frequency	PRE/POST-FFT (LUT with residual estimation errors)			PRE/POST-FFT (LUT with residual estimation errors)	
Time	Assumed to be recovered by PRE/POST-FFT acquisition and tracking			Assumed to be recovered by PRE/POST-FFT acquisition and tracking	
Channel	POST-FFT with Interpolation			POST-FFT with Interpolation	
SNIR	Ideal			Ideal	
Performance Measurements					
Target PER	**1.00E-03**			**1.00E-03**	
Estimation Performance					
Channel Estimator	Mean and variance of estimation error			Mean and variance of estimation error	
Reference PHY-RL Two-Way Scenarios					
_**IC** Ideal Channel	AWGN channel, no HPA, no Channel Impairments				
_**IE** Ideal Estimation	All channel impairments in. Perfect Channel State Information				

All these configurations have been simulated through the use of the simulation chains reported in clause I.1.5, for each of the considered cases and techniques.

I.1.5 Simulation Block Diagrams

In this clause, the block diagrams of the modules used to simulate the selected physical layer configurations are reported, for each of the considered simulation scenarios.

Block diagrams are subdivided in three sections, namely:

- **LTE FL →** Standard 3GPP Downlink Link chain, including:
  - 3GPP turbo encoder.
  - Rate matching achieved through virtual circular buffer.
  - Frame formatter, adds pilots (MBSFN or Unicast pattern), null carriers.
  - OFDM modulator, performing IFFT and adding cyclic prefix.
- **LTE RL →** Standard 3GPP Return Link chain. The differences with respect to LTE FL include:
  - Use of SC-FDMA (Localized FDMA) waveform.
  - Pilot pattern envisaging two OFDM symbols completely dedicated to reference signals.
- **WiMAX FL & RL →** Standard WiMAX chain, including:
  - LDPC coding.
  - DL or UL pilot pattern, as specified in IEEE 802.16e [i.335].

I.1.5.1 LTE FL and RL

In the following, block diagrams for the LTE forward link simulations are reported. LTE RL considers the use of SC-FDMA modulation, embedded in the OFDM MOD/DEMOD blocks.

The simulation chain considers ideal estimation and the presence of non linear amplifiers, properly compensated by a digital waveform predistorter located at the gateway, and the effect of the propagation channel, which in this case entails the presence of multiple paths due to terrestrial repeaters. Ideal channel estimation is performed at the receiver. Impairments such as phase noise and residual frequency and timing offsets are assumed to be perfectly recovered.

Regarding the multipath channel, the considered channel model depends on the scenario: for the broadcast scenario, Maestro case 5 has been considered, while for the two-way communications scenario, Maestro case 2 has been considered. For additional information on channel power delay profiles, see clause A.2.4.1.

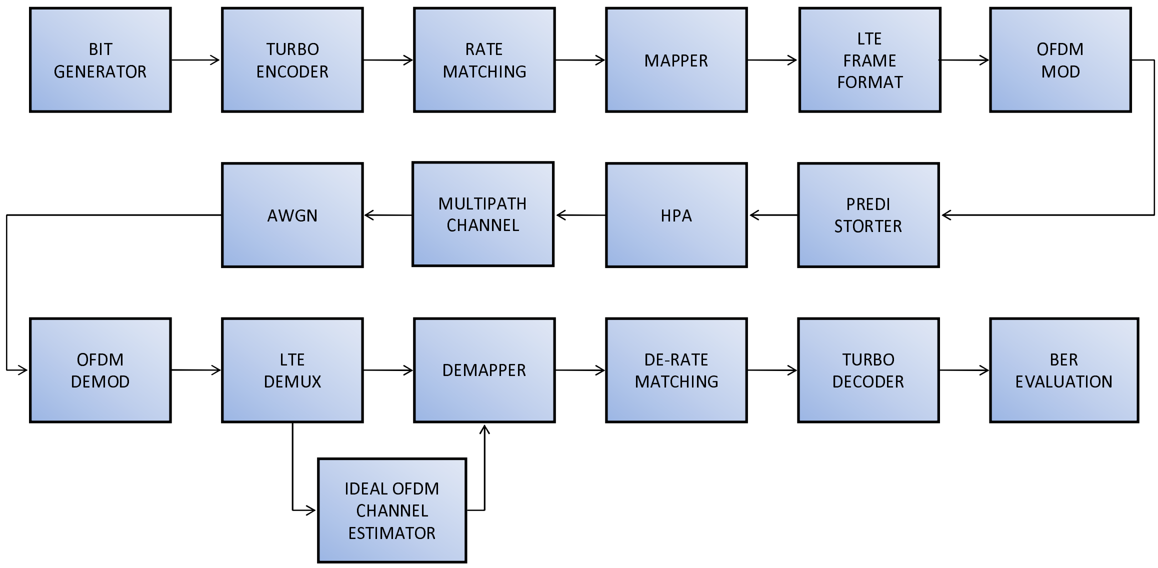


Figure I.1: LTE - Ideal estimation simulation block diagram

I.1.5.2 WiMAX FL and RL

For WiMAX scenarios, simulation blocks are analogous to the ones seen for the LTE case. In WiMAX, there are no substantial functional differences between the FL and RL chains needed to obtain physical layer data recovery performance. In particular, the return link foresees the same waveform structure as it is for the forward link. In the following, simulation chains are reported for the several analyzed cases, as seen for LTE.

I.1.5.2.1 Ideal estimation

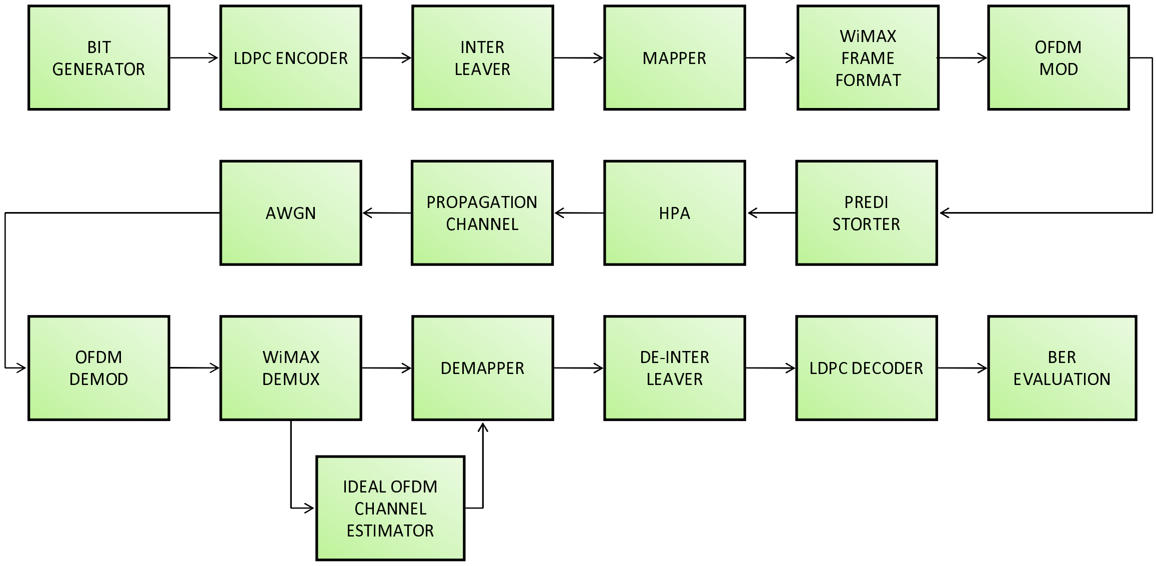


Figure I.2: WiMAX FL & RL - Ideal estimation simulation block diagram

I.1.6 PHY Time series generation for UL Simulator

Starting from the results obtained through physical layer simulations, BLER time series are extracted and then used as inputs for the system level simulator. In this way, accurate time series are used to model packet transmission for the various PHY and propagation channel configurations.

In this clause, we describe the methodology adopted in order to produce the physical layer Block Error Rate (BLER) time series considering the two-way communication scenario. In this case, the considered channel model for system level simulations is the three states Perez-Fontan model. Since fading is frequency flat and for low to medium terminal speeds time selectivity is negligible with respect to the TTI duration (channel coherence time equal to 9 ms at 30 km/h, TTI duration equal to 1 ms for LTE and 1,49 ms for WiMAX), we can assume that the SNR is constant within the whole TTI (both in frequency and in time).

Under these assumptions, BLER time series can be generated using a simplified method that does not require the actual simulation of the whole physical layer chain.

The adopted procedure is depicted in figure I.3 and is made up by the following steps:

3) **Perform AWGN simulations** (including NL distortion), to obtain the function BLER(E_b/N_0)

4) **Generate Perez Fontan model**, obtaining signal level relative to LoS

5) **Set the received C/N_0 value in LoS conditions**

6) **Map the instantaneous C/N_0 value into E_b/N_0**

7) **Generate the time series**, producing a "1" (wrong block) or a "0" (correct block) according to the following algorithm:

```
if [uniform_random_variable < BLER(Eb/N0*)]
then
        time_series_value = 1
```

```
    else
           time_series_value = 0
```

The mapping function used to translate C/N_0 in E_b/N_0 is:

$$\left[C \big/ N_0 \right] = \left[E_b \big/ N_0 \right] \cdot r \cdot \log_2(M) \cdot \left(N_{data} + \frac{E_{pilot\_tone}}{E_{data\_tone}} \cdot N_{pilot} \right) \cdot \Delta f$$

Where r is the actual code rate, M is the constellation order, N_{data} is the number of data tones, N_{pilots} is the number of pilot tones, $E_{pilot\_tone}/E_{data\_tone}$ is the ratio between the energy of a pilot tone and the energy of a data tone, and Δf is the subcarrier spacing in Hz.

For example, starting from a value of E_b/N_0 equal to 0 dB and considering QPSK modulation with rate 1/2 coding, 1,25 MHz bandwidth, $E_{pilot\_tone}/E_{data\_tone} = 1$, N_active = 72 and Δf = 15 kHz, the resulting scaling factor is 1 080 000, corresponding to a value of C/N_0 equal to 60,3 dBHz.

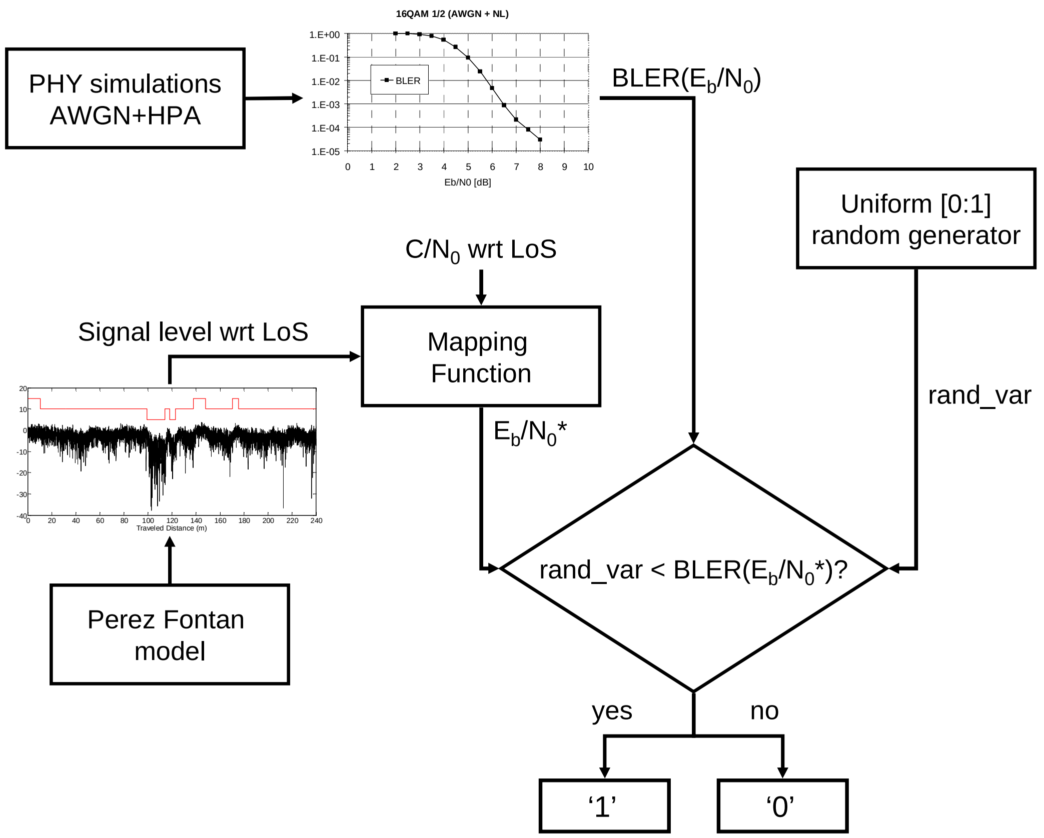


Figure I.3: Block diagram of the procedure adopted for generating the time series

In figure I.4, a channel realization obtained through the Perez Fontan model is reported, considering sub-urban environment and elevation angle equal to 40°. The adopted parameters are taken from [i.30]. The red line represents the three states foreseen by the model, namely line-of-sight, moderate shadowing, and deep shadowing, respectively.

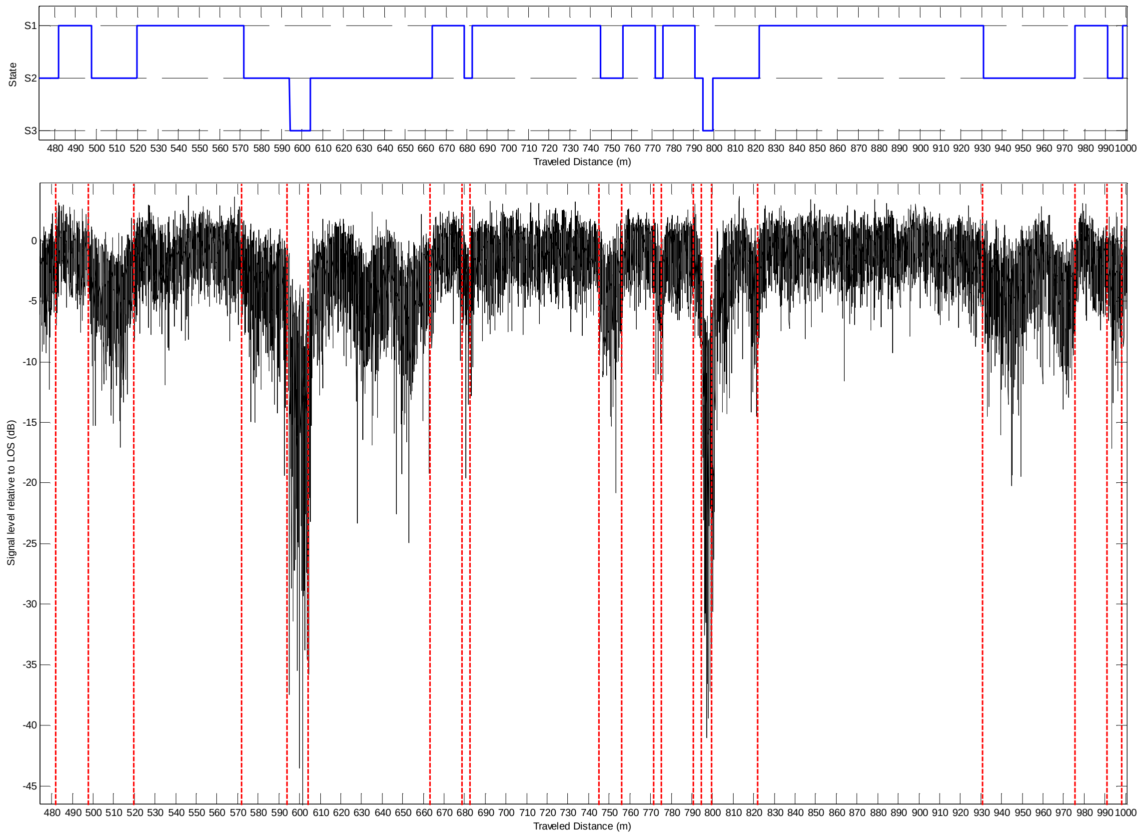


Figure I.4: Channel realization obtained with the Perez Fontan model, suburban environment, 40º elevation

The parameters of the Perez-Fontan multi-state channel model considered to generate the time series are reported in tables I.12 and I.13 for the different environments, considering an elevation of 40 degrees.

Table I.12: State characterization of Perez-Fontan channel model (elevation 40 degrees)

Environment	State 1: LOS			State 2: Shadowing			State 3: Deep shadow		
	α (dB)	ψ (dB)	MP (dB)	α (dB)	ψ (dB)	MP (dB)	α (dB)	ψ (dB)	MP (dB)
Open	0,1	0,37	-22,0	-1,0	0,5	-22,0	-2,25	0,13	-21,2
Suburban	-1,0	0,5	-13,0	-3,7	0,98	-12,2	-15,0	5,9	-13,0
Intermediate Tree-Shadow	-0,4	1,5	-13,.2	-8,2	3,9	-12,7	-17,0	3,14	-10,0
Heavy Tree-Shadow	-	-	-	-10,1	2,25	-10,0	-19,0	4,0	-10,0

Table I.13: Transition probability matrices, state frame duration and transition length for different environments

Environment	[P]			[W]	d_{corr} (m)	L_{frame} (m)	L_{trans} (m)
Open	0,9530	0,0431	0,0039	0,5	2,5	8,9	12,4
	0,0515	0,9347	0,0138	0,375		7,5	
	0,0334	0,0238	0,9428	0,125		4,0 (see note)	
Suburban	0,8177	0,1715	0,0108	0,4545	1,7	5,2	2,2
	0,1544	0,7997	0,0459	0,4545		3,7	
	0,1400	0,1433	0,7167	0,091		3,0 (see note)	
Intermediate Tree-Shadow	0,7193	0,1865	0,0942	0,3929	1,5	6,3	2,6
	0,1848	0,7269	0,0883	0,3571		6,3	
	0,1771	0,0971	0,7258	0,25		4,5	
Heavy Tree-Shadow	0,7792	0,0452	0,1756	0	1,7	-	3,5
	0	0,9259	0,0741	0,5		4,8	
	0	0,0741	0,9259	0,5		4,5	
NOTE: Extrapolated since they are not given in literature.							

I.2 PHY Techniques: Enablers

In this clause, a number of enablers are introduced with the aim of improving the suitability of the considered standards, designed for deployment over a terrestrial network, when used in the framework of a satellite network.

I.2.1 Inter-TTI interleaving through Forced Retransmission

In LTE, the encoded and interleaved bits after Rate Matching are mapped into OFDM symbols. The time unit for arranging the rate matched bits is the Transmission Time Interval (TTI), which corresponds to one sub-frame (two slots) and lasts 1 ms. In order to improve the BER performance, Hybrid Automatic Repeat-reQuest (HARQ) is applied. HARQ operation is performed by exploiting the virtual circular buffer. In particular, orthogonal retransmissions can be obtained by setting the Redundancy Version number in each retransmission, thus transmitting different patterns of bits within the same circular buffer.

HARQ allows a great performance improvement when time correlation is present because retransmission can have a time separation much greater then the channel coherence time.

Unfortunately, HARQ cannot be applied directly to the satellite case due to the exceedingly large transmission delays. Nevertheless, it is possible to devise a way to exploit the existing HARQ facilities adapting them to the satellite use. We propose a novel forced retransmission technique, which basically consists in transmitting the bits carried in the same circular buffer within several TTIs that acts as an **inter-TTI interleaving**. To do this, we can exploit the same mechanism as provided by the LTE Technical Specification for the HARQ operations with circular buffer. The block diagram of the devised technique is depicted in figure I.5. As it can be seen, 4 retransmissions are obtained by using 4 different RVs, starting from 0 up to 3. Each of the 4 transmission bursts is mapped into different TTIs, spaced by $k \cdot$TTI. k is a key parameter because it determines the interleaving depth and it should be set according to channel conditions and latency requirements.

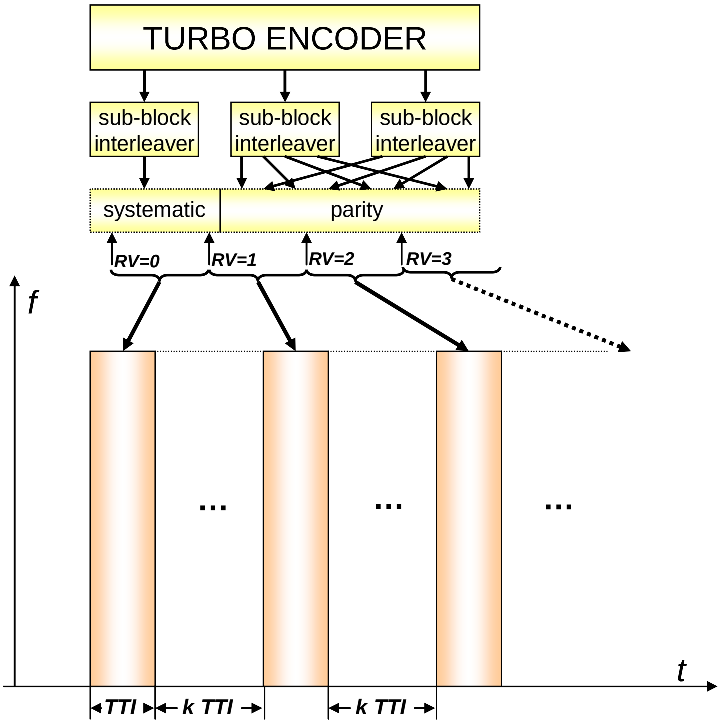


Figure I.5: Inter-TTI interleaving block diagram.

The RVs fixes the starting point from which the virtual circular buffer will be read out. For a given RV number x, the starting bit index k_0 is obtained with the following formula [i.294]:

$$k_0 = \mathrm{mod}(R \cdot (24 \cdot x + 2), K_w),$$

where R is the number of rows in each sub-block interleaver, and K_w is the size of the virtual circular buffer.

It is worth mentioning that even when RVs 0 is used, the starting bit index is not 0. This means that a puncturing of the systematic bits is performed. This is justified by the fact that at high coding rates (i.e. only a portion of the circular buffer is transmitted), the effective pattern with no systematic bits removal (i.e. puncturing parity bits only) can degrade performance [i.296]. This is because for turbo codes with well-designed interleavers, most of the Hamming weight in the minimum distance (and other terms in the distance spectrum) are due to the parity streams. As a consequence, when excessive puncturing is applied on the parity portions, the effective minimum distance of the punctured can degrade significantly, thus resulting in performance worsening.

I.2.2 PAPR Reduction

Letting $\bar{s}_l = (s_{0,l}, \ldots, s_{N-1,l})$ be the vector formed by the OFDM symbol samples (omitting cyclic prefix), the corresponding PAPR can be defined as

$$PAPR = \frac{\|s_l\|_\infty^2}{\frac{1}{N} E\left(\|s_l\|_2^2\right)}$$

The aim of the techniques presented in the following clauses is to reduce this ratio, lowering thus the dynamic range of the signal, which improves the non-distortion capability of the circuitry.

While for the simple expression of a generic OFDM signal, the PAPR Cumulative Distribution Function (CDF) can be analytically evaluated, when complex technique are applied in order to reduce the PAPR, its final CDF cannot be easily evaluated in an analytical fashion, being necessary the use of numerical simulations.

From a logic point of view, the PAPR reduction is placed just before OFDM modulator, but as it will be shown later, this process operates in an iterative fashion, which alternates modulation, processing and demodulation until a target PAPR is reached.

I.2.2.1 Active Constellation Extension (ACE)

The approach of ACE consists in dynamically extending some of the outer signal constellation points in data block towards the outside of the original constellation. The point to be extended and the direction of this extension are chosen by the transmitter in order to reduce the PAPR of the whole OFDM signal. Any point that is farther from the decision boundaries than the nominal constellation point will offer an increased margin, which guarantees a lower BER. Furthermore, there is no loss in data rate and no side information is required at the receiver. However, this technique loses its efficiency increasing the cardinality of the constellation.

The ACE working equation can be written as:

$$\min_{c} \max_{i} \left|\hat{s}_{i,l}\right|^2$$

with

$$\hat{s}_{i,l} = s_{i,l} + c_{i,l} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} (x_{k,l} + C_{k,l}) e^{j\frac{2\pi ik}{N}}$$

where c represents the time-domain signal corresponding to the set of extension vector C_l. The above written minimax problem, can be re-formulated, for a complex-baseband signal, as:

$$\begin{aligned} &\min_{C} E \\ &s.t. \left|s_{i,l} + \mathbf{f}_i \mathbf{C}_l\right|^2 \leq E \ \forall l \\ &E \geq 0, \end{aligned}$$

where $\mathbf{f}_i$ is the *i*-th row or the IFFT matrix. This problem, characterized by a linear objective function and quadratic constraints, can be considered as a special case of QCQP (Quadratically-Constrained Quadratic Programming), a class of problems for which finding the optimal solution can be very difficult. Nevertheless, there are some sub-optimal solutions that can represent a good compromise between accuracy and complexity. In the following, three algorithms are given, with a brief discussion on their potential.

I.2.2.2 Projection Onto Convex Set (POCS)

In general, the set of possible ACE vector will be convex, and a POCS algorithm can be easily written. It can be applied when the task is to minimize the peak below a given amplitude *A*. It is composed of the following steps:

Apply IFFT to data symbols $\bar{x}_l$, obtaining $\bar{s}_l$

Clip any $\left|s_{i,l}\right| \geq A$, obtaining:

$$\hat{s}_{i,l} = \begin{cases} s_{i,l} & , \left|s_{i,l}\right| \leq A \\ A\mathrm{e}^{j\cdot\arg(s_{i,l})} & , \left|s_{i,l}\right| \geq A \end{cases}$$

Obtain $\hat{\bar{x}}$ via an FFT applied to $\hat{\bar{s}}$

Enforce all ACE constraints on $\hat{\bar{x}}$ by restoring all interior points to their original values while projecting exterior points into the region of increased margin.

Return to first step and iterate until no points are clipped.

This algorithm belongs to the class of POCS algorithms for the convexity of the set of clipped vectors and the set of ACE-constrained vector. This algorithm, although easy to implement and useful in reducing the peak amplitude below a given threshold, suffers of slow convergence, making advisable the use of other algorithms, if the requirements for the PAPR reduction are not very tight.

I.2.3 Random Access Signal Detection

The Random Access Channel is a contention-based channel for initial uplink transmission, i.e. typically from mobile user to base station. This channel can be used for several purposes. The RACH function is different depending on the technology of the system. The RACH can be used to access the network, to request resources, to carry control information, to adjust the time offset of the uplink, to adjust the transmitted power, etc. It can even be used to transmit small amounts of data. Contention resolution is the key feature of the random access channel. Many mobile users can access the same base simultaneously, leading to collisions [i.308].

I.2.3.1 Up link: time and frequency structure

The physical layer access preamble structure is made up of a cyclic prefix and sequence segment:

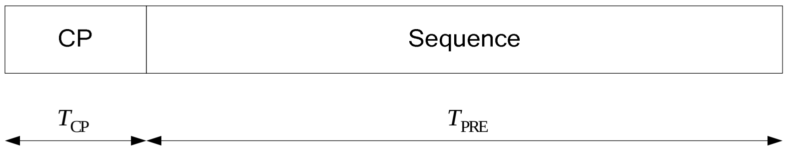


Figure I.6: Random access preamble format

The values of parameters depend on frame structure and the random access configuration controlled by higher layers.

Table I.14: Random access preamble parameters.

Preamble format (frame type 1)	T_{CP}	T_{SEQ}
0	$3152 \times T_s$	$24576 \times T_s$
1	$21012 \times T_s$	$24576 \times T_s$
2	$6224 \times T_s$	$2 \times 24576 \times T_s$
3	$21012 \times T_s$	$2 \times 24576 \times T_s$

The transmission of a preamble is restricted to fixed time and frequency resources .The resources are enumerated in increasing order of sub-frame number within the radio frame and resource blocks in the frequency domain such that index 0 is the lowest numbered resource block and sub-frame within the radio frame. Preamble format 0 to 3 has one random access for each sub-frame. The start of preamble should be aligned with the start of uplink sub-frame with timing advance null. In frequency domain the random access preamble occupies a bandwidth corresponding to 6 resources blocks for all frame structure.

I.2.3.1.1 Preamble sequence definition and generation

The random access preambles are generated from one of several root Zadoff-Chu (ZC) sequences with zero correlation zone. The u^{th} root Zadoff-Chu is defined by:

$$x_u(n) = e^{-j\frac{\pi u n(n+1)}{N_{ZC}}}, \quad 0 \le n \le N_{ZC} - 1$$

Where the length N_{zc} (preamble sequence length in samples) is equal to 839 for the considered frame type.

I.2.3.1.2 Sequence allocation for Satellite Scenario

The number of users to be allocated in each cell in the 4G reverse link via satellite depends on the system design. The scenario proposed is GEO satellite. The zero correlation zone (the size of cyclic shift) has to be larger then the maximum round trip propagation delay, depending on cell radius and multipath delay. The number of root ZC sequences and the number of cyclic shift sequences depend on cell radius and on as in table I.15 considering different geographical positions [i.309].

Table I.15: ZC-ZCZ allocation for GEO satellite scenario

Cell Radius [km]	Number of root ZC sequences	Number of cyclic shift per root sequence
150 (Near polar arctic circle)	64	1
300 (Near polar arctic circle)	64	1
500 (Near polar arctic circle)	64	1
150 (Europe)	64	1
300 (Europe)	64	1
500 (Europe)	64	1
150 (Tropical)	32	2
300 (Europe)	64	1
500 (Europe)	64	1
150 (Equator)	2	32
300 (Equator)	8	8
500 (Equator)	16	4

I.3 PHY results

This clause addresses the analysis of the physical layer performance, considering the simulation scenarios described in the previous clauses.

In particular, the following set-ups have been considered:

- **Ideal Channel:** No HPA, no fading/shadowing, no channel impairments, ideal time/frequency/phase recovery, perfect channel state information. To be considered as a reference for all other configurations.
- **Ideal Estimation:** Performance much closer to reality, including the impact of time and frequency selectivity and non-linear distortion. BLER time traces have been generated for these simulations, and are used for access and upper layer simulations.

In addition to these configurations, a number of simulation scenarios have been devised, with the aim of evaluating the impact and the achievable performance improvement deriving from the use of advanced physical layer techniques or configurations.

In particular, these scenarios include:

- Impact of IBO.
- Use of Inter-TTI transmission techniques.
- Use of PAPR techniques.
- Use of MIMO techniques.

I.3.1 Broadcasting - PHY Results

This clause reports the results of the physical layer simulations performed for the forward link of the broadcasting scenario.

I.3.1.1 Broadcast Scenario - Ideal Channel

Figures I.7 and I.8 report the performance in terms of BER and BLER for the Broadcasting LTE FL Ideal Channel simulations. Results are shown for different code rates, and for the two different coding options. The first option considers the joint coding of all the channels/services transmitted in a physical layer sub frame. The second option considers the separate coding of the different channels/services, and can be identified by the common block length of 312 bits. As it can be seen, the steepness of the performance curve is obviously larger for the aggregated coding case, as the strength of turbo coding is larger.

A phenomenon that is worth noting is the crossing of BER curves at very low Eb/N0 values, which is not observed for BLER curves. This is due to the fact that the lower the code rate the lower the energy associated to the coded bit for a given information bit energy. If on the one hand the use of lower code rates improves the steepness of the waterfall curve, on the other hand it raises the BER at very low SNRs, when channel coding cannot recover bit errors.

This phenomenon has an impact on the performance seen in the presence of quasi static fading events, as it is for the propagation channels seen in the following and terminal speeds ranging from low to medium, when time selectivity within a sub frame is negligible.

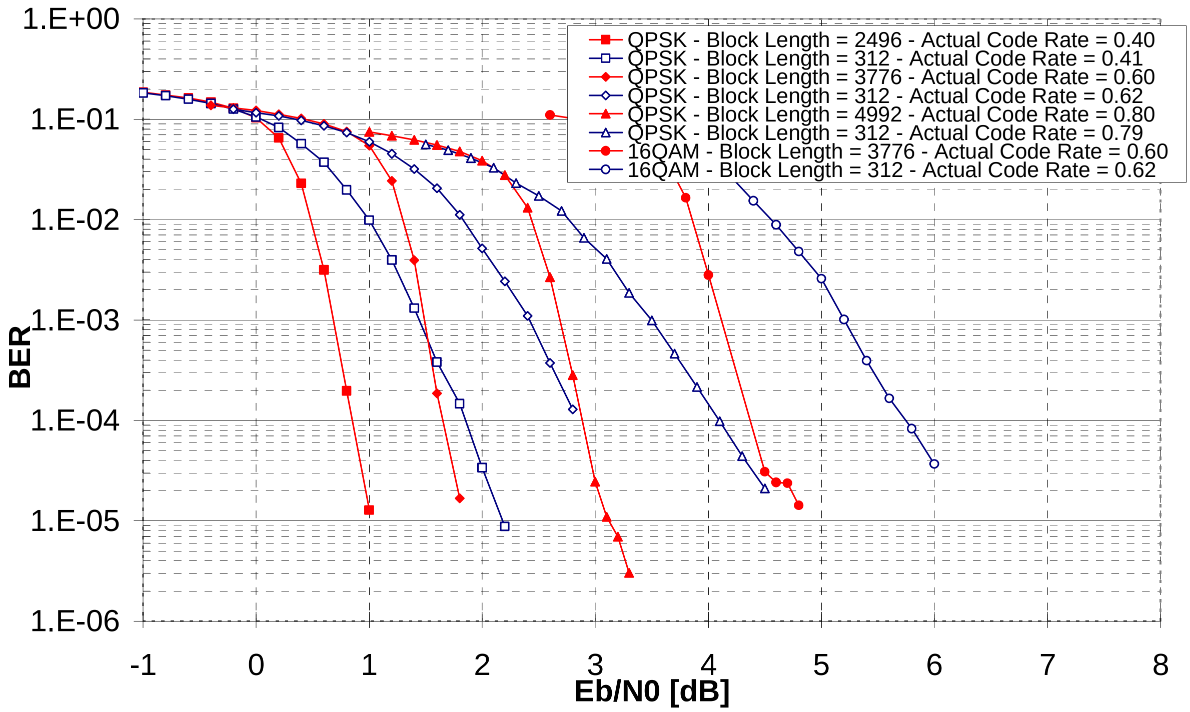


Figure I.7: BER vs. Eb/N0 [dB], LTE Broadcasting PHY configurations - Ideal Channel

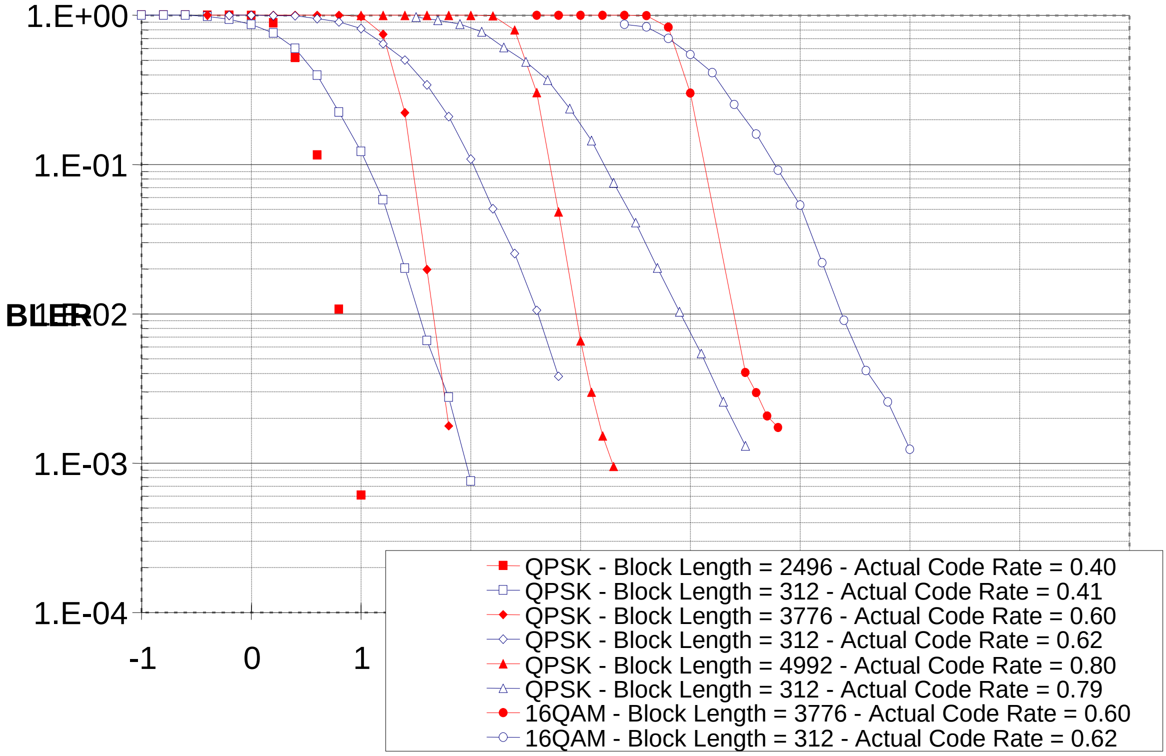


Figure I.8: BLER vs. Eb/N0 [dB], LTE Broadcasting PHY configurations - Ideal Channel

Figures I.9 and I.10 report the dual result obtained using the LDPC channel coding considered by the mobile WiMAX standard. Results are analogous to what obtained for LTE, both in terms of the gain observed when aggregating more services/channels and of the crossing at very low SNR. Due to the limited choices in terms of packet length at the input of the LDPC encoder and to the smaller block sizes, the steepness of the curves is smaller with respect to the results achieved for LTE. Even if this matters in ideal channel simulations, this effect may be less relevant or even negligible when realistic propagation channels are considered.

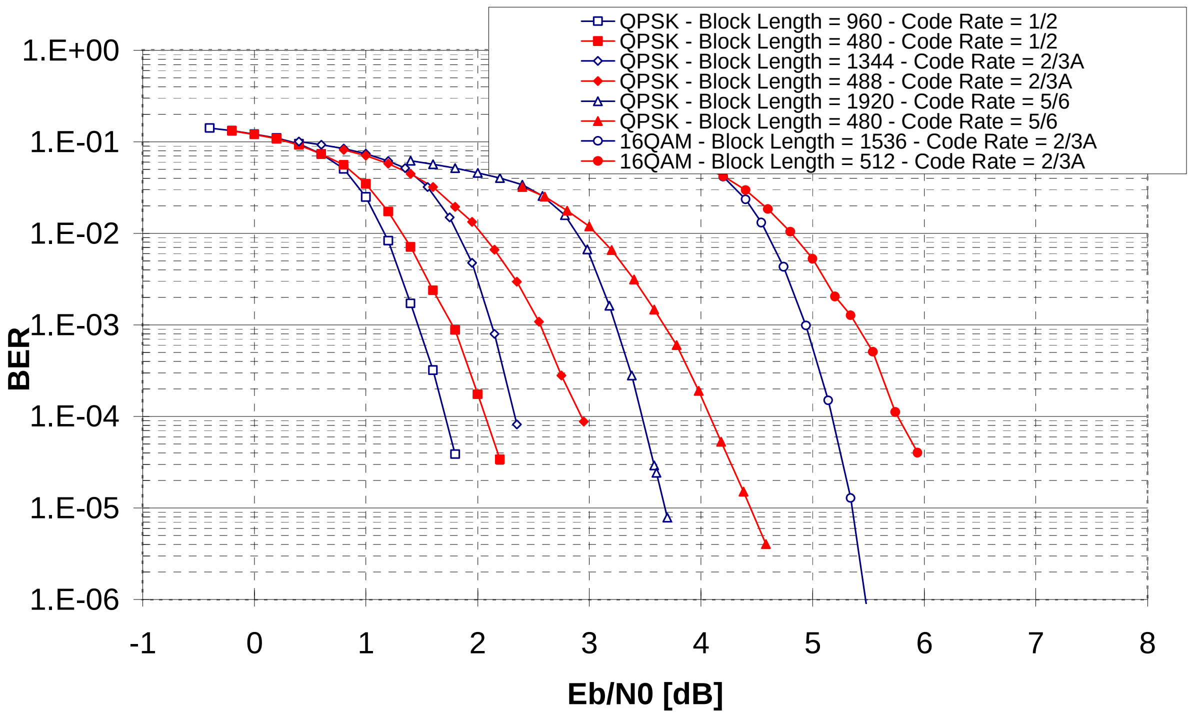


Figure I.9: BER vs. Eb/N0 [dB], WiMAX Broadcasting PHY configurations - Ideal Channel

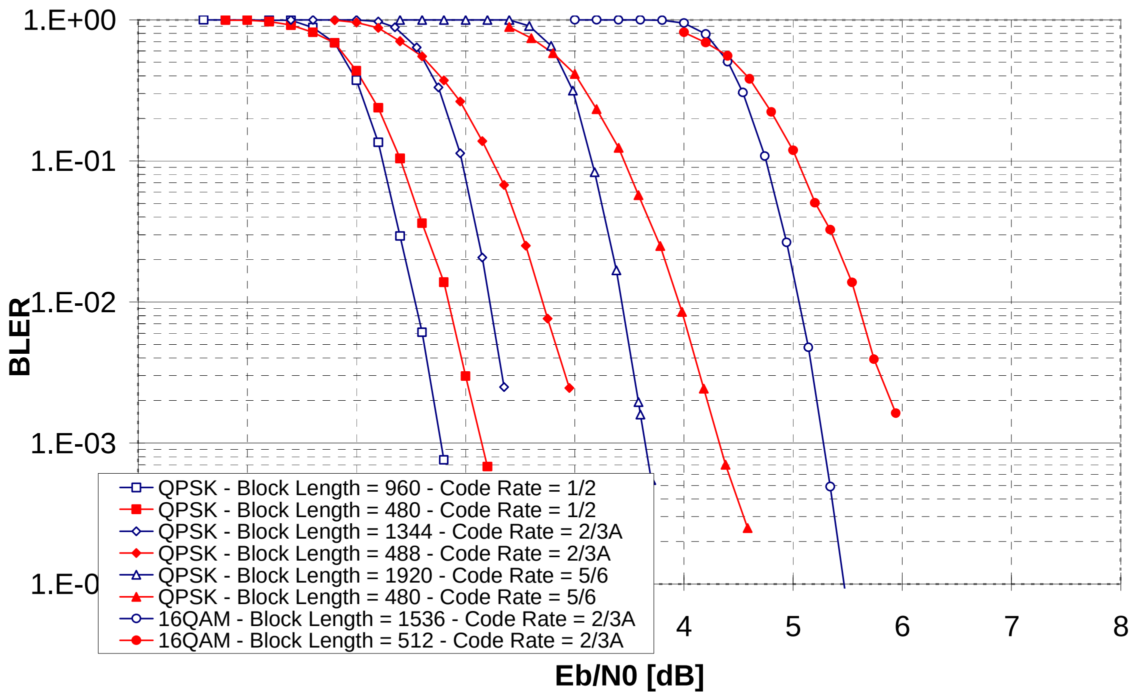


Figure I.10: BLER vs. Eb/N0 [dB], WiMAX Broadcasting PHY configurations - Ideal Channel

I.3.1.2 Broadcast scenario - Ideal Estimation

With the aim to get a close glance on the performance achievable in realistic conditions, a further step consists in introducing the effect of non linear distortion and of realistic channel conditions. Ideal channel state information has been assumed in these simulations, so as to reach the additional goal of having reference curves for the following baseline simulations.

Amongst the configurations selected for simulation in Ideal Channel conditions, a sub-set has been selected for this evaluation, considering two different code rates for QPSK and one for 16-QAM, chosen so as to allocate the full 5 MHz bandwidth to the transmission of the aggregated services. For LTE, the considered block lengths are 2 496, 4 992, and 3 776*2 (code block segmentation) for QPSK rate 2/5, QPSK rate 4/5, and 16-QAM 3/5, respectively. On the other hand, for WiMAX the considered block lengths are 960, 1 920, and 1 536 for QPSK rate 1/2, QPSK rate 5/6, and 16-QAM rate 2/3, respectively.

Figure I.11 reports the results achieved in Ideal Estimation conditions for the LTE case. Even if the required Eb/N0 @ BER 1e-3 is significantly higher with respect to ideal channel simulations, the waterfall shape is still present, as frequency diversity allows turbo decoding to converge properly. Nevertheless, the degradation with respect to the ideal channel set-up is significant, in the order of 10 dB for all configurations. This comes from the presence of non linear distortion and from multipath propagation, and from the fact that results consider 100 % of statistical cases.

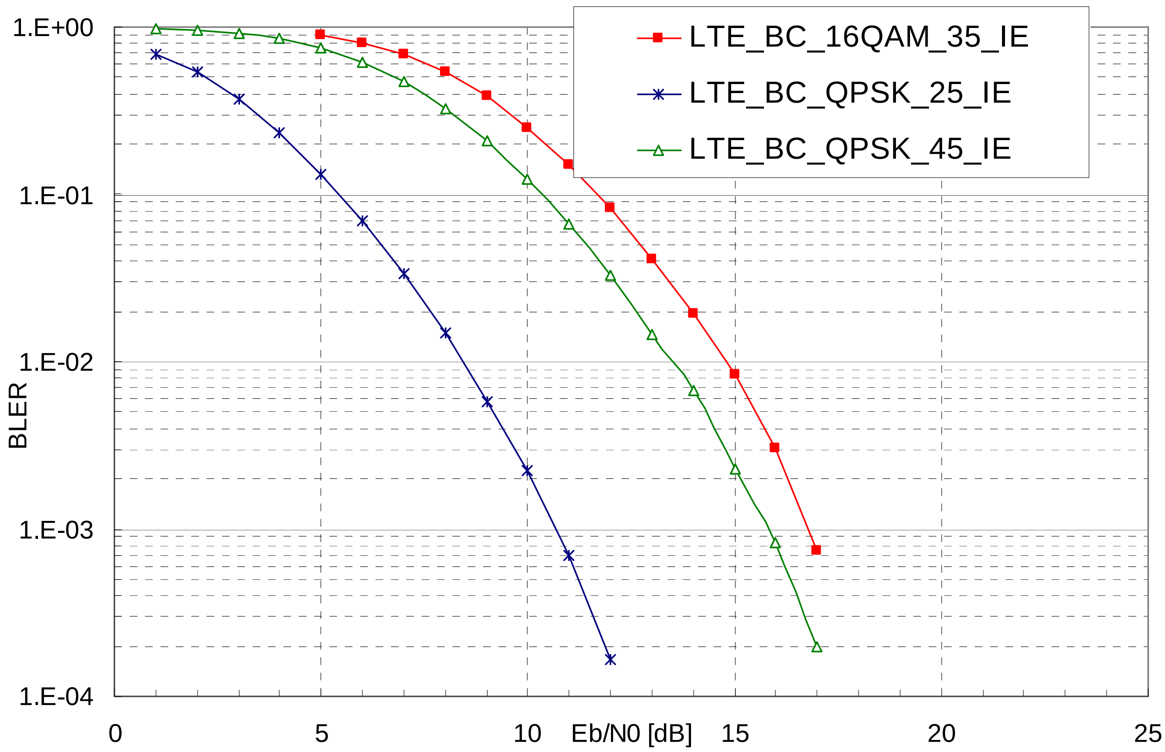


Figure I.11: BLER vs. Eb/N0 [dB], LTE Broadcasting PHY configurations - Ideal Estimation

Figure I.12 reports the results achieved in Ideal Estimation conditions for the WiMAX physical layer. Considerations analogous to what seen for LTE can be drawn. Losses of about 1 dB can be observed between the WiMAX and the LTE curves, deriving from the shorter packet length.

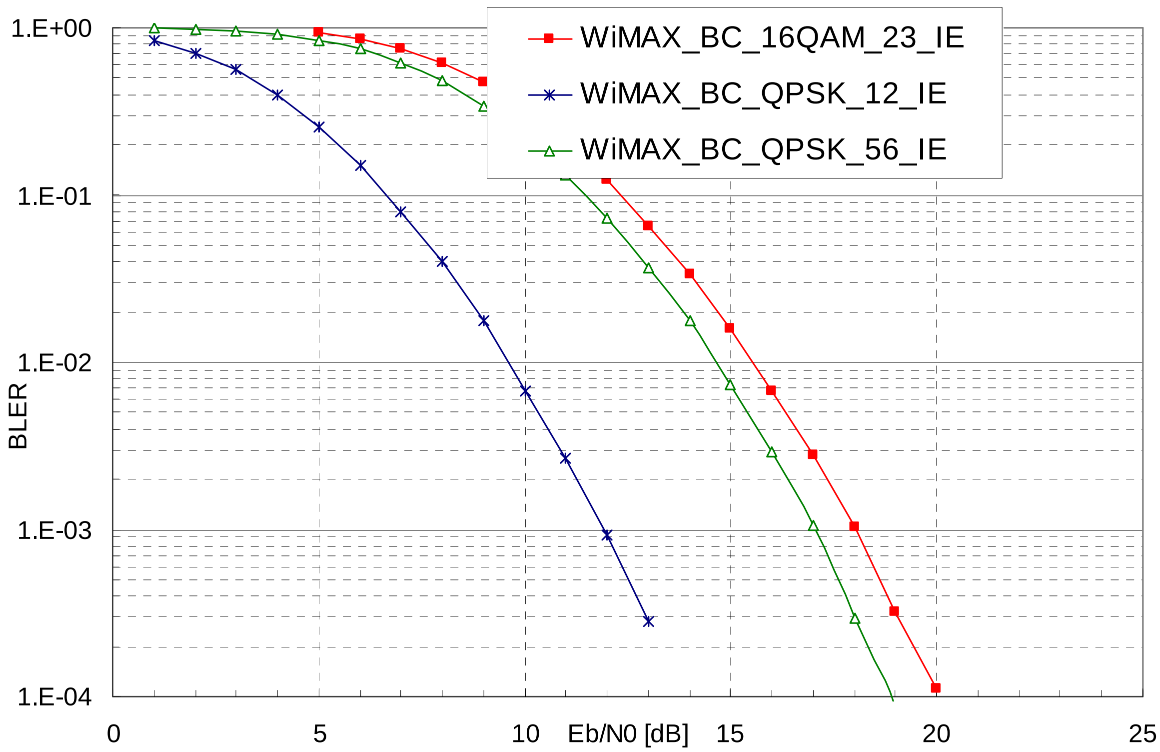


Figure I.12: BLER vs. Eb/N0 [dB], WiMAX Broadcasting PHY configurations - Ideal Estimation

I.3.1.3 Broadcast Scenario - Extended configurations

In this clause, the impact on performance of several aspects is assessed, with the aim of providing a more exhaustive view of the available system parameters and of their impact on performance.

