Power Versus Bandwidth Efficiency in Wireless Communications: The Economic Perspective

Jos Akhtman and Lajos Hanzo
School of ECS, Univ. of Southampton, SO17 1BJ, UK.
Tel: +44-23-80-593 125, Fax: +44-23-80-593 045
Email: {ya,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

***Abstract*—We carry out a comprehensive analysis of a range of wireless network efficiency considerations. Firstly, we explore the properties and the implications of the power- versus bandwidth-efficiency criteria. Secondly, we perform a detailed top-down analysis of a typical commercial wireless network, which emphasizes the inherent differences between the aforementioned two efficiency metrics, while demonstrating that the appropriate choice of the network optimization criterion can have a profound effect on the overall network performance. Finally, we address the issue of resource management and its impact on the definition of the overall system efficiency.**

I. Introduction

Wireless communications constitute a thriving trillion-dollar global industry of vast scale and socio-economic significance. The continuous investment in research and development, aimed at improving the utility and the efficiency of wireless communications networks, brings about a wealth of theoretical knowledge and practical engineering solutions. Remarkably, however, a widely accepted choice of a criterion characterizing the overall efficiency of a wireless network remains an open problem [1]. By definition, *efficiency* is a quality that characterizes the correspondence between the *consumed resources* and the *attained utility*. Clearly, therefore, the notion of efficiency is intricately related to the contextual definition of the utility as well as the specific method of attributing value and cost to the resources available. In the context of wireless communications, we may identify the two attributes of the wireless electromagnetic medium, namely the *frequency spectrum* and the *power*, as the two major resource categories.

On the one hand, many researchers consider *bandwidth efficiency* as the principal efficiency criterion [2]. This tendency is vividly reflected by the abundance of standards, such as the High-Speed Packet Access (HSPA) [3] and the IEEE 802.11n as well as the 802.16 and 3GPP LTE [4] systems complemented by a large body of theoretical studies proposing transceiver schemes exhibiting an ever-increasing bits-per-second-per-Hertz (bps/Hz) throughput, while sometimes overlooking the associated area spectrum efficiency, power consumption and complexity-related issues. The *bandwidth efficiency* criteria is largely motivated by the assumption of strict limitations on the available bandwidth, which are further stimulated by the assumption of the spectrum being a scarce resource. As pointed out in [1], [5], however, the scarcity of spectrum is an artificial constraint inflicted by the currently prevailing attitude to spectrum management.

Acknowledgements: The work reported in this paper has formed part of the Delivery Efficiency Core 4 Research Programme of the Virtual Centre of Excellence in Mobile & Personal Communications, Mobile VCE, www.mobilevce.com. Fully detailed technical reports on this research are available to Industrial Members of Mobile VCE.

On the other hand, a host of spread-spectrum methods intentionally sacrifice the bps/Hz performance for the sake of achieving a better bit-per-second-per-Watt (bps/W) *power efficiency*, typically accompanied by lower levels of spectrum contamination, as well as by an increased robustness against the interference [6]–[8]. By no means should these methods be classified as less efficient, since in the appropriate circumstances they are capable of considerably improving the overall performance of the entire network.

In order to further emphasise the importance of the problem considered, we would like to point out some of the pressing socio-economic issues, which require a prompt attention of the engineering community. Each of the major mobile carriers currently consume in excess of 50 MWatt of power to support their respective cellular network infrastructures in the UK alone. At the time of writing, the cost of the power consumed constitutes a substantial factor in the financial bottom line of all major mobile carriers. While the business model assumed by these commercial entities may still be economically viable, this might no longer be the case in the future. With the costs of energy rapidly rising as well as with the growing awareness of the associated environmental impact, the effective cost of the excessively bandwidth-efficient but interference-sensitive signalling strategies may become prohibitive.

