Adaptive Management of Network Resources

Daniel C. ONeill, David Julian, and Stephen Boyd
Department of Electrical Engineering
Stanford University
dconeill,djulian,boyd@stanford.edu

***Abstract*— This paper describes a new adaptive algorithm that smoothly and dynamically adjusts the system resources of link rates and transmitter powers to maximize the performance of the system. Performance is explicitly measured from the point of view of traffic carried by the network. Transmitter powers are subsumed in the feasible rate region for the wireless network, and are not directly involved in evaluating the network. A new adaptive algorithm, DSM, is presented. DSM seeks optimal system performance by greedily searching the rate region surface seeking link rates that best meet QoS and user demand needs and then calculates transmitter powers to support these rates. If system requirements such as the number of users or their QoS change, the DSM adapts by again exploring the now changed rate region. Changes in the wireless environment are addressed by the algorithm in a similar fashion.**

***Index Terms*— Power control, rate control, utility functions, optimization and control, adaptation**

I. Introduction

This paper describes a new adaptive algorithm that smoothly and dynamically adjusts the system resources of link rates and transmitter powers to maximize the performance of the system. The approach is new; it seeks link rates that best meet QoS and user demand needs and then calculates transmitter powers to support these rates. As the array of wireless services grows, wireless networks will need to automatically adjust for changing conditions. The network will need to smoothly adjust to changes in types of traffic carried by the network, the number and types of users, and changes in the wireless environment. The system will need to adjust the network resources of link rates and transmitter powers in response to these changes. Ideally a network or cell site must also allocate system capacity in such a way so as to maximize the the system performance.

This paper describes a new adaptive algorithm that smoothly and dynamically adjusts the system resources of link rates and transmitter powers to maximize the performance of the system. The approach is new; it seeks link rates that best meet QoS and user demand needs and then calculates transmitter powers to support these rates. The algorithm continuously searches over the set of feasible rates to find the optimal rates and consequently adapts to changes in network conditions such as the number or types of users on the system, changing QoS constraints, or link rate requirements. Different measures of performance can be chosen. Protocol based performance metrics, user utility functions, and others can be combined and used with this approach. Constraints on minimum link/user rates can be addressed and satisfied if feasible.

II. System Model

The wireless network has L links and S sources. Each link $l \in L$ uses a CDMA transmission scheme and shares a common bandwidth. Each link transmits at rate R_l, termed the link rate; this is the maximum rate at which traffic can be carried over the link. Each source $s \in S$ sends packets to its sink by injecting them into the network at rate r_s, termed the transfer rate. Packets traverse the network by travelling over a single link or hop. The approach is extended to a multi-hop wireless network in [1].

The wireless link rate is a function of the links Signal to Interference ratio, SIR,

$$\rho_l = \frac{G_{ll} p_l}{\sum_{i \neq l} G_{li} p_i}. \qquad (1)$$

where G_{ll} represents The effective gain between the transmitter and receiver on link l and includes the multiplicative spreading gain K_{s}, antenna gain, coding gain, and other gain factors. Likewise G_{lj} represents the effective gain from the interfering transmitter on link j to the receiver on link l. The gain matrix G with elements G_{ij} is assumed to be positive in what follows. The transmitter power on the lth link is denoted by p_l.

Because noise is neglected in this model, the set of powers can be arbitrarily scaled without effecting ρ_l; That is, SIR is homogeneous of order zero in the transmitter powers. By choosing to scale the sum to one, $\mathbf{1}^T p = 1$, the p_l can be interpreted as representing the relative powers of the transmitters or equivalently the percent of total power transmitted by the system. In most systems $\rho_l \gg 1$ since it represents the effective SIR after spreading gain, antenna gain, and coding gain.

0-7803-7954-3/03/$17.00

The link rate model is a simplification of an empirically based model [2]. The relationship between link rates and SIR is given as

$$R_l = \lg(\rho_l) \tag{2}$$

for a fixed bit error rate, BER. The terms $G_{ll}\ \forall l$ are scaled by a positive constant to reflect the chosen BER.