I.3.1.3.1 Different IBO

Figure I.13 shows the performance of the LTE physical layer in the broadcasting scenario, considering 16-QAM mapping and rate 3/5 channel coding, for two different IBO values, benchmarked to the performance over a linear channel. Ideal channel estimation and no channel impairments are assumed. Taking as a reference a target BLER equal to 1e-3, performance loss with respect to linear channel is of 1,8 dB for IBO = 4 dB. When operating even closer to saturation, namely at IBO = 3 dB, an additional loss larger than 2 dB has to be accounted for, which is not compensated by the increased average output power (0,7 dB from IBO = 3 dB to IBO = 4 dB, as seen in [i.318]).

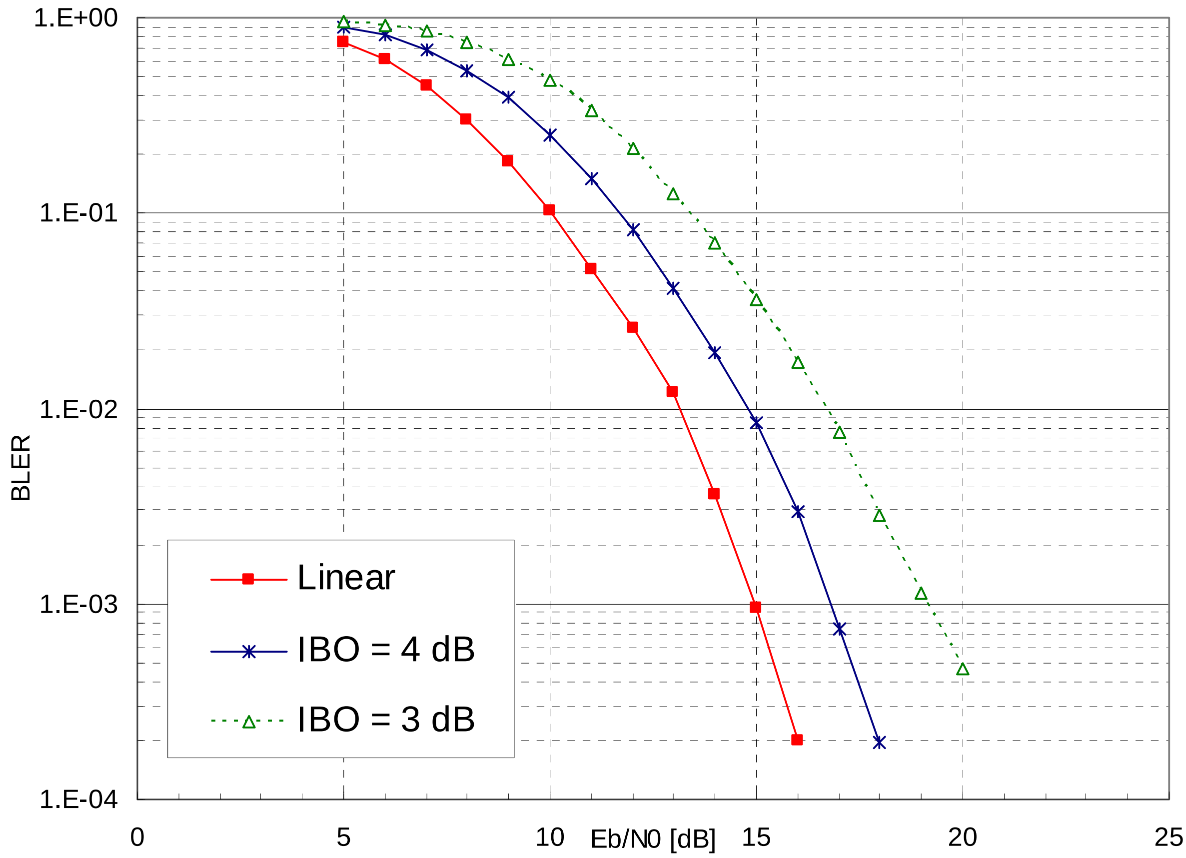


Figure I.13: BLER vs. Eb/N0 [dB], LTE Broadcasting 16-QAM - 3/5 - Impact of IBO, Ideal Estimation

I.3.1.3.2 Inter-TTI interleaving

As seen is the previous clauses, LTE and WiMAX systems are designed so as to have very short TTI intervals. The small time diversity involved with this choice is compensated by the frequency selectivity provided by the propagation channels for which these systems are designed, and by the quick channel quality estimations that can be obtained, allowing to transmit only on the tones where the received signal is strong.

In order to improve physical layer performance in the satellite broadcast scenario, a possible avenue consists in increasing the time selectivity seen by the physical layer packet through the use of inter-TTI interleaving techniques. This can be obtained in a quite straightforward way by reusing the facilities foreseen to support hybrid automatic repeat request.

In the following, simulation results are provided for two terminal speeds, namely 30 and 3 km/h, and for three different spans of inter-TTI interleaving.

Figure I.14 reports the simulation results for the LTE physical layer, when the user terminal is moving at 30 km/h, for different inter-TTI interleaving depths and sub frame size. The time diversity seen by the physical layer packets is equal to the factor K times the number of chunks in which the physical layer packet is divided. For example, when K = 4 and inter-TTI sub frame length is 1 OFDM symbol, a physical layer packet is transmitted in 12 parts once every 4 TTIs, and thus the time diversity that is seen equals 48 ms. Considering that at 30 km/h channel coherence time is equal to 9 ms, it can be seen that channel correlation is broken, allowing the turbo decode to operate properly.

This effect can be clearly seen from figure I.14 where, considering BLER @ 1e-3, a gain larger than 5 dB is achieved on the required Eb/N0 when K = 16 and transmissions of 1 OFDM symbol are used. When K = 4 and a transmission size of 3 OFDM symbols is used, the gain still exceeds 2 dB, confirming the advantages brought about by the use of this technique.

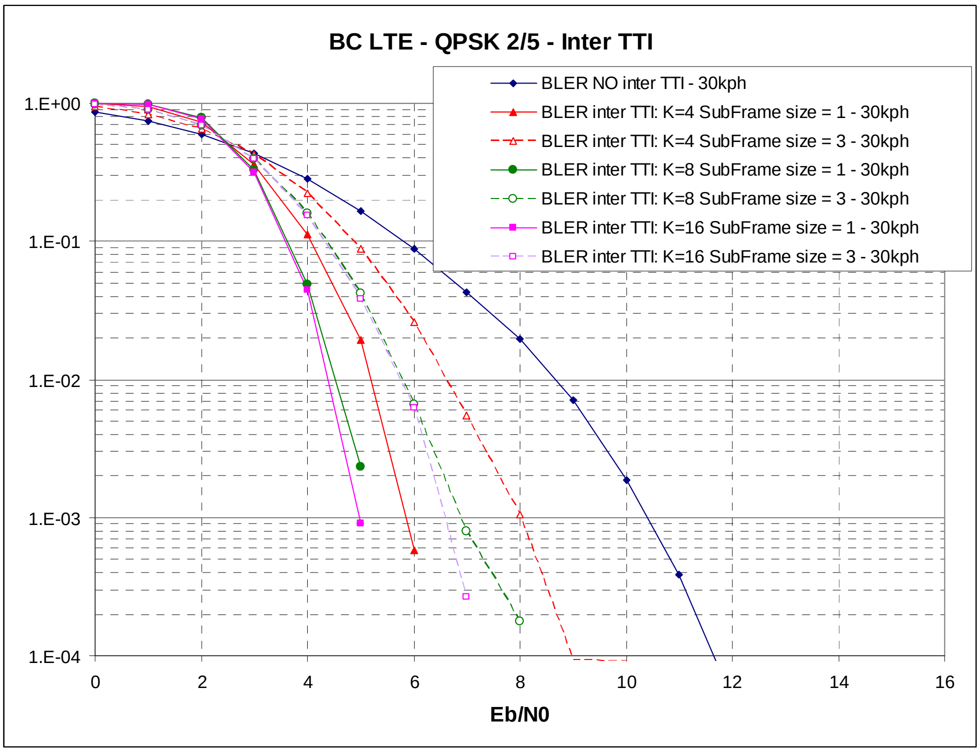


Figure I.14: BLER vs. Eb/N0 [dB], LTE Broadcasting QPSK 2/5 - 30kph - Ideal Estimation - Inter TTI.

Figure I.15 reports the results for 3 km/h. The results are analogous to what obtained for the 30 km/h case, taking into account that values of K greater of a factor 10 have been used. However, this implies an increase of a factor 10 in the required physical buffer size for a given throughput.

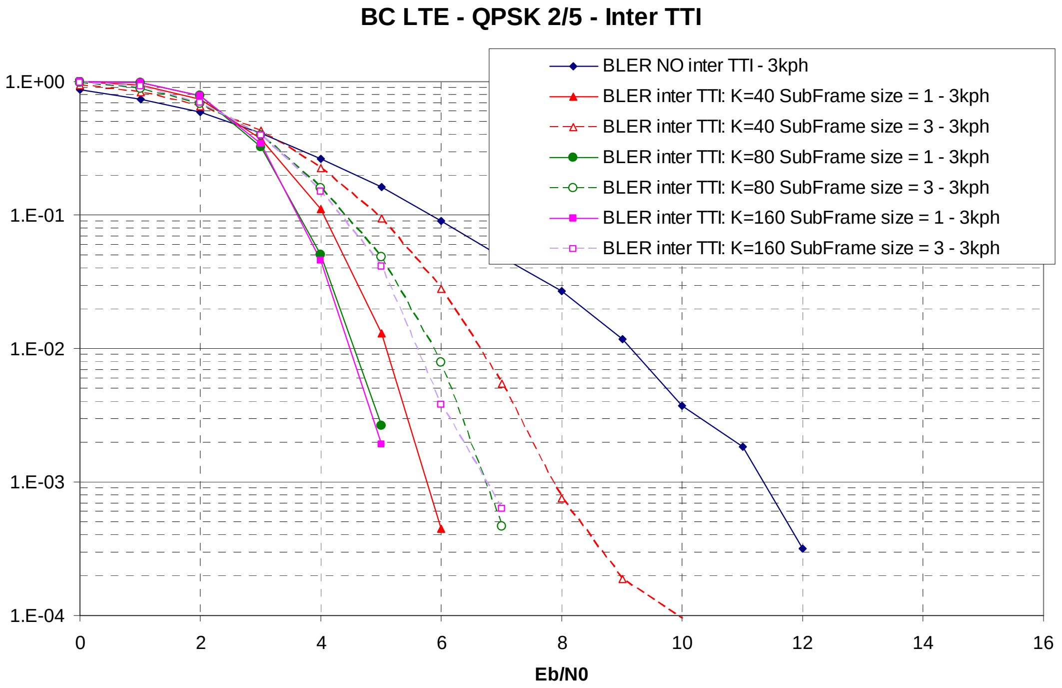


Figure I.15: BLER vs. Eb/N0 [dB], LTE Broadcasting QPSK 2/5 - 3kph - Ideal Estimation - Inter TTI.

I.3.1.3.3 PAPR Reduction

An interesting investigation which has been carried on regards the analysis of the usefulness of PAPR reduction techniques when used to mitigate the impact of non linear distortion. Figure I.16 compares the performance of the standard LTE chain with respect to a transmission chain where PAPR reduction techniques are used. As it can be seen, when operating at IBO = 4 no significant gain is obtained. Figure I.17 reports the same comparison at a lower IBO, where non linear distortion is more significant. As it can be seen, in this scenario the use of PAPR reduction techniques leads to in improvement in the BER performance. In particular, by properly choosing the target PAPR value it is possible to gain around 0,5 dB, thus reducing total degradation and allowing the HPA to be operated closer to saturation.

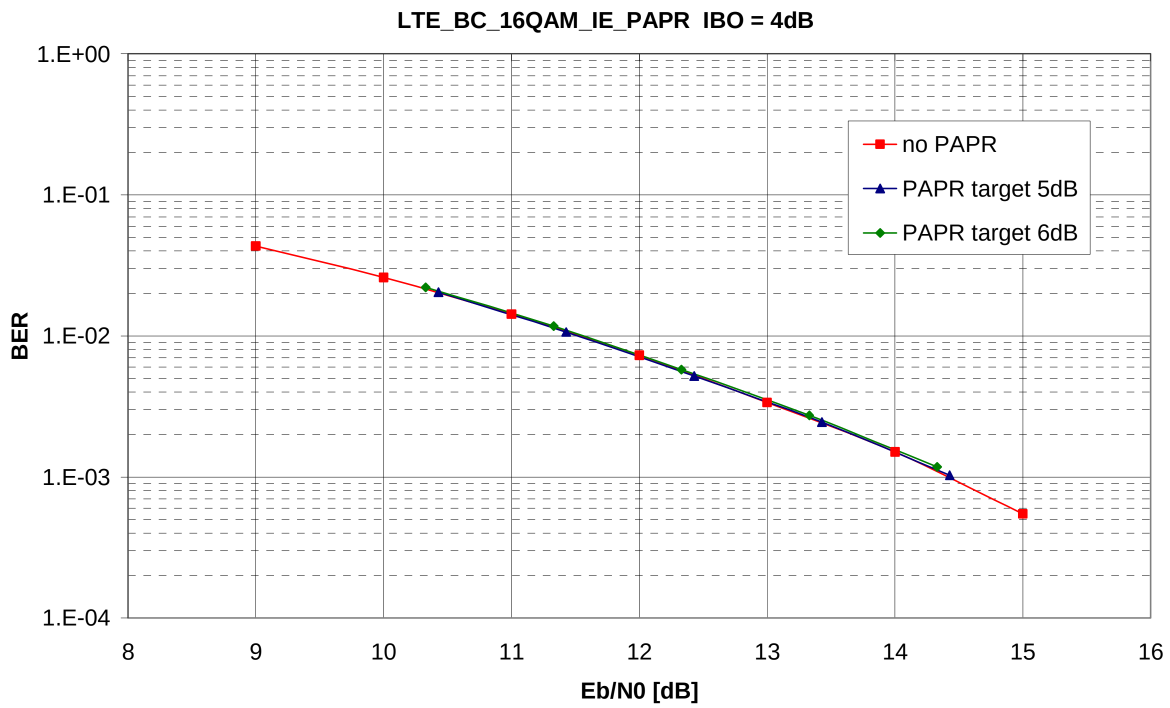


Figure I.16: BER vs. Eb/N0 [dB], LTE Broadcasting - 16-QAM 3/5 - IBO 4 dB Ideal estimation - PAPR reduction techniques

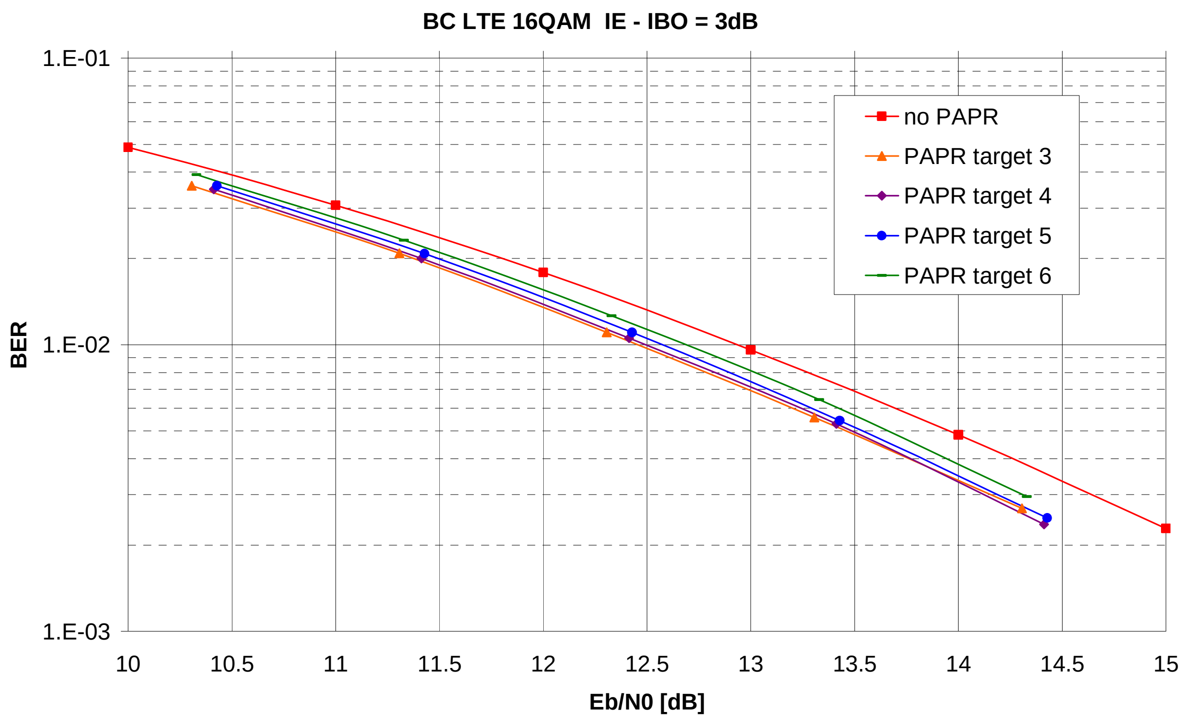


Figure I.17: BER vs. Eb/N0 [dB], LTE Broadcasting - 16-QAM 3/5 - IBO 3 dB Ideal estimation - PAPR reduction techniques

Figure I.18 reports the PAPR reduction performance for the LTE broadcasting scenario and QPSK mapping. In this case, the achieved gain is negligible, due to the higher resilience to non linear distortion of this constellation. Thus, whether to consider or not PAPR reduction techniques also depends on the mapping mix within a transmission frame.

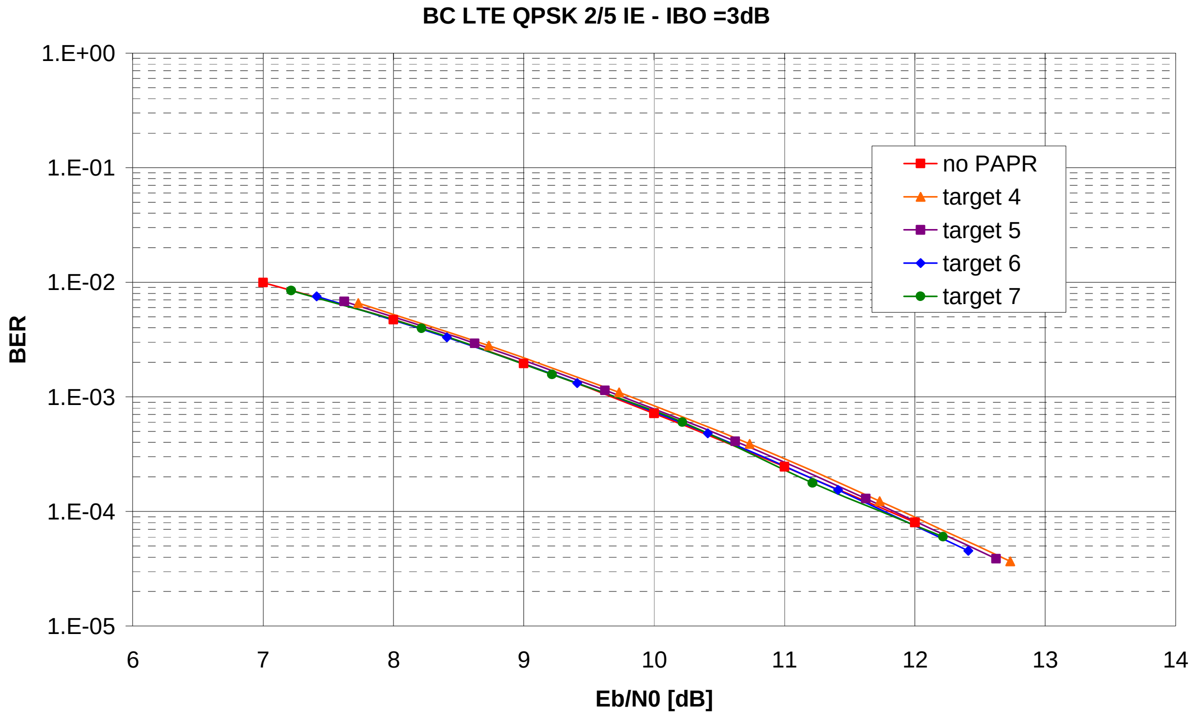


Figure I.18: BER vs. Eb/N0 [dB], LTE Broadcasting - QPSK 2/5 - IBO 3 dB Ideal estimation - PAPR reduction techniques

Finally, in figure I.19 the same comparison is carried out for the WiMAX physical layer, 16-QAM mapping with code rate 2/3. In this case the effectiveness of PAPR reduction techniques is visible when operating at IBO = 3 dB. While the larger N_{ACTIVE}/N_{FFT} ratio for the WiMAX configuration should help PAPR reduction techniques, leading to a less oversampled (and thus less correlated) signal after the transmit IFFT, the pilot boosting approach affects the attainable gain in PAPR.

Nevertheless, results for QPSK mapping with code rate 1/2 are promising, shown in figures I.21 and I.22, demonstrating the effectiveness of PAPR reduction techniques even when the HPA is driven with an IBO = 3 dB.

The results of this analysis agree with [i.317], in which the energy increase and the effectiveness of PAPR reduction is discussed. The used approach in PAPR reduction does not yield a reduction in spectral efficiency or an increased complexity at the receiver, at the expense of a slight increase in mean energy. Notwithstanding all these interesting property, when the amplifier is driven near to its linear region, the degradation due to non-linearity is small, and so the decrease in BER does not compensate the increase in signal energy. On the contrary, when the amplifier is driven near saturation, the PAPR reduction (joint with a pre-distortion technique, as suggested in [i.317]) becomes effective and the BER is decreased. This gain can be seen more clearly when the SNR is high, that is to say when the main cause of degradation is the distortion due by HPA.

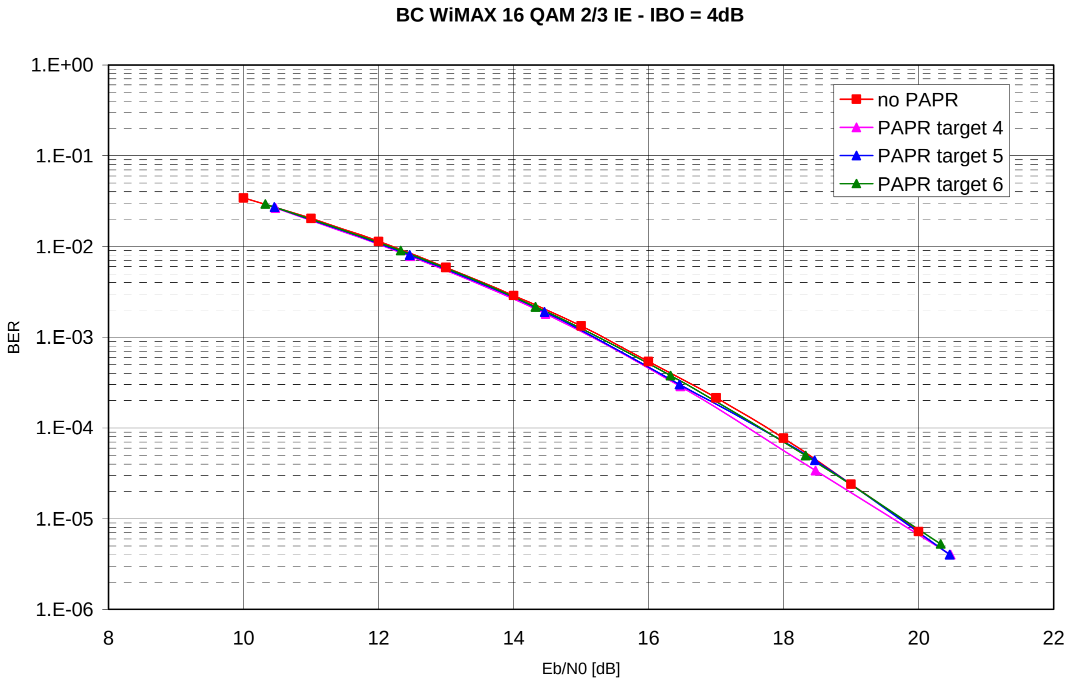


Figure I.19: BER vs. Eb/N0 [dB], WiMAX Broadcasting 16-QAM 2/3 - IBO 4 dB Ideal estimation - PAPR reduction techniques

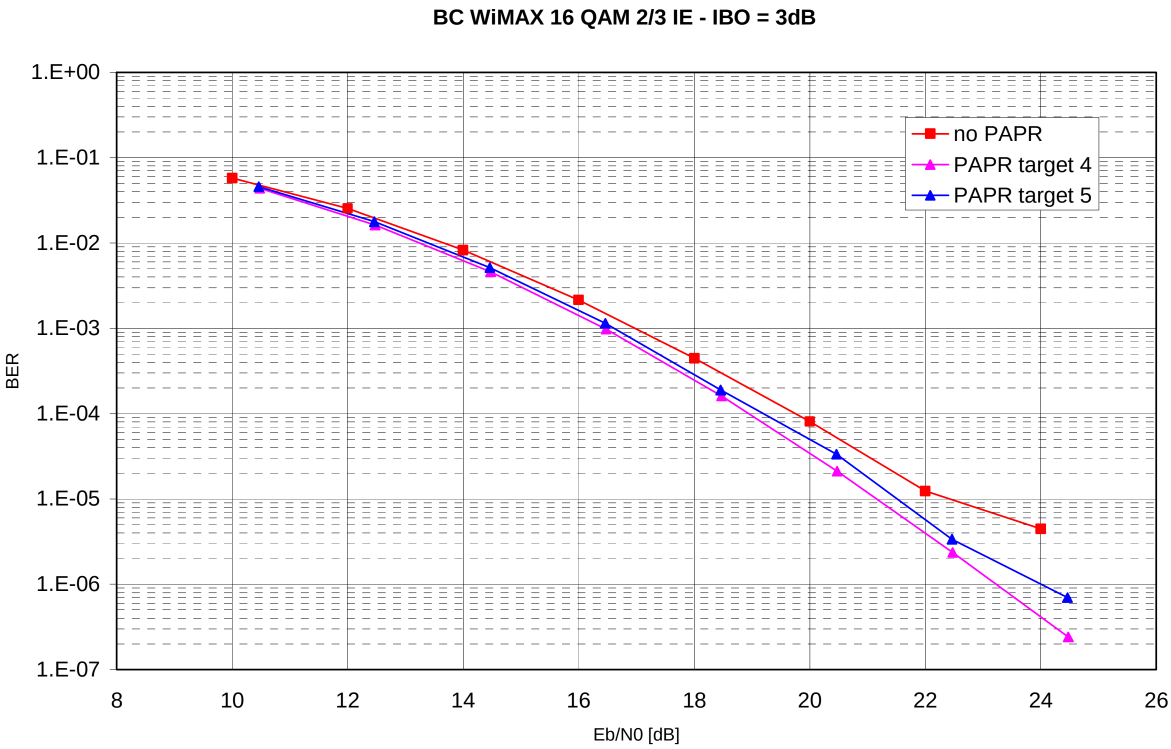


Figure I.20: BER vs. Eb/N0 [dB], WiMAX Broadcasting 16-QAM 2/3 - IBO 3 dB Ideal estimation - PAPR reduction techniques

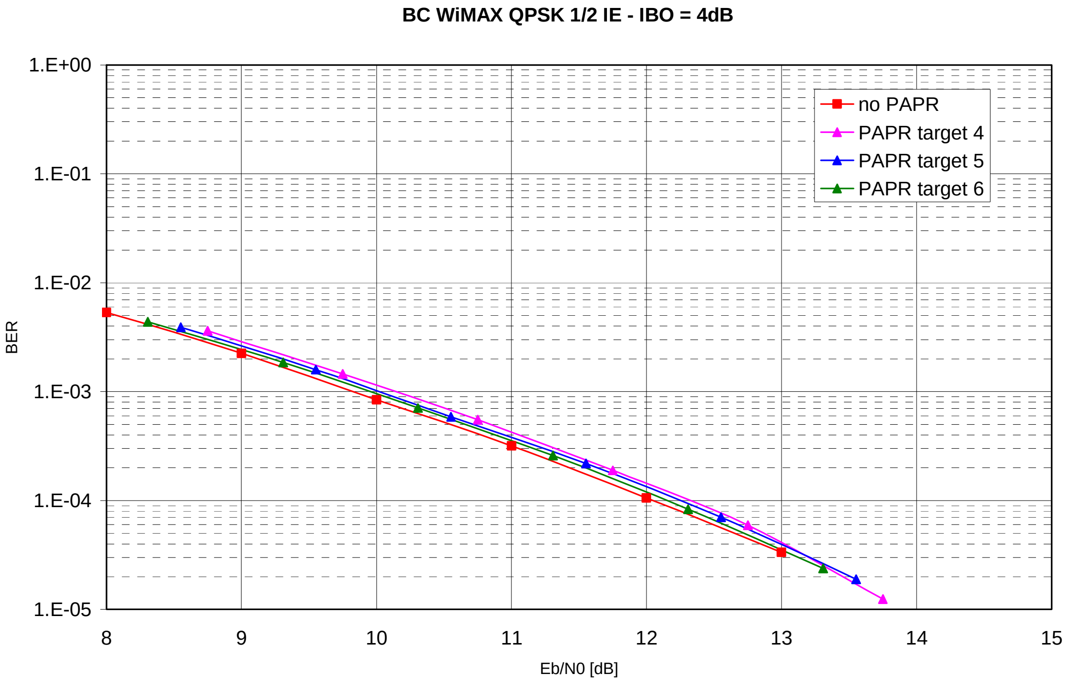


Figure I.21: BER vs. Eb/N0 [dB], WiMAX Broadcasting QPSK 1/2 - IBO 4 dB Ideal estimation - PAPR reduction techniques

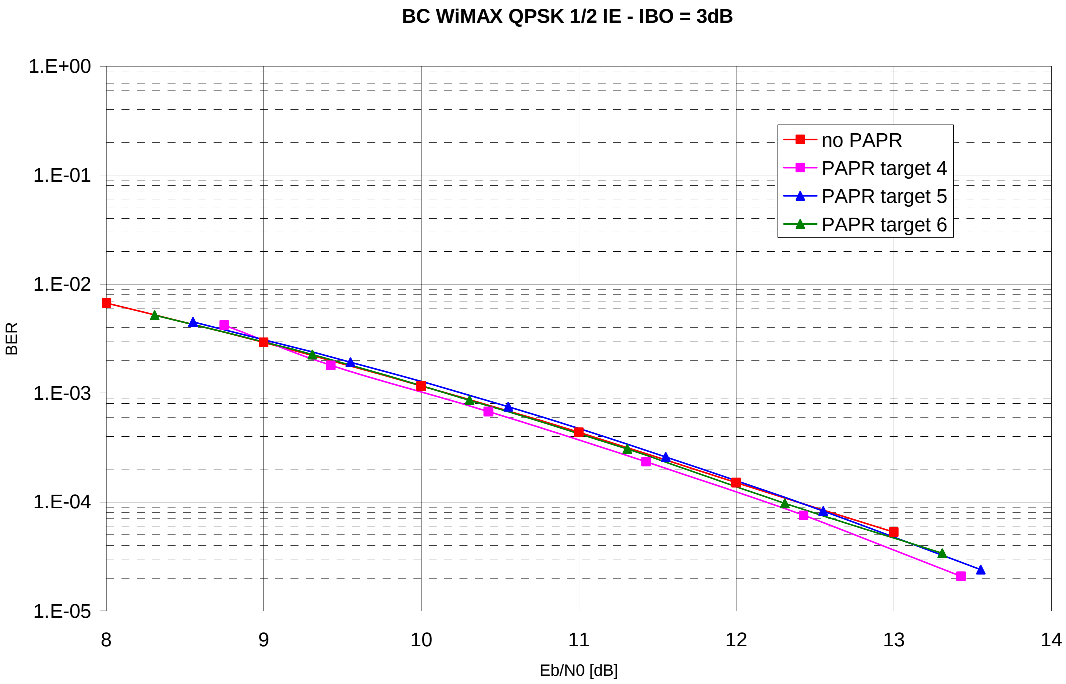


Figure I.22: BER vs. Eb/N0 [dB], WiMAX Broadcasting QPSK 1/2 - IBO 3 dB Ideal estimation - PAPR reduction techniques

I.3.2 Two-Way Communications FL - PHY Results

This clause reports the results of the physical layer simulations performed for the forward link of the two-way communications scenario. The results, reported in terms of bit error rate and block error rate, are commented and considerations are drawn.

I.3.2.1 Two-Way communications FL - Ideal Channel

Figure I.23 reports the performance in terms of BLER for the two-way communications LTE FL Ideal Channel simulations. Results are shown for three different coding/modulation pairs. Differently from what has been assumed for the broadcast scenarios where the full bandwidth was allocated to the considered services, in the two-way communications scenarios almost constant bit-rate has been assumed, resulting in a partial allocation of the available bandwidth (1,25 MHz) to the considered users. Clearly, for higher efficiency mod-cod pairs, a larger number of users can be allocated in the allocated bandwidth.

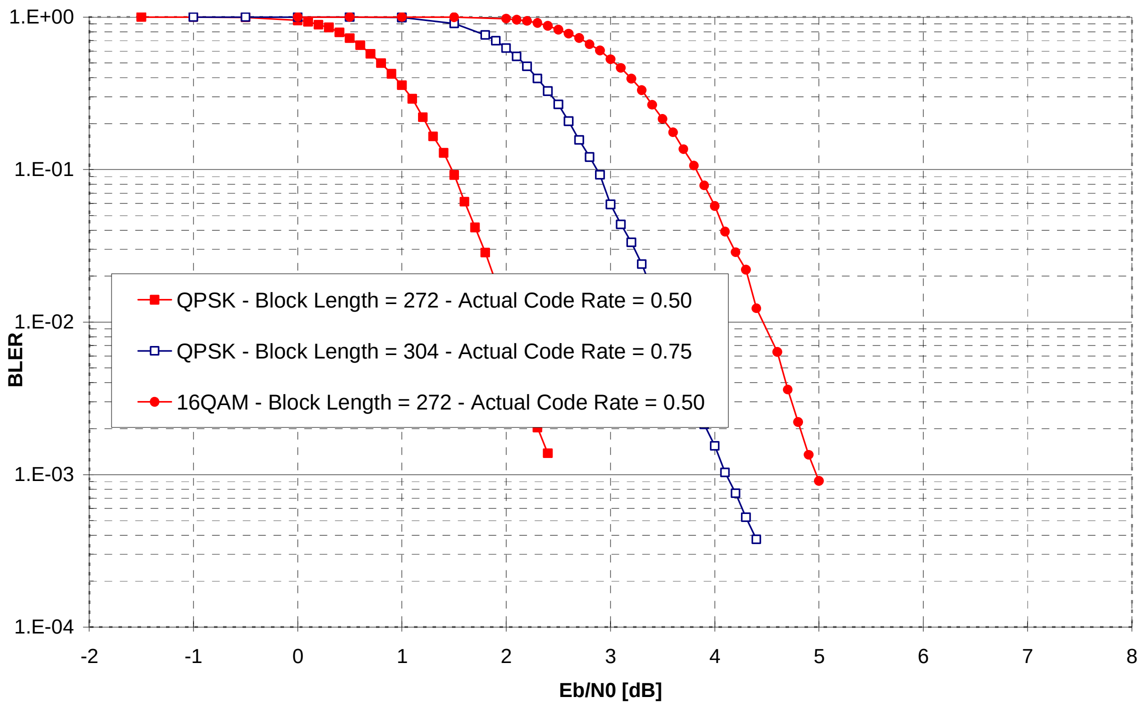


Figure I.23: BLER vs. Eb/N0 [dB], LTE Two-Way FL PHY configurations - Ideal Channel

Regarding the WiMAX physical layer results for the ideal channel configurations, results are shown in figure I.24 and are in line with what obtained for LTE. In this case, block lengths are larger than what has been chosen for LTE so as to maintain the same source data rate, as the obtained coded bits are mapped onto a TTI duration that is almost 50 % longer (1,49 ms instead of 1 ms).

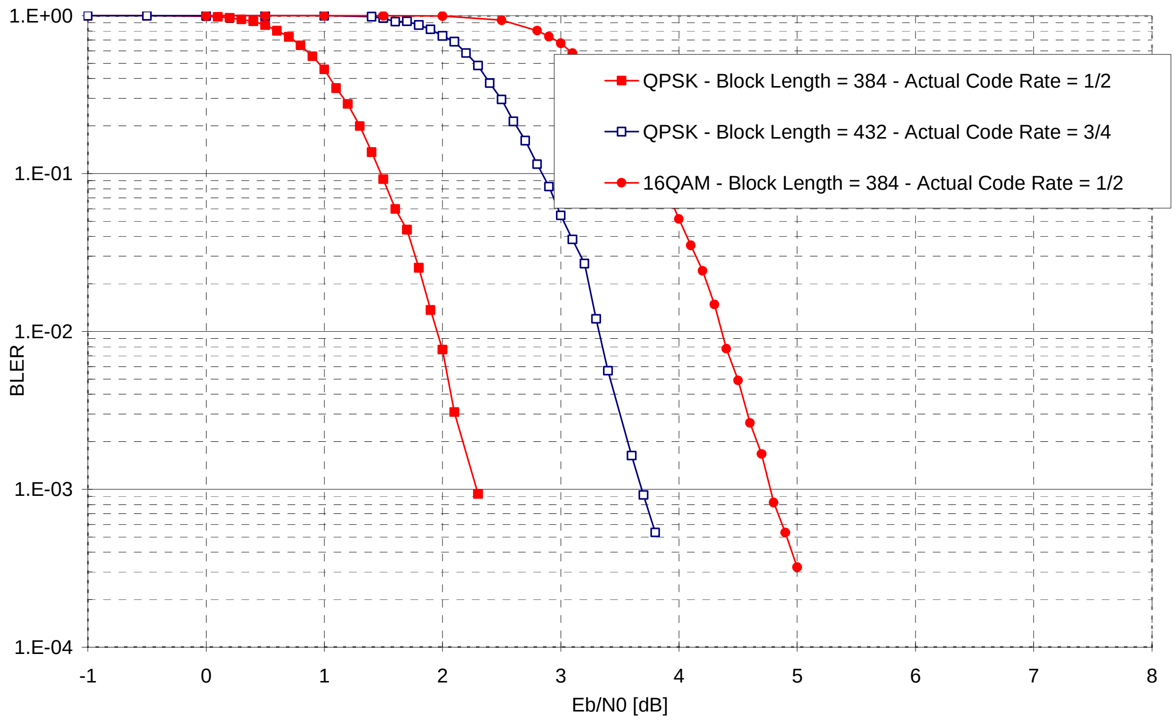


Figure I.24: BLER vs. Eb/N0 [dB], WiMAX Two-Way FL PHY configurations - Ideal Channel

I.3.2.2 Two-Way communications FL - Ideal Estimation

Figures I.25 and I.26 clearly show that in this configuration the waterfall shape cannot be achieved. This is due to the fact that frequency selectivity is not present, and channel coherence time is significantly larger than the duration of a physical layer packet, and time interleaving provided by standard configurations is not sufficient.

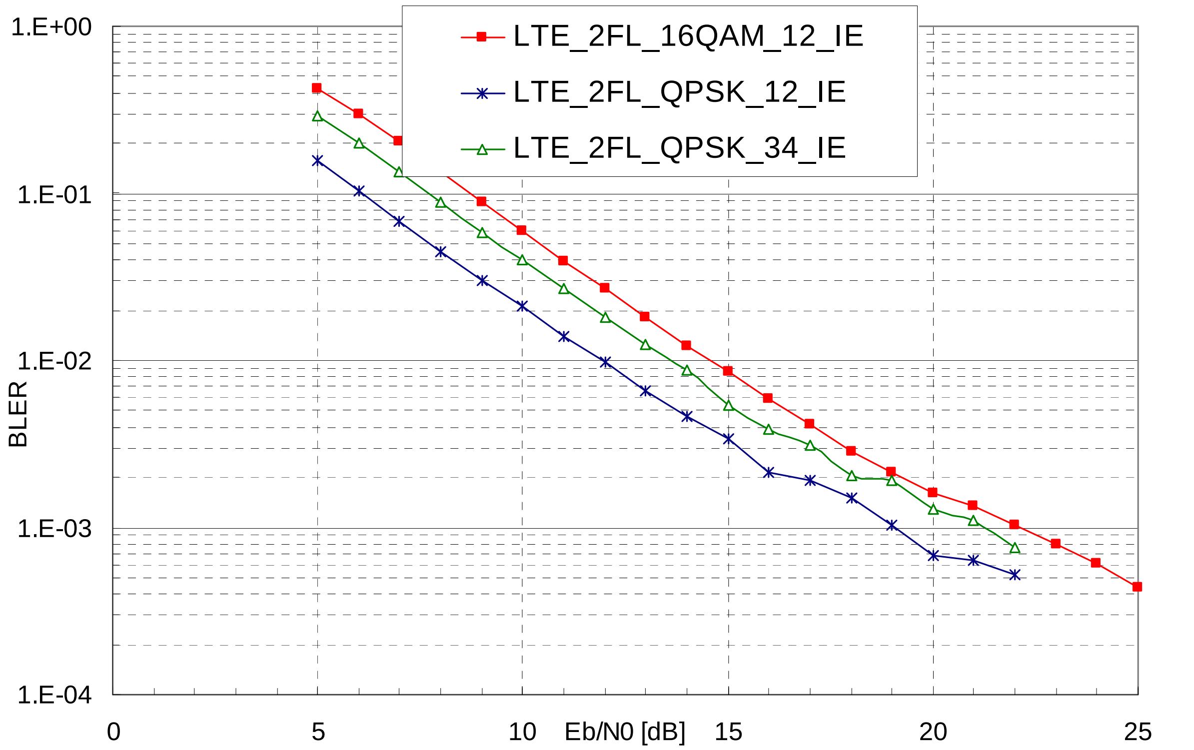


Figure I.25: BLER vs. Eb/N0 [dB], LTE Two-Way FL PHY configurations - Ideal Estimation

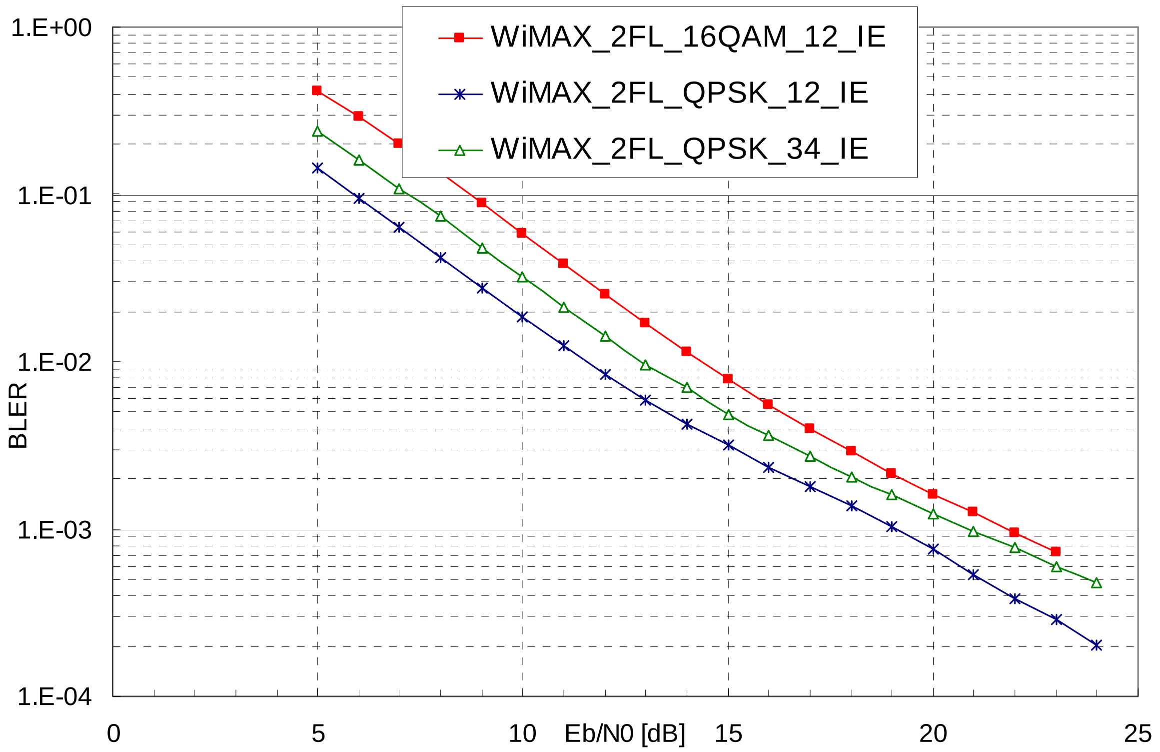


Figure I.26: BLER vs. Eb/N0 [dB], WiMAX Two-Way FL PHY configurations - Ideal Estimation

I.3.2.3 Two-Way communications FL - Extended configurations

For the two-way communications scenario, the use of MIMO techniques has been evaluated.

I.3.2.3.1 MIMO TD and SM

This clause reports the performance of MIMO Transmit Diversity and Spatial Multiplexing techniques for the LTE two-way communications scenario.

The assumptions which have been made to model satellite diversity have followed what described in [i.30] assuming to have single state channel model for each links and no spatial correlation between the two paths. As it can be seen from figure I.27, the use of MIMO Transmit Diversity techniques using two transmission satellites can greatly improve the physical layer resilience to channel fades, leading to significant improvements in terms of required Eb/No.

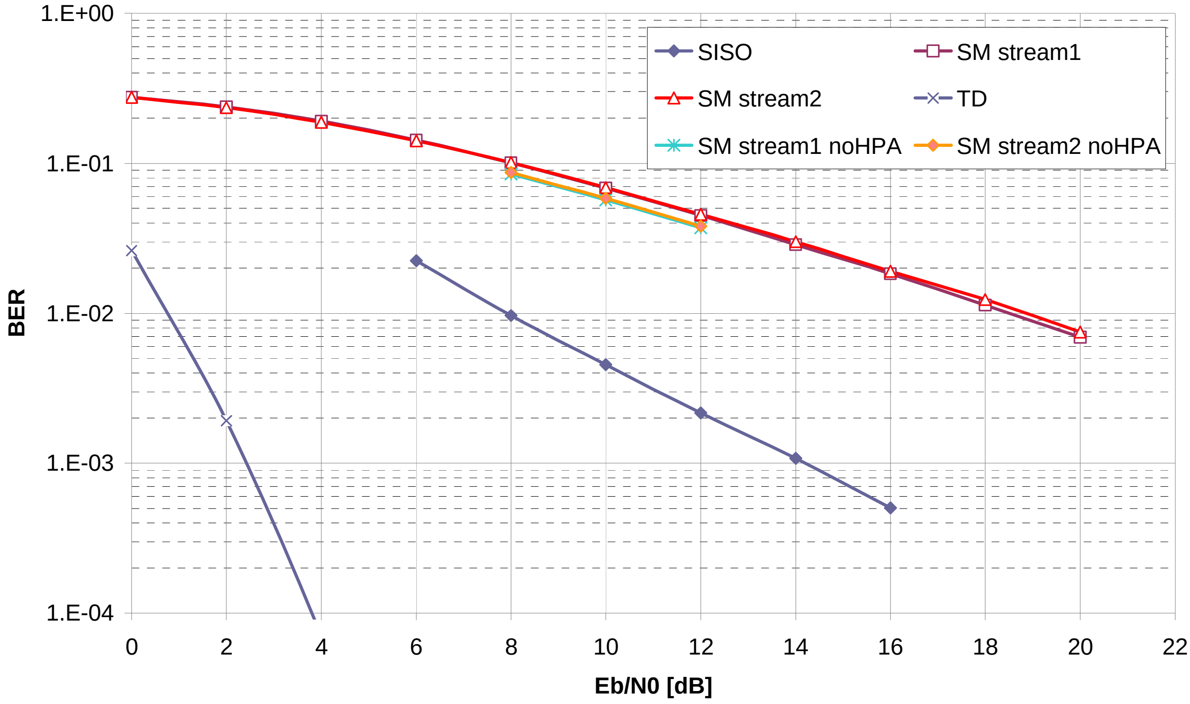


Figure I.27: BER vs. Eb/N0 [dB], LTE Two-Way FL PHY configurations Ideal Estimation - MIMO TD & SM - QPSK - Code rate 1/2

I.3.3 Two-Way Communications RL - PHY Results

In the following, results for the return link of the two-way communications scenario are reported, starting from Ideal Channel results and then extending the analysis to the more realistic Ideal Estimation results.

I.3.3.1 Two-Way communications RL - Ideal Channel

Figure I.28 report the performance in terms of BLER for the two-way communications LTE RL Ideal Channel simulations. Results are shown for three different coding rates, while only QPSK modulation is considered. As assumed for the return link, a system bandwidth of 1,25 MHz has been assumed. The different spectral efficiencies of the three configurations leads to a different number of users can be co-allocated in a physical layer sub frame, as specified in the physical layer configurations set up.