Against this background, in this paper we carry out a comprehensive analysis of a range of wireless network efficiency considerations. Specifically, in Section II we explore the properties and the implications of the aforementioned *power* versus *bandwidth* efficiency criteria. In Section III we carry out a detailed top-down analysis of a typical commercial wireless network, which emphasizes the inherent differences between the aforementioned two efficiency metrics, while demonstrating that the appropriate choice of the optimization criterion can have a profound influence on the overall network performance. Our conclusions are summarized in Section IV.

II. Power versus Bandwidth Efficiency

Let us recall the fundamental capacity upper-bound of a communication system quantified by the Shannon-Hartley theorem [9], which postulates

$$R < B\log_2(1+\gamma_s), \tag{1}$$

where B denotes the number of complex degrees of freedom per second available for communication, while γ_s is the average Signal-to-Noise Ratio (SNR) recorded at the receiver. Equation (1) describes the capacity of a Gaussian channel, assuming an *infinite* duration of the transmitted signal, as well as an *infinite* detection/decoding complexity. It is important to underline the universality of (1), which applies to both single- and multiple-antenna systems, where the number of spatial degrees of freedom constituted by the minimum between the numbers of transmit and receive antenna elements has to be taken into account in the calculation of the total number of complex degrees of freedom B. Furthermore, (1) may be readily applied to fading channels characterized by arbitrary distributions. Finally, (1) may be applied to the most generic description of wireless networks constituted by mobile ad hoc networks (MANETs) [10], where the total number of available complex degrees of freedom may be calculated as a product of the available bandwidth and the total number of transmit-receive antenna pairs. Importantly, the strict inequality sign in (1) holds in *any* realistic communication system, while the realistically achievable total rate calculated as the sum of all successfully completed session's rates attainable by such a network is typically substantially lower than the bound provided by (1).

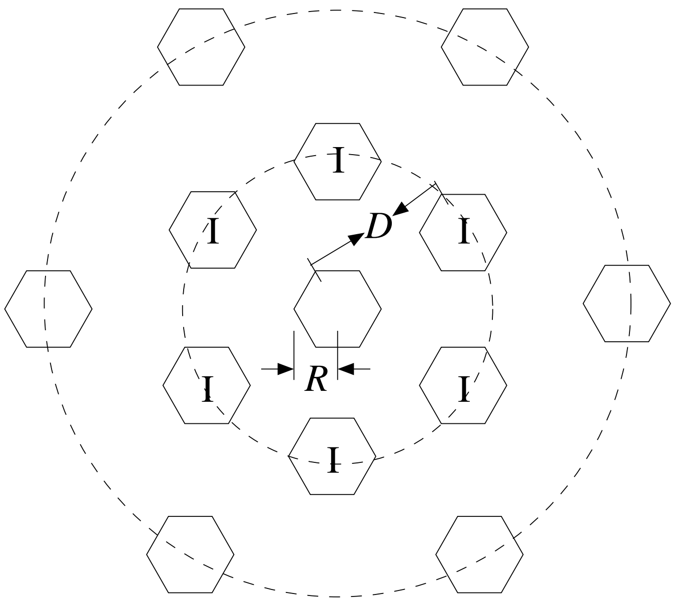

Fig. 1. Multicellular structure with interferers [11].

Subsequently, we define the *bandwidth efficiency* ν_b of a communication system as the number of bits per complex degree of freedom available for communications, which in the case of single-antenna systems corresponds to bits-per-second-per-Hertz (bps/Hz), while in multiple-antenna scenarios is equivalent to bits-per-second-per-Hertz-per-antenna, where the number of antennas, or spatial links is determined by the minimum between the numbers of transmit and receive antenna elements. Furthermore, we define the *power efficiency* ν_p as the number of bits per thermal noise energy unit (TNEU), where TNEU refers to the amount of signal energy identical to the variance of the complex-valued AWGN samples recorded at the receiver. From (1) we may conclude that

$$\nu_b < \log_2(1+\gamma_s), \tag{2}$$

while

$$\nu_p = \frac{\nu_b}{\gamma_s} < \frac{\log_2(1+\gamma_s)}{\gamma_s} \tag{3}$$

for any realistic communication system.