To be feasible the traffic carried over a link must be less than its link rate

$$\sum_{s\in\theta(l)} r_s \leq R_l. \tag{3}$$

where $\theta(l)$ is the set of sources s using link l. Combining equations 2 and 1 yields

$$\begin{array}{rcl} r \leq R & \Leftrightarrow & p_l \geq \frac{e^{r_l}}{G_{ll}} \sum_{j\neq l} G_{lj} p_j \\ & \Leftrightarrow & p \geq D(r)\tilde{G}p \end{array} \tag{4}$$

where $D \triangleq \text{diag}\left(\frac{e^{r_l}}{G_{ll}}\right)$ and

$$\tilde{G}_{ij} = \begin{cases} G_{ij}, & i \neq j \\ 0, & i = j. \end{cases} \tag{5}$$

III. Rate-Region

The rate-region is the set of feasible transfer rates $r \in \mathbf{R}^n_+$ for the system. A transfer rate $r \in \mathbf{R}^n_+$ is feasible if it is possible for the system to simultaneously transfer data over the network at the specified rates for some power vector p. By considering all power allocations the set of feasible transfer rates can be found. Analytically the rate-region can be described as

$$\mathcal{R} = \{r \in \mathbf{R}^n_+ | r \leq R(p) \text{ for some } p\}, \tag{6}$$

where $r \leq R$ for two vectors means component-wise inequality, i.e. $r_l \leq R_l$ for all l.

The rate-region $\mathcal{R}$ is convex [3]. This is shown by first defining the set of feasible transfer rate and power pairs (r,p), demonstrating its convexity, and projecting it onto the rate transfer space. The set of feasible transfer rate and power pairs, $\mathcal{M}$, is the set of (r,p) such that $r_l \leq \log(\rho_l)$ for all links l. Analytically,

$$\begin{array}{rcl} \mathcal{M} & = & \{(r,p) \in \mathbf{R}^{2n}_+ | r_l \leq \log(\rho_l), \forall l\} \\ & = & \bigcap_l \{(r,p) \in \mathbf{R}^{2n}_+ | r_l \leq \log(\rho_l)\} \\ & = & \bigcap_l \mathcal{M}_l. \end{array} \tag{7}$$

The $\mathcal{M}_l = \{(r,p) \in \mathbf{R}^{2n}_+ | r_l \leq \log(\rho_l)\}$ are convex. This can be seen by the change of variables $x_l = \log p_l$ and rewriting the set qualifier as follows:

$$\begin{array}{rcl} r_l \leq \log(\rho_l) & \Leftrightarrow & e^{-r_l} \geq \rho_l^{-1} \\ & \Leftrightarrow & e^{-r_l} \geq \sum_{j\neq l} G_{lj} e^{x_j} G_{ll}^{-1} e^{-x_l} \\ & \Leftrightarrow & 1 \geq \sum_{j\neq l} G_{lj} e^{x_j} G_{ll}^{-1} e^{-x_l} e^{r_l} \\ & \Leftrightarrow & 0 \geq \log(\sum_{j\neq l} G_{lj} e^{x_j} G_{ll}^{-1} e^{-x_l} e^{r_l}). \end{array} \tag{8}$$

It is known [4] that the function $\log(\sum \alpha_l e^{y_l})$, for $\alpha_l \in \mathbf{R}_+$ and $y_l \in \mathbf{R}$, is convex in y. Sub-level sets of convex functions always define convex sets, so equation 8 defines a convex set in the variables $\log p_l$ and r_l. Since the intersection of convex sets is convex $\mathcal{M}$ must also be convex.

The rate-region $\mathcal{R}$ is a projection of $\mathcal{M}$ onto the transfer rate space. Since linear projection conserves convexity, the rate-region $\mathcal{R}$ must also be convex.

IV. Performance

In this paper network performance is measured from the point of view of the network's input/output ports. This is the performance as seen by packets traversing the network, end to end network protocols such as TCP, or users through their individual utility functions [5], [6], [7]. Transmitter power p is not viewed as a direct measure of system performance, but rather effects system performance indirectly through link and transfer rates.

System performance is modelled by a performance metric or utility function U The Utility function is assumed to be a function of the transfer rates r, $U = U(r)$. By assumption, a higher data rate is valued at least as much as a lower data rate, so U is a nondecreasing function of r. Also, by assumption, there is a diminishing return to additional transfer data rate, so U is a concave function of r. The performance metric can be expanded to directly include measures of routing for multi-hop networks or to include individual transmitter powers, although this is not done in this paper due to space limitations.