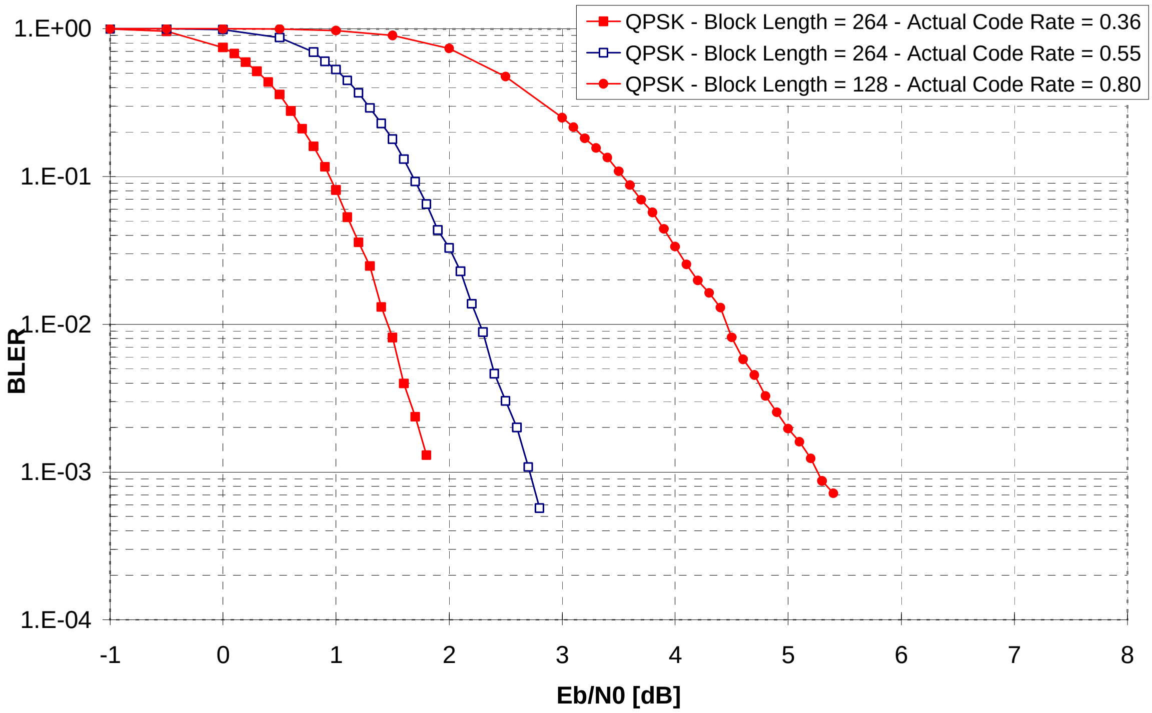


Figure I.28: BLER vs. Eb/N0 [dB], LTE Two-Way RL PHY configurations - Ideal Channel

Regarding the WiMAX physical layer results for the ideal channel configurations, results are shown in figure I.29 and are in line with what obtained for LTE.

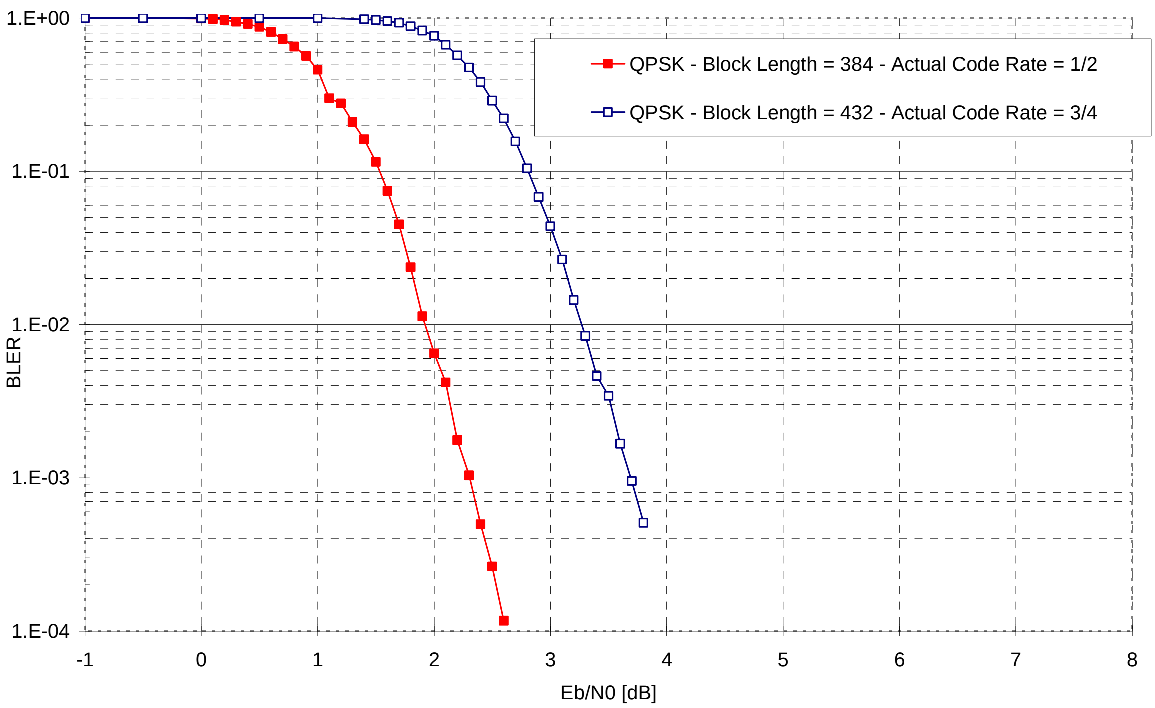


Figure I.29: BLER vs. Eb/N0 [dB], WiMAX Two-Way RL PHY configurations - Ideal Channel

I.3.3.2 Two-Way communications RL - Ideal Estimation

As seen for the forward link, when the propagation channel is inserted in the chain performance in terms of BER tends to converge for all the considered code rates. This is due to the behaviour of different code-rate configurations in ideal channel: while higher code rates start converging at higher Eb/N0 values, their BER performance at very low SNR is better than the performance of lower code rate configurations. For this reason, when average performance with respect to fading is simulated, results in terms of BER results to be very similar. This effect is not present in PER results, as the PER values at low Eb/N0 values are aligned (equal to one) for the three code rates.

These considerations hold both for LTE and for WiMAX results, which are reported in figures I.30 and I.31.

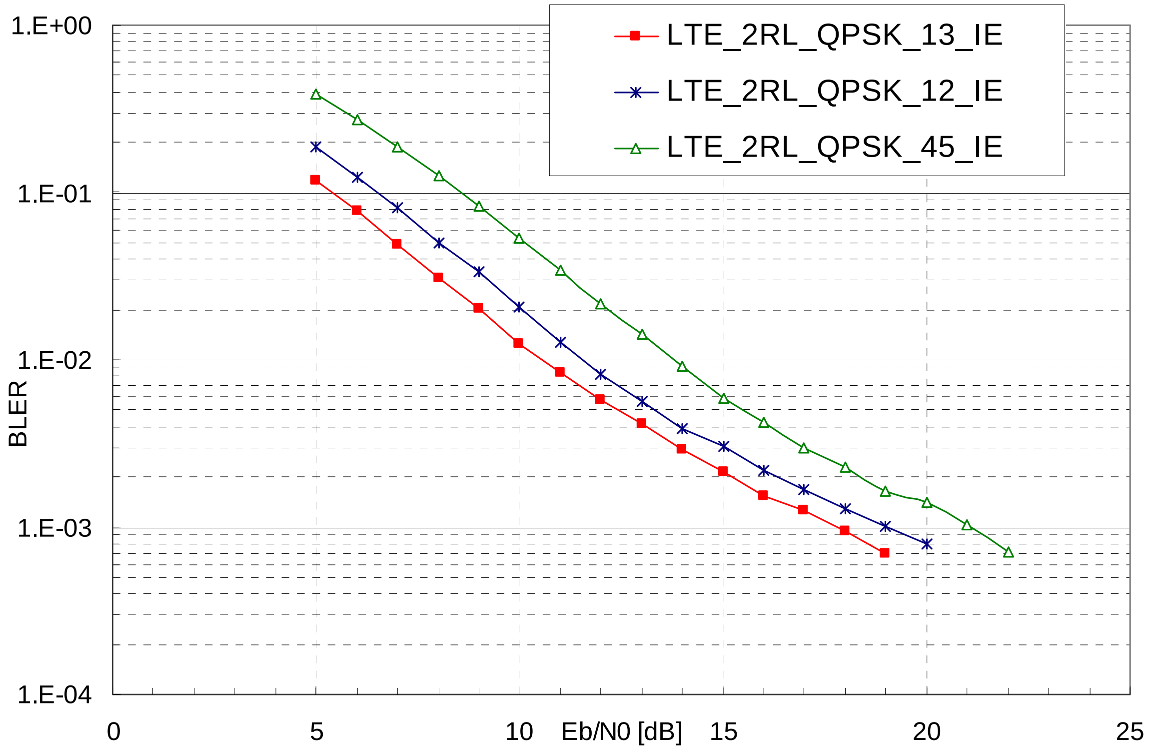


Figure I.30: BLER vs. Eb/N0 [dB], LTE Two-Way RL PHY configurations - Ideal Estimation

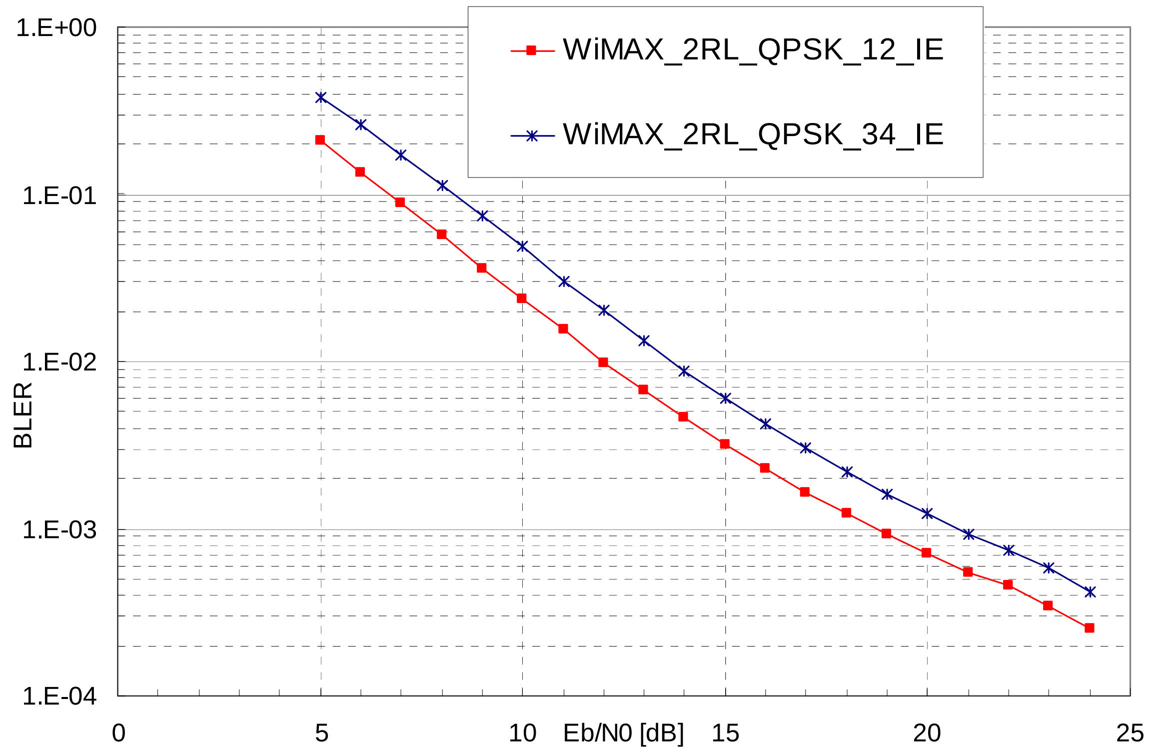


Figure I.31: BLER vs. Eb/N0 [dB], WiMAX Two-Way RL PHY configurations - Ideal Estimation

I.4 Upper Layer FEC study

This clause considers the application of UL FEC protection to increase the resilience of the PHY in urban or LMS propagation conditions.

First a description of the considered technique is given. Then, its performance is evaluated in three different cases:

- BEC channel.
- Urban SFN.
- LMS.

I.4.1 Description of the considered UL-FEC Technique

Channel coding can be performed at different layers of the protocol stack (see figure I.32). When coding is applied at higher layers with respect to the physical layer, the symbols to be encoded are group of bits, namely packets.

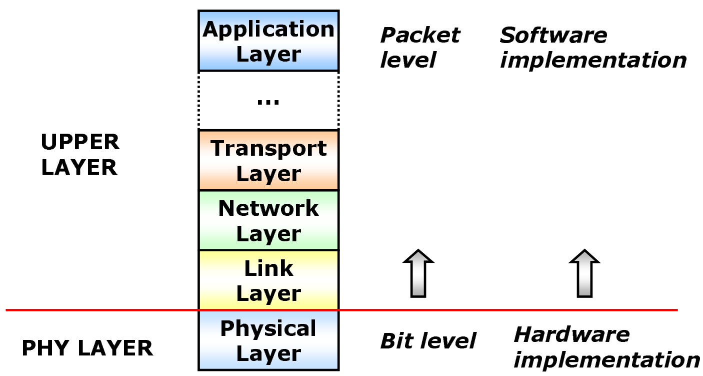


Figure I.32: Protocol stack and coding

Therefore, the upper layer coding consists in applying a forward error correction code to packets of bits. For this reason it is usually called **packet coding**. A basic block diagram of a packet coding scheme is depicted in diagram. The encoder takes *K* source packets as input, and produces *N* encoded packets that will be transmitted over the wireless channel. In this case, since we are dealing with packets, the channel can be modelled as a **binary erasure channel** (or packet erasure channel when packets are taken into account). This channel is characterized by two possible transitions between the transmitted and the received packets: a packet can be correctly received with probability 1-*p*, or can be erased with probability *p*. Indeed, since each packet is protected by a CRC field, the receiver is able to detect errors and discard the corrupted packets. If *e* encoded packets are erased by the channel the decoder will receive *n-e* packets. Depending on the amount of lost data, the decoder will be able or not to recover all the source data transmitted. Therefore the FEC schemes adopted for upper layer encoding are *erasure block codes*, since they aim at recovering losses of data, rather then correct erroneous bits. The most widely used erasure codes are the following:

- Reed Solomon (RS) Codes;
- Low Density Parity Codes (LDPC);
- Fountain codes.

The RS codes achieves ideal protection against packet loss since they are Maximum Distance Separable (MSD) code, which means they are able to decode when at least *k* packets over *n* are correctly received. LDPC and Fountain codes are less efficient in the sense that they require at least $k+\epsilon$ (ϵ is the code inefficiency) for decoding, but they are less computational demanding, allowing to operate with very large source blocks.

As already said, since the upper layer coding deals with large amount of data, it appears to be a very strong countermeasure against the long error burst due to large-scale fading.

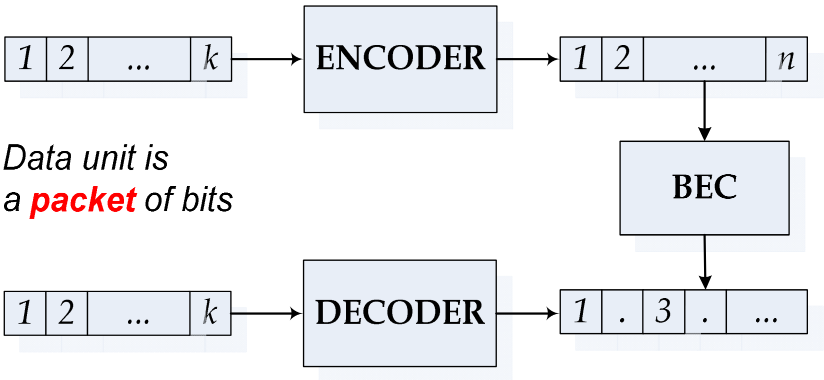


Figure I.33: Upper Layer block diagram

In the rest of the present document, we adopt the following notation:

UL Symbols: symbol @UL; each UL symbol is composed of 8 bits.

K: the UL block length; that is the number of systematic symbols to be encoded by the UL encoder.

N: the UL codeword length; that is the number of UL symbols produced by the UL encoder.

K': the actual UL-FEC block length if zero-padding is applied.

N': the actual UL-FEC codeword length if zero-padding and/or puncturing is applied.

UL-FEC Frame: the UL frame. Each UL frame represents an UL codeword; that is composed of *N* UL symbols.

N_{JCC}: Number of jointly coded channels at physical layer i.324.

S_{JCC}: channel size in bytes.

$S_{UL\text{-}CRC}$: size of the upper layer CRC in bytes.

$S_{PHY\text{-}CRC}$: size of the physical layer CRC in bytes.

K_{PHY}: physical layer block length in bytes i.324.

UL-FEC matrix: the set of systematic and parity UL-symbol organized in a matrix.

Z: number of zero padded columns in the UL-FEC matrix.

P: number of punctured columns in the UL-FEC matrix.

I.4.1.1 Transmitter Side

Figure I.34 shows a block diagram of the UL-FEC implementation at the transmitter side. In particular, the picture refers to the case in which 8 channels are gathered to compose a physical layer information block.

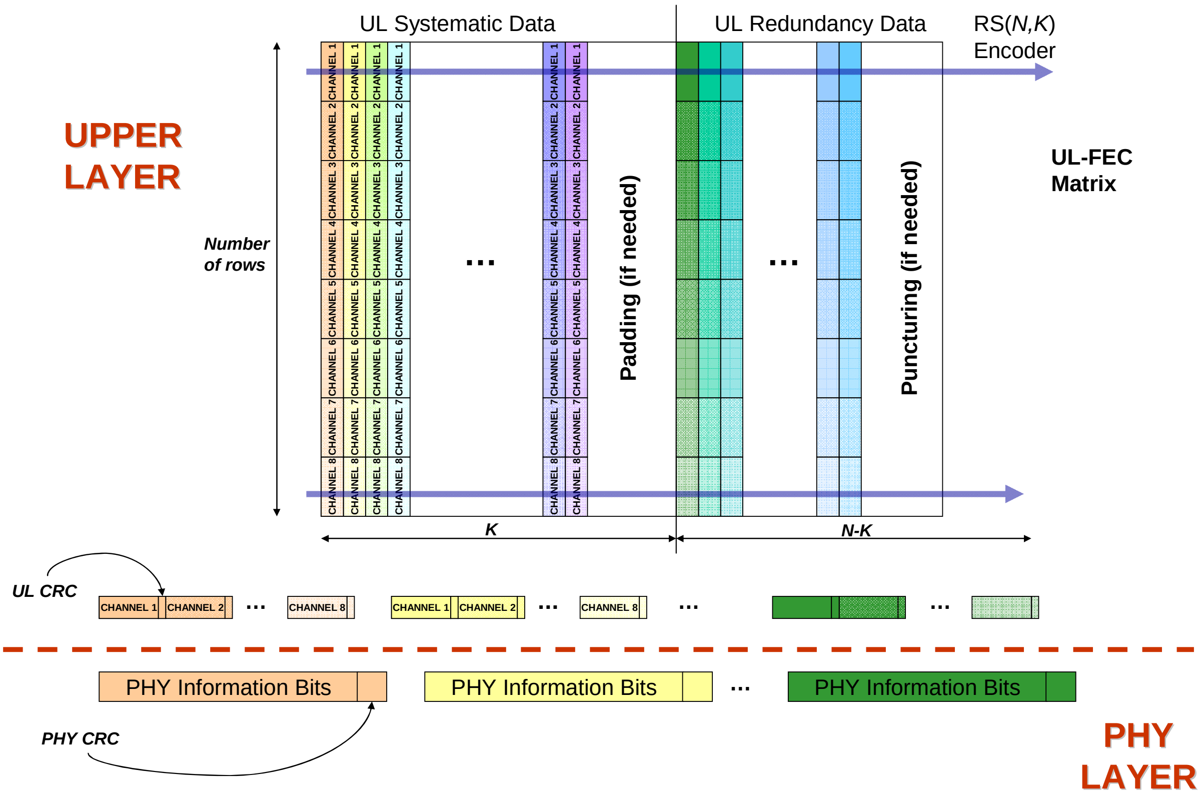


Figure I.34: Block Diagram of the proposed Upper Layer FEC technique (Transmitter side)

As in MPE-FEC, we define the UL-FEC matrix as a matrix composed of a variable number of rows (*n_of_rows*) and *N* columns. Each entry of the matrix is an UL-symbol, i.e. 1 byte. The first *K* columns represent the systematic part of the matrix and are filled with the systematic UL-symbol coming from the higher level. The last *N-K* columns carry the redundancy data computed on the first K columns. It is worth notice that *N and K* depend on the UL adopted code only, while *n_of_rows* is a parameter chosen accordingly to the PHY configuration and is set by using the following formula:

- $n\_of\_rows = K_{PHY} - S_{PHY\text{-}CRC} - N_{JCC} * S_{UL\text{-}CRC}$.
- As a consequence the number of bytes available for each channel in a given UL-FEC matrix column is $S_{JCC} = n\_of\_rows/N_{JCC}$.

With this configuration, the following operations can be performed:

- The information data coming from higher layer are written columns-wise in the systematic data part of the UL-Matrix.
- A RS(*N*,*K*) encoding is performed on each row producing the redundancy part of the UL-FEC matrix
- Data are transmitted column wise.
- In the "broadcast scenario", an UL CRC is appended after each group of S_{JCC} bytes in order to protect each channel a specific CRC (see figure I.34).

- Each group of $K_{PHY} = N_{JCC}*(S_{JCC} - S_{UL\text{-}CRC})$ bytes composes a PHY information packet.
- The PHY-CRC is appended to each PHY information packet according to the LTE or WiMAX physical layer specifications.

For sake of simplicity, we adopted the same RS mother code provided in [i.321], that is a RS(255,191) code with the following parameters:

Code Generator Polynomial: $g(x) = (x+\lambda_0)(x+\lambda_1)(x+\lambda_2)...(x+\lambda_{63})$,where $\lambda = 02HEX$

Field Generator Polynomial: $p(x) = x^8 + x^4 + x^3 + x^2 + 1$.

The code rate of the mother code is 3/4. Further code rates can be achieved by using padding or puncturing bytes. If for instance an UL-FEC rate ½ is needed, zero-padding is used in the last 127 columns of the Systematic data part of the UL-FEC matrix yielding K'=64 and N'=128. This choice allows fully compatibility with the DVB-H standard.

I.4.1.1.1 Packet Integrity check

It is important to note how the application of the CRC at UL and PHY has an impact on the overall system performance. To evaluate this impact, we distinguish to study cases:

- **CASE A:** only the PHY CRC is considered ($S_{UL\text{-}CRC}$ =0). In the broadcast scenario, the receiver is not able to check the integrity of a single channel carried within the PHY information packets. This basically means that if error is detected in the PHY information packet all the channel within that packet will be discarded. This is the default case for the 2-way communication scenario.
- **CASE B:** both PHY and UL CRC are applied. In the broadcast scenario, if the receiver is forced to ignore the PHY CRC, the integrity of each channel carried within the PHY information packets can be individually checked.

It is quite obvious that **CASE B** outperforms **CASE A**. In fact, if only a small fraction of bits are wrong after physical layer decoding, **CASE B** will be able the discard only the channel in which erroneous bits are present, while **CASE A** will discard all the N_{JCC} carried within the PHY information packets. The price to pay is an increased overhead of **CASE B** with respect to **CASE A** due to the extra CRCs appended to each channel. Finally, CASE B can allow lower complexity decoder implementation since the receiver needs to decode only one channel instead of N_{JCC}.

I.4.1.2 Receiver Side

At the receiver side dual operation is performed. In particular:

- The receiver will mark as reliable or not reliable data coming from physical layer:
  - If **CASE A** is taken into account only the CRC at PHY determines the data reliability.
  - If **CASE B** is considered the PHY CRC will be ignored and the data reliability is determined based on the UL CRC.
- The UL-FEC is filled with the reliable data:
  - If **CASE A** is taken into account an entire UL-FEC matrix column will be marked as reliable or not reliable.
  - If **CASE B** is considered UL-FEC matrix columns can be "partially" reliable.

A RS(N,K) decoding is performed on each row. If the number of reliable position in a row is at least $K=K'+Z$, the decoder is able to decode and all the unreliable position will be recovered.

In figures I.35 and I.36 are depicted the block diagrams of the receiver when CASE A and B are considered, respectively. As it can be seen the main difference is that in CASE A the columns in the UL-FEC matrix are completely reliable or not reliable, while in CASE B loss granularity (that is the minimum amount of data which can be marked as reliable/unreliable) is given by a single channel. This is due to the fact that only in CASE B the CRC over each channel is read (see figure I.36).

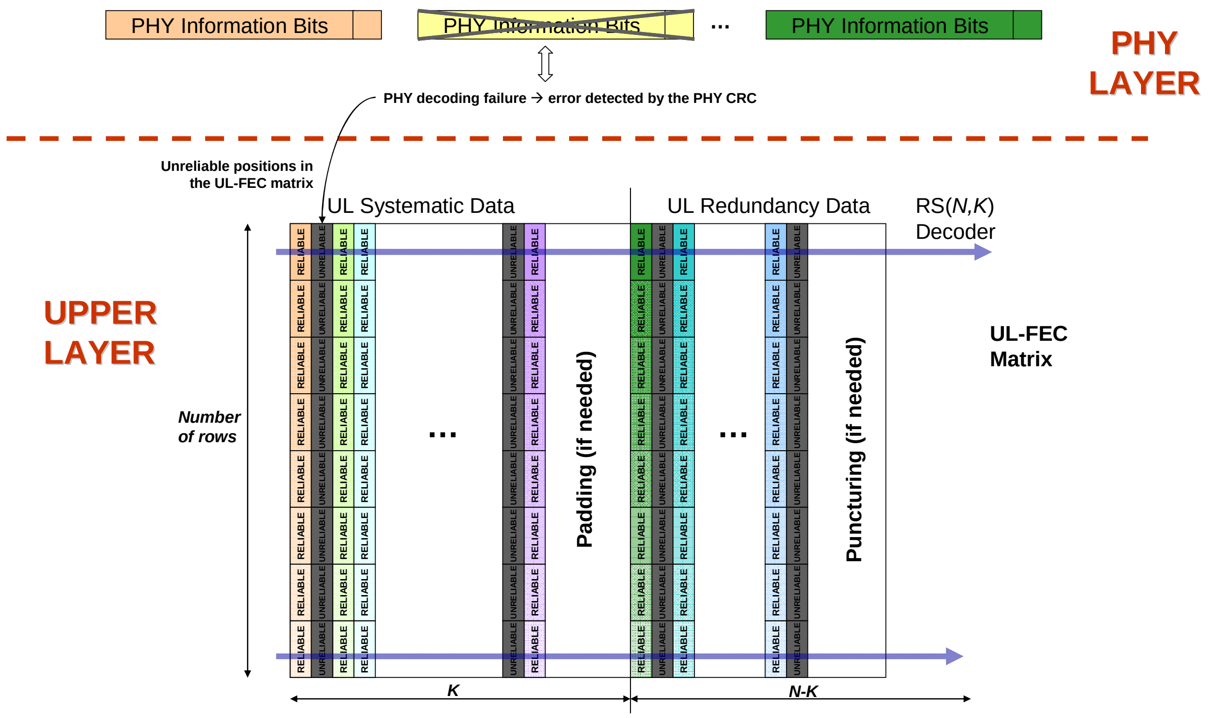


Figure I.35: Block Diagram of the receiver when CASE A is considered

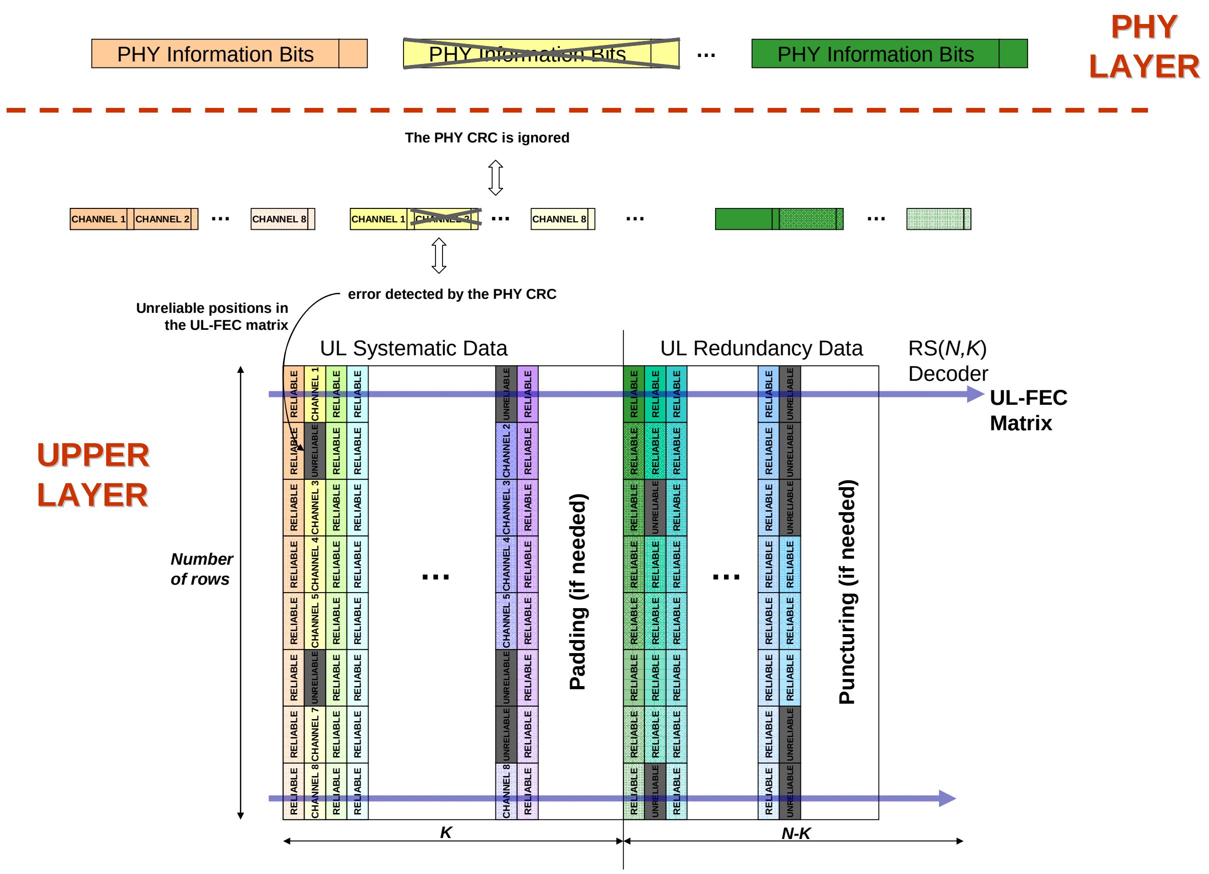


Figure I.36: Block Diagram of the receiver when CASE B is considered

I.4.2 UL-FEC Performance in BEC and urban SFN

Performance analysis has been carried out with two different approaches. Firstly, an analytical assessment has been performed to understand the behaviour of the proposed technique in the BEC channel and to assess the Maximum Tolerable Burst Length (MTBL) of the codes with reference to the LTE and WiMAX physical layer parameters. Then, numerical simulations have been performed to assess the end-to-end performance in more realistic propagation scenarios. In this latter case, physical layer error time series have been produced and used for the UL-FEC technique assessment.

I.4.2.1 Analytical assessment over the Binary Erasure Channel (BEC)

In this clause we present the performance of the proposed UL-FEC technique when the BEC is taken into account. For the sake of simplicity only the CASE A is considered. If we call p the physical layer Block Error Rate, the resulting BLER is given by the following well know formula:

$$BLER_{UL}(p) = p \cdot \left[1 - \sum_{j=K'}^{N'} \binom{N'}{j} (1-\mathrm{p})^{j} (p)^{N'-j} \right]$$

This formula holds because we are dealing with RS codes, which actually are Maximum Distance Separable (MDS) codes. Furthermore, the formula is to be considered only a benchmark for assessing the erasure correcting capability of the code. In the case of interest, where correlation between erased packets has to be faced, the formula does not apply. Finally, the factor p multiplying the square brackets takes into account the systematic code properties. As a consequence the UL frame error rate is given by:

$$FER_{UL}(p) = 1 - \sum_{j=K'}^{N'} \binom{N'}{j} (1-\mathrm{p})^{j} (p)^{N'-j}$$

Figures I.37 and I.38 show the performance over the BEC obtained with (K'=191, N'=255) e (K'=64, N'=128), respectively. Results are presented in terms of Upper Layer BLER and Upper Layer FER. Regarding the CRC configuration, only CASE A is taken into account.

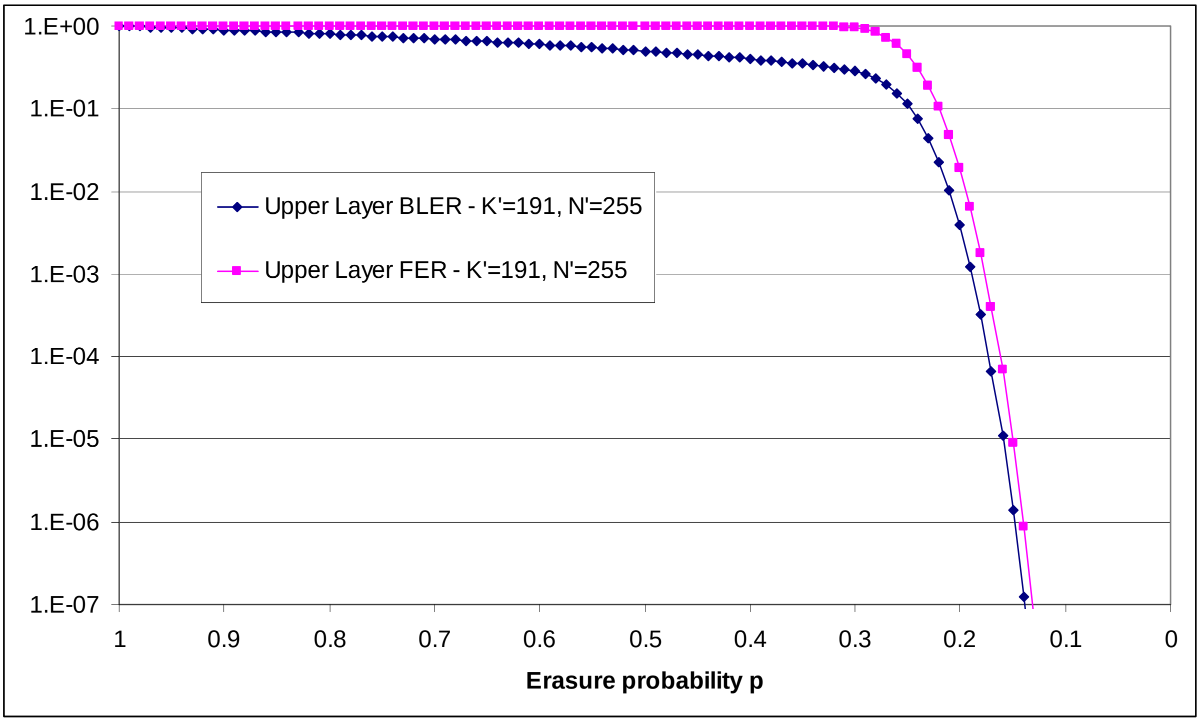


Figure I.37: Upper Layer BLER over the BEC. CASE A

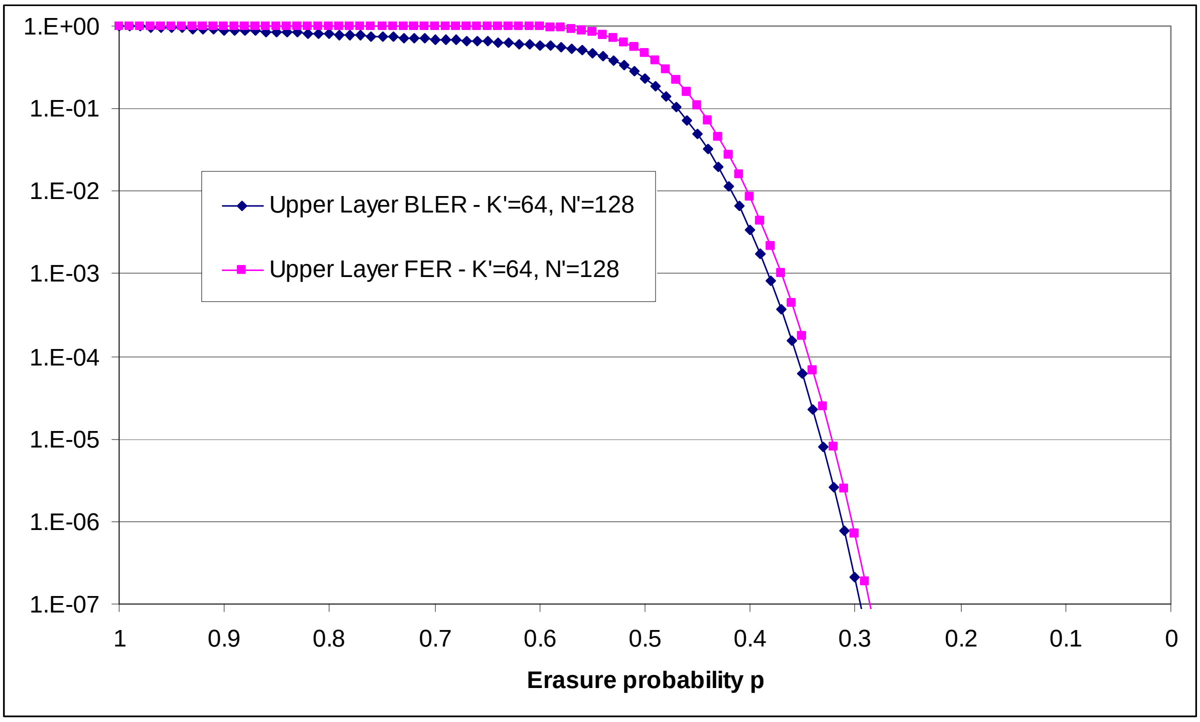


Figure I.38: Upper Layer BLER over the BEC. CASE A

I.4.2.2 Maximum Tolerable Burst Length computation

In [i.319] we propose an analytical model describing the performance of UL-FEC when periodic blockages are taken into account. Here we aim at determining the Maximum Tolerable Burst Length (MTBL), which consists in the maximum time protection that the UL-FEC technique can provide. The MTBL depends on both UL-FEC parameters and PHY data rate. In our technique one PHY information packet is mapped in one column of the UL-FEC matrix. Since we are dealing with MDS codes, the decoder will be able to successfully decode if at least K' columns are correctly received in the UL-FEC matrix. This means that the MTBL is simply given by the time taken by N'-K' columns, i.e. the duration of N'-K' information packets.

The MTBL can be increased by adopting a sliding encoding mechanism [i.320]. The sliding encoding is a UL interleaver mechanism which can be easily explained by referring to figure I.39. A UL-FEC encoder implementing sliding encoding will select the K' data columns from a window (*SW*) of UL-FEC frames and will spread the N'-K' parity sections over the same frame window (figure I.39). Basically, the same effect could be obtained by first normally encoding *SW* frames and then interleaving sections among the encoded *SW* frames.

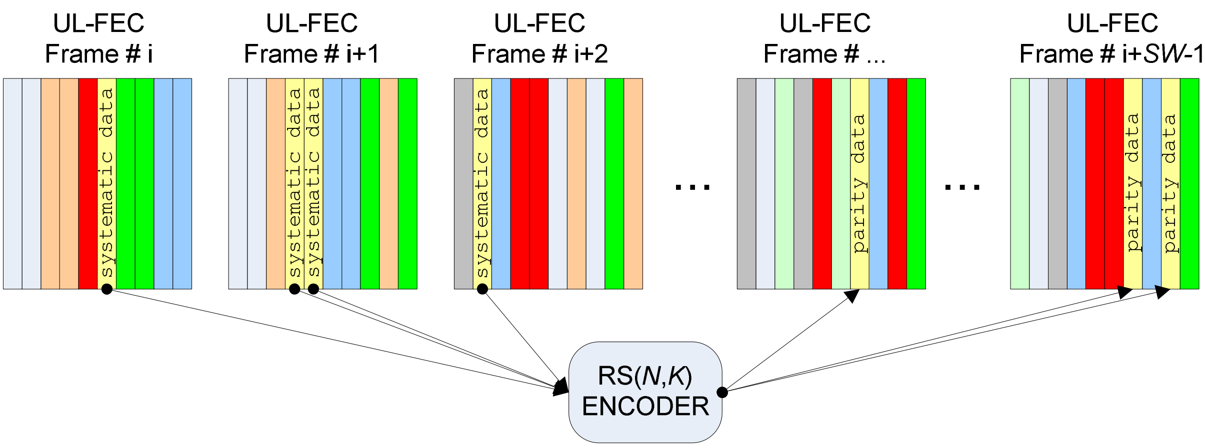


Figure I.39: UL-FEC with Sliding Window

In table I.16 the MTBL is shown considering both LTE and WiMAX standard. As it can be seen, increasing the sliding window size allows increasing the MTBL. The table refers to the case in which a perfect block interleaver is performed.

Table I.16: UL-FEC Maximum Tolerable Burst Length (MTBL)

Physical Layer	PHY information packet duration [ms]	(K',N')	SW	MTBL [ms]
LTE	1	(191,255)	1	64
LTE	1	(64,128)	1	64
LTE	1	(191,255)	5	320
LTE	1	(64,128)	5	320
LTE	1	(191,255)	10	640
LTE	1	(64,128)	10	640
WiMAX	1,49	(191,255)	1	95,36
WiMAX	1,49	(64,128)	1	95,36
WiMAX	1,49	(191,255)	5	476,8
WiMAX	1,49	(64,128)	5	476,8
WiMAX	1,49	(191,255)	10	953,6
WiMAX	1,49	(64,128)	10	953,6

I.4.2.3 Splitting the redundancy between UL and PHY

An important aspect of the design of UL-FEC technique is the definition of the redundancy split between physical and upper layer coding. Clearly, this split would be, in principle, in favour of the physical layer coding if no constraint of the physical layer interleaver is considered. However, having to deal with already well-defined standards (as LTE and WiMAX are), it is mandatory to consider the interleaver constraints imposed by the standard themselves. Therefore, depending on the considered scenario, it is possible to foreseen different redundancy splits, between physical and upper layer,. That can yield different performance in this clause, we exactly address this issue. We evaluate how to split the redundancy of the two coding schemes keeping, as constraint, the total spectral efficiency. To this aim, in table I.17 we report a subset of configurations that provide the same spectral efficiency with different redundancy split between physical and upper-layer coding. Those configurations are then compared in terms of BLER and FER.

It is worthwhile noting that in order to keep constant the UL protection time, we introduced the RS(255,128). In this way the rate 1/2 RS(255,128) and rate 3/4 RS(255,191) have the same codeword duration and are therefore better comparable in terms of protection efficiency.

NOTE 1: So far we have always considered the RS(255,199) mother code. The UL-FEC with rate 1/2 has been obtained by means of zero padding.

Table I.17: Spectral efficiency with different redundancy split between PHY and UL

Physical Layer	Scenario	Mapping	PHY Code Rate	UL Code Rate	Total Spectral Efficiency
LTE	Broadcasting	QPSK	2/5	1	4/5
LTE	Broadcasting	QPSK	4/5	1/2	4/5
LTE	2W-FL	QPSK	1/2	3/4	3/4
LTE	2W-FL	QPSK	3/4	1/2	3/4
LTE	2W-FL	QPSK	1/2	1	1
LTE	2W-FL	16-QAM	1/2	1/2	1

In figure I.40, the BLER performance is reported for the Broadcasting scenario with LTE physical layer, considering a terminal speed of 3 kph and ideal estimation conditions. The overall spectral efficiency, which takes into account modulation order, UL and PHY code rate is 4/5. As it can be seen, in the low SNR region, the curve in which all the redundancy is given to the physical layer (that is no UL-FEC is performed) outperforms the curves with both PHY and UL coding. This is due to the different waterfall region of the PHY channel codes. Nevertheless, at higher SNR, the PHY+UL FEC solution, i.e. redundancy split between PHY and UL, outperforms the case in which only PHY is considered. This can be explained considering that there is a Eb/N0 value above which the UL-FEC starts to work and removes the residual errors that PHY layer has not been able to counteract thus making the shape of the curve much steeper. In figure I.40, the crossing point is foreseen at Eb/N0=10 dB. Considering the two RS codes, it is apparent that there is only a slight difference between the RS(128,64) and the RS(255,128). This behaviour can be explained by considering the at 3 kph the channel coherence time is about 9ms and the RS(128,64) provides enough time protection against channel burstyness.

NOTE 2: In this case the MTBL is 64ms.

The considerations made referring to figure I.40 are still true when taking into account figure I.41. In this case the performance improvement obtained by splitting the redundancy between PHY and UL is even larger with respect to the case of figure I.40. In fact, the curves in figure I.41 refer to a terminal speed of 15 0kph. In this scenario, the upper layer decoder can exploit the larger time diversity within the UL-FEC frame due to the lower channel coherence time.

The advantage of split redundancy between UL and PHY layer is more evident in figures I.42 and I.43. In these curves the forward link of the 2 Way scenario is taken into account with no ideal estimation. By referring to figure I.42, the frequency estimation error makes the gap between QPSK 1/2 and QPSK 3/4 smaller, considering only physical layer. This is due to the fact the QPSK 3/4 curve can exploit a larger SNR, thus resulting in lower frequency error. As a consequence of this consideration it is convenient to redistribute the redundancy in favour of upper layer. The same behaviour can be observed in figure I.43.