Moreover, in most practical scenarios characterized by various performance-limiting factors including channel fading, interference as well as latency and complexity constraints, the actual attainable bandwidth and power efficiencies are considerably lower than those predicted by Equations (2) and (3). For instance, in the case of uncorrelated Rayleigh fading channels, (2) becomes

$$\nu_b < \frac{1}{\log 2} e^{1/\gamma_s} \mathrm{Ei}(1/\gamma_s), \tag{4}$$

where we define $\mathrm{Ei}(x) = \int_x^\infty \frac{e^{-t}}{t} dt$ [12]. Furthermore, in order to exemplify the impact of interference on the achievable data-rate upper bound, let us consider the simple interference model devised in [11], [13] and shown in Figure 1. Specifically, we assume an idealized hexagonal cell structure comprised by cells of radius R and frequency reuse cell-clusters of size N_f, which further implies having six significant interferers at a distance of $D = \sqrt{3N_f}R$ [11, Chapter 14]. Following [11], we assume a logarithmic path-loss model, as well as receiving an approximately equal power $I_k = (R/D)^\gamma S$ from all six first-tier interferers, where S denotes the signal power, while γ is the pass-loss exponent. The resultant Signal-to-Interference Ratio (SIR) may be expressed as

$$SIR = \frac{S}{\sum_{k=1}^{6} I_k + N_0} = \frac{\gamma_s}{1 + 6(3N_f)^{-\gamma/2}\gamma_s}, \tag{5}$$

where N_0 denotes the PSD of the local noise.

The efficiency upper bounds corresponding to the Gaussian and Rayleigh channel scenarios of Equations (2), (3) and (4) as well as the corresponding upper bounds for the interference-limited scenario of Equation (5) are depicted in Figure 2. It is evident that the efficiency criteria ν_b and ν_p are inherently different and may not be maximized simultaneously. Specifically, as seen in Figure 2, any transmission scheme, characterized by a bandwidth efficiency, which exceeds 1 bps/Hz/antenna exhibits a substantial degradation in terms of the corresponding power efficiency. Observe, that while the higher values of the frequency reuse cluster size N_f tend to mitigate inter-cell interference and thus result in improved "local bandwidth efficiency", the corresponding area spectral efficiency, which is inversely proportional to the cluster size N_f is substantially reduced. We would like to conjecture that an appropriate trade-off between the efficiency metrics ν_b and ν_p has to be found for the sake of maximizing the network's overall utility, as defined below. In the following section we would like to elaborate on the importance of the appropriate choice of efficiency criteria

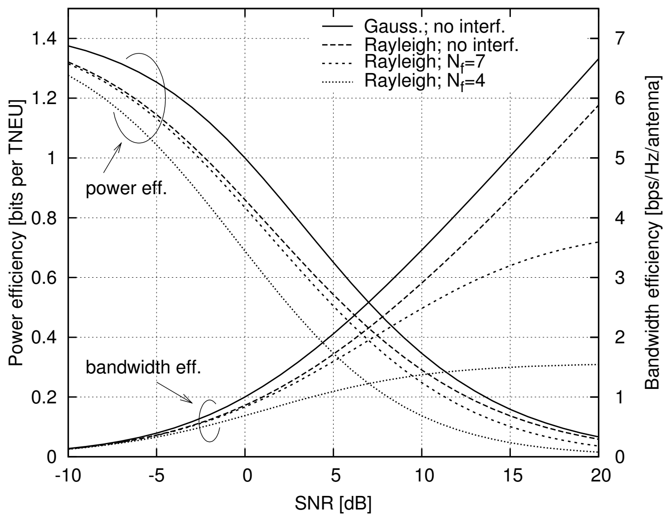

Fig. 2. Power and bandwidth efficiencies as a function of the SNR recorded at the receiver. The power efficiency is quantified in terms of the number of bits per thermal noise energy unit (TNEU), which corresponds to the number of bits communicated using a signal having the same power spectral density (PSD) as that of AWGN recorded at the receiver.

and its corresponding impact on the network's utility with the aid of an example.