The measure of system performance can be any increasing concave function, but for simplicity it is assumed to be the weighted sum of each users' utility function, $\sum \alpha_i U_i(r_i)$.

Because U is nondecreasing it can be shown that the system's best performance occurs on the surface of the rate region. This can be seen by first assuming that the best set of transfer rates are in the interior of the region, and then noting that by moving to the surface of the region, an r can be found that has equal or better performance. The surface of the region is denoted by $\mathcal{P}$.

V. Problem Statement

Formally, the problem is to find the best set of rates, r, and powers, p, such that the system performance is maximized.

$$\begin{array}{ll} \text{maximize} & \sum \alpha_s U_s(r_s) \\ \text{subject to} & p = DGp \\ & p > 0. \end{array} \tag{9}$$

The constraint is the surface of the feasible rate region and can be rewritten using Perron Frobenius theory [8] as

$$\lambda_{\mathrm{pf}}(D(r)\tilde{G}) = 1, \quad (10)$$

where $\lambda_{\mathrm{pf}}(D(r)\tilde{G})$ is the Perron Frobenius eigenvalue for the matrix DG.

VI. Optimality Criterion

By *Lagrange's Theorem* [9], at optimality

$$\nabla_r U(Br) = K\nabla_r \lambda_{\mathrm{pf}}(D(r)\tilde{G}) \quad (11)$$

where K is a constant of proportionality. In words, the gradient vectors ∇U and $\nabla\lambda_{\mathrm{pf}}$ must be parallel at the optimal rates. The gradient to the rate surface $\nabla_r \lambda_{\mathrm{pf}}(D(r)\tilde{G})$ is

$$N(r) = \nabla\lambda_{\mathrm{pf}}(D(r)\tilde{G}) = [q_1p_1, q_2p_2, \ldots, q_np_n]^T, \quad (12)$$

where p, the transmitter powers, is found to be the right Perron Frobenius eigenvector. The left eigenvector q models interference. It scales each link's transmitter power and can be thought of as the marginal effect of interference on transmitter power and link rate.

VII. DSM Algorithm

The DSM algorithm proposed in this paper is intuitive; it moves along the surface of the feasible rate region until an optimal point is reached. If the system changes, for example users are added to the network or the gain matrix G fades, the algorithm again moves to a new point that is optimal for this new demand on the system. The direction of movement is determined by comparing the vector normal to the rate region to a second vector normal to the objective function. Optimality is achieved when the two vectors are parallel. At such a point the trade-offs of moving in one direction verses another are identical based on the system performance measure. If the system subsequently changes in some way and the current operating point is no longer optimal the algorithm adapts by again seeking on optimal operating point. The transmitter powers are calculated from the vector r using $p = D(Ar)\tilde{G}p$. An alternative is the method proposed by G. Foschini and Z. Miljanic [10].

A. Feasible ascent direction

The Direct Step Method is shown in Figure 1 Let $N(r)^{\perp} = \left\{r' | (r - r')^T N(r) = 0\right\}$ be the hyper-plane that is tangent to the rate region surface at r_c. Since the rate-region $\mathcal{R}$ is convex, $N(r)^{\perp}$ is a supporting hyper-plane and lies outside of $\mathcal{R}$, except at the point r_c. The supporting hyper-plane $N(r)^{\perp}$ is a good approximation of $\mathcal{P}$ for small changes in r. For this reason a direction δr is defined to be *feasible* if it lies along $N(r)^{\perp}$, or equivalently $\delta r^T N(r) = 0$.

A small change δr is defined as an *ascent* direction if $U(r + \alpha\delta r)$ increases for small $\alpha > 0$. Thus, δr is an ascent direction if and only if $\nabla U(r)^T \delta r > 0$. A point that is both feasible and an ascent direction is termed a *feasible ascent* direction.