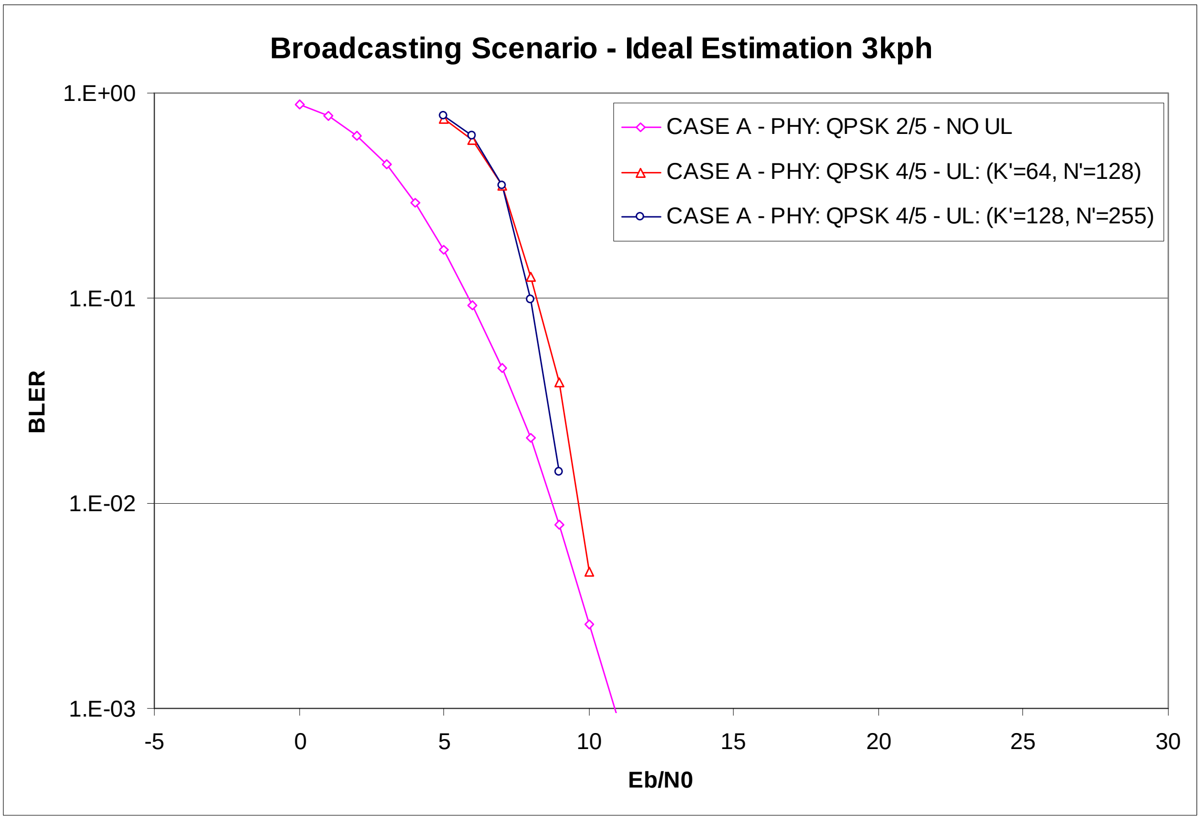


Figure I.40: BLER performance obtained by splitting the redundancy between physical layer and upper layer keeping the total spectral efficiency. Broadcasting Scenario - Ideal Estimation - 3 kph

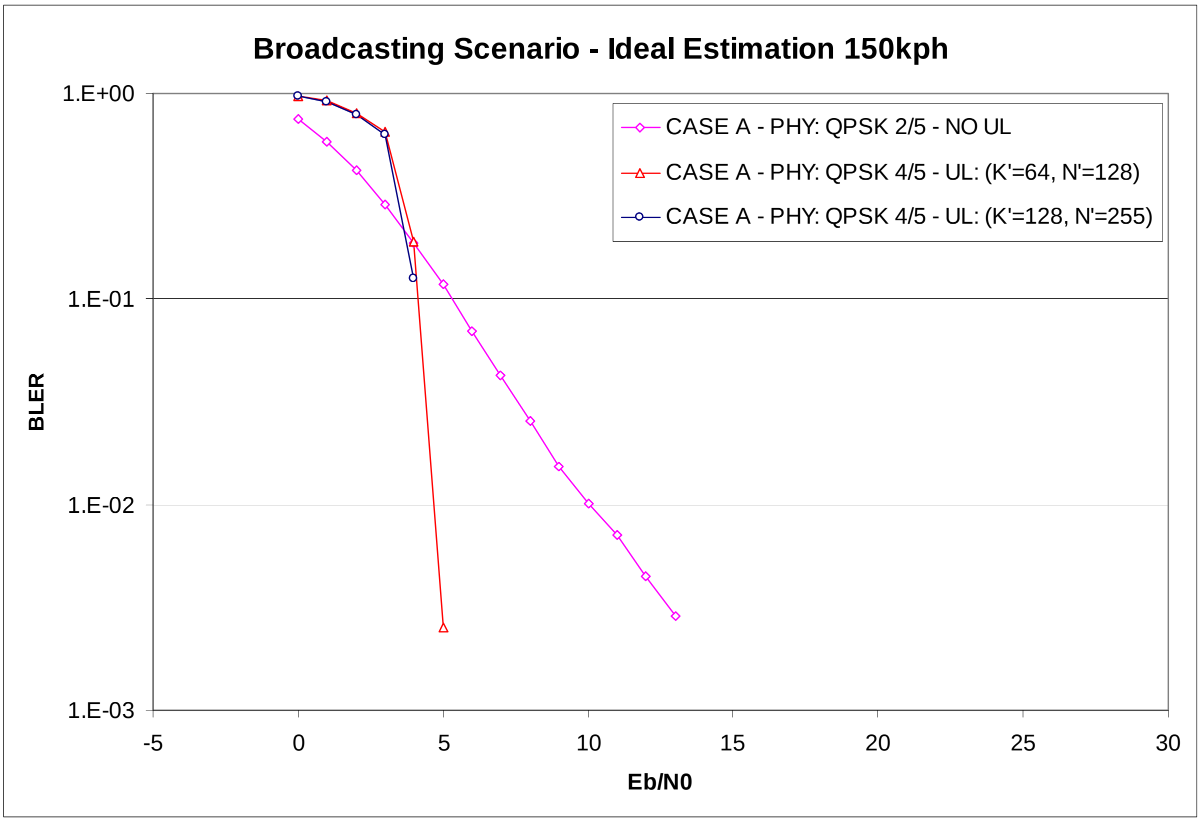


Figure I.41: BLER performance obtained by splitting the redundancy between physical layer and upper layer keeping the total spectral efficiency. Broadcasting Scenario - Ideal Estimation - 150 kph

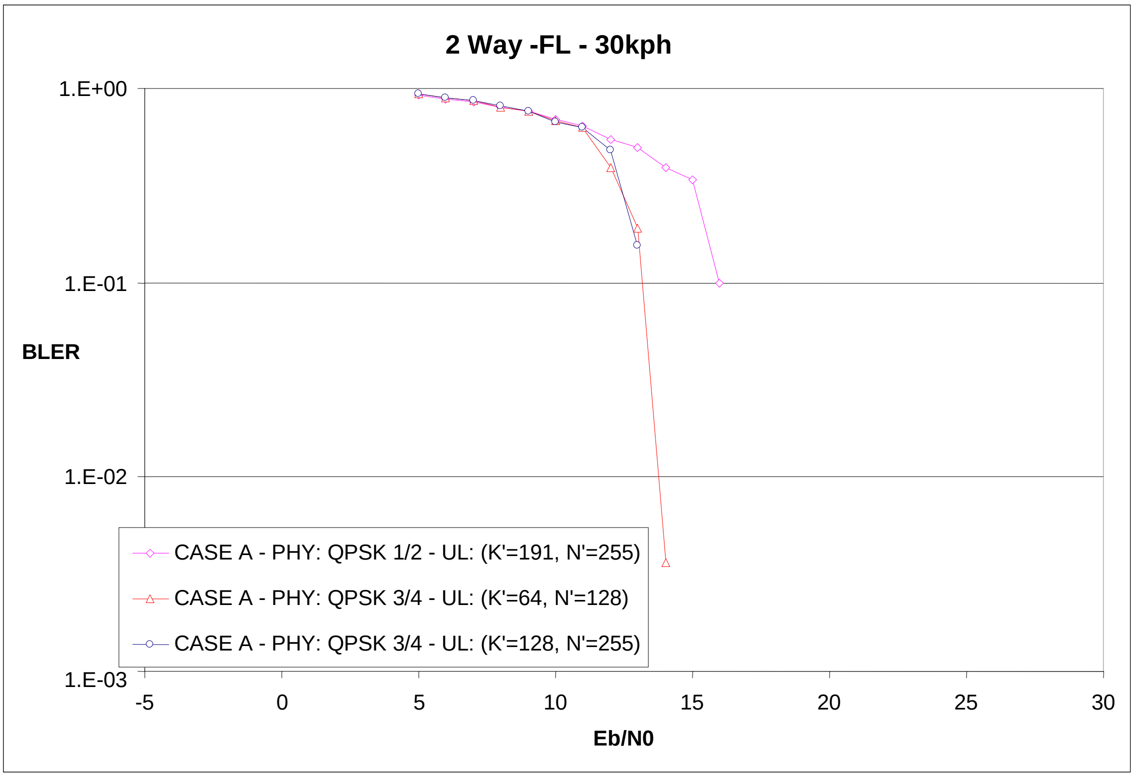


Figure I.42: BLER performance obtained by splitting the redundancy between physical layer and upper layer keeping the total spectral efficiency. 2Way FL - 30 kph

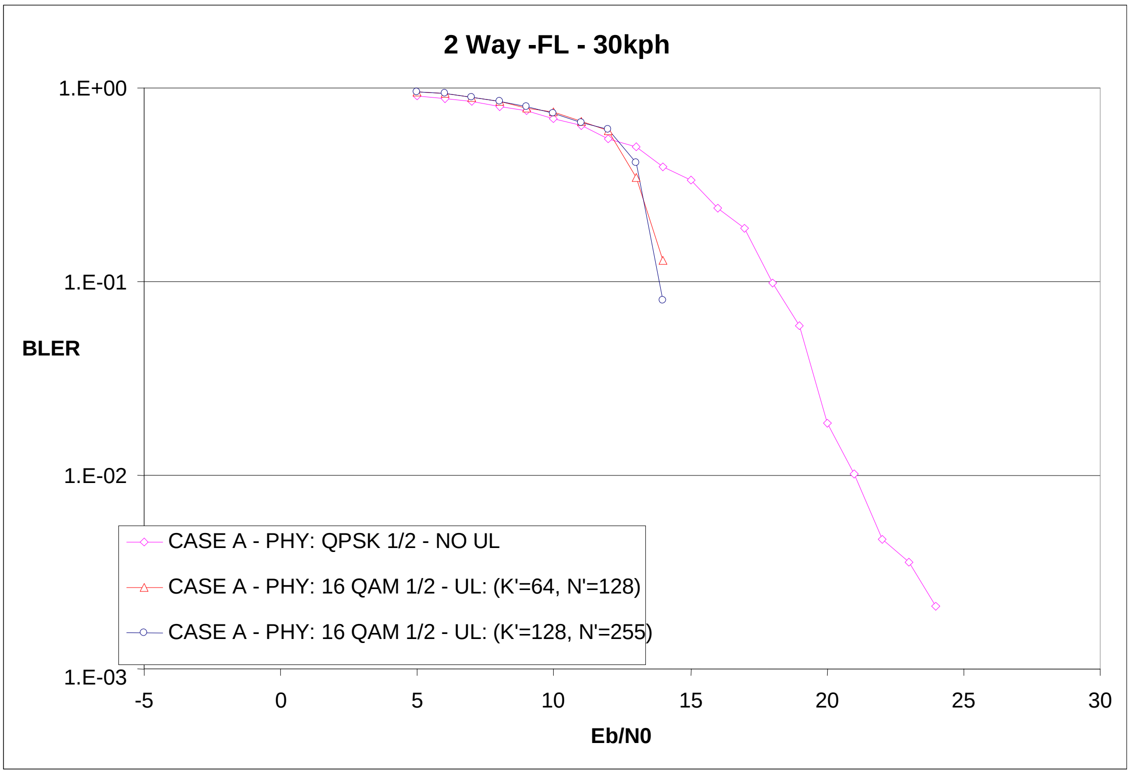


Figure I.43: BLER performance obtained by splitting the redundancy between physical layer and upper layer keeping the total spectral efficiency. 2Way FL - 30 kph

I.4.2.4 Comparison with inter TTI interleaving

In this clause we provide a comparison between the UL-FEC approach proposed in the previous clauses and the inter TTI interleaver technique proposed earlier. In order to make a fair comparison between the two techniques, in the following we keep constant the overall spectral efficiency by distributing the redundancy between UL-FEC and physical layer. Figures I.44 and I.45 show the numerical results obtained in the case of the Broadcasting scenario, assuming the terminal speed equal to 3 kphm, and ideal channel estimation. The performance is measured as BLER vs Eb/N0, and BLER Es/N0. All the reported curves have a spectral efficiency equal to 4/5 (see note). In the inter TTI case, we have considered QPSK 2/5 MOD-COD at physical layer, varying the inter TTI both the interleaver depth and the sub frame size. The UL-FEC have been obtained considering QSPK 4/5 @PHY, and a (K'64, N'=128) code @UL. Since the considered UL-FEC protection spans over N'=128 symbols, that is 128 ms, the most comparable protection time provided by the inter TTI is that obtained adopting K = 40 and Sub Frame size = 3 (in this case the PHY layer codeword spans K*4 = 160 TTI, i.e. 160 ms). From the analysis of the results, we can state that on the one end, the inter TTI techniques outperforms the UL-FEC technique, which can be justified recalling that at physical layer the decoder can exploit soft information, thus achieving much better performance with respect to the hard decoding performed at upper layer. On the other end, the inter-TTI technique requires a large memory buffer at the output of the base-band processor. A through complexity analysis needs to be carried out to this respect in order to understand the hardware feasibility of the assumption considered for the inte-TTI case.

NOTE: Here we compute the spectral efficiency as a product of the modulation order, the PHY FEC coding rate, and the UL-FEC coding rate.

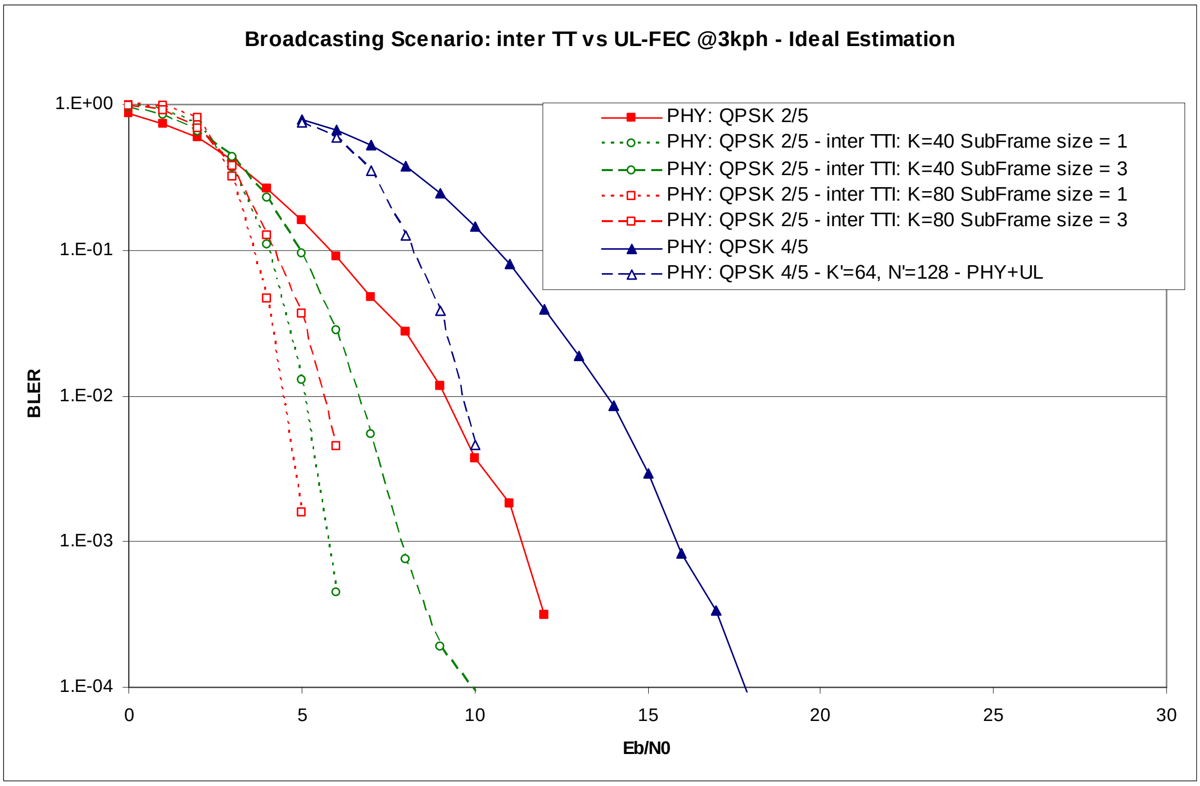


Figure I.44: Comparison between UL-FEC and interTTI. BLER vs Eb/N0

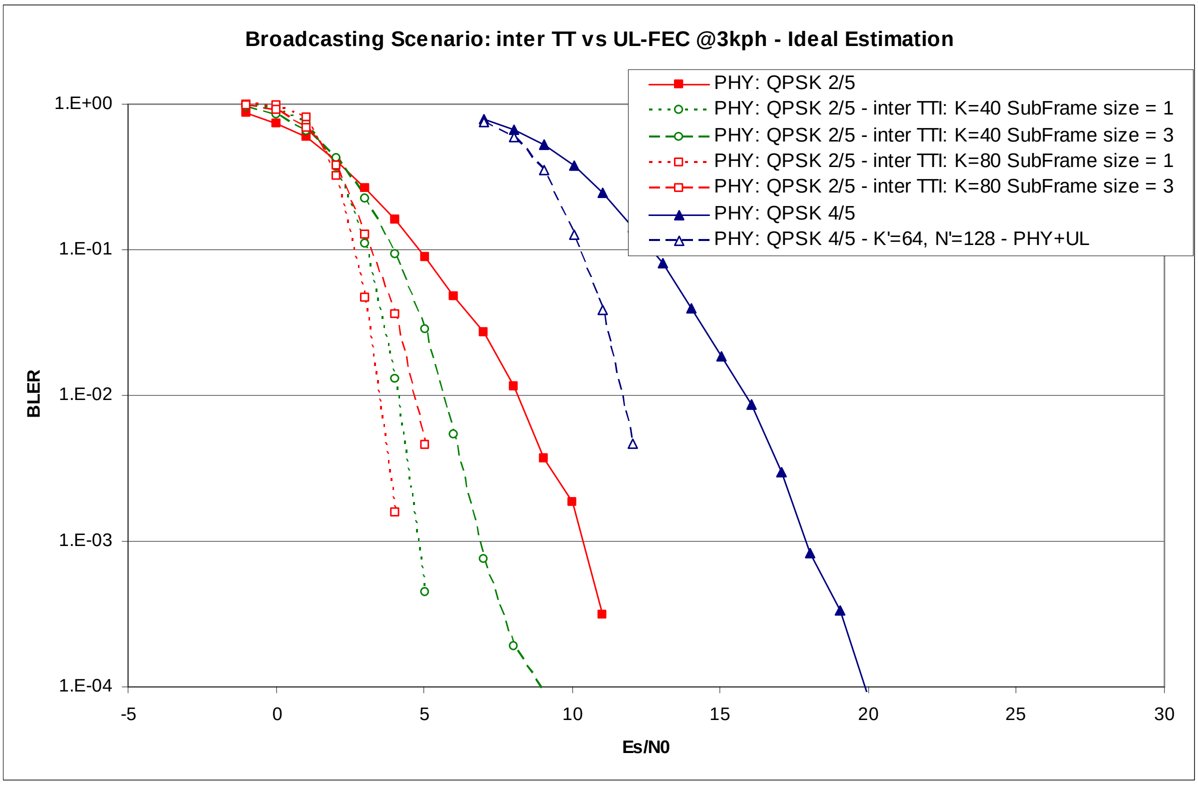


Figure I.45: Comparison between UL-FEC and interTTI. BLER vs Es/N0

I.4.3 UL-FEC Performance assessment in LMS propagation

In this clause, we report the results obtained considering the three states Perez-Fontan channel model. The time series are obtained following the procedure described in §4 of [i.323]. In our analysis we have considered an elevation angle of 40 degrees and four different environments:

- Open area [O], Suburban [S], Intermediate tree shadow [ITS], Heavy tree shadow [HTS].

Such environments are characterized by long fading events due to the shadowing. It is quite obvious that applying the proposed UL-FEC technique without any interleaver working at UL does not accommodate such fading events. Indeed, as it can be seen in table I.16, the MTBL achievable by adopting UL-FEC without sliding interleaving (SW=1) is in the order of hundreds milliseconds. To increase the MTBL we adopt the sliding window encoding technique described elsewhere in this annex. This technique basically consists in applying a block interleaver at UL. It is clear that the interleaver depth depends on the sliding window (SW) size. In particular, the total protection time achievable by means of such a technique is given by:

$$\text{Protection Time} = N * SW \ * UL_{TIME} \ ;$$

where UL_{TIME} is the duration of an upper layer symbol and N is the UL codeword length.

In order to get a synthetic analysis of the results obtained with extensive computer simulations, we have assessed the Erroneous Seconds Ratio (ESR) criterion [i.322]. ESR was also considered by the DVB-SSP group to be the most relevant performance parameter for the assessment of the impact on the video quality. In particular, we take into account the ESR5(20) criterion: ESR5(20) is fulfilled for a given time interval of 20 seconds if the percentage of erroneous seconds in the same time interval does not exceed 5 %, which corresponds to a maximum of 1 erroneous second. The percentage of time satisfying the ESR5(20) criterion represents the *"ESR5(20) fulfilment percentage"*.

The conclusions of this analysis are summarized in figure I.46 where the achievable spectral efficiency is reported as a function of the C/N required to satisfy the ESR5(20) criterion at 90 %. In this case, the spectral efficiency is computed considering the PHY configurations listed in table I.18, which constitute a subset of the configurations proposed in [i.324]. Notably, since in general in a LTE frame both information and control data are transmitted, we assumed that the equivalent of 1 OFDM symbol per TTI, i.e. 1/12 of the TTI, is completely dedicated to the transmission of control data. As a consequence, the PHY spectral efficiency resulting from table I.18 has been reduced by a factor (11/12).

Table I.18: LTE Physical Layer Configuration for Interactive Broadcasting forward link

LTE PHY Layer configuration						
Number of jointly coded channels / number of channel groups	**Information bits per packet**	**Allocated data carriers per sub-frame (MBSFN RS) [RBs x OFDM symbols]**	**Mod**	**Actual Code rate**	**Bit Rate**	**Channel Bandwidth**
8/1	2 496	3 150 [25 x 12]	QPSK	2/5	2,50 Mb/s	5 MHz
16/1	4 992	3 150 [25 x 12]	QPSK	4/5	,99 Mb/s	5 MHz
24/1	7 552 [3 776 + 3 776]	3 150 [25 x 12]	16-QAM	3/5	7,49 Mb/s	5 MHz

In figure I.46 each curve represents the performance of the QPSK constellation in a given scenario and for a given UL-FEC coding rate. The two connected markers in each curve represent the two corresponding PHY coding rates.

The single markers in figure I.46 represent the 16-QAM constellation in a given scenario and for a given UL-FEC coding rate. Notably, for the 16-QAM constellation, only one PHY FEC scheme has been considered.

Interestingly the lower UL-FEC protection, i.e. 3/4, always outperforms, at the same total spectral efficiency, the higher UL-FEC protection, with the only exception of the Heavy Tree Shadow scenario. In that case, the extremely challenging propagation conditions calls in fact for a very strong protection along with a quite demanding link budget.

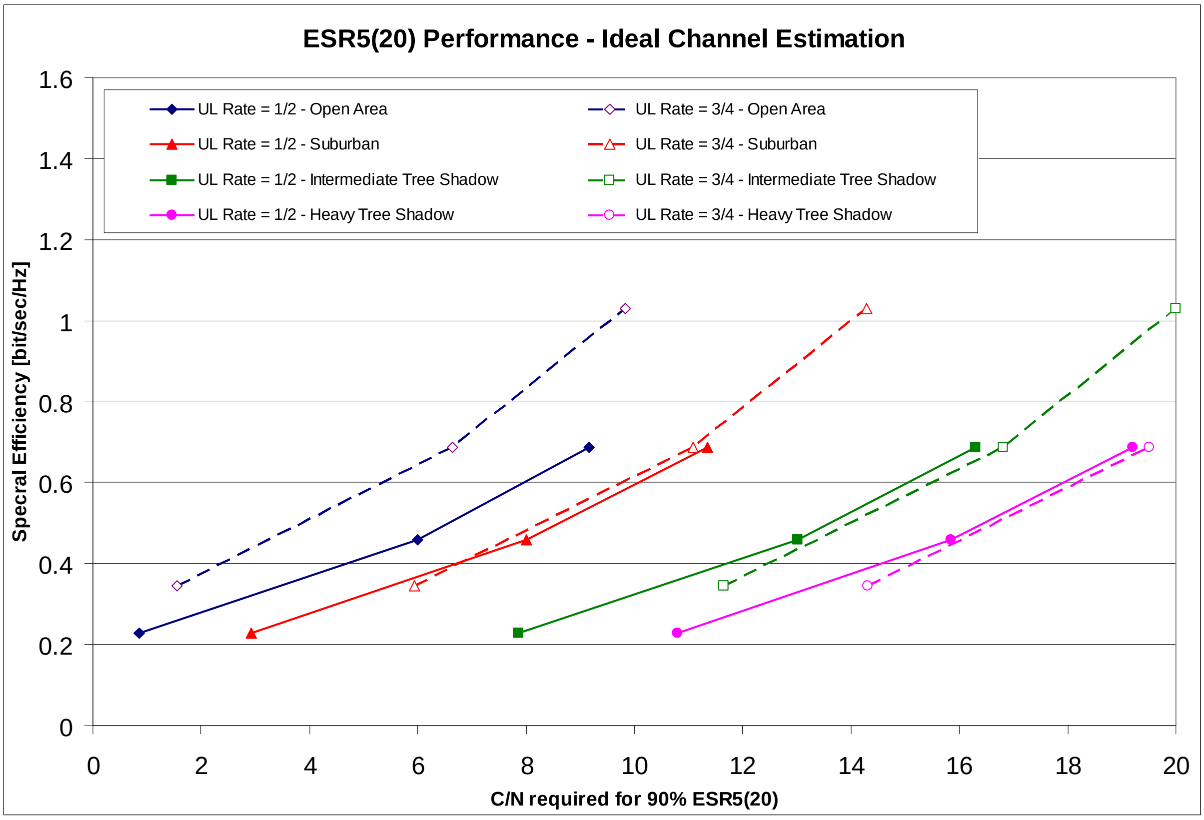


Figure I.46: Overall (PHY+UL) Spectral Efficiency vs C/N for 90 % ESR5(20)

I.5 PHY and UL FEC Study - Conclusions and Recommendations

From the analyses presented in the first part of this document, the following remarks and observations can be extracted:

1) A peculiar characteristic of LTE is to have a fixed FFT size of 2048, irrespective of the actual used bandwidth. This means that the sampling rate in the receiver is constant, and possibly largely oversized with respect to the minimum. This has an impact also over pre-FFT estimation algorithms.

2) All the selected numerologies for LTE and WiMAX, forward and reverse links, are standard compatible (except for LTE RL QPSK rate 4/5). In this sense, the results presented here are significant from the 3GPP and IEEE point of view. Specifically, for the broadcast application scenario, 8 configurations have been specified for both LTE and WiMAX forward links, while for the two-way communications scenario 3 + 3 configurations have been specified for LTE and 3 + 2 for WiMAX (rate 1/3 coding is not applicable for WiMAX).

3) Regarding time domain fade mitigation techniques: both LTE and WiMAX standards have been designed for highly optimized throughput with fast feedback and thus very accurate channel information. Fades in the time domain are circumvented through dynamic scheduling, HARQ, and adaptive coding and modulation, instead of long time interleaving. This approach is not naturally suitable for satellite communications, where fast feedback is limited by the round trip time. Therefore, some modifications are necessary. First of all, it may be useful to aggregate several channels before encoding, in order to increase the packet size and the consequent efficiency of the forward error correction coding procedure. Secondly, it is mandatory to increase the time-span of interleavers. One of the major findings consists in a way to obtain the above diversity in an almost standard compatible way. This is the inter-TTI technique, which has been shown to bring significant benefits without touching the physical layer definition. Thirdly, it may be necessary to introduce upper layer coding, to improve the performance whenever the physical layer is not sufficiently robust.

4) MIMO holds the promise of being a diversity mine, but its application to satellite communications is still to be proved. We have progressed by considering both LTE and WiMAX precoding approaches, and found no major differences between the two. This with the exception of the fact that WiMAX appears to be completely oriented towards transmit diversity, while LTE also includes spatial multiplexing as an option for increased capacity. This may be difficult to exploit in satellite MIMO, because it requires knowledge of the channel quality at the transmitter, since it needs good receiving conditions. Finally, an open issue is to characterize the performance (and possibly to design suitable countermeasures) of satellite MIMO techniques in the presence of non linear distortion.

5) PAPR reduction algorithms, coupled to predistortion techniques, are a novelty for OFDM transmission through a satellite. We have explored this architecture and our results show that the PAPR itself can be reduced by 2 dB to 4 dB (guaranteed at 99,9 %), which translates into the possibility to reduce the OBO by about 0,7 dB and to gain about 0,5 dB in E_b/N_0 for typical quality of services. All in all, we can expect a gain in total degradation above 1 dB, which is certainly not negligible.

6) Regarding frame acquisition procedures, they are quite specific for LTE and WiMAX. On this side, the design of acquisition sequences for 3GPP LTE has been performed adapting it to the different requirements set by satellite transmission involving the use of large geographic beams.

7) The rationale behind the choice of investigating the adoption of UL-FEC techniques is that, as shown in clause I.3, the LTE and WiMAX PHY coding is a suitable countermeasure to cope with small-scale fading, while it is less efficient against large-scale fading. This is mainly due to the limitation of the time diversity inherent in the necessity of adopting a physical layer interleaver of a reasonable size, so as to avoid increasing the modem complexity and the latency of all services. In terrestrial networks, this limitation is in general overcome by the adoption of hybrid ARQ techniques. Since, H-ARQ approaches are scarcely efficient in satellite scenarios, where the long satellite channel propagation delays make them almost unfeasible, we focus our attention on the UL-FEC technique.

8) Simulation results clearly show that the UL-FEC technique is a very effective solution that can drastically improve the achievable block error rate and ESR5(20) performance. In order to provide useful guidelines for the system design, the analysis of the optimum redundancy split between physical and upper layer coding has been performed. In this case, results show that in most cases it is beneficial to limit the protection at physical layer in order to ease channel estimation and to compensate the reduced performance through a stronger UL coding. The rationale behind this conclusion is that the UL-FEC benefits a larger time diversity thus performing significantly better than the physical layer coding in almost all scenarios.

9) The results reported in the previous clause are interestingly in line with those obtained in the DVB-SH framework. With all cautions necessary in comparing two different systems, it is worthwhile reading the above results considering a C/N degradation factor of about 2,5 dB for non-ideal aspects: 1 dB for non-ideal estimation losses for the QPSK case; 1 dB for the BER degradation due to non-linear distortion; 0,5 dB of implementation losses. Under these assumptions, the spectral efficiencies obtained in this analysis are consistent with those obtained with a class 1 terminal in the Ortigia field trials [i.325], which, in turn, confirms the soundness of this analysis.

Focusing on the compliance of the considered techniques and of the proposed modifications with respect to the current standards, a summary is presented in table I.19, along with an indication of which techniques are worth to be retained and considered in further analyses.

Table I.19: Summary of considered techniques in terms of performance, complexity, and compliance with current standards

Technique	Performance/Complexity/Critical aspects	Compliance with standards and required modifications
Inter-TTI interleaving	Adds robustness to small-scale fading at medium to low speed. For pedestrian speed, the required complexity to achieve satisfactory performance is probably too high, as well as if large-scale fading has to be counteracted.	Compliance with LTE at physical layer. Need for changes in HARQ handling in MAC. Need for increase in number of supported HARQ processes.
Channel estimation	Critical aspect in the presence of NL distortion, which introduces a non-removable noise floor dependent on the HPA operating point.	Receiver dependent - No impact on the standard.
Predistortion	Mitigates the impact of non linear distortion. Reduces total degradation. Its application at the user terminal requires additional complexity, as well as at the gateway where on the other hand complexity requirements are less stringent. Applicable to the single HPA per beam case.	No impact on the standard.
PAPR reduction - ACE	SNR Gain in the order of 0,5 dB to 1 dB depending on the configuration and on the PHY.	Does not require modifications to the standard. Compliance with Error Vector Magnitude test specifications to be further tested.
Frame synchronization	Interactions with pre-FFT techniques. WiMAX more robust thanks to the presence of a preamble in the frame structure. Detection in time may increase robustness to residual frequency offsets.	No modifications to the air interface.
Random access		No modifications to the air interface. Specifications regarding the number and type of sequences to be used as a function of beam footprint may be needed.
MIMO - Transmit Diversity	Large gains (6-8 dB) in mobile environment thanks to diversity to channel fades. Proper satellite separation angle has to be selected.	Compliance with LTE and WiMAX at physical layer. Need for modifications in upper layers to account for the lack of feedback from user terminals.
MIMO - Spatial Multiplexing	Increased sensitivity in mobile environment to fades due to lack of channel quality feedback from user terminals. Nonetheless, fixed reception with proper channel quality feedback may benefit.	Compliance at physical layer. Need for modifications in upper layers to account for the lack of feedback from user terminals.

I.6 Resource Allocation in Time & Frequency for LTE and WiMAX

Both LTE and WiMAX employ Orthogonal Frequency Division Multiple Access (OFDMA) as their main multiple access mechanism (although other options are also defined in the standards). The basic idea of OFDMA is to divide the available time-frequency space into a number of orthogonal subcarriers, of finite time and frequency support, each of which is assigned to a specific user. The assignment is performed by a scheduling algorithm implemented at the based station (access point). OFDMA provides a simple (albeit suboptimal) way of exploiting the channel's available degrees of freedom, both in time and frequency. In general, the scheduling algorithm may exploit all, or part, of the following available degrees of freedom for user allocation:

- Time
- Frequency
- Power
- Rate

It is noted that the active users themselves can also be viewed as an extra degree of freedom in the sense that the scheduler may optimally assign part of them based on some optimality criterion.

For serving users with the same QoS, traditional scheduling approaches divide the available signalling space into equal parts, assigning each of them to a single user. This intuitive approach can also be viewed as achieving fairness among users since the channel resources are equally shared. However, the drawback of this approach is that it is channel-blind: resources will be provided to a certain user irrespective of its channel conditions. In the case of a user experiencing severe fading (typical case in wireless communications) the transmitted signal will be lost with high probability, which translates to an inefficient use of the system's resources. If the scheduler had information on the user's channel state, it could modify (reduce) its transmitted rate to match the channel's conditions. Another option, would be to defer service to this user, due to poor channel conditions, and possibly transmit to another user which happens to have a better channel. Of course, the last option requires knowledge of channel state for all (or part) of the active users.

Clearly, knowledge of Channel State Information (CSI) can only help the scheduler in better utilization of the channel, which has led to an intensive research on this category of algorithms [i.297] and [i.310], usually named as channel-aware schedulers, in the last few years, due to the performance gains that provide compared to the non-channel-aware approached, which, in some cases, can be quite significant. Recognizing the potential of channel-aware scheduling, both LTE and WiMAX propel its employment, although explicit algorithms are not included in their description.

Essential for the success of any channel-aware scheduling algorithm is the (good) knowledge of each user's channel. The obvious approach for obtaining a channel estimate is each user transmuting his/her CSI periodically by mean of a channel quality indicator (CQI) message [i.313]. For low mobility this sort of messaging has negligible impact on the system's throughput (although it is noted that in the case under consideration the CQI is not a single SNR value, but rather a vector of SNR values over the available bandwidth). Another possibility for obtaining CSI is to exploit the channel's reciprocity by systems operating in a TDD mode. In that case the base station can measure (estimate) the user's reverse-link channel, and this estimate can be employed as the user's forward link channel estimate. Higher mobility requires more frequent updating of the CSI and therefore may have impact on the system's throughput (due to the increased overhead). In addition, feedback delays may have an impact on the scheduler's performance due to mismatch between the current channel state and its estimate employed by the scheduler [i.305].

Another important parameter of time-frequency channel-aware scheduling is the minimum Resource Allocation (RA) block. The optimal RA block from an information-theoretic perspective is a single sub-carrier. Unfortunately, the large number of subcarriers (proportional to the FFT size) and number of users served by the base station makes scheduling algorithms operating under this RA block impractical. For this reason the minimum RA blocks adopted by the standards are named sub-channels that consist of a number of sub-carriers for the duration of a few OFDM(A) symbols. When the sub-carriers are closely packed and contained within the channel's coherence bandwidth, there is no much loss in optimality. However, the standards also define sub-channels composed of sub-carriers far apart from each other (e.g. PUSC sub-channel in WiMAX) for the case when no channel information is available. Employing a channel-aware scheduling algorithm with this type of RA blocks is certainly feasible but will lead to a certain performance loss [i.316].

I.6.1 Description of channel-aware algorithms

I.6.1.1 Maximum sum rate (MSR) algorithm

It is well known from information theory that channel knowledge at the transmitter side can only help increase the transmission rate [i.310]. For point-to-point communication the advantage of CSI at the transmitter is only visible at the very low SNR regime, and is therefore of limited usage. However, in the case of a multi-user communication scenario, knowledge of the channel at the base station (scheduler) provides an essential increase of the system's sum rate by performing an *opportunistic scheduling* approach [i.310] and [i.298]. The idea is simple: Given the channel state for each user assign the available resources to the user(s) with the best current conditions. When the number of users is large and their respective channels fade independently the gains of this approach are significant.

In OFDMA-based wireless standards, the available resources are the sub-channels spanning part of the available signal space in time and frequency, since they are composed from a subset of the system's available sub-carriers. However, for ease of exposition, a per-sub-carrier allocation will be assumed in the following (generalization for su-channel allocation is straightforward). The scheduler's job is to assign the sub-carriers in an optimal manner with respect to the system's sum rate. In mathematical terms, the scheduler solves the following optimization problem at every scheduling instant (considered here to be an OFDM(A) symbol) [i.311] and [i.302].

$$\underset{\{r_{m,n}\},\{p_{m,n}\}}{\text{maximize}} \sum_{m=1}^{M} w_m \sum_{n=1}^{N} r_{m,n}(h_{m,n}, p_{m,n})$$

subject to (s.t.) $\sum_{m=1}^{M}\sum_{n=1}^{N} p_{m,n} \leq \bar{P}$,

where $m \in [1, M]$ is the user index, $n \in [1, N]$ is the (data bearing) sub-carrier index, w_m is the weight of user m ($\sum_{m=1}^{M} w_m = 1$), $r_{m,n} \geq 0$ is the (maximum) rate that can be transmitted on the n-th sub-carrier of user m that depends on each corresponding channel $h_{m,n}$ and the power $p_{m,n} \geq 0$ assigned for transmission in this sub-carrier. In general the base station has a maximum power constraint $\bar{P}$ that cannot exceed. Note that the solution of the problem provides:

1) sub-carrier (time-frequency) allocation,

2) rate allocation,

3) power allocation.

It can be proven [i.293] that the optimal sub-channel allocation is orthogonal, i.e. for the n-th subcarrier there can be only one $r_{m_n^*,n} \neq 0$ where m_n^* is the "winner user" of the sub-channel (the one with best channel conditions, taking into account the user weights), which is also intuitively satisfying.

The rate that a user's channel can "support" for a given channel realization and power allocation depends on the QoS (i.e. maximum tolerable BER). If the rate is replaced by the information-theoretic expression

$$r_{m,n}(h_{m,n}, p_{m,n}) = \log_2\left(1 + \left|h_{m,n}\right|^2 p_{m,n} \big/ \sigma_{m,n}^2\right)$$

where $\sigma_{m,n}^2$ is the noise power on the n-th sub-carrier of the m-th user, the resulting sum rate is actually the system's *sum capacity* (for the particular choice of weights), that provides the limits of the system. Note that in this information-theoretic optimization setting, the rate itself is a function of the allocated power which makes scheduling easier (i.e. only sub-carrier and power allocation is necessary). In the practical setting where the rate values $r_{m,n}$ belong to a discrete set imposed by the available AMC modes (including the zero rate mode when transmission is prohibited) there is a need for searching for the optimal rate also, making the problem somewhat more complicated. The (discrete) search is performed on tables relating the rates with the channel state (for a given QoS). Fortunately, there have been proposed algorithms for this setting with linear complexity with respect to M and N i.311.

Another simplification comes by imposing equal power allocation for all sub-carriers, e.g. setting $p_{m_n^*,n} = \overline{P}/N$. In that case both the capacity-based allocation and the discrete-rate allocation simplify considerably. It can be shown that for a large number of sub-carriers and/or users and/or available AMC modes, the restriction of equal power allocation results in small performance degradation [i.311] and [i.314].

Although the MSR algorithm for $w_m = 1, \forall m$ provides the largest sum rate, it is a highly unfair algorithm when the users' channels are not identically distributed. For example, due to the various distances of the users from the base station and shadowing effects, their corresponding channels will have different average SNR. In that setting, since the MSR algorithm always selects the user with the best channel conditions, users with poor channel conditions will be rarely scheduled, which means very low throughput and large delays that can be unacceptable by certain applications. One remedy to this problem is the proper adjustment of the weights w_m. Certain users that would otherwise be neglected by the MSR algorithm for $w_m = 1, \forall m$, get larger weights and, therefore, increased priority by the scheduling algorithm.

I.6.1.2 Proportional Fairness (PF) Algorithm

In order to partially compensate for the unfairness property of the MSR algorithm, while at the same time provide sum rates that exceed non-channel-aware approaches the PF algorithm was proposed [i.298]. The resulting algorithm manages to provide a descent amount of resources even to the users with worst channel conditions (i.e. low average SNR). In its most simple (and common) setting the PF algorithm assigns users as follows [i.298] and [i.313].

$$m_n^* = \arg \max_{m \in [1,M]} r_{m,n}(h_{m,n}) \big/ \hat{r}_m$$

where m_n^* is the winner user on sub-carrier n, $r_{m,n}$ is the same quantity defined in clause 3.1.1 and $\hat{r}_m$ is (an estimate of) the average rate r_m assigned to the m-th user from the start of transmission (summed over all sub-carriers). An estimate of the average rate is obtained by the scheduler every scheduling instant by a tracking algorithm (similar to an LMS filter) as:

$$\hat{r}_m \leftarrow \hat{r}_m + \beta \left(\sum_{n \in \mathrm{N}_m^*} r_{m,n} - \hat{r}_m \right)$$

where N_m^* is the set of sub-carriers allocated to user m, and β is a small positive constant (step size parameter). Note that the algorithm performs only sub-carrier and rate allocation and assumes fixed power allocation. The advantage of the PF algorithm is that it schedules users not when their channel has the best absolute value (compared to the other users' channels) but when it has reached *its own* relative peak. When the user channels are independent and have the same dynamics in time and frequency the PF algorithm schedules each user with equal probability (since the probability of some channel reaching its own relative peak is equal for all users). Of course, assigning a user with poor channel conditions (compared to some other user in the system) results in a lower sum rate. However, the contribution to the sum rate by the weak user assignment is the maximum possible.

Another interesting property of the PF algorithm is that for very small step sizes it converges to a rate allocation that maximizes $\sum_{m=1}^{M} \log r_m$ i.312 i.313.

Another manifestation of the proportional fairness criterion that avoids the on-line tracking and allows for non-constant power allocation is the following optimization problem [i.300]:

$$\underset{\{r_{m,n}\},\{p_{m,n}\}}{\text{maximize}} \sum_{m=1}^{M} w_m \sum_{n=1}^{N} r_{m,n}(h_{m,n}, p_{m,n})$$

$$\text{s.t.} \sum_{m=1}^{M} \sum_{n=1}^{N} p_{m,n} \le \overline{P}, \ \frac{r_1}{\beta_1} = \frac{r_2}{\beta_2} = \cdots = \frac{r_M}{\beta_M}$$

where β_m are positive constants that impose the required proportionality among user rates. In principle, a rate assignment satisfying the equality constraint can be found when continuous (capacity) rates are assumed. For the case of discrete rates this is not possible in general and some relaxation is employed. However, even in the continuous rates case, solution of the above problem is difficult. Simplified algorithms for close to optimal performance have been proposed [i.300].

I.6.1.3 Maximum Fairness (MF) algorithm

Another notion of fairness is related to maximizing the minimum rate of all users. An MF algorithm solves the following optimization problem:

$$\underset{\{r_{m,n}\},\{p_{m,n}\}}{\text{maximize}} \ \min_{m\in[1,M]} r_m(h_m, p_m)$$

$$\text{s.t.} \ \sum_{m=1}^{M}\sum_{n=1}^{N} p_{m,n} \le \overline{P} .$$

It can be shown that this problem is the same as the second version of the PF described in the previous clause for $\beta_1 = \beta_2 = \ldots = \beta_M = 1$, i.e. maximization of the minimum rate is accomplished by setting every user's rate equal to that. Therefore, the algorithm loses a large part of the available sum rate provided by the MSR algorithm, especially when the average user SNR differences are large. The optimal solution of the MF problem is difficult, and sub-optimal approaches have been proposed [i.306].

I.6.1.4 Extensions

The previous clauses described the two extremes of the channel-aware scheduling algorithms: maximize sum rate with the cost of low fairness and maximize minimum rate (fairness) with the cost of reduced sum rate. The PF algorithm lays between the two extremes, hence its popularity. However, the related optimization problems are formulated by objective and constraint (inequality) functions involving only rate and power. For continuous (capacity) rates, the resulting algorithm is the optimal from an information theoretic perspective, i.e. provides limits for any scheduling algorithm. However, this optimal performance is achieved by very long codes without any delay considerations which may not be true in practice. A more sophisticated algorithm should also take into account delay considerations such as queue lengths and minimum instantaneous rate constraints [i.307]. Of course, imposing additional constraints results in more complicated problems and sub-optimal solutions are usually pursued.

I.6.2 Partial channel state information

All the above algorithms lead to centralized scheduling algorithms, i.e. the base station compiles the information about the channel status of every user and decides based on a certain criterion. Clearly, a good channel estimate at the base station is essential for implementing channel-aware scheduling. For FDD systems, the downlink channel is first estimated by each user, using pilot symbols and then fed-back to the base station by a CQI message. In static/low mobility cases this approach is viable, as it allows plenty of time for a good channel estimate, while the channel feedback messages do not consume many resources since they are transmitted infrequently. However, in cases of moderate/high mobility, it is difficult for the base station to maintain good channel information for two reasons:

- A small channel coherence interval makes the estimate less accurate.
- The time interval required for the user to feed back its channel status may be larger than (comparable to) the channel's coherence time, making the channel information at the scheduler outdated.

In both cases the transmitter has *partial* channel state information in the sense that, although the CQI value does not correspond to the true channel status, it is related (correlated) with it. The degree of correlation between the true channel state and the channel estimate at the base station depends on the channel's dynamics and feedback time (Assuming close to perfect channel estimation). In principle, one could employ a stochastic model that relates the two quantities, e.g. denoting by $h_{m,n}$ the channel gain of the *m*-th user at the *n*-th sub-carrier and by $\hat{h}_{m,n}$ its corresponding estimate employed by the scheduler, one defines the conditional p.d.f. $p(h_{m,n} \mid \hat{h}_{m,n})$. Its shape may be obtained by various modelling assumptions or can be arbitrarily set to a convenient parameterized form (e.g. Gaussian). In the case of perfect channel estimation and negligible feedback delay, it holds $p(h_{m,n} \mid \hat{h}_{m,n}) = \delta\left(h_{m,n} - \hat{h}_{m,n}\right)$. In all other cases, the conditional p.d.f. will be concentrated around the value of $\hat{h}_{m,n}$ and its variance will reflect the level of uncertainty due to non-perfect channel estimation and non-zero feedback delay. In the extreme case of no channel information (e.g. due to a very large feedback delay) it holds $p(h_{m,n} \mid \hat{h}_{m,n}) = p(h_{m,n})$, i.e, $h_{m,n}$ and $\hat{h}_{m,n}$ are independent and the scheduler operates in a channel-blind setting (knowledge of the a-priori channel p.d.f. is not considered as channel state information).

In total, the level of partial channel information can be divided into the following regions:

- Perfect CSI ($p(h_{m,n} \mid \hat{h}_{m,n}) = \delta\left(h_{m,n} - \hat{h}_{m,n}\right)$): Use the schedulers described in the previous clauses
- Almost perfect CSI ($p(h_{m,n} \mid \hat{h}_{m,n}) \approx \delta\left(h_{m,n} - \hat{h}_{m,n}\right)$): Use the schedulers described in the previous clauses and allow for a (small) possibility of outage, i.e. the true channel value may not "support" the assigned rate. One could employ some "rate back off" in order to reduce this probability.
- Imperfect CSI: Extend the algorithms of the previous clause to incorporate the (partial) channel knowledge provided by $p(h_{m,n} \mid \hat{h}_{m,n})$. In general, the corresponding algorithms are more complicated [i.303] and [i.305].
- Unknown CSI ($p(h_{m,n} \mid \hat{h}_{m,n}) = p(h_{m,n})$): Employ a non-channel-aware scheduling algorithm (e.g. TDMA).

I.6.3 WiMAX Simulation results

In this clause the performance of the algorithms described in the previous clause is evaluated by simulations. The purpose of this analysis is to identify advantages and disadvantages of the various scheduling options in the satellite channel setting.

I.6.3.1 Channel model

A time- and frequency- selective channel model is considered for the evaluation of scheduling algorithms. The channel is essentially described by the properties of the LOS component, the delay spread and the relative movement between transmitter and receiver. For link-level simulation purposes it suffices to generate "snapshots" of the channel with a period equal to the TTI. This snapshot is to be interpreted as the mean channel value during the slot interval. This is because channel variations within the slot can not be exploited by the scheduler (it is assumed that scheduling decisions are made on a slot-by-slot basis and not within the slot). Therefore, only large-to-moderate channel variations are captured by the model (i.e. small scale channel fading, typically considered by PHY level algorithm evaluation, is ignored).

I.6.3.1.1 First Order Statistics

A first order statistical description of the channel is adequate when time-invariant conditions are assumed. If this is not the case (i.e. channel shows time-selectivity) high order statistics should be additionally specified. Frequency selectivity is observed when high bandwidth signalling is utilized at the PHY layer. Equivalently, the sampling (symbol) period T_s (sec) of the PHY layer is small enough to resolve more that one channel paths, resulting in a multipath complex-baseband, discrete-time equivalent channel $h[l];\ l = 0,1,\ldots,L_h - 1$, which captures the effect of both the physical medium and the transmitter and receiver filters. For a maximum delay of T_m (sec), the number of taps is equal to $L_h = \lfloor T_m / T_s \rfloor + 1$, where $\lfloor \cdot \rfloor$ denotes integer floor. The first tap $h[0]$ corresponds to the LOS path, which is of particular importance in the modelling of satellite channels.

In general, the LOS component is composed by two parts: A deterministic one, corresponding to the gain of the LOS between satellite and receiver and a random one which is due to multipath components whose time of arrival is smaller than T_s (and, therefore, can not be resolved from the LOS component). Under this modelling $h[0]$ can be written as:

$$h[0] = s + r\,,$$

where s is the deterministic part (real valued) and r is the multipath component (complex valued). A typical statistical model is based on the Loo distribution, which essentially specifies the dB version of s to be log-normal distributed and r to be a zero mean complex Gaussian random variable. The parameters describing these two distributions are obtained from tables generated by measurement experiments. For conditions when no LOS exists s is zero.

The excess taps $h[1], h[2], \ldots, h[L_h - 1]$ are modelled as independent complex Gaussian random variables of zero mean and of variance specified by the channel's Power Delay Profile (PDP). Typical PDPs are exponential or uniform. Again, these variance values are obtained by experimental results. Typically the energy of the excess taps is much smaller than the energy of the LOS component (if it exists).

I.6.3.1.2 Second Order Statistics

When time-selectivity is assumed the channel is now described by a time varying impulse response $h[n;l]$ where (discrete-time) index *n* denotes time (i.e. slot) and *l* denotes delay (i.e. mutlipath component). The second order statistics of the channel provide a characterization of the channel variation in time. For link-level purposes only moderate-to-large scale fading is considered. The large scale fading corresponds to changes due to gross shadowing conditions. These (very) slow changes are described by a two-state Markov model. State 0 corresponds to the "good" state, where there is a LOS between satellite and receiver (large SNR condition), whereas state 1 corresponds to the case of deep shadowing with no LOS (small SNR condition). The channel stays in a particular state for a, so called, minimum state length (typically 3 m to 5 m), which can be related to an appropriate coherence interval depending on the speed of relative movement. At the end of the coherence interval the channel either stays within the same state or changes to the other, according to transition probabilities, whose values are based on measurement experiments. In addition, the steady state probabilities can be computed based on these transition probabilities.

During each state interval (which typically spans a large number of slots), moderate scale channel variations are modelled as a (sampled) continuous process with a coherence interval corresponding to a correlation distance of, typically, 1 m to 3 m. This process is generated by filtering white noise (of appropriate first order statistics) by a low-pass filter with cut-off frequency corresponding to the coherence interval considered. Figure I.47 shows a qualitative example of a channel realization (for one path).

In order to avoid channel realizations with sharp transitions at the end of a large-scale fading coherence interval due to state change, the output of the simulator is further processed by an additional low-pass filer (implemented as a moving average) for smoothing.

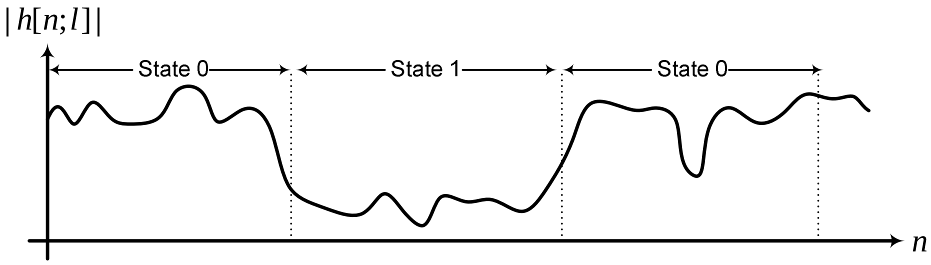


Figure I.47: One-path channel realization

I.6.3.2 Channel model parameters

The channel employed in all simulation was the one described earlier. In order to make the frequency diversity effect more visible the maximum delay spread was set to 2,7 µsec, corresponding to the Case-1 model of the MAESTRO channel (outdoor rural - satellite LOS with many rays). The large scale shadowing effects were modelled by a two state model (good/bad), with the corresponding channel parameters shown in the following table. In each simulation run the state of each user channel was randomly selected based on their steady state probabilities and assumed to remain the same for all simulation time (i.e. no state transitions were assumed) during simulation interval.