III. Top-Down Analysis

Let us consider a commercial entity X, whose main business is the provision of wireless communications services to a diverse body of end-users. We will use the average monetary profit $\mathcal{K}$ expressed in monetary units per second as the major criterion of the company's performance. Using the most basic economic principle, we may suggest that the profit $\mathcal{K}$ may be formulated as the revenue minus the actual cost of the services provided. Specifically, let us define the average revenue per second as a function $A(R)$ of the total rate R (in bits per second) of successfully communicated information over the entire network of carrier X. Likewise, we can quantify the running cost per second as

$$C = C_p P + C_r = C_p P_{\mathrm{RF}} + C_p P_{\mathrm{c}} + C_r, \tag{6}$$

where the factors C_p and P denote the cost per Wattsecond (Ws=joule) and the total average power (in Watts) consumed by the carrier's network, while the coefficient C_r denotes the cumulative rate of all additional costs not related to the power consumption, including the hardware investment and maintenance costs, as well as the spectrum licensing costs. Furthermore, the power consumption constituents P_{RF} and P_{c} correspond to the power consumed by the RF transmission equipment and the rest of the power dissipated by the network's infrastructure, including servers, air-conditioning equipment, backhaul communication equipment *etc.*

The exact correspondence between the various cost constituents in (7) is difficult to quantify and varies substantially among different systems. We may, however, safely conjecture that all three such components $C_p P_{\mathrm{RF}}$, $C_p P_{\mathrm{c}}$ and C_r are significant to the the calculation of the overall cost and may not be ignored. As the simplest general assumption, in this respect, it appears plausible to assume the following strong inequality

$$\mathcal{K} < A(R) - C_p P_{\mathrm{RF}}. \tag{7}$$

Subsequently, we may conjecture that the total power P_{RF} dissipated by the RF equipment employed by the carrier's infrastructure may be expressed in terms of the average SNR γ_s required to achieve reliable communications at the target data-rate of R. More specifically, we may formulate $P_{\mathrm{RF}} = \alpha B \gamma_s N_0$, where the coefficient α denotes the transmission chain's overall efficiency, which takes into account, for instance, the power efficiency of the RF equipment, such as the frequency conversion and power amplification efficiency as well as the average link budget, including the Tx/Rx antenna gains and path loss.

Let us now consider the properties of the revenue function $A(R)$. Firstly, we would like to suggest that the revenue of the commercial carrier X is determined by the corresponding utility of its network to the end-user. Secondly, the utility of the network to the end user is determined by the user's ability to benefit from a certain type of telecommunication services, such as, for instance, text messaging, voice communications, or mobile access to the WWW. As evidenced by the experience of the industry's ongoing transition between the 2G and 3G technologies, the multifold increase in the achievable throughput bears only a marginal increase in the network's value without the introduction of new services.

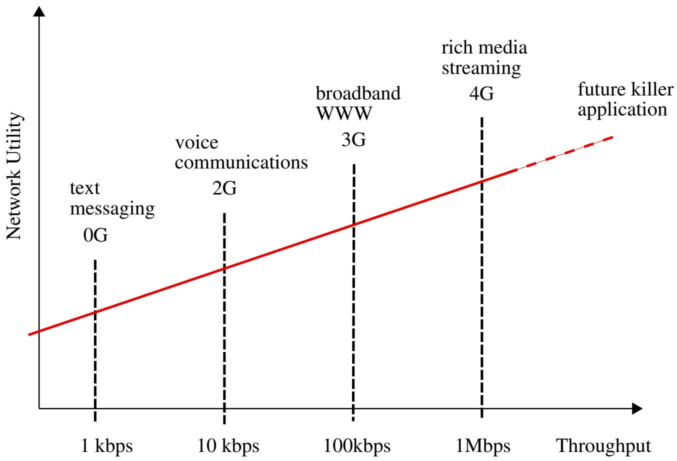

Fig. 3. Network utility to the user versus the average achievable user throughput.