B. DSM

The DSM algorithm is a two phase feasible ascent method. In the predictor phase a small feasible change or step δr is calculated. In the corrector phase this point is corrected to lie along $\mathcal{P}$. The method can be described as a simple two step algorithm. Given the current operating point $r_c(t)$

Algorithm 1: DSM

- Calculate a feasible ascent direction δr and predict a new operating point $r_p(t+1) = r_c(t) + \beta\delta r$.
- Correct this estimate by scaling it onto $\mathcal{P}$, $r_c(t+1) = \alpha r_p(t+1)$. Repeat.

End Algorithm

1) DSM: Predictor: The algorithm constructs a δr from a measure of the sub-optimality of the system. The error estimate is defined as

$$e = \left(\frac{N(r)}{\mathbf{1}^T N(r)} - \frac{\nabla U}{\mathbf{1}^T \nabla U}\right) \quad (13)$$

and compares the normal to the rate region to the normal to the performance metric. At optimality $e = 0$ and the operating point r remains fixed.

Because U is concave and λ_{pf} is convex, a rate change δr_s causes the performance metric normal and rate region normal to respond in opposite ways; an increase $\delta r_{\mathrm{ds},s} > 0$ causes the sth component of $\frac{\nabla U}{\mathbf{1}^T\nabla U}$ to decrease and the comparable component of $\nabla\lambda_{\mathrm{pf}}$ to increase. Consequently, the decision to increase the sth component of δr_{ds} can be made by comparing the two normals. If $\frac{\nabla U}{\mathbf{1}^T\nabla U}$ is greater than $\nabla\lambda_{\mathrm{pf}}$, then the associated rate should be increased. Specifically, for small rate adjustments δr_{ds} should have the same sign as $-e$. The DSM uses this information to find a δr_{ds} that is an ascent direction but which is also feasible by construction. Specifically,

$$\delta r_{\mathrm{ds}} \quad = \quad -\,\mathbf{diag}(\tfrac{1}{q_l p_l}, \ldots)e \quad (14)$$

Substituting e yields

$$\delta r_{\mathrm{ds,s}} \quad = \quad -\left(\tfrac{1}{q_s p_s}\right) \left(\tfrac{N(r)}{\mathbf{1}^T N(r)} - \tfrac{\nabla U}{\mathbf{1}^T\nabla U}\right)_s. \quad (15)$$

That δr_{ds} lies on the supporting hyper-plane can be

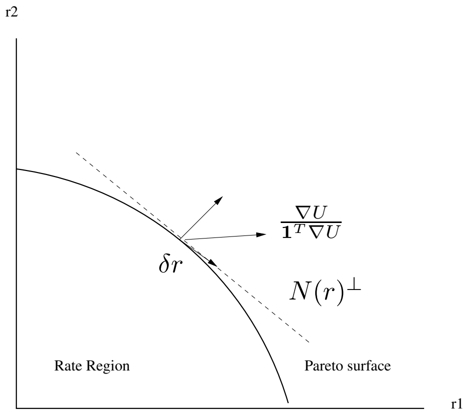

Fig. 1. DSM predictor phase. The optimality error e is used to predict a change rates δr that improves ad-hoc network performance.

seen from

$$\begin{aligned} N(r)^T \delta r_{\text{ds}} &= \sum q_i p_i \left(\frac{e_i}{q_i p_i}\right) \\ &= \sum e_i \\ &= \mathbf{1}^T e \\ &= \mathbf{1}^T N(r) - \mathbf{1}^T \nabla U / \mathbf{1}^T \nabla U \\ &= 0 \end{aligned} \tag{16}$$

where it can be shown that by construction $\mathbf{1}^T N(r) = 1$.

A new rate is calculated as $r_p(t+1) = r_c(t) + \beta \delta r_{\text{ds}}$, where $\beta << 1$. This rate lies along the supporting hyperplane $M(r)^\perp$, but, unless this is the optimal operating point, is not on $\mathcal{P}$.

2) DSM: Corrector: The estimated rate $r(t+1)$ is corrected to lie on $\mathcal{P}$ using a scaling method. The scaling method scales each element in the estimated rate vector by a constant α_a.