Table I.20: Channel parameters for "good" state

Steady state probability	0,8
Power of the LOS component (a_{dB})	-1 dB
Standard deviation of the LOS power (ψ_{dB})	3 dB
LOS scatter component power relative to LOS power ($2\sigma^2$)	-9 dB
Multipath components sum power	Linearly decreasing from -9 dB to -16 dB
Maximum delay spread	2,7 µsec
Coherence distance	10 m

Table I.21: Channel parameters for "bad" state

Steady state probability	0,2
Power of the LOS component (a_{dB})	-10 dB
Standard deviation of the LOS power (ψ_{dB})	3 dB
LOS scatter component power relative to LOS power ($2\sigma^2$)	-9 dB
Multipath components sum power	Linearly decreasing from -9 dB to -16 dB
Maximum delay spread	2,7 µsec
Coherence distance	10 m

I.6.3.3 WiMAX OFDM(A) system parameters

The OFDM(A) system parameters are detailed in the following tables. WiMAX compatible parameters were selected, although the corresponding results and conclusions hold also for the LTE scenario.

Table I.22: System parameters for the 1,25MHz bandwidth usage

Channel bandwidth	1,25 MHz
FFT size	128
Subcarrier spacing	10,94 KHz
Active sub-carriers	84
Data subcarriers	72
Sub-channelization	Band AMC 3 x 2
# of data sub-carriers per sub-channel per OFDM symbol	24
Number of sub-channels per OFDMA symbol	3
# of antennas (Tx, Rx)	(1, 1)
Channel estimation	Perfect
Feedback delay	0,5 s
Service type	best effort / full buffer

Table I.23: System parameters for the 5MHz bandwidth usage

Channel bandwidth	5 MHz
FFT size	512
Subcarrier spacing	10,94 KHz
Active sub-carriers	420
Data subcarriers	360
Subchannelization	Band AMC 3 x 2
# of data sub-carriers per sub-channel per OFDMA symbol	24
Number of sub-channels per OFDMA symbol	15
# of antennas (Tx, Rx)	(1, 1), (2, 1)
Channel estimation	Perfect
Feedback delay	0,5 s
Service type	best effort / full buffer

Note that the ratio (#of data sub-carriers)/(channel bandwidth) is larger for the 5 MHz system, which means better bandwidth utilization. For fair comparison the total power used by the scheduler is the same in both cases.

For continuous rate (capacity) scheduling and no mobility the exact value of the TTI interval is irrelevant as long as it is sufficiently large to allow transmission of long codewords generated by powerful capacity achieving codes. In the case of mobility it is assumed that the channel remains static within a TTI, i.e. a quasi-static channel is assumed. This assumption is a reasonable approximation for low mobility, but quickly degrades when high mobility is assumed. However, in that setting the following simplifying assumption is made: The channel variation with respect to its mean value (that is used by the scheduler) within a TTI can be tracked by the receiver and is sufficiently fast such as the ergodic capacity can be achieved. Note that channel variations within a TTI can not be known or taken advantage of by the scheduler.

I.6.3.4 Basic scheduling options

I.6.3.4.1 Minimum resource allocation block (sub-channelization)

For scheduling purposes the standards have defined sub-channels as the minimum resource allocation block. Each sub-channel is composed from a set of sub-carriers (48 in WiMAX). The distribution of a sub-channel's sub-carriers within the system bandwidth is important in overall performance. A sub channel with largely separated sub-carriers allows for diversity gain at the receiving end, which is advantageous in cases when no channel information is available at the transmitter. On the other hand, a sub channel composed of contiguous sub-carriers does not provide much diversity (especially when all subcarriers are within the channel's coherence bandwidth) but allow for exploiting multiuser diversity when the scheduler has channel state information available [i.316].

In order to examine the limits of sophisticated channel-aware scheduling, sub-channels with contiguous sub-carriers (referred to as band AMC sub channel in the WiMAX standard) were assumed in the simulations. In addition, it is assumed that an assigned user occupies the corresponding frequencies for all the duration of the TTI, i.e. for each TTI, the time-frequency space is divided into frequency slots (dictated by the sub-channel structure) and each assigned user gets one or more of these slots. This approach can be viewed as a manifestation of frequency diversity. Time diversity is exploited by changing scheduling decision every TTI (due to channel variations).

I.6.3.4.2 Link-to-system mapping

Based on the channel state information of the system's users the scheduler decides how to allocate the available resources. However, since every sub-channel is composed from a number of sub-carriers, each of which has its own, different SNR value, a mapping is required the maps the individual sub-carrier SNRs to a single measure for the corresponding sub-channel that will be used by the scheduler. For continuous rates (capacity) scheduling a typical option is the mean instantaneous capacity (MIC), which is defined as [i.313].

$$C(\text{subchannel}) = \frac{1}{N_{sub-channel}} \sum_n \log_2 \left(1 + |h_n|^2 \, p_n \big/ \sigma_n^2\right)$$

where index n runs over the sub-carriers belonging to the sub channel, h_n is the sub-carrier channel gain, p_n is the power of the transmitted sub-carrier (to be assigned by the scheduler), $N_{sub-channel}$ is the total number of sub-carriers within a sub-channel and the user index is dropped for simplicity.

For discrete-rate scheduling various link-to-system mappings have been proposed in the literature that map the sub-carrier SNRs to a single *effective* SNR [i.313].

I.6.3.5 Performance in static environment (no mobility)

In this clause the performance of the scheduling algorithms is evaluated for the case when all users are static. This test case reveals the ultimate limits of the algorithms since the introduction of mobility can only decrease performance (in the long run). Another viewpoint of the results of this clause is that they correspond to the (artificial) case when feedback delay is zero, and can therefore serve as a benchmark for the non-zero feedback case.

The algorithm performance is evaluated for the following setting:

- Number of users: variable (3 to 30).
- Average user (channel) SNR per subcarrier: variable (0 dB to 20 dB), with uniform probability.

In all cases there is a "best" user which has always average sub-carrier *channel SNR* equal to 20 dB and a "worst user with average channel SNR 0 dB.

Continuous rate scheduling was considered with constant power loading

The "channel SNR" of each sub-carrier is defined by the ratio $E\{|h_{m,n}|^2\} \big/ \sigma_m^2$, where $h_{m,n}$ is the complex gain of the n-th sub-carrier of the m-th user, E denotes the statistical average and σ_m^2 is the noise power of the *m*-th user (assumed to be the same for all sub-carriers). It is noted that this SNR definition assumes that the user is in the good channel state. If the users happens to be in severe shadowing (bad channel state) the "true" channel SNR will be 8 dB lower on the average.

Since continuous rate is assumed the rates assigned for a particular channel realization can be viewed as the maximum rate that can be transmitted assuming capacity achieving coding. By performing this computation over the ensemble of channel realizations, the resulting averaged rates are the *ergodic rates* achieved by the scheduler. In order to examine the scheduling performance both in terms of channel usage efficiency (reflected by the sum rate) and fairness the following metrics were computed:

- Sum (aggregate) rate.
- Best user rate.
- Worst user rate.
- Average user rate.

All rates are depicted in units of bits/sec for various values of system load (number of active users).

The scheduling algorithms examined are:

- Maximum Sum Rate (MSR)
- Proportional Fairness (PF)
- Maximum Fairness (MF)
- Time Division Multiplexing (TDMA)

The first three algorithms were briefly described earlier and serve as a representative example of OFDMA scheduling algorithms. The TDMA scheduler is essentially a non-channel-aware scheduler that assigns the whole bandwidth to each user periodically every TTI, irrespective of its channel conditions. The purpose of examining TDMA is to study the gains provided by the sophisticated channel-aware algorithms. It is noted that for the TDMA case, the *ergodic* achievable rate is depicted (i.e. infinite length coding is considered with no outage events) which serves as a performance upper bound in the case of small length coding.

I.6.3.5.1 Results for the 1,25 MHz system parameters

The results for the 1,25 MHz system parameters (table I.22) are shown in figures I.48 to I.51. In all cases constant power allocation was employed, where the power of each assigned subcarrier was normalized to $p_{m_n^*,n}=1$ (n is the sub-carrier index and m_n^* is the winner user for this sub-carrier). The following observations can be made:

The MSR algorithm clearly outperforms all others in terms of sum rate, as expected, by a factor of two. PF and MF have similar sum rate performance, which is a little higher that the simple TDMA scheduling. However, the cost for the superior sum rate performance of the MSR algorithm is that it completely neglects the worst user and consistently schedules the best user, in order to exploit his good channel conditions. On the other hand, the other three algorithms (including TDMA) do assign a non-negligible part of the resources to the worst case user (which is also reflected by the reduced sum rate). Both PF and MF outperform TDMA for the worst user rates. Interestingly, the worst user performance of the MF algorithm is outperformed by the corresponding one of the PF algorithm. This can be justified by the use of a sub-optimal algorithm for the solution of the min-max problem associated to the MF scheduler and the fact that for the 1,25 MHz system only 3 sub-channels are available per frame. In terms of average rate, all algorithms outperform TDMA, with MSR achieving the best rates.

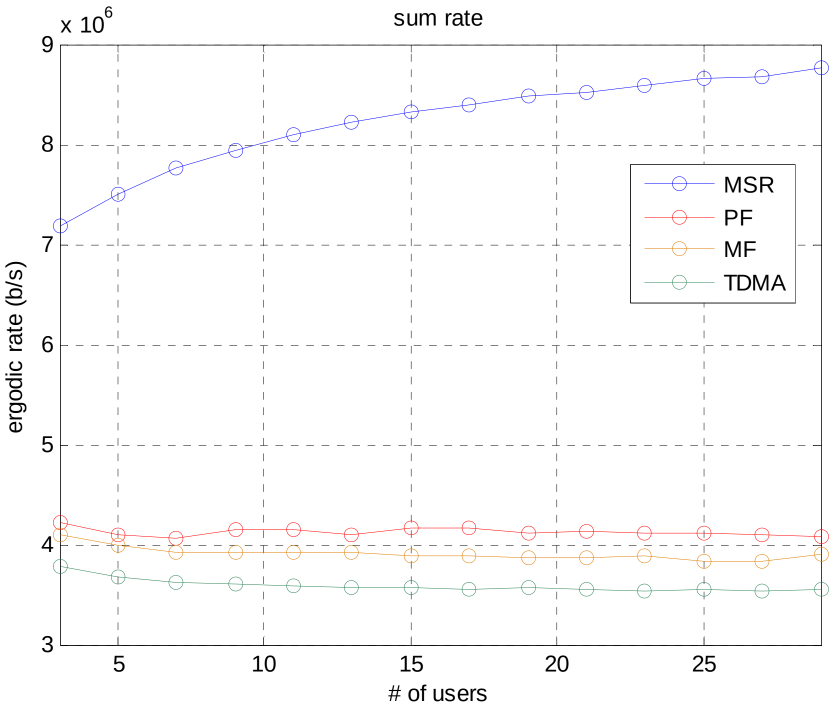


Figure I.48: Sum rate performance (1,25 MHz system)

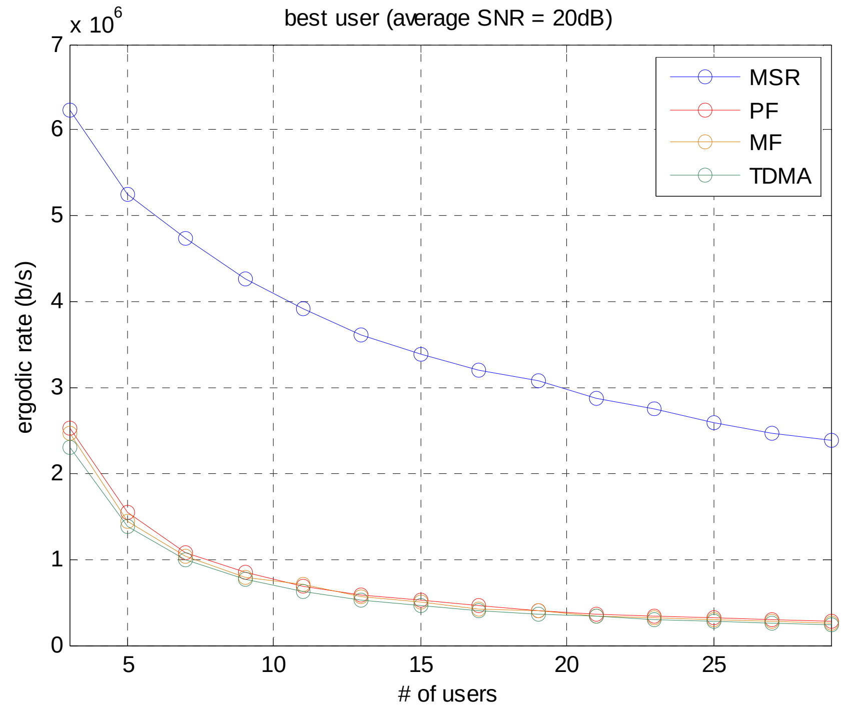


Figure I.49: best user rate performance (1,25 MHz system)

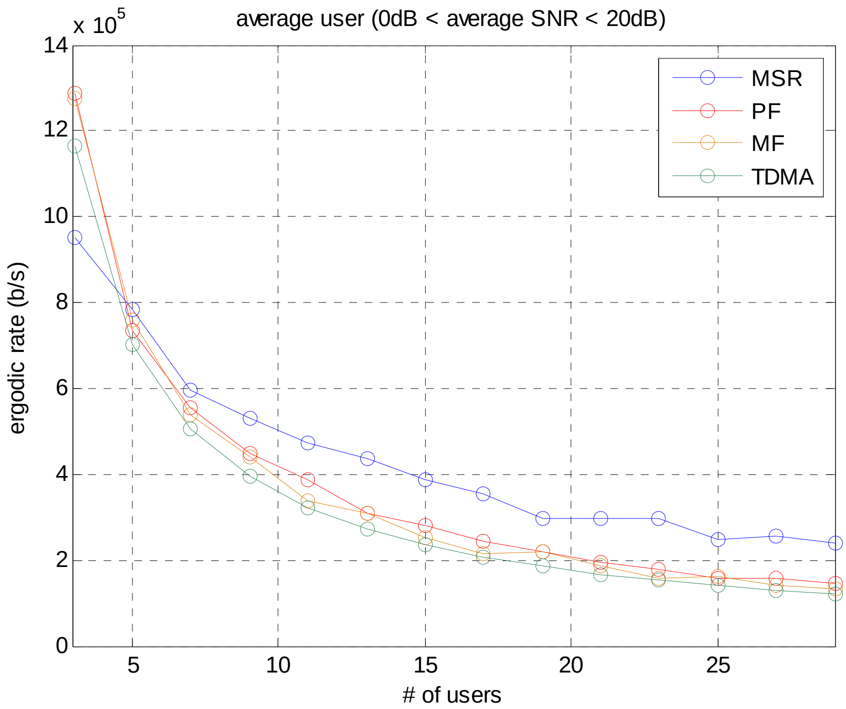


Figure I.50: average user rate performance (1,25 MHz system)

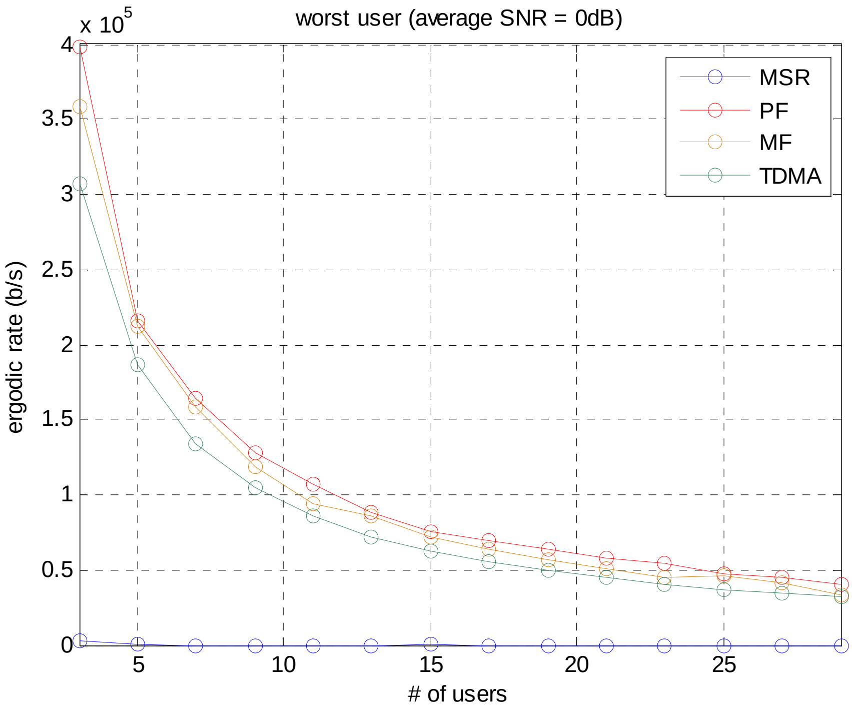


Figure I.51: worst user rate performance (1,25 MHz system)

I.6.3.5.2 Results for the 5 MHz system parameters

The results for the 5 MHz system parameters (table I.23) are shown in figures I.52 to I.55. In order to have a fair comparison with the 1,25 MHz system the same total power was used, i.e. $p_{m_n^*,n} = 72/360 = 1/5$ (the nominator and denominator correspond to the active number of sub-carriers for the 1,25 MHz and 5 MHz systems, respectively). In addition, the case of transmit diversity when the base station employs two transmit antennas and the receiver one is also examined. All other simulation parameters are the same as in the previous clause.

It can be seen that the rates have increased in all cases at least by a factor of two, compared to the 1,25 MHz system. This is due to the increased bandwidth which translated to increased frequency diversity (note that the available sub channels have increased to 15 compared to 3 in the previous clause) and better bandwidth utilization (as stated in clause 3.3.3). However, the general trends of the algorithms' performance are, in most part, the same. Clearly, the increase of available degrees of freedom leads to improved performance. The most notable difference here, is that the gains of channel-aware algorithms are more pronounced compared to the TDMA performance and the MF algorithm now provides the best rate for the worst user.

The use of transmit diversity increases achievable rates, as expected. However, this performance gain is in most cases minimal with the exception on the performance of MF algorithm. This is due to the fact that transmit diversity offers SNR (and not multiplexing) gain [i.310], which is beneficial for low SNR communications (and therefore advantageous for the users with bad channel conditions). This means that the MF algorithm which promotes the worst user performance is greatly enhanced. On the other hand, the use of transmit diversity leads to performance degradation in some cases. For example, the sum rate of the MSR algorithm is reduced in the regime of many active users. This can be explained by noticing that transmit diversity has an "SNR hardening" effect, i.e. users SNR fluctuates less around its mean value. But those fluctuations are essential for multiuser diversity [i.298] and, consequently, larger sum rates. Fortunately, the loss is rather small. Therefore, it seems that incorporation of transmit diversity is beneficial in terms of enhancing the service experience of users with poor channel conditions.

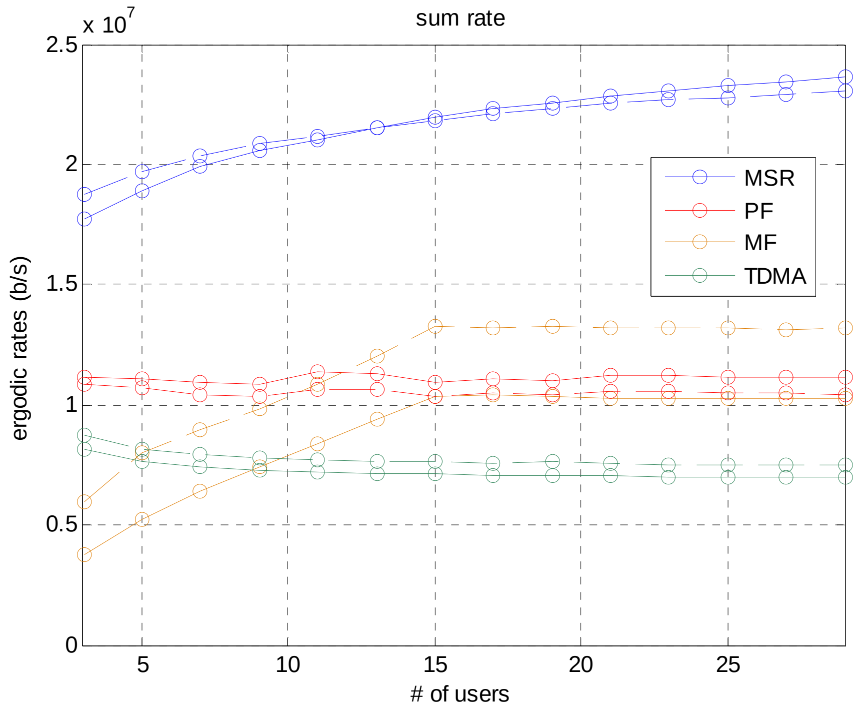


Figure I.52: Sum rate performance (5 MHz system). Solid lines: (1, 1), Dashed lines: (2, 1)

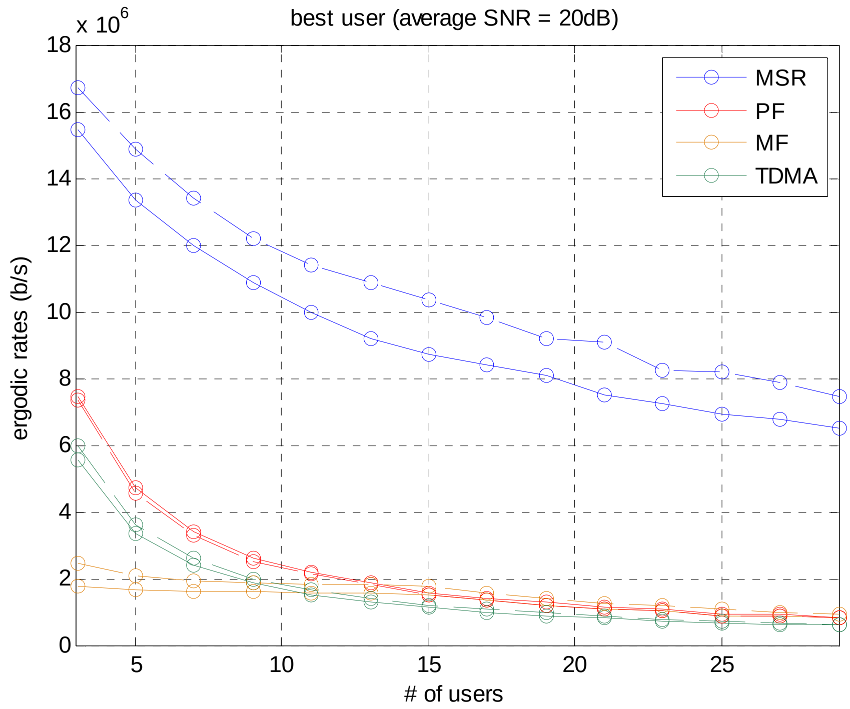


Figure I.53: best user rate performance (5 MHz system). Solid lines: (1, 1), Dashed lines: (2, 1)

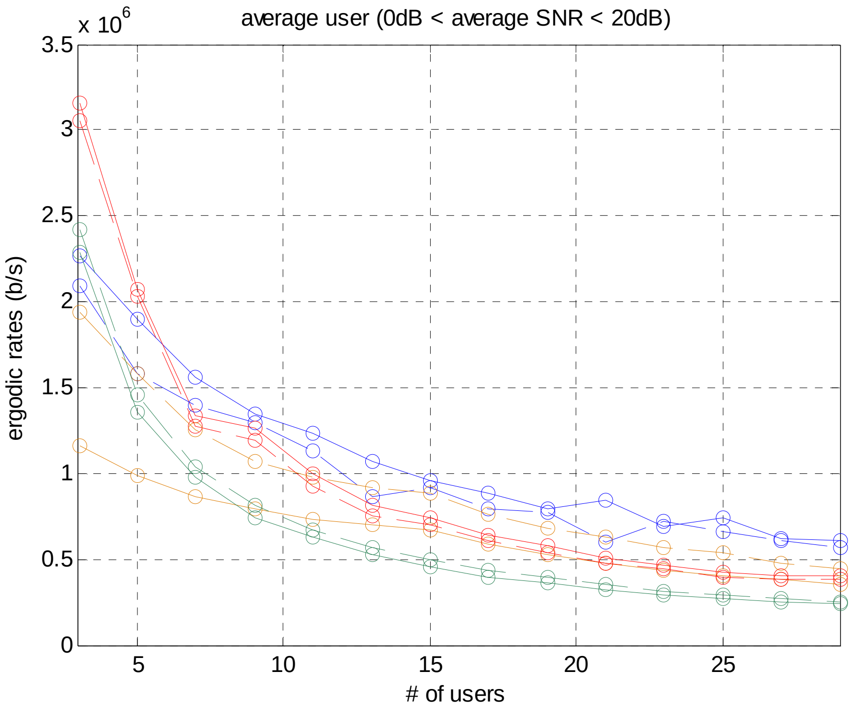


Figure I.54: average user rate performance (5 MHz system). Solid lines: (1, 1), Dashed lines: (2, 1)

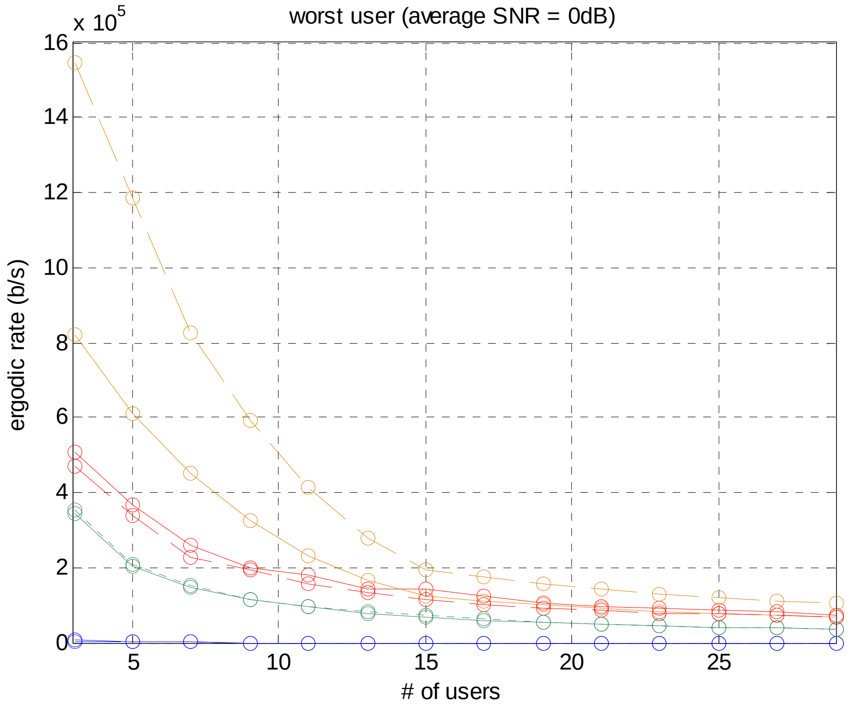


Figure I.55: worst user rate performance (5 MHz system). Solid lines: (1, 1), Dashed lines: (2, 1)

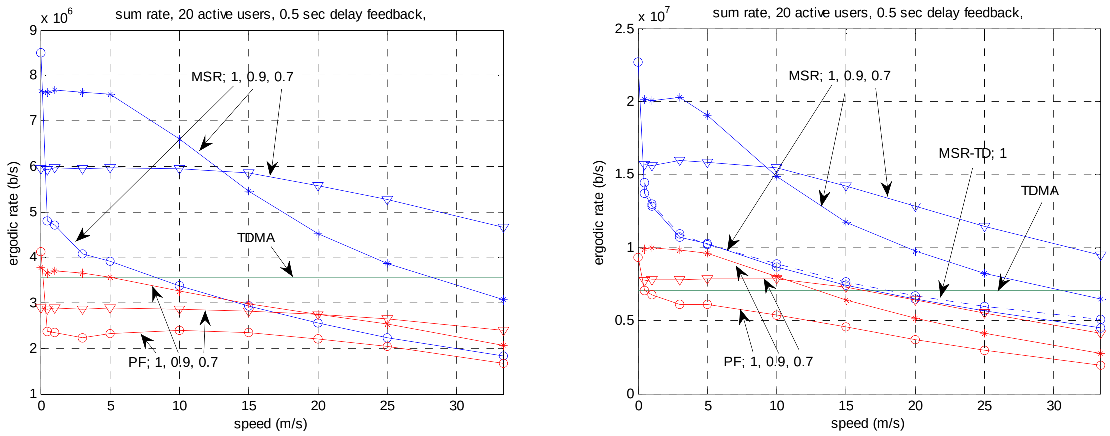
I.6.3.6 Performance under mobility

In this clause the performance of scheduling in the case of user mobility is examined, and the performance degradation due tot feedback delay is quantified. The channel parameters are those described earlier. Users are assumed to have the same speed (mobility). This is unrealistic in practice but can serve as a worst case scenario (when high mobility is considered for all users). For each simulation 20 users were assumed active.

Figures I.56 to I.59 depict the sum, best user, average user and worst user rates, respectively, as a function of the user speed, for the MSR, PF and TDMA schedulers. Results for both 1,25 MHz and 5 MHz system bandwidth are shown. In some cases the performance with 2x1 transmit diversity (TD) is shown. For the MSR and PF algorithms the scheduler assumes that it has perfect information about the user channels and proceeds exactly as in the no mobility case (i.e. the scheduler ignores the error in the channel state information due to feedback delay). Note that, due to channel uncertainty, it is possible that the assigned rate based on the scheduler's channel's knowledge does not match the true channel, i.e. the rate exceeds the true channel capabilities for a given QoS. In that case there is an outage event, and the assigned recourses go wasted. A simple remedy to this effect is to reduce the "nominal" transmission rate, i.e. instead of transmitting at the rate dictated by the channel estimate used by the scheduler, the rate is reduced by a certain amount. In the simulations, a rate reduction of 10 % and 30 % was examined (corresponding to a multiplication of the nominal rate by 0,9 and 0,7 respectively). All figures depict the successfully transmitted rate (goodput).

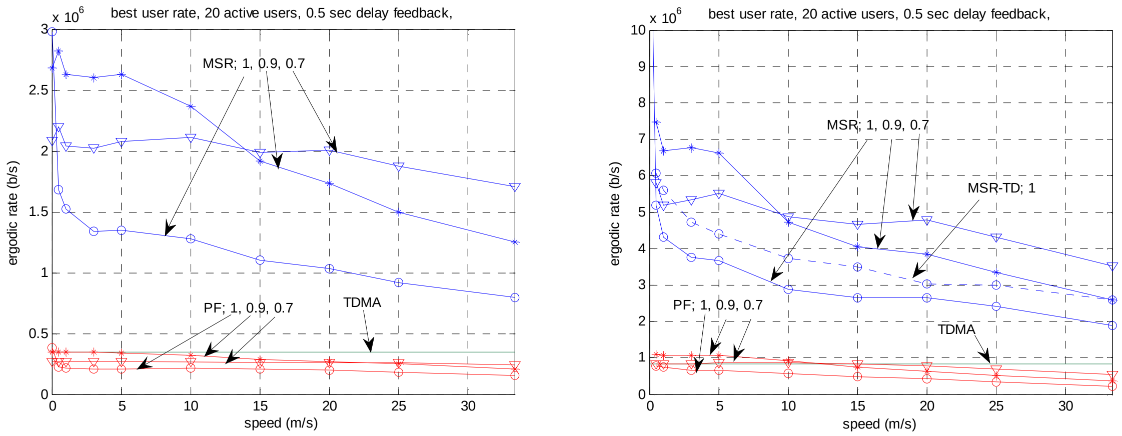
In terms of sum rate, it can be seen that transmitting at the nominal rate (corresponding to curves label with "1" in the figures) leads to severe performance degradation of the MSR algorithm as the mobility increases. Actually, even for small user speeds the successfully transmuted rate is reduced almost by half. This can be explained by noting that for low mobility, and therefore small channel uncertainty, the true channel capacity given the channel knowledge at the transmitter side is well approximated by a Gaussian p.d.f. centred at the nominal rate [i.312]. Therefore half of the time the true channel capacity will exceed the nominal rate and the other half will not be able to support it. Transmitting at a reduced rate (curves labelled 0,9, 0, and 7, in the figure) manages to reduce the outage probability and leads to an increase in sum rate. Clearly, the larger the rate is reduced the smaller the outage probability, but the cost is that for low mobility this rate reduction is unnecessary and wasteful. In any case, the MSR significantly outperforms TDMA in terms of sum rate, by proper adjustment of the transmitted rate back off. The PF performance also seems to be severely affected by the channel uncertainty, which leads to eliminating the (small) performance gain exhibited over TDMA for the no mobility case in terms of sum rate. The same observations more or less hold for the rates of the best and average users. MSR can retain its performance advantage from TDMA by proper rate back off, whereas PF achieves smaller rates, and for the case of the 1,25 MHz system is inferior to the TDMA scheduler. Only for the worst user case the situation changes, since PF shows the best performance compared to TDMA and MSR (the latter practically giving no resources).

Comparing the performance with respect to the system bandwidth, it can be seen that larger bandwidth is advantageous both from an absolute rate perspective and a maximum user speed below which channel-aware scheduling outperforms TDMA. The PF algorithm seems to be mostly benefited by a large system bandwidth. Table I.24 summarizes the sensitivity of the MSR and PF algorithms with respect to user mobility, based on the numerical results shown in this clause. Note that the results of this clause are generated assuming a shadowing coherence distance of 10 m, typical of a rural scenario. For an urban scenario where the correlation distance increases (50 m is a typical value), the results of this clause can be interpolated in order to find the algorithms performance for this case also.



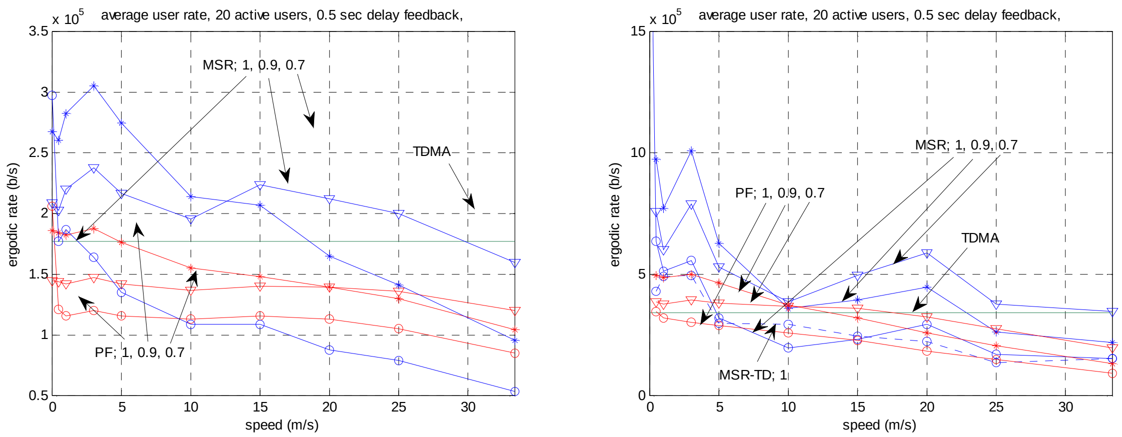
a b

Figure I.56: sum rate performance, a) 1,25 MHz system, b) 5 MHz system



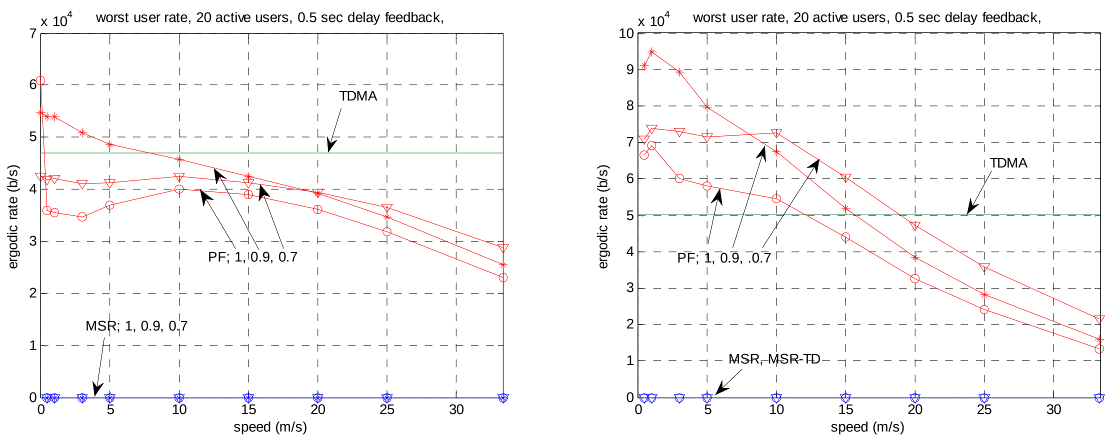
a b

Figure I.57: best user rate performance a) 1,25 MHz system, b) 5 MHz system



a b

Figure I.58: average user rate performance a) 1,25 MHz system, b) 5 MHz system



a b

Figure I.59: worst user rate performance a) 1,25 MHz system, b) 5 MHz system

Table I.24: Maximum speed for which channel-aware scheduling (MSR/PF) outperforms channel-blind TDMA. Channel coherence length = 10 m

Performance metric/System BW	1,25 MHz	5 MHz
Sum rate	MSR: > 35 m/s PF: 5 m/s (limited gain)	MSR: > 35 m/s PF: 16 m/s (limited gain)
Best user rate	MSR: > 35 m/s PF: 0m/s (limited gain)	MSR: > 35 m/s PF: 10 m/s (limited gain)
Average user rate	MSR: 30 m/s (moderate gain) PF: 5 m/s (limited gain)	MSR: > 35 m/s (moderate gain) PF: 16 m/s (limited gain)
Worst user rate	MSR: 0 m/s (zero rate) PF: 8 m/s (limited gain)	MSR: 0 m/s (zero rate) PF: 19 m/s (moderate gain)

I.6.4 Resource allocation - Conclusions

Various scheduling algorithms were examined for WiMAX OFDMA downlink signalling, for a satellite channel model. The trade-off between sum rate maximization and fairness was discussed. The effect of user mobility, and, in particular, the effect of outdated channel state information at the scheduler was examined by extensive simulations. It was shown that for the feedback delay considered in the satellite scenario, the performance of channel-aware schedulers significantly degrades, even for relatively slow mobility (which would pose no serious problem in a terrestrial signalling scenario). For acceptable operation, the assigned rate is reduced compared to the nominal rate suggested by the channel estimate at the scheduler in order to reduce the outage probability and increase the successfully transmitted rate (goodput). However, the advantage of channel-aware scheduling in this case can be moderate to low, compared to a channel agnostic scheduling, such as TDMA, which necessitates a careful tradeoff analysis between gains and complexity in order to have them adopted in practice in a satellite link.

It is noted that the recommendations drawn from the examination of the forward link also hold for the return link. This is because both links are "dual" in the sense that the objective functions to be maximized by the scheduling algorithms remain the same with the exception that instead of having a total power budget that can be (optimally) distributed among users, as in the case of the forward link, in the return link each user has its own power that can not be changed. This modification will lead to different results in terms of absolute values (typically return link performance is worse due to smaller transmit power capabilities of user terminals) but qualitative conclusions with respect to the performance trends of the various algorithm alternatives or the effect of feedback delay will remain the same.

I.7 End-To-End Assessment of WiMAX OFDMA scheduling over satellite

I.7.1 Forward Link

I.7.1.1 System description

This clause presents system level simulations in order to quantify the performance of channel-aware scheduling algorithms for a downlink OFDMA system (as in WiMAX or LTE) operating under a satellite channel environment. The system parameters employed in the simulations are shown in table I.25. These parameters correspond to a WiMAX system; however, results for LTE systems are expected to have similar performance trends.

Table I.25: System parameters

Channel bandwidth	5 MHz
FFT size	512
Subcarrier spacing	10,94 KHz
Active sub-carriers	420
Data subcarriers	360
Sub channelization	Band AMC 3 x 2
# of data sub-carriers per sub-channel per OFDMA symbol	24
Number of sub-channels per OFDMA symbol	15
Number of AMC modes	12 (0 to 5 b/s/Hz efficiency; see table I.26)
# of antennas (Tx, Rx)	(1, 1)
Channel estimation	Perfect
Feedback delay	0,5 s
Service type	best effort / full buffer
Scheduling algorithm	Maximum sum rate (MSR) Proportional Fairness (PF) Maximum Fairness (MF) Time Division Multiplexing (TDMA) (All algorithms assume perfect CSI at each scheduling instant, regardless of the user(s) mobility; constant power allocation over the subcarriers)

The system adapts its transmission based on channel state information (CSI) for each of the active users. Based on the CSI, the scheduler assigns the system resources, namely, power, rate and bandwidth (subcarriers) to each user in order to optimize some objective function. For simplicity, the scheduler employed here uses constant power allocation for every data subcarrier, which does not pose major performance degradation, especially when the system operates at high-to-moderate SNR levels and the number of users is large [i.310]. The rate the system allocates to a user is chosen among a finite set of rates, usually referred to as adaptive modulation and coding (AMC) modes. These AMC modes correspond to specific combinations of modulation (symbol constellation) and code rate that achieve different levels of spectral efficiency, suitable for the various levels of channel state each user is likely to encounter. For the simulations, the set of AMC modes employed are shown in table I.26. Note that the modulation-code rate pairs 16-QAM-3/4 and 64-QAM-1/2 have the same spectral efficiency. For scheduling purposes the mode with lowest SNR threshold (see discussion below) would typically be employed. In addition to the modes of table I.26, the "null" mode is also employed, corresponding to no transmission at all. This mode will be selected when the channel state of a user is not capable of supporting any of the available AMC modes.

Table I.26: AMC modes employed by the simulated system

Modulation	Code rate	AMC mode rate (b/s/Hz)
QPSK	1/2 2/3 3/4 5/6	1 1,333 1,5 1,667
16-QAM	1/2 2/3 3/4 5/6	2 2,667 3 3,333
64-QAM	1/2 2/3 3/4 5/6	3 4 4,5 5

For a single (point-to-point) link, the rate is adopted based on the instantaneous SNR of the link [i.314] (this statement is not entirely accurate for the frequency selective channel case as discussed below). The optimal strategy is to select the largest rate possible that can deliver the specified quality of service (QoS), e.g. block error rate. Of course, this strategy assumes that the channel remains essentially the same for the transmission of the codeword. Typically, for a specific QoS metric, a table is constructed that matches each available mode to an SNR threshold, i.e. the lower value of the SNR above which the mode can be transmitted satisfying the QoS requirements. This table can also be viewed as specifying a mapping of link SNR values to rates.

Figure I.60 shows this mapping for the ideal/non-realistic QoS of no transmission errors, assuming capacity achieving coding schemes. The solid curves show the capacity (maximum spectral efficiency) that can be achieved by a modulation constrained transmission scheme employing either 4-QAM, 16-QAM, or 64-QAM symbols (blue, red, and black curves, respectively), when there is no limitation in code rate selection (infinite number of modes). For comparison, the optimal spectral efficiency achieved by Gaussian signalling is also depicted [i.310] as it serves as an upper bound for the efficiency of any signalling scheme. As can be seen employing discrete constellations results in a certain performance loss compared to Gaussian signalling (about 2 dB in the large SNR regime) but this approach is preferred in practice due to its simplicity. For each of the solid curves corresponds a dotted curve obtained by discretization of the solid curve, that can be viewed as the SNR to rate mapping given that a finite set of code rates are supported. The SNR-to-rate mapping for the system is obtained by superposition of these curves. Clearly, this discretization operation is suboptimal, e.g. for the specific case considered in figure I-60 and an SNR value of 20 dB, only 5 b/s/Hz efficiency is achieved by the available AMC modes (specifically, by the 64-QAM-5/6 mode) whereas flexibility in the code rate would allow for about 5,7 b/s/Hz (using 64-QAM) and flexibility in the constellation (Gaussian signalling) would allow for about 6,6 b/s/Hz spectral efficiency. It is therefore of crucial importance to identify a system's operational SNR and have a sufficient discretization of the corresponding rate region. However, one should have in mind that employing many modes complicates matters in terms of mode selection and generation and, also, that the gains of using more modes are diminishing after a particular number [i.311].

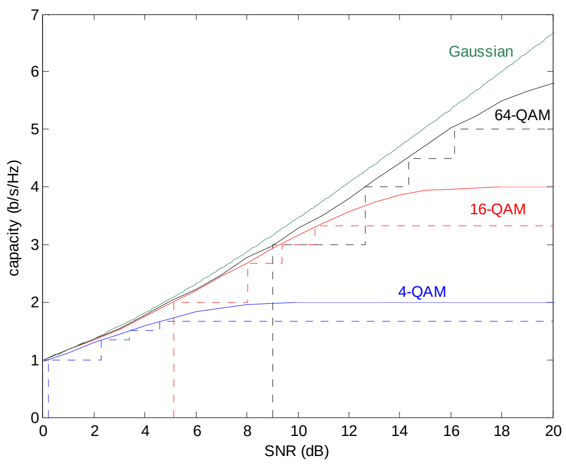


Figure I.60: SNR-to-rate mapping

The AMC mode thresholds of figure I.60 are unrealistic since they assume capacity achieving coding and no complexity limitations. In practise, a system employs coding schemes with finite length codewords and sub-optimal decoding algorithms that lead to performance degradation with respect to the optimal performance (capacity). This degradation is conveniently viewed as an SNR gap Γ, i.e. the thresholds of the modes are increased by Γ, compared to thresholds of capacity achieving modes [i.314].

Based on link-level simulations for AMC codes employing turbo codes, the SNR gap was shown to be equal to Γ~2 dB for a BLER of 10^{-3} using the largest length codewords. The SNR thresholds employed in the simulations are shown in table I.27. Note that employing smaller length codewords would result in higher SNR thresholds. However, the quantitative results and performance trends are the same for both long and short codeword lengths.

Table I.27: SNR thresholds for turbo-coded AMC modes

AMC mode rate (b/s/Hz)	SNR threshold (dB)
1	2,1485
1,333	4,2731
1,5	5,3831
1,667	6,5868
2	7,1430
2,667	10,03
3 (64-QAM)	11,3935
3,333	12,685
4	14,6356
4,5	16,3591
5	18,1745

I.7.1.2 Effective SNR

Initial studies on adaptive modulation and coding assumed the setting of a flat fading channel, i.e. small delay spread compared to the signal's baud rate [i.314]. This assumption directly suggested the link's SNR level to be the key quantity for appropriate mode selection. For systems with high transmission rate and considerable delay spread, the flat fading assumption no longer holds, i.e. the channel is frequency selective over the signal's bandwidth. In that case, the SNR level cannot be directly applied as an appropriate AMC metric since the SNR may change significantly over the system bandwidth. OFDM (A) is a modulation technique that "transforms" the frequency selective channel to a set of parallel flat fading channels. Therefore, in principle, AMC could be employed on a sub-carrier level, independently of the other sub-carriers, based on the SNR value of the sub-carrier only. This strategy is actually optimal with respect to optimizing the system's capacity [i.293] and [i.311]. However, this approach has two major drawbacks:

Frequency selectivity can be exploited in order to decrease the probability of transmission errors (outage) with non-capacity achieving signals [i.312].

For a system with many sub-carriers it is not practical to perform AMC on a sub-carrier level, particularly with a high number of users.

For this reason current OFDMA-based standards (WiMAX, LTE) have adopted the notion of sub-channel. The available frequency space (spanned by the data sub-carriers of the system) is divided into smaller frequency intervals, each containing a few sub-carriers that form the so-called sub-channel (there frequency intervals may not be contiguous). For example, for the system considered in the simulations (table I.5), the 360 data bearing sub-carriers are assigned to 15 sub-channels. Typically, the sub-channels span a few OFDMA symbols. Each sub channel (with all its sub-carriers) is assigned to one user, and a single AMC mode is transmitted over it [i.313]. Note that there are two options for selecting the set of sub-carriers for a sub-channel:

- Use a set of contiguous sub-carriers.
- Use a set of sub-carriers with large separation in frequency.

The first choice is the most preferable in a setting where the transmitter (scheduler) has knowledge of the (user's) channel over the subcarriers, since it allows for maximum *multi-user diversity [*i.316*]*, due to the fact that the SNR level of the sub-carriers is pretty much the same due to correlation. The second choice is most preferable in a setting where the transmitter does not have channel state information (except, probably, for the average SNR level over the frequency), since it achieves maximum *frequency diversity* due to the sub-channel's carriers being spaced far apart, and experiencing (almost) independent fading. (A contiguous set of sub-carriers will typically have much smaller SNR variation.) Of course, both cases of sub-carriers sets could be employed in a setting where the transmitter has channel information or not.

In the case where the transmitter does know the channel over the sub-carriers (the situation considered in the simulations), there is a problem on how to chose the AMC mode that will be selected for a particular sub-channel [i.313]. This problem is more difficult for the second choice of sub-carriers due to the significant SNR variation over the sub-carriers. Choosing an AMC mode according to the best SNR over the sub-channel's sub-carriers will result in wasting sub-carriers with lower SNR as they will not support this mode. Working based on the lowest SNR, will also result in decreased utilization of the highest SNR sub-carriers.

The approach employed by recent research studies and widely accepted is based on the concept of *effective SNR* (ESNR) [i.315]. ESNR is a scalar that represents the sub-channel capabilities in terms of transmitting a particular mode. In essence, the set of SNR levels over the sub-carriers are mapped onto a single ESNR value which is then employed in an identical manner as if the ESNR was the SNR value of a flat fading transmission. This approach makes the AMC mode selection much easier and compatible with "conventional" AMC in flat fading channels. The difficulty is to define this ESNR mapping of the sub-carriers SNRs. There have been various proposals for ESNR calculation [i.315] and [i.313] including:

- Received Bit Mutual Information Rate (RBIR).
- Mean Mutual Information per Bit (MMIB).
- Mean Instantaneous Capacity (MIC).