Naturally, each specific type of service contributed an additional value to the user. For instance, the user of the voice communication network might be willing to pay an extra price for the ability to access the WWW. The user having the ability to carry out voice communications and access the WWW may be willing to pay a further premium for the additional ability to stream high-quality video from the on-line server using his/her mobile communication device. Although, each such additional service may entail as high as a ten-fold increase in the required throughput, the network's users typically see these additional capabilities as incremental increases in the network's value.

This trend is illustrated in Figure 3. Specifically, we would like to emphasise that the emergence of each subsequent generation of mobile communications technology has been accompanied by the introduction of a new class of services, which would attempt to justify such transition in term of the network's increased value to the consumer. Furthermore, while the increase of the attainable throughput facilitated by each successive generation of mobile technology has roughly followed Moore's law, the perceived value of the network per user appears to have increased roughly linearly.

In order to reflect the behaviour depicted in Figure 3, we would like to model the attainable revenue $A(R)$ in the following form

$$A(R) = N_u \log_2 \left(1 + \frac{B_u}{\bar{R}_u}\right), \tag{8}$$

where we conjecture that the revenue $A(R)$ is linearly proportional to the number of active subscribers N_u, while being logarithmically proportional to the average data-rate attainable for each user. Consequently, taking into account the aforementioned fact that the utility $\mathcal{K}$ is defined up to multiplicative currency-related factor and using the Shannon-Hartley theorem of Equation (1), we may formulate the following expression

$$\mathcal{K} = N_u \log_2 \left(1 + \frac{B_u}{\bar{R}_u} \log_2(1+\gamma_s)\right) - C_p \cdot \alpha\gamma_s B N_0 \tag{9}$$

$$= \frac{B}{B_u} \log_2 \left(1 + \frac{B_u}{\bar{R}_u} \log_2(1+\gamma_s)\right) - C_p \cdot \alpha\gamma_s B N_0, \tag{10}$$

where we define the average bandwidth per active user as $B_u = B/N_u$ and the baseline data-rate per user as $\bar{R}_u$. Let us elaborate a little further by defining the utility-per-channel-use as $\kappa = \mathcal{K}/B$, which yields

$$\kappa = \frac{1}{B_u} \log \left(1 + \frac{B_u}{\bar{R}_u} \log(1+\gamma_s)\right) - C_p \cdot \alpha\gamma_s N_0. \tag{11}$$

Note that both the per-channel-use quantities γ_s and κ constitute bandwidth-normalized versions of the per-second quantities P and $\mathcal{K}$. Using Equations (10) and (11) we would like to explore the relationship between the network's utility $\mathcal{K}$ and three major network characteristics, namely the total consumed power P, the total effective bandwidth B, as well as the average user throughput R_u. More specifically, Figure 4 depicts the utility-per-channel-use κ evaluated from Equation (11) as a function of the average SNR γ_s. Observe that the resultant function $\kappa(\gamma_s)$ of Figure 4 exhibits a single global maximum. Although the specific result depicted in Figure 4 was calculated using the set of assumptions based on the UK's GSM network statistics detailed in [14]–[16], it may be readily verified that any sensible choice of network characteristics has a limited impact on the general shape and on the properties of the functions $\kappa(\gamma_s)$ and $\mathcal{K}(P)$. We may thus surmise that the utility achievable by any bandwidth-limited communication network is ultimately upper-bounded and there exists an average SNR as well as a corresponding power consumption point, where the network's utility is maximized.