The scaling method multiplies each element in the rate vector r_p by a fixed scalar $r_c = \alpha_p r_p$, $\alpha_p > 0$ to find a rate vector $r_c \in \mathcal{P}$. Increasing α increases all rates $r_c = \alpha r_p$, and in turn increases the elements of $D(r_c)\tilde{G}$. By the monotone property for the Perron Frobenius eigenvalue, $\lambda_{\text{pf}}(D(\alpha r_p)\tilde{G})$ also increases and is monotonic in α.This leads to a bisection algorithm to find α_p.

The bisection algorithm increases α linearly until $\lambda_{\text{pf}}(D(\alpha r_p)\tilde{G}) \geq 1$, so α_p lies between zero and α. Next $\lambda_{\text{pf}}(D(\frac{1}{2}\alpha r_p)\tilde{G})$ is computed and compared with one; if it is greater than one then $\alpha_p \in [0, \alpha/2]$ while if it is less than one then $\alpha_p \in [\alpha/2, \alpha]$. If $\alpha_p \in [\alpha/2, \alpha]$, then $\lambda_{\text{pf}}(D(\frac{3}{4}\alpha)\tilde{G})$ is computed and compared with one to again reduce the range containing α_p by half. The segment that α_p lies in is reduced through repeated bisections until α_p is known to the desired number of decimal points.

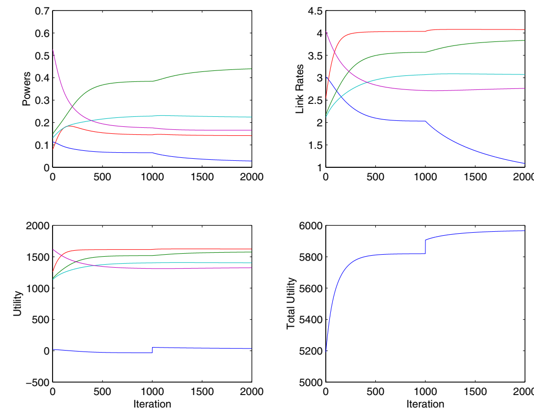

Fig. 2. Performance with rate constraints.

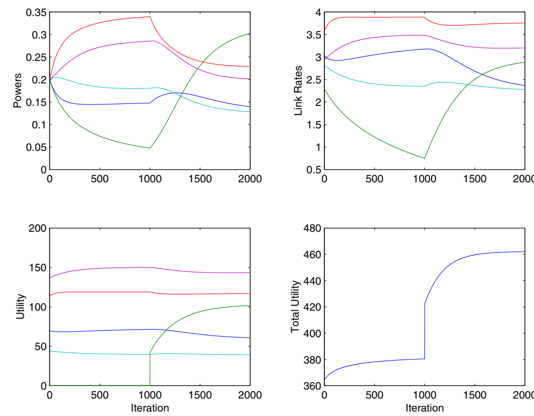

Fig. 3. Performance with a new user.

VIII. Simulation

This section uses the DSM to find optimal link rates, transmitter powers and system performance measures as the system adapts to changes in its operating requirements. Similar behavior can be shown for changes in the wireless transmission environment such as Rayleigh fading or log-normal shadowing.

Two different simulations are presented. In the first simulation, the network seeks optimal performance subject to a rate floor for a particular link. This link might be thought off as a voice only link that requires a minimum data rate irrespective of the impact this might have on other links. The second simulation explores the adaption of the DSM with the introduction of a new user to the system.

The model is of a 5 link single hop network. The performance metric corresponds to the sum of individual user utility functions and is given by

$$U(r) = \sum_{s=1}^{5} a_s \log(r_s) + \Delta \log(r_s - r_{\text{th},s}), \tag{17}$$

where the first term in the sum is the utility associated with a given rate r_s and the second term acts as a barrier limiting this rate to $r_s > r_{\mathrm{th},s}$. For the utility function, the natural log is used. The a_s are scale constants associated with different users, and the constant $\Delta \ll a_s \; \forall s$. The barrier portion of the individual utility functions is negligible for $r_s > r_{\mathrm{th},s}$ but dominates for $r_s \sim r_{\mathrm{th},s}$, preventing the link rate from dropping below the threshold. The scale factor $\Delta = 0.0001$. The DSM algorithm is used with parameter $\beta = .0001$.