These approaches try to provide an ESNR value that will reflect the sub-channel capabilities with the highest accuracy. All these approaches are based on the instantaneous mutual information of the set of sub-carriers for a particular channel realization. While the first two methods consider the constellation-constrained mutual information, the third assumes Gaussian signalling, making it less accurate but at the same time more easy to compute. Specifically, the ESNR provided by the MIC approach is

$$ESNR(\text{sub-channel}) = 2^{C(\text{sub-channel})} - 1,$$

where

$$C(\text{subchannel}) = \frac{1}{N_{sub-channel}} \sum_n \log_2 (1 + |h_n|^2 \, p_n / \sigma_n^2) \ \text{(b/s/Hz)}$$

is the averaged instantaneous mutual information of the sub-channel (based on the SNR levels $|h_n|^2 \, p_n / \sigma_n^2$ of its $N_{sub-channel}$ sub-carriers). For simplicity of the simulation, it was assumed that the ESNR value as provided by the MIC approach is valid. No major differences are expected if the effects of an inaccurate ESNR value were also considered.

Since the current analysis aims to examine the (possible) gains of employing a channel aware scheduler, sub-channels composed of contiguous sub-carriers were selected since they are better matched for exploiting multiuser diversity. These sub-channels are called band-AMC sub-channels within the context of the WiMAX standard.

I.7.1.3 Simulation results

I.7.1.3.1 Channel Model

This clause provides simulation results assuming a channel model employed was representative of a satellite channel, as described earlier, composed by two states (good/bad) reflecting the case of severe shadowing. Tables I.28 and I.29 provide a detailed description of the channel parameters.

Table I.28: Channel parameters for "good" state

Steady state probability	0,8
Power of the LOS component (a_{dB})	-1 dB
Standard deviation of the LOS power (ψ_{dB})	3 dB
LOS scatter component power relative to LOS power ($2\sigma^2$)	-9 dB
Multipath components sum power	Linearly decreasing from -9 dB to -16 dB
Maximum delay spread	2,7 µsec
Coherence distance	10 m

Table I.29: Channel parameters for "bad" state

Steady state probability	0.2
Power of the LOS component (a_{dB})	-10 dB
Standard deviation of the LOS power (ψ_{dB})	3 dB
LOS scatter component power relative to LOS power ($2\sigma^2$)	-9 dB
Multipath components sum power	Linearly decreasing from -9 dB to -16 dB
Maximum delay spread	2,7 μsec
Coherence distance	10 m

I.7.1.3.2 Static Channel

In this clause the performance of the scheduling algorithms is evaluated for the case when all users are static. This test case reveals the ultimate limits of the algorithms since the introduction of mobility can only decrease performance (in the long run). Another viewpoint of the results of this clause is that they correspond to the (artificial) case when feedback delay is zero, and can therefore serve as a benchmark for the non-zero feedback case.

The algorithm performance is evaluated for the following setting:

- Number of users: variable (3 - 30)
- Average user (channel) SNR per subcarrier: variable (0 dB to 20 dB), with uniform probability

In all cases there is a "best" user which has always average sub-carrier *channel SNR* equal to 20 dB and a "worst user with average channel SNR 0 dB.

Continuous rate scheduling was considered with constant power loading.

The "channel SNR" of each sub-carrier is defined by the ratio $E\{|h_{m,n}|^2\}/\sigma_m^2$, where $h_{m,n}$ is the complex gain of the n-th sub-carrier of the m-th user, E denotes the statistical average and σ_m^2 is the noise power of the m-th user (assumed to be the same for all sub-carriers). It is noted that this SNR definition assumes that the user is in the good channel state. If the users happens to be in severe shadowing (bad channel state) the "true" channel SNR will be 9 dB lower on the average (see tables I.28 and I.29). Note that the worst case user channel SNR is less than 2,1485, which is the minimum SNR value for transmission (see table I.27). This means that based on its average SNR the worst case user would never be assigned any rate. However, in the opportunistic scheduling setting considered care, transmission can be admitted to the user taking advantage of a good channel realization,

Figures I.61 to I.64 show the:

- sum (aggregate) rate;
- best user rate;
- worst user rate;
- average user rate;

respectively.

Figure I.61 shows the sum rate performance of the selected scheduling algorithms. As expected, the MSR provides the largest sum rate, which actually increases with the number of users since it exploits optimally the multi-user diversity. PF and MF algorithms have about a 7 % to 40 % reduced sum rate compared to the MSR performance depending on the number of users. Interestingly, the sum rate provided by either PF or MF is not sensitive to the number of active users. The performance of the TDMA scheduler is by far the worst, since it fails to exploit any channel information about the users for scheduling purposes.

MSR also clearly outperforms the other options in terms of best user rate. This is because the best user is mostly favoured by the MSR algorithm in order to exploit its good channel conditions and increase the sum rate. MF and PF have similar performance but still significantly better than the TDMA scheduler. Same observations hold for the average user's rate, where now the MSR performance is comparable with the corresponding rate provided by the PF and MF algorithms. However, the MSR algorithm is clearly not a good choice for the worst user, since it does not assign any rate at all. Actually the MSR performance for the worst user is inferior to the TDMA one, since there is a (small) possibility the worst user's channel to be in a sufficiently good state when assigned by the TDMA scheduler. On the other hand, both PF and MF outperform TDMA, i.e. they are better matched to serving users with bad channel conditions.

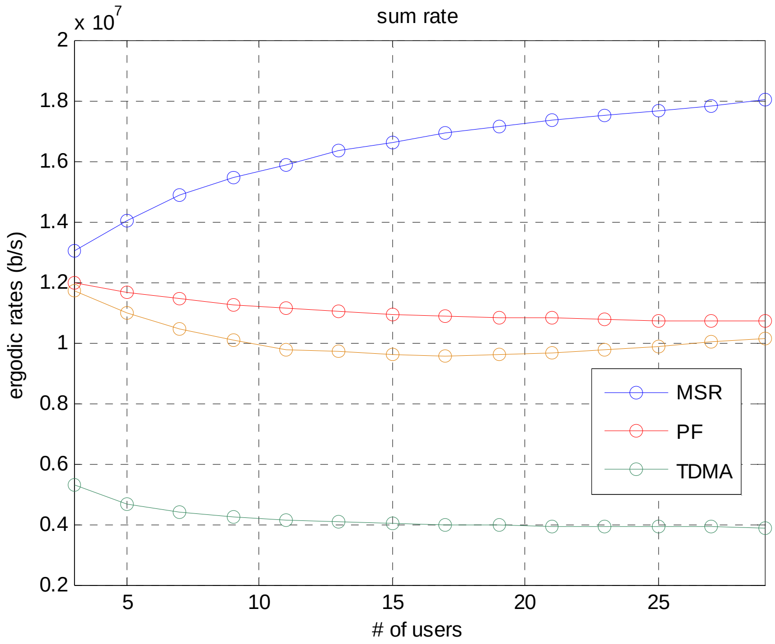


Figure I.61: Sum rate performance

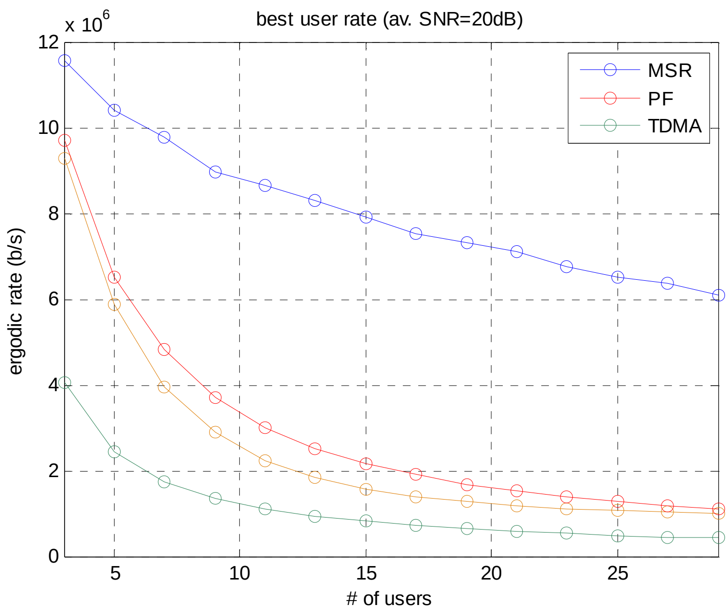


Figure I.62: Best user rate performance

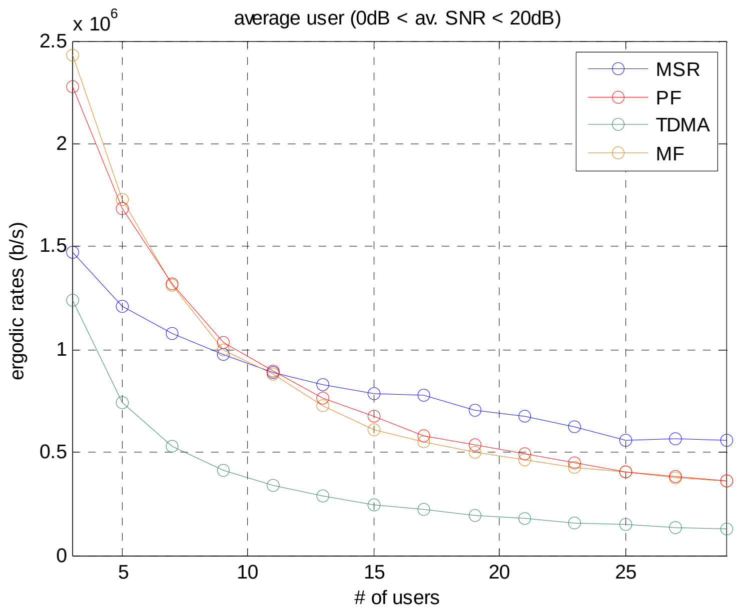


Figure I.63: Average user rate performance

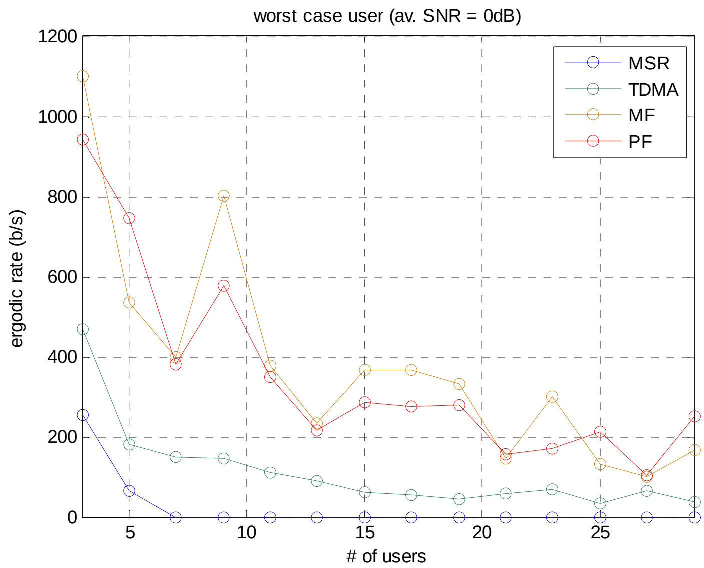


Figure I.64: Worst user rate performance

I.7.1.3.3 Performance under mobility

In this clause the performance of scheduling in the case of user mobility is examined, and the performance degradation due tot feedback delay is quantified. The channel parameters are those described earlier. Users are assumed to have the same speed (mobility). This is unrealistic in practice but can serve as a worst case scenario (when high mobility is considered for all users). For each simulation 20 users were assumed active. The rates assigned by the scheduler and their corresponding (E) SNR threshold were those in table I.27.

Although perfect channel estimation is assumed, the channel state information is received at the scheduler site with a significant delay due to the large round trip time of the satellite system. This makes the channel estimate used by the scheduling algorithm erroneous, i.e. above a certain mobility level, the channel information of the scheduler becomes (more or less) outdated. The effect of this phenomenon is that the nominal rate imposed by the channel information at the scheduler may exceed the true channel capabilities, resulting in an outage event (i.e. reliable transmission, with respect to the chosen QoS metric, is not achieved). Various options can be adopted in this setting, including modelling the channel information error in a statistical manner and devising optimal scheduling strategies based on this model, completely ignoring channel state information at the scheduler and employing a channel-blind scheduling algorithm or reduce the nominal rate implied by the channel state information by a certain amount in order to reduce the probability of outage. In the simulations presented in this clause, the third option was selected due to its simplicity. Note that the rate reduction should be chosen such as the outage event is minimized but at the same time a reasonable amount of rate is successfully transmitted. Simulation showed that gains are observed by transmitting the mode with the largest rate that is smaller by the nominal mode implied by the channel state information.

Figures I.65, I.66, I.67 and I.68, show the performance (successfully transmitted rate, averaged over channel realizations) of the MSR, PF and TDMA algorithms for a setting of 20 active users with average channel SNRs within the range of 0 to 20 dB, chosen with uniform probability, and for various levels of mobility (parameterized by the terminals' speed). Two of the users are named "best" and "worst", similar to the scenario examined in the static channel case. Both options of transmitting at the nominal mode and at the next smaller rate mode was simulated (the latter option denoted by MSR(B), PF(B), TDMA(B) curves; "B" for rate back off).

Figure I.65 shows the sum capacity of the system, where it is seen that the MSR algorithm is the algorithm of choice for slow mobility levels using the nominal rate. However, this algorithm shows significant degradation when moderate mobility exists, due to the large probability of outage events. Employing rate back off manages to significantly compensate for this loss, with the cost of reduced throughput for low mobility. PF seems to have a comparable performance with respect to the MSR algorithm, while employing rate back off does not change performance drastically. Interestingly, both PF and MSR outperform TDMA for the mobility range considered, implying that for the system setting considered, it is advantageous to incorporate (partially outdated) channel information for scheduling purposes rather than rely on an average SNR information for the channel.

Figure I.66 depicts the performance of the best user. As expected, MSR algorithm provides the best performance, whereas PF significantly reduces best user's throughput, which, however, is still better with respect to TDMA performance. Figure I.67 shows the performance of the average user. It seems that PF is more preferable for moderate to high mobility whereas MSR is superior only for low speeds. The worst user's performance is shown in figure I.68. The MSR does not provide any rate to the worst user, since in most cases either there will be another user with better SNR that will get assigned instead, or the worst user's SNR will not be able to support any of the rates of table I.27. Also TDMA fails to deliver non-trivial rates, as the probability that the channel can support a non-trivial rate when the worst user is assigned is small. PF seems to be able to transmit a non-trivial rate to the user (but not its back off rate version), and therefore is preferable. Note that the large deviations shown by the PF curve is due to the small number of simulation runs for accurate estimation of the worst user's performance.

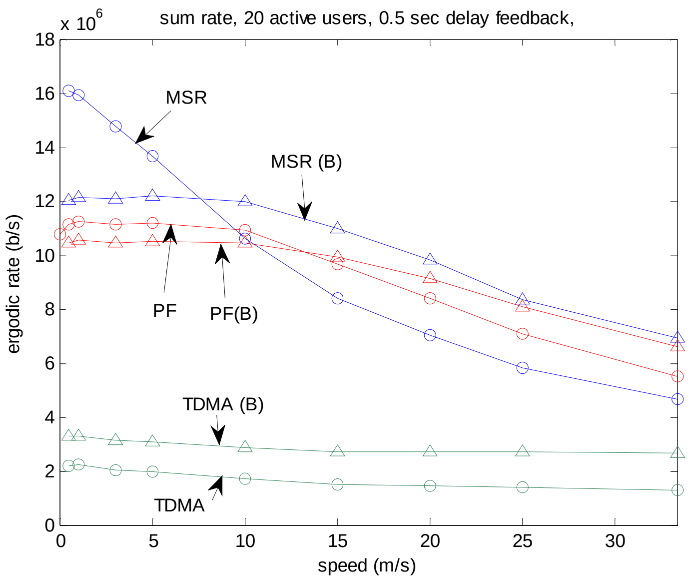


Figure I.65: Sum rate performance versus mobility

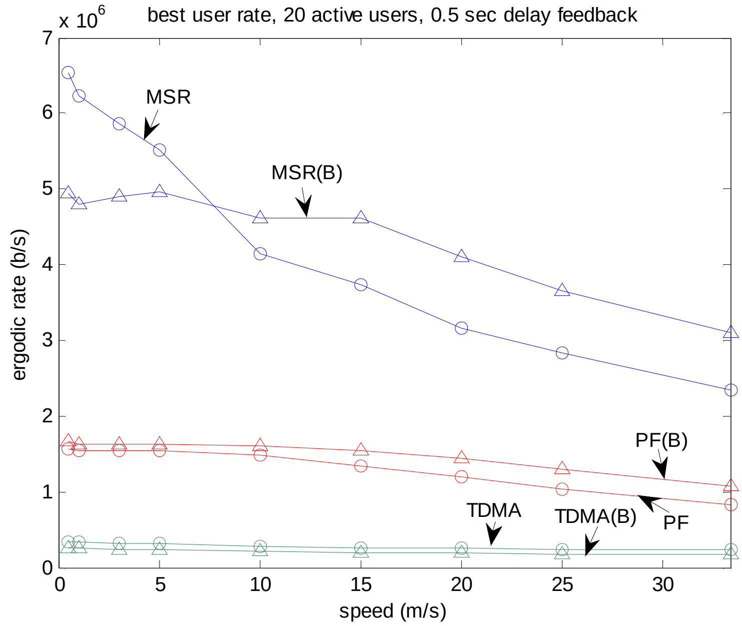


Figure I.66: Best user rate performance versus mobility

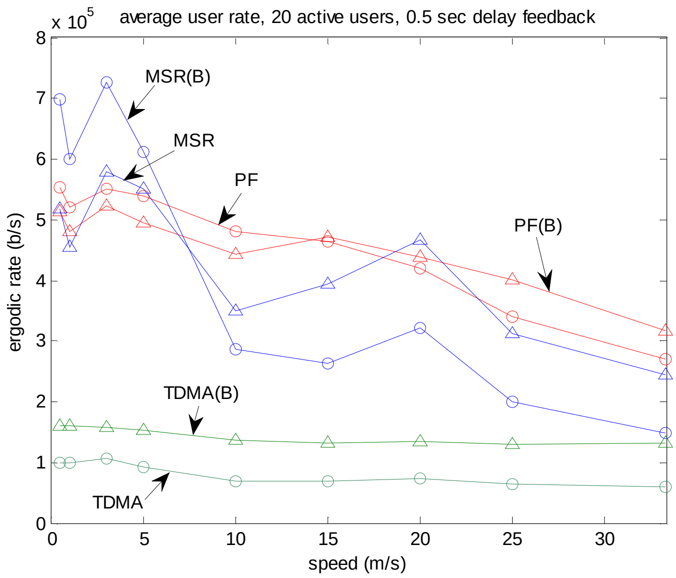


Figure I.67: Average user rate performance versus mobility

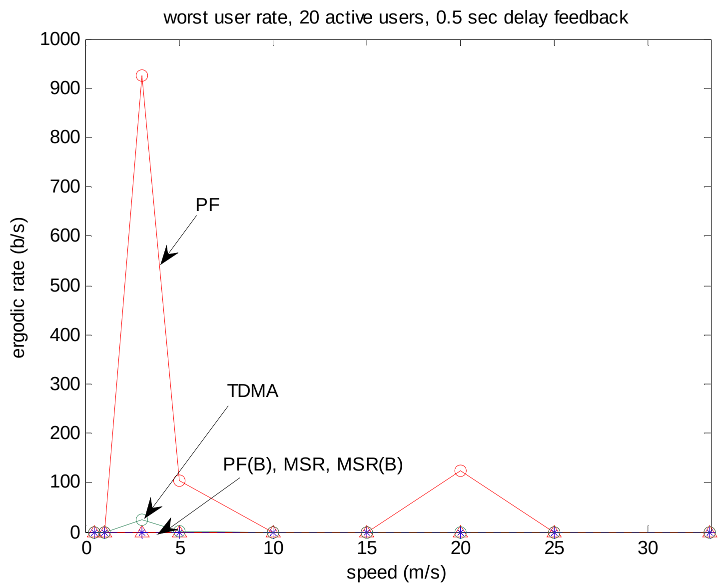


Figure I.68: Worst user rate performance versus mobility

I.7.2 End-to-end assessment - Conclusions

Simulations were performed under a satellite channel model for a WiMAX system in order to quantify the performance of channel-aware scheduling algorithms under ideal channel information conditions (no mobility) and under outdated channel information (significant mobility). The results for the non mobile case strongly suggest that employing sophisticated channel aware schedulers can significantly enhance system's performance in terms of both aggregate and individual user throughput. The cost is that channel information is fed back to the scheduler which presents a small overhead in static/slow mobility conditions.

When mobility is significant, performance of scheduling algorithms degrades as expected, since the channel information becomes outdated, and there is a finite probability of outage. The simple scheme of rate back off was examined as a means to decrease the probability of outage events and increase the successfully transmitted rate. Simulations indicate that sophisticated channel aware schedulers can provide still gains compared to non-channel aware scheduling. Various algorithms are possible that allow for trade-off between maximum aggregate rate and fairness. However, in the case of large mobility the overhead of channel information feedback may be significant, which, under certain circumstances will make the employment of channel-aware scheduling not attractive.

Scheduling and resource allocation for the satellite component is obviously a hard task as compared to the terrestrial case. Nonetheless, various scheduling algorithms were examined for OFDMA downlink signalling: max-SNR, proportional fair, max-sum-rate, max-fair. The trade-off between sum rate maximization and fairness is discussed, including the effect of user mobility and outdated channel state information. It is shown that for the feedback delay considered in the satellite scenario, the performance of channel-aware schedulers significantly degrades, even for relatively slow mobility (which would pose no serious problem in a terrestrial signalling scenario). For acceptable operation, the assigned rate is reduced compared to the nominal rate suggested by the channel estimate at the scheduler in order to reduce the outage probability and increase the successfully transmitted rate (goodput). However, the advantage of channel-aware scheduling in this case can be moderate to low, compared to a channel agnostic scheduling, such as TDMA.

Annex J: Review of Cognitive Radios

J.1 Formal Definitions and Characteristics of Cognitive Radio

A formal definition of cognitive radio has been given by Haykin in [i.174]:

> *"Cognitive Radio is an* ***intelligent*** *wireless communication system that is* ***aware*** *of its surrounding environment (i.e. outside world), and uses the methodology of understanding-by-building to* ***learn*** *from the environment and* ***adapt*** *its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g. transmit power, carrier frequency, and modulation strategy) in real-time, with two primary objectives in mind:* ***highly reliable*** *communication whenever and wherever needed; and* ***efficient*** *utilization of the radio spectrum"*

In addition to the adaptive facility of the hardware given in Haykin's definition, implying short term changes in transmission parameters, a CR also possesses reconfigurability, which is provided by the software defined radio platform upon it is built. Reconfigurability endows a CR the following features [i.175]:

- Adaptation of the radio interface so as to accommodate variations in the development of new interface standards.
- Incorporation of new applications and services as they emerge.
- Incorporation of updates in software technology.
- Exploitation of flexible heterogeneous services provided by radio networks.

Some of the cognitive capabilities of CR (and CN), as taken from the definitions of Mitola and Haykin are:

- ***Implementation of the cognitive cycle:*** This states that cognitive behaviour includes the ability to learn, change behaviour, and assess the effectiveness of the chosen action (as well as the actions of others). The cycle stages defined by Mitola are "*orient*", "*plan*", "*decide*", "*act*", "*observe*" and "*learn*". The cognitive cycle is shown in figure J.1. In the cognitive cycle, a radio receives information about its operating environment (Outside World) through direct observation or signalling. This information is then evaluated (Orient) to determine its importance. Based on its valuation, the radio determines its alternatives (Plan) and chooses an alternative (Decide) in a way that presumably would improve the evaluation. Assuming a change was deemed necessary, the radio then implements the alternative (Act) by adjusting its resources and performing the appropriate signalling. These changes are then reflected in the interference profile presented by the cognitive radio in the Outside World. Throughout the process, the radio is using these observations and decisions to improve the operation of the radio (Learn), perhaps by creating new modelling states, generating new alternatives, or creating new valuations [i.176]. Although Mitola's definitions are intended for a cognitive radio terminal, they could also be applied to a Cognitive Network.

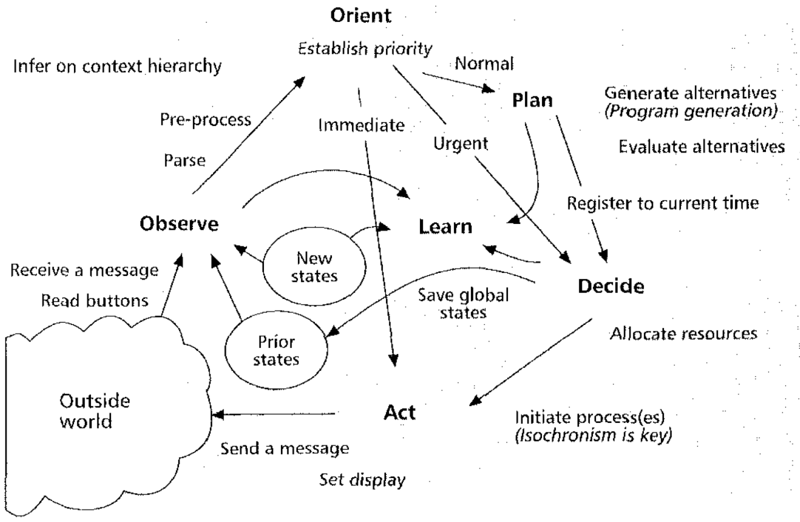


Figure J.1: The Cognitive Cycle

- ***Interaction between terminals in an intelligent manner:*** This gives rise to topics of study such as the degree of "competitiveness" and "cooperation" between cognitive terminals.
- ***Memory of past behaviour:*** This is in order to plan optimally for future requirements (this planning may be applied to cognitive networks, as well as cognitive radios). An example of this would be a terminal remembering that a user always commutes between 8:00 am and 8.30 am weekdays via a predictable route and thus will have access to the same set of wireless networks.

Some examples of enabling characteristics of CRs are:

- ***Reconfigurable and adaptable hardware:*** This will permit selection of air interfaces and/or frequency on a short or long term basis. A short term selection may be made for the length of a particular connection, in order to take advantage of a particular spectrum opportunity, whereas a long term selection could be appropriate in the case where a user terminal is taken to regulatory domain where only use of a single air interface is permitted.
- ***Ability to sense the local environment:*** For example the propagation channel and interference in order to assist the choice of the best method and time of transmission. The application for this sensing capability could be to measure interference, detect other transmissions already using a channel, detect vacant spectrum bands or to determine the characteristics of the propagation channel. These enabling features are also known as Radio Scene Analysis and Channel Identification (this also includes estimation of the channel state information CSI and prediction of channel capacity for use by the transmitter) [i.174].
- ***Smart antennas:*** For adaptable network coverage, or to establish point to point links between nodes in the wireless network.
- ***Positioning capability:*** In order to provide information about the terminal's own location. This will allow access to location based services and will facilitate the operation of cognitive networks (e.g. to maintain a database of networks available in a particular location and to assist learning of previously successful channel parameters in a given location). Location information can also be used to define the regulatory domain including information about transmission power limits, permitted physical layer waveforms and unavailable channels.

Additional capabilities of the infrastructure when applied at a network level include:

- The ability intelligently to allocate resources (such as frequency channels, multiple access codes or timeslots) to cells on a dynamic basis.
- The use of network architecture such as those involving relaying, or formation of ad hoc networks.
- Sharing of cognitive information between nodes, such as advertisement of the availability of free spectrum and negotiation of transmission parameters.

The concept of CR can be naturally extended beyond the cognitive process within the user terminal in order to encompass other aspects of network functionality; these include the processes that allocate network resources on long term basis, and those that control access to the channel on a connection by connection basis. The extension of cognitive techniques from the radio terminal domain (as envisioned by Mitola and Haykin) to the channel access and networking domains has given rise to the need for a new area of study: *Cognitive Networks (CN)*.

J.2 Research Challenges in Cognitive Radios and Networks

The current state of physical and upper layer technologies and techniques allows developing some first basic CR-based systems for dynamic spectrum sharing (see example B) in clause 6.2.2). However, developing highly dynamic and flexible CR-based systems that are able to exploit spectrum opportunities over wide spectral regions and over diverse system architectures, requires the development of new efficient solutions to a wide range of system design challenges.

Some of the research topics that are currently being addressed in scientific articles and paper are:

- Achieving reliable sensing of spectrum opportunities over wide spectral regions and with manageable computational complexity.
- Optimum usage of past spectrum measurements for improving the reliability of spectrum exploitation decisions.
- Development of new waveforms to best exploit spectrum availability patterns (e.g. OFDM waveforms with variable sub-carrier spacings and symbol durations).
- Efficient implementations of signal processing techniques for interference avoidance and cancellation, e.g. beamforming, multi-user detection, precoding for interference cancellation, dirty paper coding.
- Allocation of spectrum, either on an ad hoc basis or by a controlling authority, and achieving fairness in resource allocation for ad hoc implementations.
- Channel access and selection of transmission parameters on a co-operative basis between terminals.
- Implementation of reconfigurable techniques in hardware.
- Optimum network architectures.
- Use of location information for assisting cognitive algorithms.
- Information sharing between network nodes.
- How cognitive algorithms can be implemented and what benefits can be offered by computer science techniques such as machine learning, game theory and nature inspired algorithms.
- Implementing regulatory constraints on the CR nodes.
- Stability of the cognitive algorithms: Because of the recursive operation of the cognitive cycle, concerns are introduced as to under what conditions the recursions settle down to a steady state and what is this steady state.
- Limiting bandwidth requirements of signalling associated with information sharing between radio nodes.

J.3 Literature Review of Research Topics in Cognitive Radios and Networks

J.3.1 Interference Sensing and Identification

Interference sensing is the most fundamental task that needs to be carried out by CRs, within the framework of decentralized opportunistic spectrum access; as part of the spectrum opportunity identification process. The spectrum detection process will be repeated periodically, and the frequency sweep time should be short enough to be able to rapidly identify free spectrum. Long detection duration will lead to inaccurate spectrum estimation, which can cause interference to both the primary (licensed) and secondary (unlicensed) systems.

Several general tradeoffs should be considered in the design of spectrum sensing algorithms:

- Finding the optimal point between the fast frequency sweep time and an appropriate detection resolution bandwidth.
- Achieving short detection process sweep intervals and high sensing "signal processing gains" with low detection hardware complexity.
- In cooperative sensing scenarios, the sensing cooperation gain needs to be balanced against the number of cooperative users (and in particular the associated scheduling complexities and signalling overheads), their distance spreads, and energy consumptions.

As discussed in [i.181] and [i.182], the technical challenges of spectrum sensing are:

> *The demands of highly reliable detection probability with low false alarm/miss detection probabilities in regards of stringent requirement on RF front-end sensitivity for up to multi-gigahertz wide bandwidth and improving sensing processing gain through robust digital signal processing technologies from single terminal/user viewpoint. From multiple users' viewpoint, power control in cognitive radios, their cooperation sensing and fairness issue in cooperation are also the valuable topics for future research.*

In the remaining of this clause single/cooperation spectrum sensing techniques as these are related to RF front-end issues, robust signal processing, cognitive radios power control, cooperation among cognitive radios, and fairness issue in cooperation, are reviewed.

J.3.2 RF front-end design issues

The RF front-end design issues that need to be addressed for allowing the practical application of cognitive radios coincide to a very large extend to the issues in designing Software Defined Radios (SDR).

One important issue is the high dynamic range of received signals and resolution levels the receiver should be able to cope with (between diverse levels of signal strength, noise and interference). This means that a high number of bits (e.g. 12 or more bit) are required for quantization levels in analog-to-digital (A/D) converter within stringent timing constraints. Wideband sensing also leads to the demand of high sampling rate (up to multi-GHz) in A/D converter and high power consumption.

According to [i.210], SDRs, in order to be able to cope with worst-case combinations of signal conditions (in terms of dynamic range, SNR, power imbalance between desired and interfering signals, etc.) the ADC converter should be able to cope with a dynamic range in excess of 100dB and speed of the order of several hundred mega-samples per second. ADCs that achieve such performances and also consume reasonable power levels (suitable for handheld device applications) do not exist. Presently available 12 bit converters can handle a dynamic range in the order of 60 dB to 70 dB, and given that resolution improvements have been rather slow in the past, then a sustained long-term research effort is required in order to achieve the predicted performance requirements.

In the short term reconfigurability can be achieved by designing converters that target the most useful applications and usage scenarios. Research into dynamic-range reduction methods can also provide solutions. Two dynamic range reduction methods that limit the strong in-band primary user signals (which are unnecessary when sensing weak signals in noise), in order to reduce the bits in the A/D converter, have been proposed in [i.182] and [i.207]. One technique uses active cancellation processing, realized by a tuneable notch filter, a linear prediction filter and reconstruction D/A converter. The drawback of this approach is its high hardware complexity and it can also lead in distortions. The second technique is based on multiple antennas processing by using beamforming (e.g. coefficients-controlled phased antenna arrays) to select or suppress unnecessary signals for further detection. Several novel sampling methods have also been proposed recently in [i.208] and [i.209].

A different research issue on the RF-level design of SDRs, according to [i.210] is the design of direct (single-stage) receiver architectures for the down-conversion of the RF signal to baseband. Present signal down-conversion architectures are mainly based on Heterodyne architectures that consist of two or more stages (RF-IF, and IF-Baseband). For SDR receivers there is a high interest towards direct conversion architectures (RF-Baseband), mainly because such approach solves the image signal problem efficiently. Homodyne architectures also reduce the overall receiver complexity since image rejection and channel selection filters would no longer be required. Research challenges in realizing Homodyne receivers are the maintenance of the I/Q balance, achieving low second order distortion and DC offset cancellation.

Superheterodyne receiver architectures require only a single (RF-level) analogue-domain down-conversion, with the remaining IF to baseband down-conversion stages being implemented digitally (DSP). Such hybrid down-conversion architectures represent the best choice for a SDR receiver in the short-medium term, given the difficulty of implementing direct conversion architectures with the available set of technologies. The main challenge in realizing Superheterodyne architectures is maintaining high linearity in the receiver hardware in order to allow performing channelisation in DSP.

An approach that attempts to combine the advantages of Homodyne and Superheterodyne architectures is the Low IF architecture. It gives the advantage of simplifying the image rejection problem but also avoids the DC offset problem of the Homodyne architecture (since the LO signal is not at the same frequency with the wanted signal).

J.3.3 Signal Processing Techniques for Spectrum Sensing

Energy-detection is regarded as the simplest but least optimal spectrum sensing technique. The detection can be done without a priori knowledge of detected signals (such as pilots, synchronization channels, periodical features in modulations or frequency hopping sequence). Two architectures of current spectrum analyzers can be implemented in energy-detection scheme:

a) The Scan-Based Super-Heterodyne Energy Detection Architecture:

As shown in figure J.1 after rejecting the image signal through a band-pass filter, a combination of voltage ramp, voltage controlled oscillator and mixer transfer the frequency-separated signal back to baseband. A fixed sharp channel selection filter is then utilized to suppress adjacent channels, followed by a power detector to obtain the power spectrum.

The processing is done sequentially channel by channel by sweeping the tuning voltage, which results in several limitations: Firstly, it requires a long sweep time (t_{sweep}) to fulfil one spectrum sweep, and this delay is proportional to Frequency Span(f_{span}) /Resolution Bandwidth($B_{res.}$):

$$t_{sweep} = k \frac{f_{span}}{(B_{res.})^2} \tag{J-1}$$

where k is filter proportionality factor with typical values between 2 and 3 [i.183]. This means a good $B_{res.}$ is obtained at the price of long sweep time. Secondly, because the detector works serially, it is possible that after finishing the entire spectrum sweeping, the status of the initial part of spectrum has changed, which will result in wrong spectrum allocation decision. Thirdly, when facing non-stationary signals, the inter-bin time taken to do power measurement processing might increase miss detection probability on bursty signals.

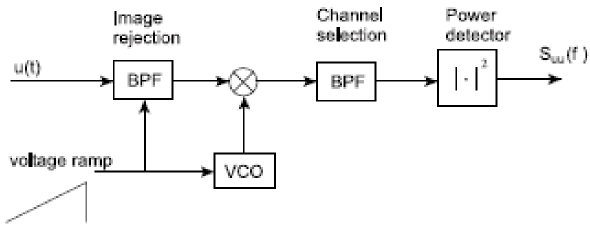


Figure J.1: Scan-Based Super-Heterodyne Energy-Detection Architecture

b) FFT-based Sensing Energy Detection Architecture:

Figure J.2 illustrates the FFT-based sensing architecture. In this architecture, the fixed voltage VCO is followed by a wider-band selection band-pass filter. ADC and FFT blocks are utilized, which make parallel processing possible. This means entire spectrum sweep time is equal to the time required to obtain a single channel's status and all the channels' results are obtained at the same time. Processing gain is proportional to FFT size *N*, which improves frequency resolution [i.181]. However, the drawback of this architecture is also on ADC/FFT parts, which require a higher dynamic range when sensing multi-channels resulting in much more power consumption and increase in hardware complexity, compared to scan-based architecture. The combination of these two architectures may be adopted.

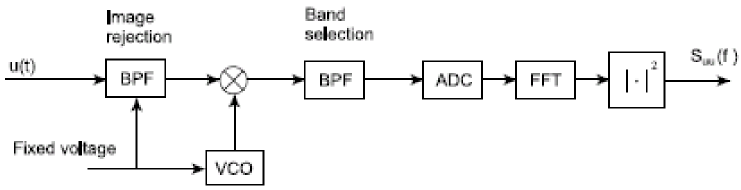


Figure J.2: FFT-based Energy-Detection Architecture

Due to non-coherent processing, the complexity (in terms of number of samples required to achieve a target probability of missed detection/false alarm) of the energy detection depends on the SNR according to: $O(SNR^{-2})$ i.184. Energy detection also suffers from poor performance at low SNR (e.g. when detecting spread spectrum signals), and due to not having any prior knowledge of the signals to be detected, the technique does not allow discriminating between modulated signals, noise and interference, and thus fails to benefit from adaptive signal processing methods [i.181]. Another limitation of the energy detection based methods, occurs because of the uncertainty of the exact noise power level, which can be variable with time. This uncertainty introduces a minimum SNR threshold (referred to as the "SNR wall"). Figure J.3 illustrates the SNR wall in energy-based detectors as a function of the noise power uncertainty.

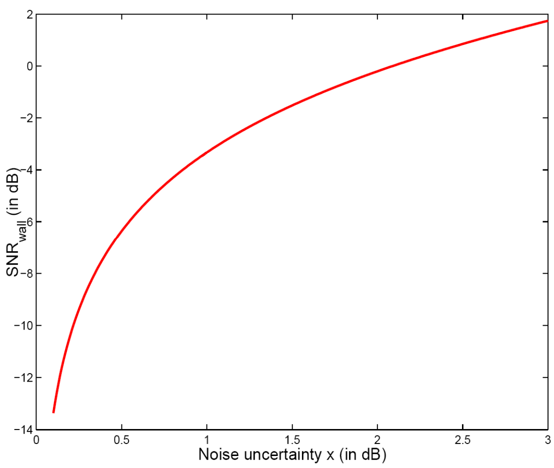


Figure J.3: SNR Wall in Energy Detection Methods as Function of Noisy Power Uncertainty

Another difficulty in energy-based detectors is setting accurately the threshold for detecting channel occupancy. This difficulty arises due to the very dynamic interference variation conditions especially in mobility environments. An adaptive reduced-rank method for setting the occupancy threshold dynamically according to the varying interference levels is proposed in [i.185]. In this method, noise floor is firstly estimated through adaptive noise level estimation, using an adaptive subspace algorithm. Based on the assumption that incident signals and noise are uncorrelated, this subspace algorithm calculates the noise floor for each sub-band in the spectrum. When the noise floor is obtained, an adaptive "Constant False Alarm Rate" algorithm, based on the number of non-coherent integrations at the fast time scale to match the noise probability density function, is applied to compute the occupancy threshold iteratively. Then, the channel entropy and predictability as critical criteria are introduced to decide the most stable spare frequency channel for cognitive radio access.

Improved spectrum estimation performance can be achieved through matched-filtering techniques, which make use of pilots or training sequences transmitted by the primary system. The sample complexity of matched-filtering techniques depends on the received SNR according to $O(SNR^{-1})$, which in practice translates to much reduced sweep times, relative to energy detection techniques. Matched filtering techniques also achieve improved estimation performances at low SNR, since their estimation accuracy improves consistently with the number of training/pilot symbols. Moreover the noise power uncertainty does not introduce any practical limitations, since the SNR wall also reduces proportionally with the estimation dwell time. In practical systems the performance of matched filtering detection techniques is limited by the maximum length of the training/pilot symbols and the associated estimation complexity and latency.

The increased interest in spectrum estimation in the context of CR systems (i.e. with requirements for high estimation accuracy and small sweep times at low SNR conditions) has stimulated research interest in novel approaches (alternative to the classical techniques reviewed above). In particular wavelet-based estimation techniques have been recently reported in [i.188] and [i.189].

J.3.4 Power control

Cognitive radios should not only be able to make binary decisions about the status of channels (e.g. "black", "white", "grey") but they should also be able to control their transmission power according to the network topology and propagation environment. In particular the transmit power level depends on several factors, including the distance between secondary cognitive users and primary users, different propagation fading/decay situations and the power aggregation effects from multiple secondary users.

Cooperation between users can provide a good solution to deal with multiple secondary users' power limitation, as well as in different propagation fading/shadowing situations. An analysis is presented in [i.218], where the primary system's SNR can be regarded as an appropriate metric for calculating the safety distance. In this analysis, secondary cognitive users' power control scheduling are discussed, based on the impact of the interference caused by single-pair cognitive user's transmission power (which is limited in a certain value or can be adjusted according to the distance away from primary users), through to considering the impact from a sea of secondary cognitive users.

J.3.5 Cooperation sensing

The requirements of high detection and low false alarm probabilities have already pushed the single node sensing techniques close to the estimation limits that can be practically realised. Single node sensing can perform unsatisfactorily mainly due to time-varying shadowing and multipath fading effects. Recently, cooperation sensing, where secondary users collaborate in spectrum sensing, by sharing their sensing information has been proposed as a collaborative approach for improving spectrum detection performance [i.186] and [i.187].

The degree of cooperation can depends on factors such as the complexity of sensors, the available signalling control channel bandwidth and distributed sensing strategies. The scope of cooperation also depends on the type of the primary system. For instance, if secondary nodes operate in a broadcasting primary system environment, then cooperation between nodes in the very near vicinity will yield limited benefits. On the other hand the cooperation benefits can be significant in a wireless LAN/PAN due to the short effective communication ranges.

It is reasonable to assume that secondary cognitive radios experience independent multipath fading. However, log-normal shadowing can exhibit high correlation depending on the distance spread of the secondary nodes, and also on the type of the physical environment. Another design issue is the number of users in cooperation. As it is illustrated in [i.190], the sensitivity threshold will asymptotically rise to a limit when increasing the number of cooperative users under a certain distance spread assumptions. This means that is not always beneficial to poll many users into sensing cooperation, since estimation performance saturates while the bandwidth and power overheads increase.

From the above it becomes clear that efficient cooperative sensing is a complex and dynamic design problem, in which the number of cooperative users needs to be chosen according to detection performance requirements, but also taking into account the users' distance spreads, their energy consumptions, scheduling complexity and signalling overheads.

J.3.6 Spectrum Access Control in Cognitive Radios/Networks

Spectrum access is defined as the mechanism by which access to a radio resource is controlled to meet a user's service requirement (for the duration of that requirement). The longer term allocation of radio resources to networks, cells or layers is considered as being "resource allocation" and is separate design issue. In a conventional network, a Medium Access Control (MAC) protocol is required to allocate radio resources (which could include a frequency, timeslot or multiple access codes) to a user. The environment in which such a system operates (propagation, interference from the same and different systems, available resources) and the operating parameters of the radio terminals themselves are assume to be known and predictable.

SAC is a complex mechanism that takes into account various levels of the system such as the radio segment, the network segment and the equipment segment (physical layer capabilities) as well as external factors such as the economic requirements and the regulatory requirements. The SAC entity also involves a close interaction between elements across the layer stack, hence the need to properly understand the impact of cross layer operations in order to characterise this entity.

In CR systems the SAC will be functioning with an expanded set of informational inputs and action outputs; as it shown in figure J.4. In summary, the SAC algorithm acts in response to a need for a terminal to transmit, and an allocation of appropriate radio resources is made in the form of instructions to the relevant radio terminals. These terminals may include not only the original requestor, but also the destination and any intermediate terminals (in the case of networks incorporating relaying). The allocation of resources considers any hardware and signal processing abilities (e.g. interference cancelling) of the terminals involved.

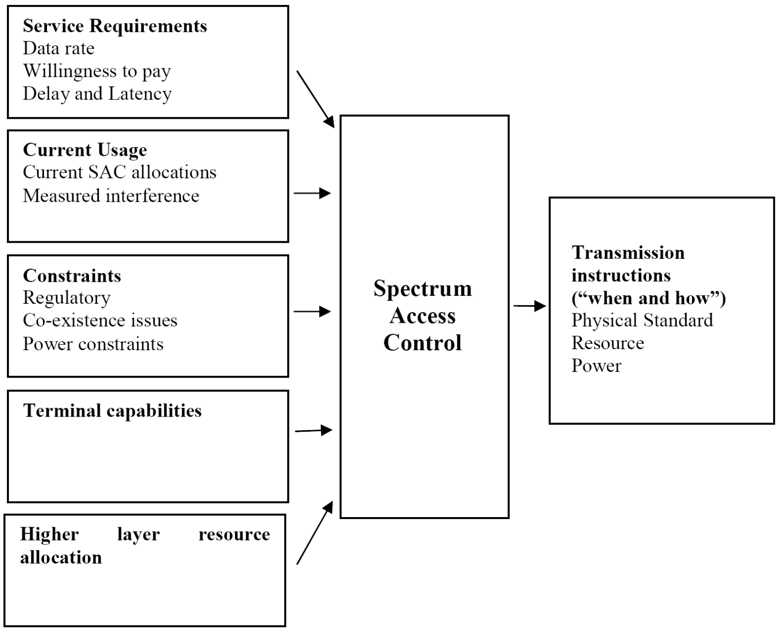


Figure J.4: SAC Operation in Terms of Input Parameters and Action Decisions

Specific research questions relating to the design of the SAC include:

- How the SAC should be implemented; via central control or through a distributed mechanism?
- How the spectrum occupancy information should be shared between the network nodes (if at all).
- Should the SAC allow multiple co-existing transmissions; exploiting simultaneous channel use that will be available with CDMA, OFDMA, etc?
- What time frame is required for spectrum sharing; on demand or longer term?

- How should the SAC algorithm achieve fairness?
- Is there a requirement for on-going monitoring during a transmission? In certain cases (particularly primary/secondary sharing), this may be necessary in case a primary user appears during transmission.
- How can feedback be incorporated to improve the effectiveness of the selected cognitive algorithm?
- What does a node do when it "boots-up" and knows nothing about the environment?
- How will hidden nodes affect the behaviour of the system?

The rest of this clause provides a literature review on the progress on some of the above topics.

J.3.7 Specific implementations of Spectrum Access Algorithms

One of the most comprehensive implementations of cognitive spectrum access is considered as part of the DARPA XG programme. As this is a military project, there is not a significant amount of information available, but presentations available in the public domain state that the project is about spectrum sensing and organising access based on available resources and policies. Some of the work is about codifying the policy requirements (these include available frequencies, transmit power allowed and transmit times) into an XML language. Press releases also suggest that a new waveform has been developed for CR.

Reference [i.211] presents an opportunistic MAC algorithm for use with 802.11 ad hoc networks. The algorithm described is named Opportunistic Packet Scheduling and Auto Rate (OSAR), a modified version of the CSMA protocol which uses channel probing with Request To Send (RTS) packets to determine the optimum time and rate of packet transfer (including selection of the appropriate 802.11 physical layer standard). The channel probing is designed to build up a list of SNRs (and thus transmission quality) for the neighbouring nodes. This is most applicable to ad hoc networks, as it is assumed that each node has a pool of nearby neighbours. The analysis is used for an access point which sends data via one or more hops (although this is not clear in the paper) to a receiving node. Transmission control is via Distribution Coordination Function (DCF). One good feature about this work is the fact the operation of the algorithm has been assessed through simulation. The technique of using the RTS/Clear To Send (CTS) probing is most of interest for this work.

Another implementation of a shared spectrum system is given is [i.212], which describes a HiperLAN/2 system which is able to occupy unwanted channels of a GPRS network. The OFDM subcarrier spacing of 200 kHz means that each subcarrier could occupy a single vacant GPRS channel. The spectrum access procedure described explains how the sharing nodes schedule transmission based on packet arrivals generating resource requests. MAC protocols are also used to send sensed spectrum availability info between themselves (the sensing algorithm itself is not described). The decision to schedule packet transmission is made by the HiperLAN/2 access point. An event driven simulator has been implemented.

One practical implementation of a practical of a cognitive radio test bed is given is in [i.213]. This comprises a lot of FPGA boards programmed as front ends, inter-connected via fast data connections, with sufficient length to allow their distribution over a wide area. The aim is to implement various sensing algorithms in order to test their effectiveness, although MAC/PHY interference avoidance strategies could presumably also be investigated. The primary system for implementation is 802.11b/g, incorporating a number of fixed systems to model primary users. The paper describes the effectiveness of sensing primary users.

J.3.8 Cognitive Radio Enabling ideas

Part of the CR literature deals with the idea of a meta-language to exchange information. This concept is important for cognitive systems due to the increased amount of information to exchange for facilitating cognitive operation. The meta-language provides a way of both sharing and categorising the information. Much of the work presented in [i.218] is concerned with the development of a meta-language (RKRL) designed to define and share the vast array of information needed for cognitive operation (e.g. radio capabilities, user behaviour, propagation conditions). Such a language finds application in many aspects of CR (e.g. exchanging information concerning hardware capabilities or physical layer standard compatibility). One considered application is that of the ability to codify spectrum access rules by use of a meta-language. This (amongst many others) is presented in [i.214].

Specific meta-languages that have been proposed include XML and DAML (DARPA Mark-up Language). The Software Communications Architecture Reference Implementation (SCARI) protocol can also be used to address interoperability issues. Channel probing as suggested in [i.211] (and others), probing packets are a way of obtaining information about channel conditions to a number of neighbouring nodes.

The IEEE 802.11k task group has the remit of performing radio resource measurements to enhance the capabilities of 802.11 networks. Measurements include [i.348]:

- The channel load report, the fractional duration over which the channel is sensed to be free.
- The clear channel assessment, the instantaneous measurement which detects whether a channel is free.
- The noise histogram, the detection of non-802.11 signals in the in-use spectrum.
- The medium sensing time histogram.