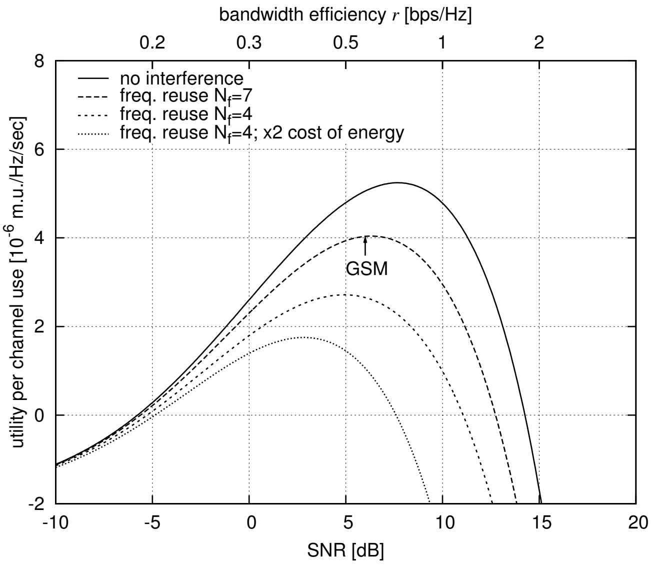

Fig. 4. Network utility per channel use of Equation (11) versus average SNR based on AWGN Shannon capacity and GSM network economy statistics.

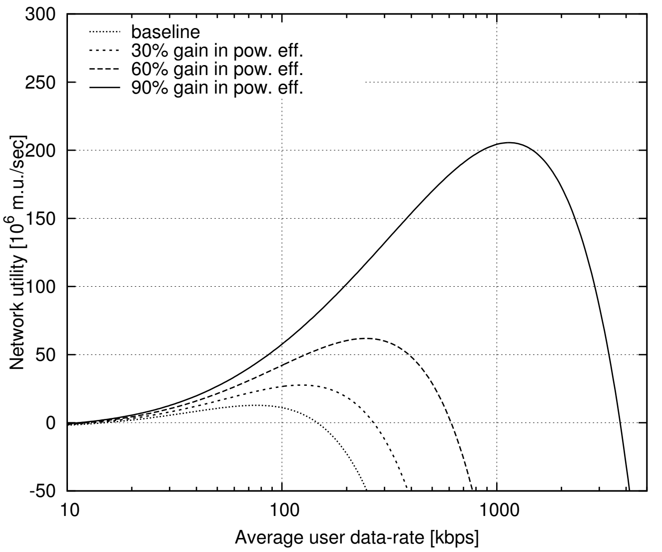

Fig. 5. Network utility of Equation (10) versus the average user data-rate based on the AWGN Shannon channel capacity and on the UK's GSM network statistics [14]–[16].

Subsequently, we would like to characterize the attainable network utility $\mathcal{K}$ as a function of the power efficiency coefficient α in Equation (10). More specifically, Figure 5 depicts the network utility $\mathcal{K}$ from Equation (10) versus the average user rate R_u for the baseline scenarion based on the UK's GSM network statistics [14]–[16] as well as for $30, 60$ and 90% improvements in the overall power efficiency. Once again, in Figure 5 we stipulated the aforementioned assumption of logarithmic data-rate-related user-base expansion. From the results depicted in Figure 5 we may conclude that both the maximum attainable network utility $\mathcal{K}$ as well as the corresponding average user data-rate R_u may be improved virtually indefinitely by improving the network's power efficiency, while gradually expanding the total effective bandwidth and the corresponding size of the user-base.