One technique that has been considered for use in 802.11k (although it appears that this has not been adopted by the 802.11k working group) is the medium sensing time histogram [i.215]. This permits an 802.11 station to measures the busy/idle traffic activity from other 802.11 stations and derive a statistical analysis to be used for channel access. This therefore fulfils the need to have a distributed measurement of traffic, rather than interference sensing being considered elsewhere in the project. Reference [i.215] presents a simulation of the implementation of this technique, although it does not actually show how the results could be put to use in the operation of a network.

A way of distributing information about existing channel activity is given in [i.216]. The central topic of this paper is spectrum sharing for OFDM systems; this is achieved by a implementing a secondary system that can take advantage of non-contiguous spectrum availability by inserting sub-carriers in the available spectrum bands. Spectrum availability is determined by use of a technique called the boosting protocol. This procedure involves the secondary terminals detecting which primary system spectrum are free and then, at a pre-allocated time, transmitting a complex OFDM symbol at maximum power on the free sub carriers. This avoids the necessity for the spectrum availability information to be coded and transmitted by the MAC layer (in contrast to the system described in [i.212]). A pre-requisite for success is that the subcarriers of the secondary system remain silent for the duration of the sensing period.

Annex K: Review of Specific Interference Mitigation Approaches

K.1 Multi-User Detection for CDMA Co-existing Systems

Conventional MUD techniques (reviewed in the interim deliverable) can be applied for CDMA based systems that share the same spectrum resource, provided some degree of co-operation is allowed (i.e. sharing knowledge of the user spreading signature sequences). However, even if such knowledge sharing is not provided, blind MMSE MUD can be applied, which only requires knowledge of the signature of the user to be demodulated [i.192]. Improved detection performance in this non-cooperation framework can also be achieved through a "group-blind" MUD approach, where available (inter-system) knowledge is exploited to the maximum degree, and only intra-system interference is mitigated blindly [i.193]. Spectral co-existence of non-cooperating CDMA systems can also be achieved using low-complexity Parallel-Interference Cancellation (PIC) structures, when the two systems operate at different transmission rates [i.194]. In this case, the additional degree of freedom introduced in the receiver design, due to the out-of-rate transmission of the interfering system, is exploited for performing blind MUD.

In the spectral co-existence of CDMA systems, it should be expected that total number of users will be exceeding the spreading factor (processing gain) of the user spreading signatures. Therefore intra-system users may end-up using highly correlated spreading signatures, and this generally leads to poor interference suppression performance. Near optimal interference suppression, under such overloaded system conditions, can be achieved by Turbo-MUD techniques, as these have been reviewed in the interim deliverable.

A different promising approach, under such challenging conditions, has been recently developed by the University of Surrey [i.196], where the spreading sequences are designed according to the low-density principle, as this applies in LDPC codes. More specifically, user signatures are designed to have only very few non-zero chips, and each non-zero chip position is allowed to overlap (in time) with very few other chips form other user signatures. In this way, every received chip sample contains interference only from a very small group of interfering users, and each user spreads its signals over very small number of chips. This structure allows to model the association between users and chips by a Tanner graph (as with LDPC codes), and choose the connections between nodes using the parity check matrices of practical LDPC codes. This construction also allows the use of a Belief-Propagation (BP) algorithm in order to estimate APP probabilities for each user bit. Simulation results show that under BP decoding single-user performance can be achieved even when the system is uncoded and overloaded by a factor of 3.

K.2 Linear Precoding in MIMO Systems

Adaptive beam-forming techniques (see interim deliverable) allow filtering interference across the angle of arrival dimension and are thus considered a key technology for enabling dynamic spectrum sharing models. In practice, multiple antennas are not usually employed in the terminal side due to the cost and power consumption. Therefore it is of higher practical significance to consider interference suppression in the downlink of a Multiple Input Single Output (MISO) (i.e. where the multiple antennas are only used in the base station side).

The design and optimization of MISO precoding techniques depends on the assumption of co-operation (i.e. information sharing) between the co-existing systems. In practice the assumption of cooperation between systems does not typically apply, and therefore MISO precoding techniques should be designed accordingly. However co-operative scenarios can be relevant in some types of systems (e.g. co-operative satellite and terrestrial systems operating in the same band), and therefore a related framework of MISO precoding techniques should be considered.

In the non-cooperative setting, the Maximum Ratio Transmission (MRT) method [i.197] can maximize the received signal-to-noise ratio, but it does nothing to limit the interference it is causing to the co-existing systems. The Zero-Forcing (ZF) method, which is an extension of MIMO-MUD techniques [i.198], can perfectly mitigate the interference to other radio systems, but it may degrade the power of desired signals and lose some of the diversity gain of the channel.

Figure K.1 illustrates the block diagram of two linearly precoded (non-cooperative) co-existing MISO systems (A and B). System A is the primary system and system B is the secondary system. The base stations of system A and system B have N_A and N_B antennas respectively. In the transmitters, the messages of system A and system B, S_A and S_B , are multiplied by the vectors $\mathbf{v}_A$ and $\mathbf{v}_B$ respectively, then transmitted over the frequency non-selective radio channels. In the receivers, the received signals are multiplied by the complex scalars g_A and g_B respectively. Assuming the average signal attenuation between the base station and receiver is unity for both systems, then the cross-interference "attenuation factors" $0 \le r_{AB}, r_{BA} \le 1$ model the average attenuations of interfering signals. When $r_{BA} = 0$ then a spectrum overlay model is implied, but when both $r_{AB}, r_{BA} > 0$, then an open spectrum sharing model is implied. The receivers estimate their corresponding data channel and interference channel information, and according to different applications, feedback this channel information back to the transmitters.

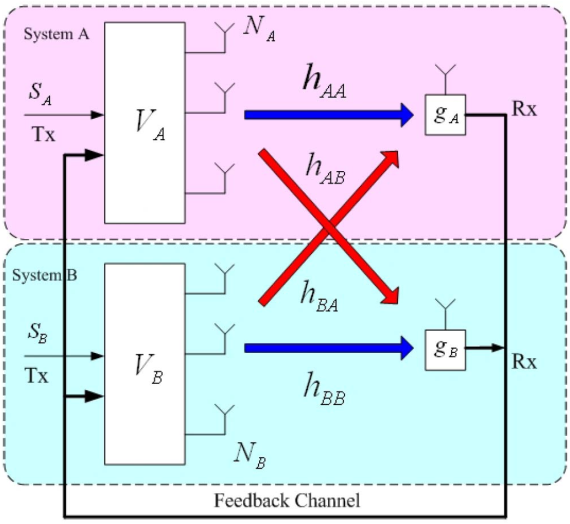


Figure K.1: Block diagram of MISO Linear Precoded Co-existing Systems

The received signals y_A and y_B for the n^{th} sample time can be written as:

$$\begin{aligned} y_A(n) &= \mathbf{h}_{AA}(n)\mathbf{v}_A(n)s_A(n) + r_{BA}\mathbf{h}_{BA}(n)\mathbf{v}_B(n)s_B(n) + z_A(n) \\ y_B(n) &= \underbrace{\mathbf{h}_{BB}(n)\mathbf{v}_B(n)s_B(n)}_{I} + \underbrace{r_{AB}\mathbf{h}_{AB}(n)\mathbf{v}_A(n)s_A(n)}_{II} + \underbrace{z_B(n)}_{III} \end{aligned} \tag{K-1}$$

where part I is the desired signal power, part II is the Co-Channel Interference (CCI), and part III is the additive Gaussian white noise. The scalars $z_A(n)$ n and $z_B(n)$ are the i.i.d. complex Gaussian random variables with zero means and variances σ_A^2 and σ_B^2, respectively. The vectors $\mathbf{h}_{AA}, \mathbf{h}_{BB}, \mathbf{h}_{BA}, \mathbf{h}_{AB}$ are the $1\times N_A$ and $1\times N_B$ row channel vectors. The total transmitter power constraint for system A and system B are denoted P_A and P_B, and the constraints for precoding vectors are $\|\mathbf{v}_A\|^2 = \|\mathbf{v}_B\|^2 = 1$.

The received Signal-to-Interference-Noise Ratio (SINR) for system A (and correspondingly for system B) is given as:

$$SNIR_A = \frac{P_A(\mathbf{h}_{AA}\mathbf{v}_A)^H\mathbf{h}_{AA}\mathbf{v}_A}{r_{BA}^2 P_B(\mathbf{h}_{BA}\mathbf{v}_B)^H\mathbf{h}_{BA}\mathbf{v}_B + \sigma_A^2} \tag{K-2}$$

and the normalized MSE is defined as:

$$MSE_A = \frac{E\left(\left\| s_A - g_A y_A \right\|^2\right)}{P_A} \tag{K-3}$$

The MRT method maximizes the received desired signal's power subject to limited transmitter power [i.197]. For system B, the solution of this approach is:

$$\hat{\mathbf{v}}_B = \frac{(\mathbf{h}_{BB})^H}{\left\|\mathbf{h}_{BB}\right\|} \tag{K-4}$$

This method also provides the maximum SNR and minimum MMSE solutions for system B, but without considering the system A. If we fix the system A's chosen precoding vector, the MRT solution can achieve best performance for system B and maximum diversity gain in a Rayleigh fading channel is also obtained. In addition, only the data channel information for system B is needed. The main drawback of MRT in coexisting environments is that it may greatly degrade the performance of system A due to the interference coming from system B.

The ZF [i.198] method can perfectly cancel the interference to system A. The key idea of this method is to find a vector that is orthogonal to the interference channel. If the system B interference self-correlation matrix is defined as $\mathbf{F}_B = (\mathbf{h}_{BA})^H \mathbf{h}_{BA}$, it can be seen that $rank(\mathbf{F}_B) = 1$ and the only non-zero Eigen value takes the value $\left\|\mathbf{h}_{BA}\right\|^2$. The eigenvectors corresponding to the zero Eigen values of the interference self-correlation matrix form the solution of the ZF method. The data channel information for system B is not needed for this approach. If there are more than two antennas in base station B, the solution of to the ZF algorithm is not unique. If an arbitrary choice of the ZF vector is used, some diversity gain may be lost due to not exploiting the data channel information.

The MRT achieves the best performance for the secondary system, but will degrade the performance of primary system. In a practical system, this may not be acceptable. So the optimal precoding vector will need to be derived under the constraint that no interference is caused to the primary system. The optimization problem under such constraints can be formulated as:

$$\hat{\mathbf{v}}_B = \arg\min_{\mathbf{v}_B}\{MSE_B(\mathbf{v}_B)\}$$
$$\text{s.t. } \left\|\mathbf{h}_{BA}\mathbf{v}_B\right\|^2 = 0 \text{ and } (\mathbf{v}_B)^H \mathbf{v}_B = 1 \tag{K-5}$$

In the case the two systems function co-operatively joints optimal precoding method can be derived. The co-operation assumptions require the two systems (i.e. their base stations) to be aware not only of the information of its own data and interference channel, but also the other's system channel information as well. In non-cooperative precoding schemes, the increase of one system's performance is likely to lead to the decrease of the other system performance. The key point of the joint optimal method is to find the best trade-off for the two systems and to maximise a joint performance criterion. This criterion may be the maximal sum capacity, minimal sum mean square error, etc. So the generic joint optimization precoding problem can be posed as:

$$\{\hat{\mathbf{v}}_A, \hat{\mathbf{v}}_B\} = \arg_{\{\mathbf{v}_A, \mathbf{v}_B\}} \{\textit{optimal}\ \text{criterion objective function}\}$$
$$\text{s.t. } \left\|\mathbf{v}_A\right\|^2 = \left\|\mathbf{v}_B\right\|^2 = 1 \tag{K-6}$$

K.3 Dirty Paper Coding Techniques for Co-operative Systems

In co-operative systems, which can have access to each other information, the DPC [i.195] based techniques provide improved capacities relative to classical resource sharing techniques (e.g. TDM), and also have the advantage that can be applied in SISO systems as well.

In the original work on DPC ([i.195]) it was proved that when the transmitter has non-causal access to the interfering signal over an AWGN channel, then through a suitable coding technique (DPC), the capacity of the interference channel is the same as of the interference-free AWGN channel: $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$, where P is the average transmit signal power and N the AWGN noise power.

Figure K.2 illustrates the "Dirty Paper" channel, which consists of a transmitter wishing to transmit a message *w* through an AWGN with additional additive interference which is known non-causally to the transmitter (i.e. the interference time sequence is known even before it occurs in the channel). In the original DPC paper, the interfering signal was assumed to be zero mean Gaussian, but the "no capacity-degradation" result has been generalized for any distribution of the interfering signal in [i.199].

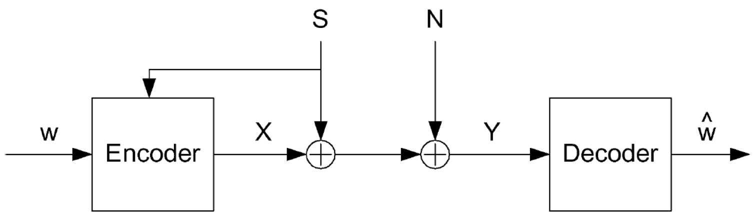


Figure K.2: The Dirty Paper Channel

The general idea in DPC is to infinitely repeat the constellation and adopt a message-based binning approach, where channel codewords corresponding to the same message are grouped into a bin, and within each bin, the codeword chosen according to the state of the interfering signal. In other words, the codeword to be transmitted should depend on the interfering signal state. This is analogous to adapting to the "dirt" when writing on dirty paper. The generic DPC is based on unstructured code designs and is therefore not practical for real system implementations. Some practical DPC techniques, which are based on nested lattices and the theory of coset codes, have been proposed [i.199] to [i.201], and more recently DPC technique based on superposition coding between shaping and channel codes has been reported in [i.202].

DPC has recently attracted much attention because it achieves the capacities of MIMO broadcast channels (see [i.203] and references therein). Practical DPC designs for this type of system have been proposed in [i.203], and in [i.204] for the multi-user MIMO broadcast channel case. These techniques require that the transmitter has available the channel state information of sub-channels in the MIMO channel. The impact on the achievable capacities of DPC with imperfect channel estimation at the receiver and no channel knowledge at the transmitter are analyzed in [i.205].

The perfect interference removal property of DPC gives it the potential to play an important role in spectrum sharing models, where different systems operate co-operatively. One such example could be the spectrum overlay (the downlink transmission of) a Wireless MAN network over a hybrid satellite-terrestrial S-DMB system. If the W-MAN system has prior information of the broadcasting transmission, which is a practically feasible assumption, then according to [i.191], it can act in two extreme ways, according:

- Selfishly, i.e. treat the S-DMB signal as interference to the W-MAN receiver and apply DPC in order to cancel this interference (in practice this approach would probably require the W-MAN receiver to feedback the channel state information of the interference channel to the W-MAN base station). This would however cause interference to the S-DMB receivers.

- Selflessly, i.e. if local S-DMB receivers are detected (in some manner) then recognize the priority of the S-DMB system and enhance its capacity by acting as a relay. This approach however would nullify the capacity of the W-MAN system.

In [i.206] an achievable region capacity region for the two systems was demonstrated, that smoothly interpolates between these two extreme configurations.

The resulting achievable region in the presence of additive white Gaussian noise is plotted as the "cognitive channel region" in figure K.3. The figure shows four capacity regions the co-operative systems can jointly achieve: The time-sharing region (1) displays the result of pure time sharing of the wireless channel. These points would be amenable to the proposals on secondary spectrum licensing. The interference channel region (2) corresponds to the best known achievable region of the classical information theoretic interference channel. In this region, both senders encode independently, and there is no message knowledge by either transmitter. The cognitive channel region (3) is the achievable region proposed [i.203] (secondary system has non-causal knowledge of primary system's transmission), which uses a coding scheme that combines interference mitigation with relaying the primary system's message. The modified MIMO bound region (4) is an outer bound on the capacity of this channel: the 2 × 2 multiple-input multiple-output (MIMO) Gaussian broadcast channel capacity region.

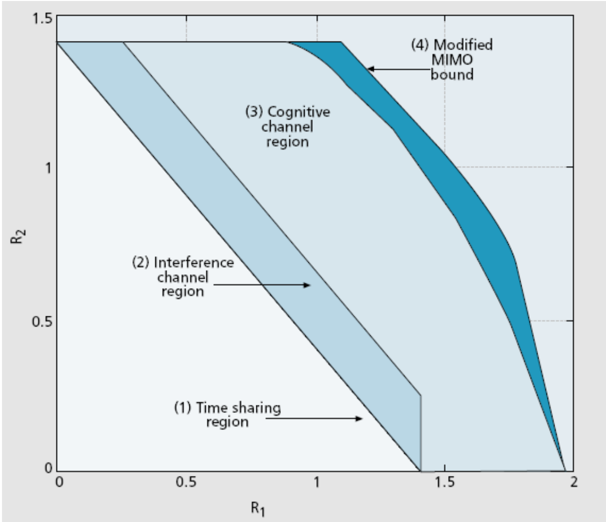


Figure K.3: Rate regions (R1, R2) for different two-sender, two-receiver wireless channels

Annex L: Cooperation through Relaying and Distributed MIMO Techniques

L.1 Introduction

Relaying is the earliest proposed type of cooperation between radio nodes and in its simplest form consists of point-to-point multi-hopes in order to combat signal attenuation in long-range communication; particularly in types of environments where the channel attenuation exponent takes big values (in free-space propagation the distance attenuation exponent is 2 but this can get up to 6 in shadowed urban environments.). A distinguishing feature of multi-hopping is that each node in the chain communicates only with the one before and the one after in the chain, or nodes that are one "hop" away. In a wireless environment, it may be possible for a node to receive or transmit its signal to other nodes that are several hops away, but such capability is ignored in multi-hopping, making it a simple and extremely popular (but suboptimal) mode of user-cooperation.

Despite its conceptual and implementation simplicity, multi-hopping does not yield any diversity benefits in fading channels. The investigation into alternative (and more complex) co-operation strategies dates back to the late 60's and early 70's, where bounds for the capacity of the "general" single-relay channel (depicted in figure L.1), were derived [i.220]. As seen in figure L.1 the relay channel is the three-terminal communication channel with the terminals labelled as: the Source (**S**), the Relay (**R**), and the Destination (**D**). All information originates at **S**, and travels to **D**. The relay aids in communicating information from **S** to D without actually being an information source or sink. The signal being transmitted from the source is labelled *X*. The signal received by the relay is *V*. The transmitted signal from the relay is *W*, and the received signal at the destination is *Y*.

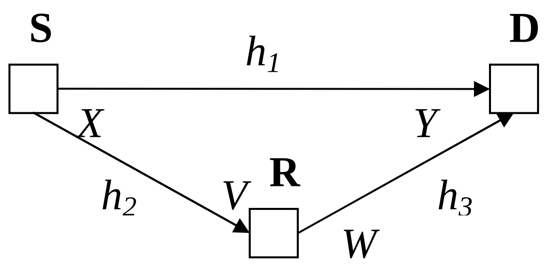


Figure L-1: General Single Relay Channel

Conceptually, information is relayed in two phases or modes: first, when **S** transmits and (**R, D**) receive, commonly called the broadcast (BC) mode; and second when (**S, R**) transmit and **D** receive, also known as the multiple-access (MAC) mode. This differentiation is only conceptual since it is possible for communication in both modes to take place simultaneously. Four different models of relaying that can be derived based on the above two modes.

1) **S → (R, D) ; (S, R) → D** (Most general form of relaying);
2) **S → R ; (S, R) → D** (**D** ignores signal from S in first mode);
3) **S → (R, D) ; R → D** (**S** does not transmit in second mode);
4) **S → R ; R → D** (Multi-hop communication).

The 2nd and 3rd models have been proposed mainly as simplifications of the 1st most general model in order to achieve analytical tractability in capacity derivations.

A relay is said to be half-duplex when it cannot simultaneously transmit and receive in the same band. In other words, the transmission and reception channels are orthogonal. Orthogonally between transmitted and received signals can be in time-domain, in frequency domain, or using any set of signals that are orthogonal over the time-frequency plane. If a relay tries to transmit and receive simultaneously in the same band, then the transmitted signal interferes with the received signal. In theory, it is possible for the relay to cancel out interference due to the transmitted signal (since it knows perfectly the transmitted signal). In practice, however, any error in interference cancellation (due to inaccurate knowledge of device characteristics or due to the effects of quantization and finite-precision processing) can be catastrophic because the transmitted signal is typically 100 dB to 150 dB stronger than the received signal as noted in [i.221]. Due to the difficulty of accurate interference cancellation, full-duplex radios are not commonly used. However, advances in analog processing could potentially enable full-duplex relaying.

Although early literature on information theoretic relaying was based almost entirely on full-duplex relaying (e.g. [i.220]), in recent years a lot of research, and especially research directed towards practical protocols, has been based on the premise of half-duplex relaying (see [i.221] to [i.223]).

L.2 Relay Protocols

So far there has not been proposed a single relay transmission protocol (for the general relay channel), that performs best under all conditions (as these classified with respect to the relative quality (SNR) of the three links in the single relay channel). In practice the selection of relay strategy also depends on implementation-complexity constraints. There are three main types of relay protocols, as these are summarized below.

a) Amplify and Forward

Amplify-and-Forward (AF) relays do not need to decode and regenerate the source node's data. Instead, they simply amplify what they receive subject to their power constraint and forward it to the destination node. Besides requiring less complexity, AF can also be advantageous in terms of BER with respect to other more complex techniques, especially when the **S**-**R** link is not reliable enough, and any attempt of the relay to decode the signal will introduce many errors.

The simplest practical AF algorithm divides transmissions into two blocks of equal duration; one block for the source transmission and one block for the relay transmission (alternative AF protocols have been considered in [i.224]). This type of two-phase slot allocation requires the relay to be able to synchronise to the used time-slot structure. This means that the AF processing should be done at baseband, and the received baseband signal should be buffered to be forwarded in the next allocated time slot.

In the single antenna case, processing is straight forward; an AF gain calculation block first measures the received signal power and then adjusts the amplifier gain in order to boost the signal to the maximum transmit power level. Although no channel estimation needs to be performed at the relay, the destination node should be able to estimate the response of the equivalent concatenated **S**-**R**-**D** channel. This estimation can be done using the pilot symbols transmitted by the source.

The AF technique becomes more involved when the relay is quipped with more than a single antenna. One possible approach is to first perform receive antenna diversity combining, followed by amplification, and then apply space-time transmit processing. This type of AF requires the relay to estimate the **S**-**R** channel responses, and is therefore significantly more complex than in the single antenna case. With respect to space-time processing it is noted that although the complex baseband signals do not belong to a finite alphabet, still popular techniques such as ST orthogonal linear block codes and space multiplexing can be applied.

When more than a single relay is available, the relays can all transmit at the same time in order to retain efficient usage of the time resource. However simultaneously transmitting relays will interfere and they should thus properly format their transmit signals to allow reliable and efficient detection by the destination node. In practice this can be achieved though a distributed Space-Time coding, since although the non-regenerated signals will not be the same, they will have significantly high correlation if obtained by reliable **S**-**R** links. Each relay can participate to form an overall $M \times T$ ST code, where M is the number of relaying nodes and T is the time-size of the code.

b) ***Decode and Forward***

Decode and Forward (DF) involves decoding of the source transmission at the relay. The relay then retransmits the decoded signal after possibly compressing or adding redundancy. The DF protocol is close to optimal when the source-relay channel is excellent, which practically happens when the source and relay are physically near to each other. When the source-relay channel becomes perfect, the relay channel becomes a 2×1 multiple-antenna system.

Considering the single relay case, the simplest multiplexing approach again is to divide transmissions into two blocks of equal duration; one block for the source transmission and one block for the relay transmission. More elaborate DF algorithms are considered in [i.225] and [i.226]. The most basic form of DF relay cooperation is repetition coding, i.e. the relay simply repeats the regenerated source signal. Alternatively, the relay can encode the source message using a codeword that is correlated with, by not necessarily identical to, the source codeword. Within the context of the simple algorithms, this corresponds to a form of parallel channel coding. When multiple relays are involved, they can all employ repetition coding or a more general space-time code to transmit information jointly with the source to the destination. Repetition coding in separate blocks has the advantage of low complexity, but it requires more complex scheduling protocols to support it.

The destination node needs to receive and combine signals over the time phases (**S**-**D**, **R**-**D**) before it can proceed with decoding and detection. The combining stage depends on the transmission functions of the source and the relay. For example if only the FEC encoding is common between the source and relay nodes, then the destination node will need to carry out two different detection chains up to the FEC decoder stage, at which point the soft reliability information will need to be combined. It is also possible that the source and relay differ even by the type of FEC code they employ, in which case the joint decoding will need to employ some Turbo-type of detection between the two chains (i.e. the two non-identical channel decoders exchange soft reliability information in order to recover the common data bits).

The Turbo-type of detection/decoding between the two receive branches (one demodulating the **D**-node signal and the other the **R**-node signal), points the way to distributes Turbo coding; for example the **D**-node encodes the signal through a first branch recursive convolution encoder and the R-node uses an interleaver and a second recursive encoder, then the overall code can be jointly decodes as classical (serial) Turbo code.

When the relay is equipped with more than a single antenna it can benefit in both the reception of the signal as well in the transmission to the destination. All the traditional multi-antenna diversity reception and multi-antenna transmission coding techniques could be used in such case. In cases where the relay needs to communicate to more than a single destinations, the multi-antenna capability can be exploited through transmit beam-forming techniques. Depending on the availability of channel state information of the **R**-**D** link by the relay node, adaptive and non-adaptive schemes could be envisaged for the **R**-**D** link transmission. For example a typical adaptive multi-antenna transmission technique is Eigen beam-forming where several streams of data are transmitted over the eigen modes of the **R**-**D** channel. With respect to non-adaptive techniques, Space-Time Coding and Trellis Coding can be used.

c) ***Compress and Forward***

Compress and Forward (CF) [i.220](this strategy is also known as Estimate and Forward or Quantize and Forward) is important when the **S**-**R** and the **S**-**D** channels are comparable, and the **R**-**D** link is good. In this situation, the relay may not be able to de-code the source signal, but nonetheless it has an independent observation of the source signal that can aid in decoding at the destination. More specifically, the relay quantizes and compresses the received source signal, and the destination combines the compressed and source signals prior to decoding. Another advantage of the CF protocol is that, in contrast to DF, it always outperforms the non-cooperative two-node system configuration.

If the received signal (at the relay) belongs to a discrete alphabet, then it has been shown in [i.227] that the received signal can be compressed at certain compression rate without any distortion. On the other hand, for continuous alphabets it has been shown in [i.228], that the signal can be compressed at a rate *R(D)* with an average distortion $d \le D$.

Figure L.2 illustrates the architecture of a practical CF relay model. The source transmits the signal X_S to both the relay and the destination. The relay first quantizes the continuous noisy received signal and subsequently encodes the quantized signal into a binary sequence U. Compression is performed by extracting the syndrome of an LDPC code with a parity check matrix of dimensions $n \times m$, where $m < n$. The compression rate is thus m/n. Decompression at the destination is carried out in a sequential manner (i.e. quantization bit by bit of the binary word *U*) using a Message Passing decoder which accepts information both from the syndrome-word forwarded by the relay and the information binary word transmitted by the source.

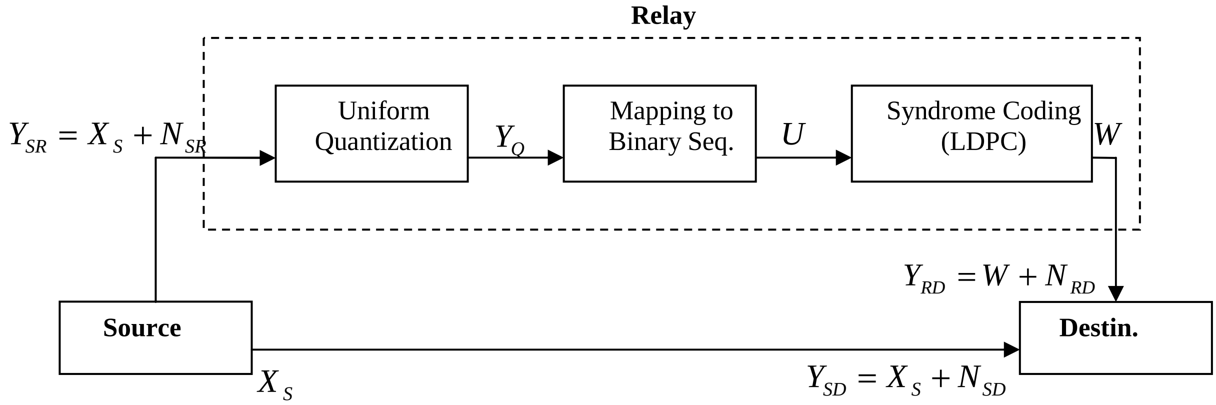


Figure L.2: Practical Compress and Forward Relay Model

L.2.1 Adaptive Relay Protocols

Similar to point-to-point links, availability of channel knowledge at all or some cooperating transmitting nodes will allow applying some type of adaptation in order to improve the performance of the cooperative transmission. The exact type of adaptation will depend on the type of channel information that is made available to the transmitting nodes, and the also on the exact nodes which can be assumed to have access to this information. Obviously the ideal case is when all nodes have full knowledge of all links, but in practice this assumption is difficult to satisfy and it would also incur large transmission overheads.

Another (more realistic) assumption that could be "affordable" for low-mobility TDM systems is that transmitting nodes only have feedback channel information of the link they transmit upon. Such information can allow adapting transmission parameters, such as block-size, FEC rate, power, modulation, sub-carrier allocation in OFDM systems, etc. A different type of adaptive techniques do not make use of exact channel state information but instead rely on acknowledgements of decoding failure/success fed-back by the destination (i.e. similar to ARQ/HARQ schemes).

This clause introduces some of the adaptive relay protocols that are being considered in the literature.

a) Selection and Dynamic Relaying

Fixed DF is limited by direct transmission between the source and relay. In a dynamic fading environment, where the channel parameters can be assumed to be quasi-static, the fading coefficient of the **S**-**R** link can be measured to high accuracy by the cooperating terminals, and this can allow them to adapt their transmission format according to the channel attenuation. More specifically, if the measured **S**-**R** link attenuation falls below a certain threshold, the source simply continues its transmission to the destination, in the form of repetition or more powerful coding. If on the other hand it lies above the threshold, the relay forwards what it received from the source, using either AF or DF, in an attempt to achieve diversity gain.

In a sense, selection relaying of this form offers diversity because, in either case, two of the fading coefficients are small in order for the information to be lost. Specifically, if the coefficient of the **S**-**R** is small, then **S**-**D** is also to be weak for the information to be lost when the source continues its transmission. Similarly, if **S**-**R** is strong, then both **S**-**D** and **R**-**D** are be weak for the information to be lost when the relay employs AF or DF.

A further improvement is dynamic DF [i.225] and [i.226]. In dynamic DF, the relay starts by receiving from the source and does not begin transmitting until it is sure it has correctly received the source transmission. Because of quasi-static conditions, the reception time at the relay can be modelled as a random variable, and the coding scheme takes this into account.

b) Incremental Relaying

Fixed and selection relaying can make inefficient use of the degrees of freedom of the channel, especially for high rates, because the relays repeat all the time. Incremental relaying protocols can exploit limited feedback from the destination terminal, e.g. a single bit indicating the success or failure of the direct transmission. These incremental relaying protocols can be viewed as extensions of incremental redundancy, or HARQ, to the relay context

In ARQ, the source retransmits if the destination provides a negative acknowledgment via feedback. In incremental relaying, the relay retransmits in an attempt to exploit spatial diversity. As an example, consider the following protocol utilizing feedback and AF transmission. First, the source transmits its information to the destination. The destination indicates success or failure by broadcasting a single bit of feedback to the source and relay, which we assume is detected reliably by at least the relay. If the source-destination SNR is sufficiently high, the feedback indicates success of the direct transmission, and the relay does nothing. If the source-destination SNR is not sufficiently high for successful direct transmission, the feedback requests that the relay amplifies and forwards what it received from the source. In the latter case, the destination tries to combine the two transmissions. Protocols of this form make more efficient use of the degrees of freedom of the channel, because they repeat rarely, and only when necessary.

L.3 Cooperation through Virtual Antenna Arrays

An extension of co-operative transmission through the classical relaying approach, is the formation of a Virtual Antenna Array (VAA) by a collection of neighbouring mobile or fixed terminals. The idea builds on the works of co-operation through radio-relaying and on the MIMO channel capacity results, and related coding techniques (introduced in the late 90's). Recent results have shown that popular space-time coding (e.g. space-time block coding) and diversity combing techniques can be adapted for VAA systems and yield similar capacity and diversity benefits as in standard MIMO systems. A series of theoretical performance results and practical system architectures based on the UMTS (W-CDMA) standard, have been studied by Dohler et. al. in [i.234] to [i.239]. The proposed techniques are therefore directly relevant and can potentially offer significant performance improvements in S-UMTS systems.

Similar to relay-based cooperative techniques, VAAs can be applied within a diverse range of systems and usage scenarios; including satellite and hybrid system architectures.

For example within cellular systems, base stations consisting of several antenna elements can apply space-time coding and forward the encoded data stream to the associated mobile terminals, which can form several independent VAA groups. Each mobile terminal within a group receives the entire data stream, extracts its own information and concurrently relays further information to the other mobile terminals. It then receives more of its own information from the surrounding mobile terminals and, finally, processes the entire data stream. The wired links within a traditional receiving antenna array are thus replaced by wireless links.

The same principle is applicable to the uplink, where a synchronised space-time encoded data stream is emitted from the VAA group. In this situation, the VAA accomplishes a special type of network which bridges cellular and ad-hoc concepts to establish a heterogeneous network with increased capacity. It calls for intelligent synchronisation, relaying and data scheduling algorithms, the exact realisation of which depends on the access scheme, choice of main link technology, choice of relaying technology, technological limits, number of antennas within a given geographical area and other factors, e.g. the ability of the cellular system to synchronise users, etc.

L.3.1 Survey of VAA Techniques

In [i.234] the original signal stream at the base station (with two transmitting antennas) is encoded using the Alamouti code, and then transmitted to the target mobile terminal and the relaying mobile terminal. The relay terminal retransmits the original information in a non-regenerative way to the target destination, and acts as a virtual antenna of the receiver, as shown in figure L.3. Comparing this distributed scheme with the classical 2X2 MIMO channel, the diversity gain (assuming fast fading only), though degraded due to the additional noise in the relaying mobile and double-Rayleigh channel conditions of the relaying channel, it is still near to 4. Nevertheless, as other relay-based architectures, the scheme can yield substantial diversity benefits against slow fading processes, since the paths between the BS and the two terminals can be highly uncorrelated.

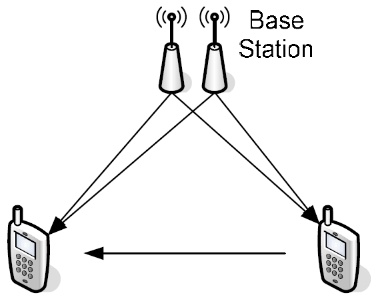


Figure L.3: Virtual 2- Antenna Reception for Achieving Receive Diversity

In [i.235] the authors propose a more detailed analysis of applying space-time block codes in VAA. The UMTS W-CDMA air interface is utilized to accomplish the orthogonal relaying. Due to the property of RAKE receivers, any paths differing by more than 260 ns/80 m will be distinguishable by the receiver. Thus, distributed relaying terminals that are spatially close to each other are considered to be in-chip range, and the relaying of these terminals performs different from the relay terminals which are out of range. Various in-chip and out-of-chip range configurations are studied in this paper. All the schemes are found to operate advantageously in comparison with the Alamouti scheme with one receive antenna, and in some cases (with a sufficient number of relaying terminals) the performance is better than for the Alamouti scheme with two receiving antennas.

In [i.236], the authors analyse the performance of VAA as a system, and compare its performance with UTMS W-CDMA without VAA applied. Numerical analysis and simulations show that the overall system performance is dramatically improved by the application of VAA. The link capacity for VAA is analysed in [i.237], which follows the "capacity behaviour" of traditional MIMO systems. The main results are summarized as follows:

- The capacity saturates very fast with the number of transmit antennas exceeding the number of receive antennas.
- The capacity exhibits a logarithmic increase with an increasing number of receive antenna elements.
- The capacity shows a linear behaviour for an equally increasing number of transmit and receive antennas. Thus, maximum capacity in form of diversity and coding gains can be achieved in downlink and uplink when the number of transmit and receive antennas are equal.

VAA is further studied in [i.238] under more general system assumptions. The various factors that may influence the detailed design of the VAA system (such as the access technology, and the choice of relaying technology) are taken into consideration. A general conclusion is that the VAA systems yield significant benefits as long as the number of relaying terminals (antennas) matches the number of transmitting antennas. Furthermore, link-level simulators prove an increase in link QoS and data rates with VAA, and system-level simulations show an improvement in performance of VAA system.

In [i.239], the same author evaluates the performance of space-time trellis codes (STTC) deployed together with a VAA, and a 10 dB gain over a system without VAA is shown. In [i.240], the authors evaluate the performance of a VAA system over traditional MIMO system. It is shown that when there are correlations in the receiver antennas in traditional MIMO system, VAA outperforms MIMO. A generalization of VAA into multiple stages is called VAA multistage system. Both the source and target terminals cooperate with their neighbouring terminals in order to form "source and target VAAs". The intermediate relaying nodes (which are assumed to be close to each other) form a relaying VAA, and there can be multiple relaying VAA groups which form a multi-hop communication system.

Another contribution on distributed MIMO techniques is the one reported in [i.241], where an extended Alamouti coding scheme is applied to a multi-hop distributed MIMO system. The paper studies the performance of such a system considering both frequency-flat fading and frequency-selective fading, and both non-regenerative and regenerative relaying scheme. The study can be viewed as an implementation of extended Alamouti code on to multistage VAA communication system, as shown in figure L.4. It is shown by simulations that the proposed scheme provides diversity gain and more improvement of frame error rate as the number of hops increases. The results of Signal to Interference Ratio (SIR) performance have shown that space-time coded cooperative relaying approach has potential to provide higher spatial frequency reuse.

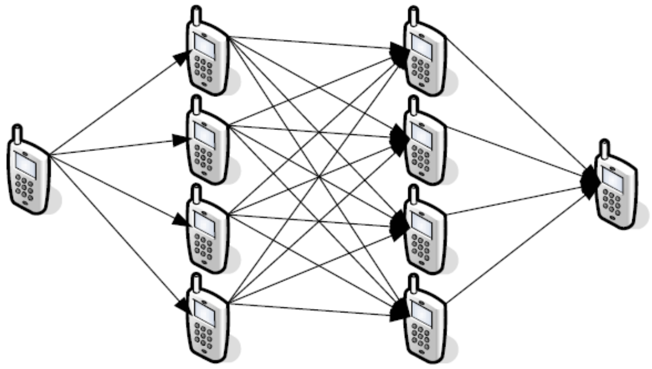


Figure L.4: Multi-hop Distributed MIMO Configuration

In [i.242], the authors discuss a one-hop distributed MIMO system. Detailed outage probability of both distributed MIMO with STBC and repetition encoding are analyzed and simulated. It is shown that distributed MIMO with STBC provides same diversity as repetition code but with better multiplexing gain. The authors also discuss issues related to space-time code design for these protocols, emphasizing codes that readily allow for appealing distributed versions.

L.4 Detailed Description of Some Practical Cooperation Techniques

This clause reviews several practical cooperative transmission schemes that have been proposed in order to realize the theoretical benefits of relay and VAA channels.

L.4.1 A cooperative Transmission Protocol for CDMA Systems

A CDMA based user-cooperation strategy has been proposed in [i.229] to [i.243], which was one of the first implementations of user-cooperation to have been proposed, and it was designed keeping in mind the realities of cellular communication.

The proposed scheme assumes that each user has a single spreading code, which is orthogonal to the spreading codes of all other users, and also that the coherence time of the channel equals L symbol periods, i.e. the channel does not change for L symbol periods. The technique is demonstrated for the simple case where $L = 3$.

If the sources were not cooperating they would transmit:

$$\begin{aligned} X_1(t) &= a_1 b_1^{(1)} c_1(t), a_1 b_1^{(2)} c_1(t), a_1 b_1^{(3)} c_1(t) \\ X_2(t) &= \underbrace{a_2 b_2^{(1)} c_2(t)}_{1}, \underbrace{a_2 b_2^{(2)} c_2(t)}_{2}, \underbrace{a_2 b_2^{(3)} c_2(t)}_{3} \end{aligned} \quad \text{(L-1)}$$

where $b_j^{(i)}$ is the i^{th} bit from user j, $c_j(t)$ is the spreading code used by user j, and $a_j = \sqrt{P_j / T_S}$, where P_j is the power used by user j and T_S is the symbol period. For fairness, any cooperative scheme developed in the same framework satisfies some basic criteria. The total number of spreading codes used by the two users remains the same as in the non-cooperative scheme, and the cooperative strategy should be of comparable complexity to the non-cooperative scheme. Under the proposed cooperative scheme, the users transmit:

$$\begin{aligned} X_1(t) &= a_{11}b_1^{(1)}c_1(t), a_{12}b_1^{(2)}c_1(t), a_{13}b_1^{(2)}c_1(t) + a_{14}\tilde{b}_2^{(2)}c_2(t) \\ X_2(t) &= \underbrace{a_{21}b_2^{(1)}c_2(t)}_{1}, \underbrace{a_{22}b_2^{(2)}c_2(t)}_{2}, \underbrace{a_{23}\tilde{b}_1^{(2)}c_1(t) + a_{24}b_2^{(2)}c_2(t)}_{3} \end{aligned} \quad \text{(L-2)}$$

where $\tilde{b}_j^{(i)}$ is the partner's estimate of user j's i^{th} bit. The parameters a_{ij} control the amount of power allocated to a user's own bits versus the bits of the partner; while maintaining an average power constraint of P_j for user j, over L symbol periods.

The way to interpret the above is as follows. In Period 1, each user sends data to the base station only. In period 2, users send data to each other in addition to the base station. After this, each user estimates its partner's data and constructs a cooperative signal that is sent to the destination in Period 3. This cooperative signal is a superposition of spreading codes of both users.

In order to generalize the above scheme to arbitrary number of symbol periods L, another parameter L_C needs to be defined. The two users cooperate for $2L_C$ periods, whereas the remaining $L - 2L_C$ periods are used for sending their own information. For example, in (3.10), $L = 3$ and $L_C = 1$, whereas in (3.9), $L = 3$ and $L_C = 0$. By changing the value of L_C over time, it is possible to achieve different points on the capacity region. The a_{ij} parameters are chosen to satisfy the power constraints:

$$\begin{aligned} \frac{1}{L}\left(L_n a_{11}^2 + L_C(a_{12}^2 + a_{13}^2 + a_{14}^2)\right) &= P1 \\ \frac{1}{L}\left(L_n a_{21}^2 + L_C(a_{22}^2 + a_{23}^2 + a_{24}^2)\right) &= P2 \end{aligned} \quad \text{(L-3)}$$

This cooperative scheme is depicted in figure L.5 for the case of $L = 6$, $L_C = 2$. The performance of the above scheme and the design of optimal receivers for this type of user-cooperation are discussed in [i.243].

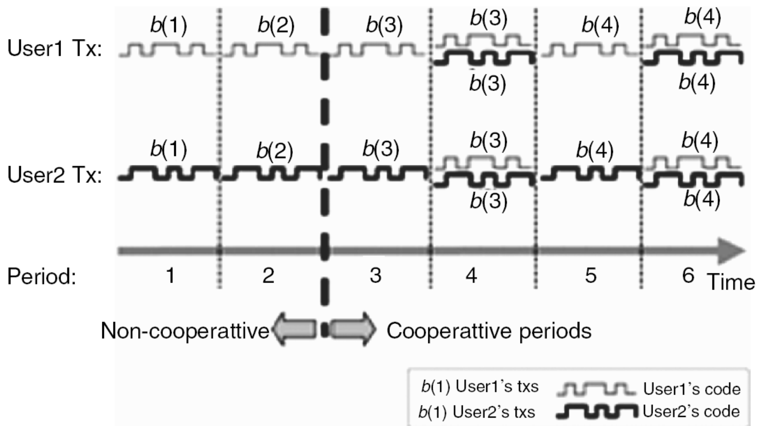


Figure L.5: Practical Cooperative Scheme for CDMA Systems Proposed in [i.229] to [i.243]

Clauses L.4.2 and L.4.3 present relay code designs using LDPC component codes that have been proposed for full-duplex and half-duplex relays, in [i.244] and [i.245], respectively.

L.4.2 LDPC Coding Scheme for Full Duplex Relaying

Reference [i.245] offered on of the first attempts on practical full-duplex relay code design, and although the proposed code designs are not optimal in an information-theoretic sense, they perform well.

Two specific protocols are proposed in [i.244]. The first is called the simple protocol, where transmission from the source occurs in B blocks of length N. A pair of consecutive blocks uses a pair of jointly designed constituent codes. Odd blocks use one of the constituent codes, and even blocks use the other. The source sends new information in each block. At the end of each block, the relay finds the codeword that is closest to its received signal, and retransmits it without re-encoding. The second protocol, which is called the DF protocol (inspired by the decode-and-forward scheme), and is somewhat similar to the simple protocol. Again, transmission from the source occurs in B blocks of length N. In each block, the source sends a superposition of a new codeword and a repetition of the previous codeword with an appropriate power ratio. In the first and last blocks, only one codeword is sent. At the end of each block, the relay decodes the new codeword from the received signal and retransmits it the same way as in the simple protocol. The constituent codes used in the above protocols are irregular LDPC codes proposed, chosen for their capacity-approaching performance.

The signal received by the destination in each block is a superposition of two codewords. This complicates the decoding process since optimal decoding at the destination needs to be based on the entire set of B blocks. It is extremely complex to find optimized LDPC code profiles for the entire factor graph since it requires joint optimization of B matrices. Therefore, as a practical alternative, only pairs of codes are optimized at a time.

Two algorithms were proposed for decoding the received signals at the destination, called the forward and the backward decoding algorithms. Note that the first and the last transmissions in the above coding scheme use only a single code, whereas any intermediate received signal is a superposition of a pair of codes. Therefore, decoding may either commence from the first or the last received codeword, corresponding to forward and backward decoding respectively. Forward decoding has a minimal latency of two blocks, and also performs better when the relay is near the destination. Backward decoding, in contrast, is better when source and relay are close to each other; however, it has a decoding latency of B blocks.

L.4.3 LDPC Coding Scheme for Half Duplex Relaying

LDPC code designs for the half-duplex relay channel were proposed in [i.245]. The code designs are based on the information theoretic random-coding scheme for half-duplex decode-and-forward relaying. Although the relay channel is commonly visualized as a combination of a broadcast and a multiple-access channel, it is shown that the achievable rate of DF relaying can be approached by using single-user codes decoded with single-user receivers. The single-user decidability of these codes supports the practicality of half-duplex relaying.

In i.245 it is shown that the gains of relaying are significant only in the low to medium SNR range. At high SNRs, the throughput of relaying is not a significant fraction larger than that of a direct link. Also in the low to medium SNR range, binary modulation on each channel dimension (QPSK) approaches the capacity of the AWGN channel. This justifies the use of binary codes. Another challenge in code construction is that the implementation of source-relay correlation in multiple-access mode introduces an added level of complexity. In contrast it is simple to devise coding schemes where this correlation is either 0 or 1. Empirical results in [i.245] show that the loss in throughput is negligible when the better of $\rho = 0,1$ is chosen instead of the optimal correlation.

When $\rho = 1$, **S** and **R** transmit identical signals in MAC mode. For this case, the following scheme is used. In the beginning of BC mode, **S** encodes information bits using a code $LDPC_{SR}^{BC}$ to yield a codeword of length tN bits. This codeword is transmitted by **S**. At the end of BC mode (which is also the beginning of MAC mode), both **R** and **D** receive the BC mode source signal. This signal is successfully decoded by **R**. However, **D** cannot decode the received signal, and stores a copy of it. In the beginning of MAC mode, the tN variable bits from BC mode are compressed. Compression is done at both **S** and **R**, by multiplying with the same parity matrix. These compressed bits, acting as parity together with the parity bits of $LDPC_{SR}^{BC}$ form a composite code $LDPC_{SD}^{BC}$ that can be decoded at **D** at the end of MAC mode. In order to communicate the compressed bits to **D** reliably, **S**, and **R** treat them as information bits for MAC mode, and re-encode them using a code $LDPC^{MAC}$ to yield a codeword of length $(1-t)N$, which is then transmitted synchronously from **S** and **R** with appropriate powers. The structure of the code is shown in figure L.6.

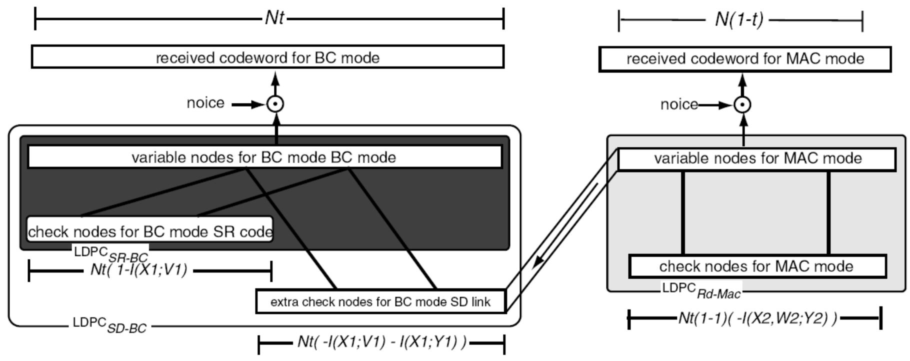


Figure L.6: LDPC Code Structure for $\rho = 1$

For $\rho = 1$, decoding is performed as follows. **R** decodes $LDPC_{SR}^{BC}$ at the end of BC mode using belief propagation like any single-user LDPC code. **D** waits for both BC and MAC mode signals to arrive before it commences decoding. $LDPC^{MAC}$ is decoded like a single-user LDPC code, from which side information in the form of additional parity bits is obtained about the BC mode signal. Using knowledge of the single-user BC mode source-relay code, and with the help of these additional parity bits, $LDPC_{SD}^{BC}$ is decoded. This final decoding also is performed using belief propagation.

For $\rho = 0$, the BC mode is the same as before. In MAC mode, however, **S** and **R** transmit independent (therefore uncorrelated) information using codes $LDPC_{SD}^{MAC}$ and $LDPC_{RD}^{MAC}$ respectively. As before, **R** compresses the information bits received in BC mode to produce additional parity bits, which serve as relay information bits in MAC mode. These bits are re-encoded by **R** using $LDPC_{RD}^{MAC}$ to yield $(1-t)N$ coded bits. The source, in MAC mode, sends bits of new information encoded using $LDPC_{SD}^{MAC}$ to yield another set of $(1-t)N$ coded bits. Thus, $(1-t)N$ coded bits each from **S** and **R** are transmitted simultaneously with appropriate power allocation, so that the two (uncorrelated) signals appear superimposed at **D**. The structure of the code is shown in figure L.7.

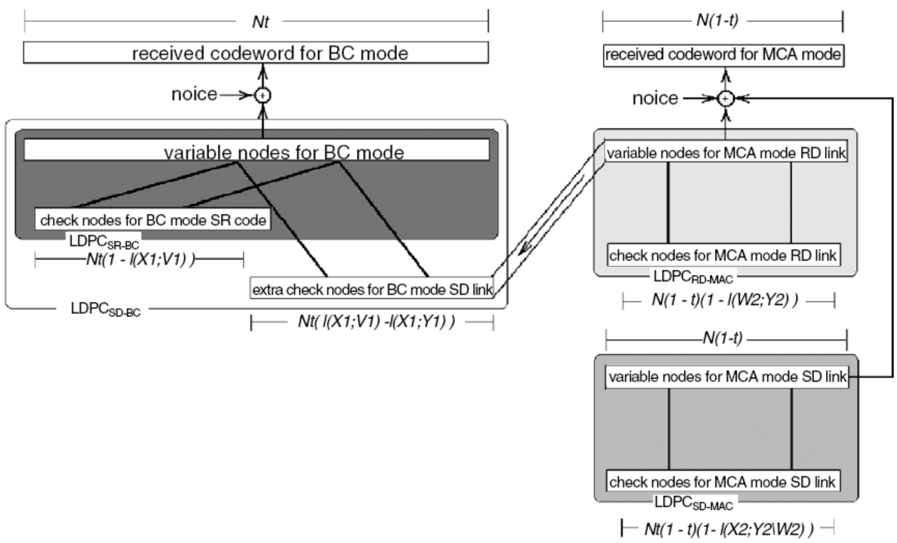


Figure L.7: LDPC Code Structure for $\rho = 0$

For $\rho = 0$, decoding proceeds as follows **R** decodes the BC mode signal like a single-user LDPC code. **D** waits for both BC and MAC mode signals. The MAC signal is successively decoded to first reveal the relay codeword, treating both noise and interference from **S** as noise. Next, the relay codeword is subtracted out to reveal the source codeword in the presence of noise alone, which is then decoded. The MAC mode source information is new information, whereas the relay information provides additional parity bits to aid in decoding the BC mode codeword.

The main challenge is the design of codes $LDPC_{SD}^{BC}$ and $LDPC_{RD}^{BC}$, which are to be jointly optimized, since the factor graph of the latter is a subgraph of the factor graph of the former. It is also note that these codes are of different rates, and although the received codeword is same for both **R** and **D**, the received SNRs are different. To avoid confusion, neither **S**, nor **R** actually uses $LDPC_{SD}^{BC}$ to encode information. It is merely a convenience to visualize the side information received by **D** in MAC mode as extra parity bits in addition to the actual parity bits of $LDPC_{RD}^{BC}$, and call the composite a code $LDPC_{SD}^{BC}$. The optimization of code profiles is performed using a modification of the density evolution algorithm. In the implementation of density evolution, the messages have been approximated as Gaussians to speed up the optimization, the cost being usually small inaccuracy in threshold determination.

L.4.4 Cooperative OFDM Architecture

A space-time cooperative system based on orthogonal frequency division multiplexing (OFDM), which is referred to as a COoperative (CO)-OFDM system, has been designed in [i.246]. This clause briefly outlines the main features of the CO-OFDM system and some performance results. More details can be found in [i.246].

Figure L.8 illustrates a block diagram of the CO-OFDM transmitter and receiver. The structure is similar to that of the IEEE 802.11a standard [i.350] except for the use of space-time cooperation. Transmit symbols are encoded according to a form of time-division cooperative diversity. The transmission of each frame involves two subsequent phases with fixed duration: the listening phase and the cooperation phase. In the listening phase, the source broadcasts a listening sub frame to the relays and destination. Space-time coding is not employed in this phase, since the source is equipped with only one transmit antenna. If the destination succeeds in decoding the listening sub frame, the following cooperation phase is ignored at the destination. Otherwise, the destination attempts to decode the succeeding cooperation sub frame.

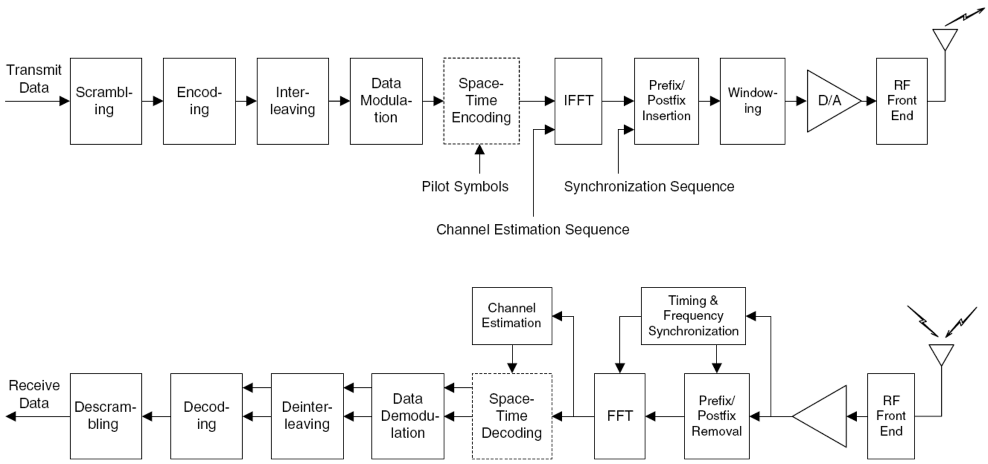


Figure L.8: Block Diagram of the CO-OFDM Architecture proposed in [i.246] (dotted blocks are used only in the cooperation phase)

The relays and destination can realize whether decoding of each sub frame is successful or not by computing the checksum of the frame check sequence. In the cooperation phase, the source constructs and transmits a cooperation sub frame, which corresponds to a portion of the space-time coded version of the listening sub frame. The behaviour of the relay depends on whether it has succeeded or not in decoding the preceding listening sub frame. If a relay has succeeded in decoding, the relay also constructs and transmits a cooperation sub frame, which corresponds to another portion of space-time coded signal.

Then the destination may receive the complete space-time coded signal from the source and relay, enabling the reliable decoding of the cooperation sub frame. Otherwise, if the relay has failed to decode the listening sub frame, it is silent in the cooperation phase. The listening and cooperation sub frames are allowed to be transmitted at different transmission rates. For the case of a single relay node, [i.246] has also devised a frame structure including preamble sequences, and provided simple and effective timing and frequency synchronization algorithms and a channel estimation algorithm.

Figure L.9 shows the overall FER performance of the CO-OFDM system, when the synchronization and channel estimation algorithms proposed in [i.246] are adopted. The performance of a Single-Antenna OFDM (SA-OFDM) system and a Double-Antenna OFDM (DA-OFDM) system without cooperation is also presented for comparison. The geometric gain G is assumed to be 10 dB. It is observed that the C-OFDM system achieves significant performance improvements over the SA-OFDM system. At a FER of 10^{-2} , for example, the energy gain of the CO-OFDM system over the SA-OFDM system is as much as 6,7 dB for channel A, and 2,5 dB for channel B, where the channel models are given in the ones defined be ETSI (see [i.246]). From the slopes of the FER curves, it can be seen that the CO-OFDM system achieves a diversity order comparable to that of the DA-OFDM system, as predicted by the theory.

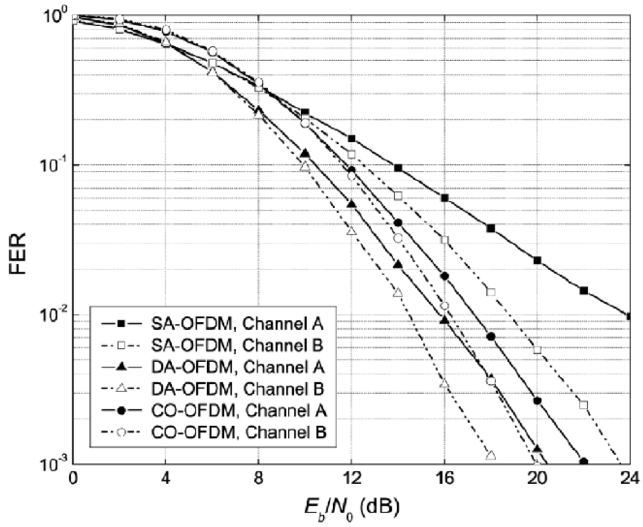


Figure L.9: The overall FER performance of the CO-OFDM system

L.5 Other Research Challenges in Realizing Cooperative Systems

L.5.1 Antenna Design Considerations

This clause discusses some antenna design consideration that need to be taken into account in scenarios where terminals are expected to simultaneously support a cellular/satellite links and the short range support links (it is assumed that these have close proximity operation - possibly with the same transceiver).

From the transceiver point of view, operation in neighbouring bands with very high power level differences (much more than 20 dB) is very problematic. Consequently, it would be beneficial if the antenna system itself could provide some sort, of duplexing action (possibly up to about 20 dB). This would reduce the stress on the receiving part of the terminal in terms of return link energy feedback (the problem is less severe if the major link and the short range link operate in a coordinated time-sharing fashion).

For access point terminals, spatial separation provides a solution to this problem. However, for more compact terminals where links are achieved through very closely spaced antennas (having different transceivers or capability of splitting the radio signal in two), or on the same antenna element (if we assume single transceiver operation), then the elements can help to provide some sort of separation between the ports. Here there aretwo possibilities for this dual port discrimination:

- Polarization: for example with dual port patch antennas.
- Mode excitation: by invoking different modes on different ports on the same physical patch.

Both solutions are realizable for "free space" terminals (as in note-book terminals), but become much more difficult for near field loaded terminals such as handhelds. With this in mind, new antenna designs are to be sought with respect to particular casing and handling.

Another very important issue that affects hand-helds at 1 800 MHz is a 7 dB to 10 dB drop in antenna efficiency with respect to free space, due to near field loading effects in normal handling [i.230]. At 5 GHz this effect might be worse. For the major link (cellular link), this loss appears at one end of the link, but it appears at both ends at links between mobile terminals. Thus, the short range links for this sort of terminals are particularly power handicapped, possibly loosing practical achievable gains (the power amplifier on the support links would need to be "cranked up").

The link budget threshold for the short range link needs to be large enough to absorb this. On top of the terminal antenna inefficiency, body blockage/shadowing can be very severe and abrupt, and it can completely dominate short range person to person links, when the users are using hand-helds. In this case even more link margin is required, further diminishing the gain potential of cooperation.

From an antenna point of view, more free space operated terminals such as note-book computers, may appear better suited for the first beneficial cooperative antenna deployments. From the diversity and capacity point of view, the more spread the antenna elements are within the environment, the higher the expected gain potential. Here there is intuitively a large potential for cooperative operation. Most personal terminals have compact antenna systems that might provide micro diversity against short term fading. When it comes to long term/shadowing diversity and beneficial capacity gain through spatial multiplexing, both access point and user terminal antenna systems need to be in each other's "near field" [i.231]. However, when exploiting multiple different user terminals as one large antenna system, we get wide spatial spread antenna system to provoke "near field" situations for shadowing diversity and capacity gain. This though requires similar average power on all links. The difficulty is how to coordinate and share the power in the overall system operation. Particular difficulties appear when the objective is to increase the capacity, because that requires instant CSI from the complete system to provide decomposed Eigen state information for all the links. Therefore it is very likely that heavy practical limitations will appear for such operation.

L.5.2 Routing Protocol Design Considerations

In multi-hop networks without central node (e.g. Base-Station, Access-Point (BS/AP)) the main issue is establishing the connectivity, i.e. finding a route from each source node to the corresponding destination node. The design of such routing protocols has been studied extensively within the context of ad-hoc networking.

In infrastructure-based networks, multi-hop communications are facilitated through the use or mobile relays. When fixed relays are used the routing problem becomes comparable to that in wired networks (which is an easier design problem to solve). Even when mobile relays are used, routing is still an easier task in comparison to infrastructure-less networks. This is firstly because the BS/AP can assist the mobile terminals in the routing process, and secondly because the BS/AP constitutes a common source or sink. Therefore the issue is such type of network is finding the best route (based on some criteria), rather than a route.

In [i.247] various routing algorithms are proposed for maximizing the network throughput in TDMA/TDD multi-hop networks which have central nodes to facilitate the scheduling (orthogonal resource partitioning among hops in a route; in a way that that no additional bandwidth is used due to relaying). It is demonstrated that if the routes are established by taking into account the potential gains due to adaptive modulation and coding, as well as diversity, significant increases in throughput can be achieved. In [i.248] routing is considered in a multi-hop network supported by infrastructure and communication relations limited to a few hops only. Multiple simultaneous routes become possible and this makes the choice of the routing algorithm important. An algorithm ensuring that no queue at a relay node explodes for the largest possible set of packet arrival rates is called throughput-optimal [i.248].

Routing becomes more challenging when considering mobile relays. In the MANET subgroup of IETF, several routing algorithms for mobile ad hoc networks have been investigated. Studies of these algorithms have shown a high routing overhead and low efficiency in network throughput. Based on this observation it has been proposed that routing in the multi-hop network be supported by an area wide cellular overlay network [i.249]. There a hybrid routing scheme called cellular based multi-hop routing has been studied where route requests are sent to the BS of the overlying cellular network. The central entity determines the route and responds with a packet comprising a series of mobile nodes willing to relay the data traffic between the source and the destination. The service and route discovery is performed by the overlay cellular network and the packet transmission in the micro-range multi-hop network. This approach exploits both the ability of the macro-network to communicate with all the nodes, and the throughput efficiency of multi-hop transmission in the micro-range layer.

Results have shown that this overlay-assisted routing approach leads to low packet drops due to wrong route information and adds little overhead to the network traffic [i.250]. Moreover it allows fast packet delivery because of quick route establishment and the routing overheads increases only linearly with the number of nodes, which allows scalability with the network size.

This very effective cooperation between mobile ad-hoc networks and wide coverage overlay networks points to the utilization of satellite networks for coordinating the routing processes.

L.5.3 Radio Resource Management Design Considerations

Radio resource management deals with the assignment of BS, channel, transmit power, etc. In view of this, the sensitivity of radio coverage to the selection of the relay, relay channel and relay power control are investigated in [i.250] for a cellular TDMA system where two-hop mobile relaying is employed whenever necessary. Whenever relaying is performed, an additional time or frequency channel is required for the second hop. In [i.250] an aggressive strategy that does not require any new channels for relaying is adopted: the relay channel is always selected from among the already used channels in the adjacent cells.

Various selection schemes for the relay and the relay channel, from random to smart selection, with and without power control, are considered in [i.250]. It is observed that with the proper selection of relay, relay channel and relay power, a significant enhancement in high data rate coverage can be attained through two-hop mobile relaying. The observed trends and corresponding conclusions are:

- Performance gains due to relaying increase as the number of wireless terminals in the system increases.
- Employing power control in both hops further enhances the performance, especially as the cells get smaller; the returns to power control become substantial for interference-limited cells.
- The maximum relay transmit power level is an important factor only in large cells. In small cells most of the benefits are gained with relatively small relay transmit power levels.
- The performance gains are quite sensitive to the relay selection criterion. If the relays are chosen randomly, the performance gets worse in comparison to the no-relaying case (this is analogous to the case where a user is connected to a wrong BS). Yet, highly suboptimal (i.e. with minimal intelligence) but feasible relay selection schemes (e.g. relay selection based solely on proximity through the use of the GPS data available at the BS) still yield significant coverage improvements.
- Once a good relay is selected the performance gains become fairly insensitive to the relay channel selection criterion. Therefore, in systems with limited resources for monitoring and control purposes, the priority should be given to proper relay selection, rather than proper relay channel selection.

It is also worth noting that a relay's energy consumption will increase linearly regardless of the multiple access scheme uses, as more and more terminals' signals are relayed. The increased energy consumption is not critical for fixed relays, however this increase will constitute major for mobile relays. The change in transmit power of a relay with respect to the load will depend on the multiple access scheme used. In a TDMA system, no additional power will be needed, since a relay will transmit signals to and from terminals in a time division manner. In a CDMA system, on the other hand, a linear increase in the transmit power will be necessary as a result of the simultaneous transmissions.

For infrastructure based systems with fixed relays, the selection of relays is much simpler and predefined. For this case, possible concepts can be based on the centrally controlled heuristic methods for relay channel selection within a single multi-hop cell [i.251]. Selection criteria involve the mutual interference between relay channels.

Annex M: Design Considerations in Ad-hoc Networks

M.1 Introduction

Stand-alone (isolated) ad-hoc networks are the simplest form of ad-hoc networks. These are established by a certain number of nodes in order to provide, or use, services between themselves. No communication links to other networks are assumed to exist. Although it is possible to envisage very large ad-hoc networks, a more realistic estimation is that networks will usually be small to medium size (up to 100 nodes) with the longest routes not longer than 10 hops.

Mobility of nodes participating in the network will be low (people sitting in a bus) to medium (emergency services personnel in a field) and throughput available in the network should be large enough to accommodate various multimedia applications. Since the network is established spontaneously, available services are not known in advance, and appropriate mechanisms have to be provided for service discovery. Ensuring integrity and privacy of information is especially important, bearing in mind that other, not necessarily trusted, nodes relay data. Energy efficiency of all protocols is important due to the limited available power resources of devices.

IP and WLAN 802.11 based ad-hoc networks have been the most frequent focus of ad-hoc networking research so far. The most important research issues in regard to this ad hoc network scenario are described in the following clauses.

M.2 Network Organization

The actual level of the self-organization capability of an ad-hoc network will have a profound impact on the user's experience and satisfaction. Depending on the wireless technology used, various tasks have to be performed in this phase. If Bluetooth is used, devices have to detect each other, establish a communication link and to determine if they have a compatible set of supported services. In WLAN 802.11 networks, the broadcast nature of the communication channel is used to advertise presence of nodes. Once a new node is recognized, address allocation, service discovery, routing protocols, etc. are invoked. Other MAC protocols might need to allocate timeslots, frequencies or codes for transmission and perform network clustering before any network and application level communication can start.

As ad-hoc network topology changes frequently and nodes are joining, leaving or changing position in the network, network organization protocols have to ensure a smooth and uninterrupted functioning of the network. Nodes running out of battery should be dynamically freed from routing assignments before they switch-off. If dedicated nodes are used (ad-hoc cluster controller), such a node leaving the network can cause problems. The functionality of the leaving node has to be adopted by another node (assuming that there exists a node which supports this functionality). The configuration becomes easier if overlay nodes (no mobility, fixed routes) are used to span an overlay network, but this cannot always be assumed. In WLAN 802.11 based networks, routing protocols take care of some aspects of network maintenance through route maintenance procedures.

M.3 Address Assignment

Stand-alone ad-hoc networks are usually considered to be IP-based networks, and therefore each node is designated with an IP address, assigned to the node in advance or dynamically by an addressing allocation server. However, certain applications can use other addressing schemes: wireless sensor networks are usually based on attribute-based naming schemes, where nodes do not have unique addresses, but are designated by their capabilities; in Bluetooth networks, each node has a unique Bluetooth address that can be used on higher communication layers as well.

Focusing on IP address allocation protocols, since ad-hoc networks do not have a central authority responsible for address allocation coordination, this functionality has to be distributed across all network nodes. A protocol for addressing auto-configuration in IPv4 ad-hoc networks is proposed in [i.257]. Addresses are randomly selected from a special part (169.254/16) of the network address space. Duplicated Address Detection (DAD) is used to eliminate duplicated addresses; this approach uses route discovery messages from a reactive routing protocol like AODV [i.258] or DSR [i.259]. DAD is performed only once per node. Hence, the uniqueness of addresses cannot be guaranteed after merging two networks. This approach is not suitable for large ad-hoc networks. It should be considered that mobile nodes could have more than one interface to different networks, and therefore may require multiple IP addresses.

Another approach, called Dynamic Registration and Configuration Protocol (DRCP), tries to modify DHCP to an auto-configuration protocol for wired and wireless networks. Therefore, each node represents a DRCP client and server and owns an IPv4 address pool. The Dynamic Address Allocation Protocol (DAAP) is responsible for the distribution of the address pools. Each node requesting a pool gets half of the pool of a neighbouring node. This results in a lot of unassigned addresses in an already scarce IPv4 address space. Network merging is not considered either.

A promising method based on Mobile IP [i.260], consists of a home address and a care of address that is built by using a distinct prefix for each subnet [i.261]. The locally assigned address could be used as the care of address, whereas the unique home address could enable the authentication, authorization and, hence, the accounting, similar to the International Mobile Subscriber Identity (IMSI) in GSM. IPv6 Stateless Address Autoconfiguration (SAA) is another proposed approach. It is a hierarchical solution that works together with the LANMARK routing protocol [i.262].

M.4 Service Discovery

Ad-hoc networks are organized "on-the-fly", opportunistically, and therefore, available services and service providers are not known in advance. It is the responsibility of service discovery protocols to provide that information. Obviously, as a permanent central service information database does not exist, this protocol has to be distributed across all network nodes.

Service discovery protocols have to enable not only discovery of services available in the 1-hop range, but all services available within a multi-hop ad-hoc network. An efficient protocol should also ensure that services and infrastructure are not underutilized. It should first identify the existence of a service, and then decide if the existent technology can bind to it, and finally establish a session successfully. In order to do that, it should be capable of giving the ability to the devices to announce their presence to the network and describe their capabilities. It should also be independent of the transmission protocol.

Several service discovery protocols have been developed and proposed, primarily for wired networks. In centralized pull protocols, clients pull the services whenever needed from a central component (called the central registry) where all the existent services in the network are registered. Distributed pull protocols pull services from the network itself, and in distributed push protocols, service providers push information concerning the services to the network.

Some popular service discovery protocols for wired networks are:

- ***"Universal Plug "n Play"*** uses the distributed pull method and relies on HTTP and TCP/IP. Its main drawbacks are that it supports only known devices, and that it does not support many network configurations.

- ***"Jini"*** is a Java based service discovery protocol that uses the centralized pull method. Its main disadvantages are that it does not support many network configurations, and that its centralized service discovery is not suitable for ad hoc networks.

- "***Salutation"*** supports both centralized and distributed service discovery, transport-independent addressing, and device capability exchange. It is designed to function in pervasive and heterogeneous networks, and above most of the network protocols. Its main problems are lack of leasing functionality and complex addressing.

- ***IBM DEAPspace*** supports the distributed push method and the service description is fulfilled with strings and XML. Its main drawbacks are its word view of services and its emphasis on devices.

These wired network proposals have to be adapted for the wireless world and, especially, for the highly dynamic ad hoc networks. Appropriate strategies have to be investigated, which may include hierarchies for service distribution and announcement.

M.5 Routing and Relaying

The design of robust and efficient routing protocols is one of the most critical technical issues associated to the design of ad-hoc networks, particularly when these consist of large numbers of nodes. This clause reviews the main challenges in the design of routing algorithms and candidate solutions that have been proposed in the last few years.

In ad-hoc networks, a direct communication between any two nodes is possible, subject to adequate radio propagation conditions and transmission power limitations of the nodes. However, only in rare cases will direct communication with all nodes in a network exist. Usually, multi-hop communication paths will have to be used. Care of communication route establishment and maintenance is taken by all network nodes using adequate routing protocols.

Ad-hoc network routing is a very challenging task for several reasons:

- High mobility of nodes which can join and leave the network at any time, thus causing network topology changes and making routing tables obsolete.
- The bandwidth of the wireless channels is limited and has to be used carefully, thus requiring the routing overhead to be kept at a minimum.
- The wireless channel is susceptible to various interferences, low throughput and other problems; the limited energy resources of network nodes impose severe constraints.

A number of ad-hoc routing protocols have been proposed, primarily as part of the lETF's MANET (Mobile Ad-hoc NETworks) working group activities. These protocols are designed for IP-based, homogenous, mobile ad-hoc networks, and focus on fast route establishment, re-establishment and maintenance with a minimum overhead. Each node in the network is assumed to have identical capabilities (wireless communication interface and ability to perform functions from the common set of services) and a unique IP address. The number of hops is used as the only route selection criterion. Other parameters, like: ***route delay***, ***energy usage***, ***fair distribution of power usage among terminals***, ***load balancing*** or ***quality of service*** are not considered.

The two main groups of the proposed protocols are ***proactive*** and ***reactive*** protocols. Proactive protocols continuously update the topological view of the network by exchanging appropriate information among the network nodes and, thus, immediately have a route to a destination when required. A typical example of the proactive group of protocols is Optimized Link State Routing (OLSR) [i.263]. It is an optimization of the classical link state algorithm, tailored to the requirements of a mobile wireless LAN. The main problem of the proactive approach stems from the fact that topology of ad-hoc networks is changing continuously. Hence, a frequent dissemination of topology information is required, which causes a large routing overhead. Also, depending on the traffic pattern in the ad-hoc network, it is possible that only a small fraction of routes is used, which leads to a waste of already constrained wireless and computing resources.

AODV [i.258] and DSR [i.259] are examples of reactive, or "on demand", routing protocols. These protocols do not maintain the overall network topology, but instead maintain only those routes that are in use. When a route is not used anymore it is removed from routing tables. When a new route is required the network is flooded with "route request" messages. When the destination or a node, which has a route to the destination, receives a "route request" message, a "route reply" message is generated and sent back to the source node.

The limitations of the so called topology-based routing protocols discussed above are eliminated by using position-based routing protocols [i.264] and [i.265], which utilize the physical position information of the participating nodes. In these methods, each node determines its own position through the use of Global Positioning System (GPS) or some other type of positioning service. This position information is then included in the packet's destination address. The routing decision at each node is then being made based on the destination's position contained in the packet header and the position of the forwarding node's neighbours in such way that a performance metric is maximized. This performance metric indicates the efficiency of the routing algorithm in terms of the length of the route between the source and the destination and/or the transmission delay.

M.5.1 Routing in multi-hop infrastructure-based network

For infrastructure based multi-hop wireless networks, the stationarity (or low mobility) of the infrastructure nodes motivates the utilization of topology-based proactive methods. In this case, the routing information corresponding to the users within the coverage area of an access-point can be stored in and maintained by that access point. Reactive routing methods can also be considered as a part of a hybrid method especially for providing ubiquitous network coverage for inter-system interconnection.

Routing techniques for multi-hop infrastructure-based networks should exploit the inherent characteristics of this network architecture, namely:

- ***Network-oriented processing:*** Part of the routing in an infrastructure-based multi-hop network can be implemented in the infrastructure entities as these entities have more processing power. Having a network-centric routing technique not only simplifies the routing process but also provides the opportunity of performing routing jointly with other layers' functionalities.

- ***Position information and data flow direction:*** The position information and flow direction in both uplink and downlink are available. This information can be utilized for developing efficient position-based routing mechanisms.

- ***Cooperation incentive:*** Referring to the fact that the infrastructure deals with the charging issues, there could be a network coordinated framework, which promotes users' participation in cooperative communication schemes. Users' cooperation can also be very helpful in the process of routing particularly in the case of mobile relays.

In multi-hop infrastructure-based networks, selecting a particular route and transmission on it can also be envisaged as a part of the resource management mechanism. Therefore, routing might be implemented jointly with or as a part of other radio resource control mechanisms ([i.266] and [i.267]).

M.5.2 Performance Metrics in Routing Protocols

Routing in MANETs has traditionally focused on finding out solutions that minimize hop-count and provide fast adaptation in the case of highly dynamic (mobile) networks. One of the problems with the minimal hop-count approach is that it does not take the link-quality into account. Especially in the case of IEEE 802.11 based networks that are deployed into large area, the difference between link qualities can be very large. As a result, it is not rare case that the minimum hop-count based routing schemes chose routes with significantly less capacity than the high-quality paths available in the network. This issue has been pointed out in details, e.g. by [i.269].

A number of different performance metrics, such as the Expected Transmission Count (ETX) metric in [i.268] (expected transmission count metric), per-hop Round Trip Time (RTT) metric [i.269], link-quality dual (SNR, BER), and per-hop packet-pair matrix [i.270], that characterize the quality of the wireless link have emerged in the recent years. For example, ETX finds high-throughput paths using per-link measurements of the packet loss in both directions of the wireless links. In the per-hop RTT approach, the nodes probe periodically their neighbours measuring the RTT. The RTT samples are averaged using TCP-like low-pass filter and the path with the least sum of RTT is selected. The per-hop packet-pair technique, on the other hand, uses two two-back-to-back periodic probings to the each neighbour. The receiving node measures the arrival delay between the two probes and reports it back to the sender. The sender averages the delay samples and the finally the route with the least delay is chosen. Both the per-hop RTT and the PckPair metric implicitly take into account the load, the bandwidth and the loss rate of the wireless link.

M.6 Air Interface

The air interface requirements for ad-hoc networking are many and varied [i.253]. They can range from very low-power, low-data rate telemetry and sensor requirements, to very high data rates for high-quality multimedia distribution in the home. Common requirements include coexistence between multiple instances of the same air interface (in the same, or collocated ad-hoc networks) and any other air interfaces (ad-hoc or deployed). To reach the requirements, techniques for Dynamic Frequency Selection (DPS), link adaptation and power control have to be included, as in the recent standardized Hiper-LAN/2 or IEEE 802.11 [i.269] systems. Furthermore, techniques to support QoS have to be added. Current efforts for QoS support in pure ad hoc networks lead to establishment of a central controller (e.g. IEEE 802.l1e). Techniques to address this issue include dynamic resource allocation, spectrum sharing and spectrum overlay.

M.7 MAC Layer

The MAC layer has to provide efficient and fair access to the wireless medium for all devices, and to ensure reliable data transmission. Current MAC protocols for ad-hoc networks could be classified in to three groups, depending on their channel access strategy.

Contention protocols

Like ALOHA or CSMA, are based on asynchronous communication models. Collision avoidance is an important feature of these protocols that is realized through some form of control signalling. It has been shown that contention protocols are simple, but tend to degrade as the traffic load increases, as the number of collisions rises. In overload situations, a contention protocol can become unstable as the channel utilization drops. This can result in an exponential packet delay increase and network service breakdown, since few, if any, packets can be successfully exchanged.

The Multiple Access with Collision Avoidance (MACA) protocol uses a handshaking dialogue to reduce the *hidden node interference* and minimize the number of exposed nodes. Further enhancements are introduced by the MACAW [i.271] protocol, which includes positive acknowledgements and carrier sensing to avoid collisions. Improvements are also made to the collision resolution algorithm to ensure a more equitable sharing of the channel.

A very similar approach to MACAW is used in the Distributed Coordination Function (DCF), in the IEEE 802.11 standards, with improved collision avoidance [i.269] and [i.273]. Nodes deliver data packets of arbitrary lengths (up to 2 304 bytes), after detecting that there is no other transmission in progress. However, if two nodes detect the channel as free at the same time, a collision occurs. IEEE 802.11 [i.269] defines a Collision Avoidance (CA) mechanism to reduce the probability of such collisions.

Allocation protocols

Allocation protocols employ a synchronous communication model and use a scheduling algorithm that generates a mapping of timeslots to nodes. The mapping results in a transmission schedule that determines in which particular slots a node is allowed to access the channel. This effectively leads to a collision-free transmission schedule. It turns out that the allocation protocols tend to perform well at moderate to heavy traffic load, but these protocols are disadvantaged at low traffic, due to the artificial delay induced by the slotted channel.

Hybrid protocols

Hybrid protocols can be loosely described as any combination of two or more protocols. IEEE 802.15.3 MAC draft standard [i.274] is one such protocol. It is defined for narrowband 2,4 GHz WPAN applications. The emerging UWB physical layer draft standard IEEE 802.15.3a, adopted to be compatible with the IEEE 802.15.3 MAC standard [i.274], will possibly have a few adaptations due to the inherent specificities of the UWB physical layer.

The IEEE 802.15.3 MAC protocol is centrally coordinated, with a PicoNet Coordinator (PNC) that synchronizes the devices and allocates the resources. Even the MAC protocol is a centralized one, the topology is ad-hoc and communication is established in a peer-to-peer mode. The PNC can be chosen dynamically, i.e. it is autoclaimed each time a new piconet is created. The main part of the processing power is concentrated in the PNC's hands, but if the PNC disappears, another station can take on its role, which is an advantage over static centralized management.

M.8 Radio Resource Management (RRM)

The development of efficient algorithms for RRM is critical from a network point of view, since such functionalities have a significant impact on the fulfilment of QoS requirements, and on attaining higher degrees of spectral efficiency. Radio resource management activities encompass a number of functions:

- *Admission control* ultimately decides whether a new flow can be granted, while preserving overall QoS requirements. The admission control would be invoked at each node to make a local accept/reject decision in the framework of cluster-oriented architectures.

- *Congestion control* mechanisms are invoked whenever network overload leads to unfulfilled QoS requirements for the admitted users (for a fraction of time). When in congestion, some users could experience a reduced QoS margin, not beyond, though, an agreed percentage of time.

- *Packet scheduling* schemes determine how different flows are forwarded in a specific network node (mechanisms such as priority queues, timers, etc. are used). Priorities can be service-dependent and, for a specific service, transient QoS needs can also be considered. Despite the existence of some degree of flexibility in the choice of the scheduling policy, any sensible approach should target optimizing the overall network performance.

In the context of self-configurable radio networks, RRM functionality can no longer be centralized in a specific node. Conversely, a new distributed RRM architecture has to be envisaged, where RRM functionalities are implemented in every single network node or mobile station.

M.9 Cross-layer Strategies

Traditionally, in MAC protocol design, little or no attention has been paid to the underlying physical layer features. Thus, most MAC protocol enhancements were proposed with the common idea to suitably manage and avoid collisions. However, the advent of sophisticated signal processing techniques (array processing, multiuser detection, channel coding strategies, etc.) that are able to extract useful signal(s) from noise, interference and unwanted signal replicas, could change many of the underlying assumptions in the conventional MAC schemes.

For example, the assumption that more than one simultaneous transmission over the same radio resource (e.g. identical frequency, time and spreading code assignment) inevitably leads to a collision, should be revisited. In other words, making MAC, RRM and upper-layer functionalities aware of the physical layer state information (for example, in terms of diversity-based component status, channel response or interference indicators), could boost system efficiency in terms of resource reutilization, by allowing each mobile terminal to transmit so that an optimal usage of the spectrum available was attained. This strategy departs from those in conventional MAC schemes, where packet collisions should always be avoided and, hence, more than one user is not allowed to share the same radio resource. Given the time-varying nature of those parameters, the envisaged MAC schemes are most likely to be adaptive.

Accordingly, other inter-layer dialogues can be established in the OSI stack. For example, investigations have shown drawbacks of 802.11 MAC protocol in multi-hop communication. In particular, the optimization in terms of routing could be improved, by providing some of the information available (SNR information, packet acknowledgments, etc.) at the MAC layer to the ad hoc routing protocol on the network layer.

M.10 Security

High-level security requirements for ad-hoc networks are basically identical to security requirements for any other communications system, and include: *authentication*, *confidentiality*, *integrity*, *non-repudiation*, *access control* and *availability*. However, similar to wireless communication systems creating additional challenges for implementation of the above-mentioned services when compared to fixed networks. Ad- hoc networks represent an even more extreme case, requiring even more sophisticated, efficient and well designed security mechanisms [i.275] to [i.278]. These additional challenges are caused by two basic assumptions of an ad-hoc system: a complete lack of infrastructure, and a very dynamic and ephemeral character of the relationships between the network nodes.

The lack of infrastructure implies that there is no central authority that can be referenced when it comes to making trust decisions about other parties in the network, and that accountability cannot be easily implemented. The transient relationships do not help in building trust based on direct reciprocity, and give an additional incentive to nodes to cheat.

Ad-hoc networks rely on cooperation of involved nodes in order to build and maintain the network. Current versions of mature ad-hoc routing algorithms only detect if the receiver's network interface is accepting packets, but they otherwise assume that routing nodes do not misbehave. Whereas such an assumption may be justified when single domain networks are concerned, it is not easy to transpose it on a network consisting of nodes unknown to, and not trusted by, each other.

Since ad-hoc networks use multi-hop routing protocols, where each of the nodes, in addition to its own packets, has to forward packets belonging to other nodes, selfish behaviour may represent a significant advantage for a node, saving its battery power and reserving more bandwidth for its own traffic. However, if a large number of nodes start to behave non-cooperatively, the network may break down completely, depriving all users of the services. Non-cooperative behaviour in multi-hop routing protocols may also result in a denial of service attacks on the network, where malicious nodes join the network for the sole reason of misbehaving and depriving all other nodes of legitimate services. Such denial-of-service focused misbehaviour may consist of dropping (not forwarding) the packets, injecting incorrect routing information, replaying expired routing information or distorting routing information in order to partition the network. Also, bogus nodes may try to attract as much traffic as possible to them in order to be able to analyse it. In general, attacks on a routing protocol can be classified as dropping of data packets, route modifications, dropping of error messages and frequent route updates.

Another challenge is metadata protection, including confidentiality of identity (pseudonym and anonymity), confidentiality of location (traceability) and traffic analysis. The confidentiality of this metadata will gain in importance in the ubiquitous computing environment, where the ubiquitous computing infrastructure could potentially become a tool for a powerful surveillance, making us involuntary participants in a worldwide "Big Brother" show.

A dangerous attack in civil applications, typically using an open ad-hoc environment, may consist of so-called "sleep deprivation torture". In this type of denial-of-service attack, the attacker is trying to deprive a device of battery power by keeping it awake and engaging in communication all the time. Strong authentication of communication peers, or some kind of accountability, based on either expensive pseudonyms and reputation mechanisms or micropayments, could be used to prevent, to some extent, this kind of attack [i.276] and [i.277].

M.11 Interoperability with Fixed/Overlay Networks

Although stand-alone ad-hoc networks provide support for many interesting applications, in many scenarios a connection to fixed/overlay networks will be required. This can be primarily achieved by connecting one or more ad-hoc network nodes, wirelessly or using a wired link, to a fixed network. These nodes then act as fixed network access points.

Wireless nodes are distributed over a given area. Some of the nodes connect directly to a wired access point. Because of the missing infrastructure, not every access point can be hard-wire connected. Therefore, virtual access points are introduced. Virtual access points are connected directly, or over multi-hop with the wired access points. Wireless nodes can connect to any of the access points, depending on their location and/or signal strength. If a node cannot connect directly to any access point connection can be established using other nodes as routers for multi-hopping based access.

Research issues applicable to stand-alone ad-hoc networks are also valid in this scenario, but there are several specific issues as well:

- ***Gateway node role:*** a gateway node can act as a bridge or a router device. Depending on the role, different addressing and protocol translation mechanisms have to be proposed.
- ***Authentication, authorization and accounting:*** when accessing a fixed network, ad-hoc network nodes have to be authenticated and granted appropriate access rights. The gateway node can either take responsibility for all underlying ad-hoc nodes, i.e. can be the only node seen and authenticated by the overlay network, or can just tunnel the ad-hoc network nodes' traffic to an authentication server in the overlay network.
- ***Addressing:*** as in the stand-alone ad-hoc networks scenario, the assigned node address has to be unique. Now, however, at least the gateway node has to have a public address along with the private one. If other nodes are to be accessible from outside of the network, then they need public addresses too.
- ***Node mobility:*** how is the mobility of ad-hoc networks supported, i.e. is it possible for an ad-hoc network to change its point of access to the fixed network without interrupting current communication, or for an individual node to transfer between ad-hoc networks without losing the connection to the fixed network?
- ***Gateway service:*** how does a mobile node find a gateway? are the gateways advertising themselves, or only responding to requests from nodes? If the network is stable, gateway advertisements produce unnecessary load; but if it is unstable and the nodes have primary internet traffic, the nodes can profit from the gateway advertisements.

M.12 Integration of Ad-hoc Networks into Cellular/Satellite Networks

Unlike systems providing an ad-hoc mode, cellular systems rely on an infrastructure (BS) and require network planning and operation in licensed radio spectrums. UMTS provides cumulative data rates of up to several Mbit/s, which might still not be enough for hot spot areas, where the number of Mobile Nodes (MNs) per area is very high. To increase the individual data rate of users, WLAN systems are introduced at these places, which can provide transmission rates of at least 54 Mbit/s. Nevertheless, transmission power in such communication systems is limited, hence the coverage is limited as well, and interference between such systems is difficult to predict and to control.

Taking into account the advantages, potential and drawbacks of cellular networks, WLAN, and self-organizing network architectures with respect to, e.g. *coverage*, *capacity*, *mobility*, *cost* of *infrastructure* and *flexibility*, it becomes obvious that a combination of them is the logical consequence for the next generation network concepts.

In situations where cellular networking capability and ad-hoc networking capability coexist in the same devices (MN), **it is possible to utilize a cellular network to assist ad-hoc networking**. This kind of hybrid network could include centralized servers in a fixed/cellular network to handle ad-hoc network topology to assist routing and authentication, but can also be considered as an extension of cellular networks.

While the defining goal of the ad-hoc networks is the ability to function without any infrastructure, the goal of these multi-hop-augmented infrastructure-based networks is the almost ubiquitous provision of very high data rate coverage and throughput.

As an example of a Hierarchical Multi-hop Cellular Network (HMCN) a Land Mobile Satellite system, could be the basis of the proposed network architecture. The possible connection of each MN to the satellite guarantees full coverage, and this connection can always be taken as a fall-back solution in the case where an MN loses connection to any other kind of network it might be connected to. This requires interoperability of the existing and the future networks and the support of vertical handover, i.e. handover between different wireless access networks (intersystem handover). The satellite provides access to the backbone, which is most likely to be based on the TCP/IP protocol suite.

The next evolutionary step towards a hierarchical multi-hop network structure is to introduce Multi-hop-Capable Nodes (MHN), which can be fixed or mobile. With fixed MHNs, the coverage of the satellite can be extended to urban/indoor and other harsh propagation environments. At the same time, a fixed MHN can be connected to a power supply to offer more potent services. Subcells can be established in a self-organizing manner. This means the MHN can takes over control functionality within the subcell. A typical control function comprises the management of the medium access within the subcell. Furthermore, it provides connections between MNs in the subcell, which can directly communicate with each other by means of a direct mode. Moreover, the satellite can provide to the MHN and MNs useful signalling information, like routing information for example.

Of great interest too, is the case where the routing in the subcells is assisted by the overlying satellite (or 3G) systems. Due to the hierarchical structure, an optimum control of resource allocation can be organized.

Besides fixed MHNs, mobile MHNs are also considered in a further evolutionary step. In the case that the required data throughput cannot be provided any more by the satellite link or established fixed MHN, a MN can become an MHN and can establish a sub-cell on demand. These cells can use the same, or different, frequencies. In this case, subcells can be adaptively established. The spectral efficiency of the system can be enhanced when reusing the same frequency in different subcells. In the case of using different frequencies in different subcells, the available transmission rate within the considered cell can be increased.

There are many issues to be investigated on the path towards a successful integration of the multi-hop capability in conventional wireless networks:

- The advantages and disadvantages of having fixed versus mobile repeaters (routers).
- The advantages and disadvantages of relaying in analogue (amplify-and-forward) versus digital (decode-and-forward) form.
- The load balancing capability by diverting the traffic with repeaters as necessary.
- The signalling overhead.
- Relaying interference.

- A possible cap on the number of hops, incurred latency and its impact on QoS.
- Complexity and functionality of relay devices.
- Scheduling.
- Radio resource management.
- Novel diversity techniques (macro diversity for example).

A detailed description of this scenario and the above requirements is given in [i.252].

Annex N:
Example of beyond 3G satellite services for Korea

N.1 Introduction

IMT-Advanced systems allow mobile services that are able to have an always-on access to a wide rage of broadcasting and telecommunication networks including advanced mobile service with transparent devices. In order to provide the seamless service over a global coverage, the satellite component of IMT-Advanced should play an important role and be able to bring new services without which will not be possible.

The objective of this annex is to introduce potential IMT-advanced satellite services considered by ETRI which represents an example of beyond 3G satellite services in Korea.

N.2 Potential Services

The two IMT-Advanced satellite services considered by ETRI are as follows:

- Fill-in Service.
- Two-way S-DMB Service.

Fill-in service is defined as a service that guarantees always-on access to voice and data with transparent devices through satellite fill-in coverage for blind spots. The handsets use a terrestrial network when they are within the terrestrial reach, but they communicate directly with a satellite, otherwise

Two-way S-DMB service is a low cost satellite digital multimedia broadcasting service to permit seamless availability of a return-link for interactive television.

Both services are expected to be able to contribute to bridge the digital divide significantly and to provide advanced broadcasting and telecommunications services to rural or under-served areas. In particular, these services are compelling not only to customers to travel along the countryside or stay in mountainous national parks, which could not be covered by any terrestrial network, but to public safety and security personnel who need ubiquitous coverage in times of emergencies. Figure N.1 depicts the two potential services that are to be provided by IMT-Advanced satellite systems

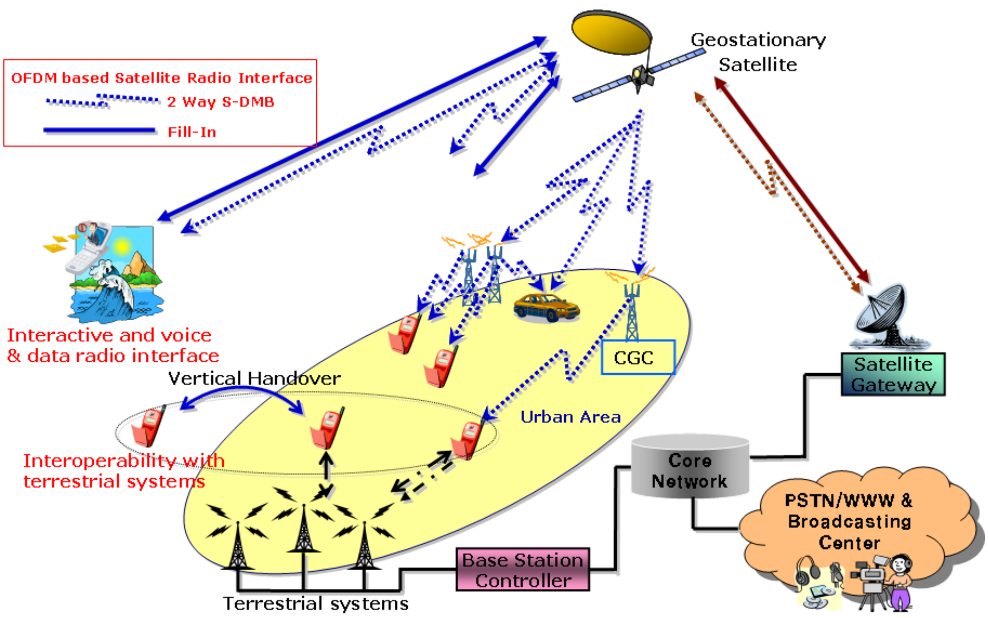


Figure N.1: Two potential services provided by IMT-Advanced satellite systems

Skyterra (formerly MSV) and Globalstar have obtained licenses from FCC to provide Ancillary Terrestrial Component (ATC) service which is very similar to Fill-in service mentioned above. The major difference lies in that Skyterra has its own terrestrial network resulting in single-band/single-mode handsets while Fill-in service corporates with the existing terrestrial networks resulting in dual-band/single-mode handsets. Thuraya and AceS handsets have a dual-mode feature that allows them to operate in both satellite network and terrestrial mobile network. The mode and band features of Satellite/ Terrestrial Integrated mobile Communication System (STICS) being developed in Japan are not known yet.

N.3 Economic Assessment

N.3.1 Demands

Below are assumptions made and demand forecasting of the two services. More detailed surveys and analyses are on-going. Demands based on the assumptions are forecasted as:

- Numbers of subscribers of the voice fill-in service:
  - Cellular phone subscribers who want to stay connected anytime anywhere even in case of an emergency : Telephone survey(sample size:1 000) reveals that 85 % of the cellular phone subscribers are willingly to pay 100 KW (7 US cents) of monthly minimum for ubiquitous connectivity. These subscribers are willing to get insurance of being able to establish emergency calls.
  - Cellular phone blind spot residents: 50 000 households (0,3 %) are estimated to reside in cellular phone blind spots, and the numbers will be reduced by 0,01 % each year. Tariff is exactly same as existing cellular phone.
  - National park visitors: 38 Millions of man-days per year are exposed to blind spots during 19,8 % (a newspaper field study) of their stays. Tariff is exactly same as existing cellular phone.
- Current S-DMB service provider in Korea, Tu-media, continues to provide two-way S-DMB for minimizing churn rate into T-DMB, therefore, current S-DMB subscribers (1,7 million) will be two-way S-DMB subscribers without paying any premium. According to a survey (e-daily, 2008), there will be 4 % CAGR for the advanced DMB market [i.326].
- Data Fill-in service is not included in this stage.

N.3.3 Conclusion

In this annex, potential IMT-advanced satellite services considered by ETRI are introduced, and their potential revenues are assessed. The two services proposed are fill-in service and two-way S-DMB service. Subscribers of fill-in service consists of 85 % of cellular phone subscribers, blind spot residents in remote areas and national park visitors while two-way S-DMB service inherits existing S-DMB subscribers without any premium charge.

ETRI estimates that total revenue from both of the services starts from US $119M at the beginning year and amounts up to US $167,7M at the final year. In the cost side, satellite system cost is estimated as US $272M, and other costs are projected from profit and loss statements of last four years of a Korean cellular phone company. In addition government subsides as another potential source of revenues can be added.

It is important to note that there will be many other intangible benefits created by the two services such as bridging the digital divide, reducing the cost of recovering from disasters.

History

Document history		
V1.1.1	March 2010	Publication