Based on the implications of Equations (10) and (11) augmented by the results depicted in Figures 4 we would like to summarize the following conjectures:

1) The utility achievable by **any** realistic communication network characterized by a fixed power efficiency is ultimately upper-bounded.
2) For any set of three values selected from the network parameters $\{B, P, N_u, R_u\}$ there exists a unique value of the fourth parameter, for which the network utility $\mathcal{K}(B, P, N_u, R_u)$ is maximized.
3) In any bandwidth-limited communication network a substantial network utility gain may only be achieved by improving the underlying power-efficiency.
4) **The only sustainable way of increasing the network's utility is constituted by a combination of improved power efficiency accompanied by bandwidth expansion.**

It is important to emphasis that the 90% power efficiency gains hypothesised in Figure 5, might not be realistically achievable in the context of the currently prevailing cellular network topology. In contrast, we would like to speculate that a gradual transition towards Mobile Ad Hoc Networks (MANETs), facilitating the employment of short-range ultra-wideband low-power and low-complexity reconfigurable co-operative networking [17], has the potential to sustain the substantial utility gains suggested in Figure 5 and thus guarantee the successful long-term development of the wireless communications industry.

IV. Conclusions

In this paper we have emphasised the importance of power efficiency considerations in the context of the design and optimization of wireless communication networks. We provided formal definitions of power as well as bandwidth efficiencies and explored their relative importance using a detailed top-down analysis of a typical commercial wireless network. Our results suggest that no substantial area spectral efficiency gains beyond those exhibited by the 2G cellular technology may be economically achieved. Moreover, no significant economic gains may be realized by simply increasing the bps/Hz bandwidth efficiency. In contrast, manifold utility and throughput gains as well as substantial Quality of Service (QoS) enhancements may be attained by the appropriate combination of improved power efficiency, bandwidth expansion as well as the conclusive evolution of the networking paradigm.

References

[1] G. Gow and R. Smith, *Mobile and Wireless Communications: An Introduction*. Open University Press/McGraw-Hill, 2005.

[2] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218–1230, 1997.

[3] L. Hanzo, J. Blogh, and S. Ni, *3G, HSDPA, HSUPA and FDD Versus TDD Networking: Smart Antennas and Adaptive Modulation*. IEEE Press-John Wiley, February 2008.

[4] L. Hanzo, J. Akhtman, and M. Jiang, *MIMO-OFDM Turbo-Transceivers for LTE, WIFI and WIMAX*. IEEE Press-John Wiley, 2008, under Review.

[5] K. Werbach, "Open spectrum: The new wireless paradigm," New America Foundation Working Paper, October 2002. [Online]. Available: http://werbach.com/docs/new_wireless_paradigm.htm

[6] A. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading MA, USA: Addison-Wesley, June 1995, iSBN 0201633744.

[7] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.

[8] L. Hanzo, L.-L. Yang, E.-L. Kuan, and K. Yen, *Single- and Multi-Carrier DS-CDMA*. John Wiley and IEEE Press, 2003, 430 pages.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, June and October 1948. [Online]. Available: http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[10] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, December 2008.

[11] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. John Wiley & Sons, 2004, 1036 pages.

[12] J. Li, A. Bose, and Y. Zhao, "Rayleigh flat fading channels' capacity," in *Proceeding of the 3rd Annual Communication Networks and Services Research Conference*, May 2005, pp. 214–217.

[13] W. C. Y. Lee, "Spectrum efficiency in cellular [radio]," *IEEE Transactions on Vehicular Technology*, vol. 38, no. 2, pp. 69–75, May 1989.

[14] R. Steele, C.-C. Lee, and P. Gould, *GSM, cdmaOne and 3G Systems*. John Wiley and Sons, 2001.

[15] Ericsson, "Sustainable energy use in mobile communications, White paper," December 2007. [Online]. Available: http://www.ericsson.com/technology/whitepapers

[16] DBIRR, "Quarterly energy prices," Department of Business Interprise & Regulatory Reform, A National Statistics Publication, Tech. Rep., March 2008. [Online]. Available: http://www.berr.gov.uk/files/file45393.pdf

[17] K. Werbach, *Radio Revolution: The Coming Age of Unlicensed Wireless*. New America Foundation, 2003. [Online]. Available: http://werbach.com/docs/RadioRevolution.pdf