For the simulation, the gain matrix is

$$G = \begin{bmatrix} 144.1 & 0.217 & 0.311 & 0.068 & 0.617 \\ 0.469 & 83.0 & 0.307 & 0.125 & 0.269 \\ 0.537 & 0.053 & 120.5 & 0.166 & 0.221 \\ 0.563 & 0.229 & 0.954 & 144.3 & 0.713 \\ 0.511 & 0.167 & 0.131 & 0.136 & 108.2 \end{bmatrix}. \tag{18}$$

The initial conditions are

$$\begin{aligned} a &= [\; 0.1 \quad 1 \quad 3 \quad 5 \quad 7 \;]^T \\ r_{\mathrm{th}} &= [\; 2 \quad 0 \quad 0 \quad 0 \quad 0 \;]^T \\ p &= [\; 1 \quad 0.01 \quad 0.01 \quad 0.01 \quad 0.01 \;]^T \end{aligned} \tag{19}$$

where a are the performance measure weights, r_{th} the threshold values, and p the initial link transmitter powers. The value of $a_1 = 0.1$ might be thought of as a user whose evaluation of the performance of the system is relatively unimportant. This might happen for pricing or priority reasons.

Figure 2 depicts the first simulation. The four panels show, in clock wise order, link transmitter powers, the associated link rates, the total utility of the system, and the performance of the system as seen by each of the five sources.

During the first 1000 time periods the network evolves from the arbitrary initial conditions and seeks the best set of link rates and transmitter powers to maximize system performance but subject to the constraint that the first link has a minimum rate of 2, that is $r_{\mathrm{th},1} = 2$. The initial rate on link 1 is in excess of the rate that is optimal for the network; by reducing this rate other links can increase their rates, improving overall network performance. Consequently the DSM is shown reducing the transmitter power of this link and increasing the others. When the link rate on link 1 reaches 2, this trade-off ceases and the system remains in rate equilibrium.

At time period 1001 the rate floor is removed. The DSM adapts by further reducing the rate on link 1 and increasing it on the remaining links. Overall system performance and utility improves as intuition would suggest.

In the second simulation the number of users changes. This is shown in Figure 3. At time 1 a data source leaves the system only to return at time 1001. The system evolves from initial conditions similar to those of equation 19 , but with all rate floors set to zero. As can be seen, the DSM adapts to the departure of user 1 at time 0 by decreasing the rate and power on link 1 and increasing power on other links to increase their link and transfer rates. The weighting of the performance measures prioritizes the allocation of additional power at each step of DSM's evolution of the system. Performance improves monotonically with this reallocation of resources. At time 1001 the user returns to the network. The DSM adapts to this change by incrementally increasing the rate on this link until a new optimum is obtained. As this rate is increased the SIR of the other links worsens, and the algorithm reallocates the powers among these links to maximize system performance. As can be seen in the lower two panels, overall system performance improves monotonically, but the performance of links 2-5 actually decreases as the performance of user 1 increases. Rate floors can be added to this formulation, reflecting more realistic constraints on reallocating bandwidth among users.

IX. Summary

This paper presents a new approach to adaptively manage the resources of link rate and transmitter power in a wireless network. The algorithm smoothly moves along the feasible rate surface in order to find the best set of rates to meet the changing demands on the system, but consistent with and QoS established for each user. The associated powers are then found from Perron Frobenius or several other power calculating algorithms.

References

[1] D. ONeill, D. Julian, and S. Boyd, "Optimal routes and flows in congestion constrained ad hoc networks," in *Submitted to Infocom 2004*, march 2004.

[2] D. Julian, M. Chiang, and D. ONeill, "Qos and fairness constrained convex optimization of resource allocation for wireless cellular and ad-hoc networks," in *Proc. IEEE Infocom*, June 2002, pp. 477–486.

[3] D. ONeill, D. Julian, and S. Boyd, "Seeking foschini's genie: Optimal rates and powers in wireless networks," *submitted to IEEE Trans. Vehicular Technology*, 2003.

[4] S. Boyd and L. Vandenberghe. (2003) Convex optimization. [Online]. Available: http://www.stanford.edu/ boyd/cvxbook.html

[5] S. Low, L. Peterson, and L. Wang, "Understanding vegas: A duality model," *Journal of ACM*, vol. 49(2), pp. 207–235, Mar. 2002.

[6] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communications networks: Shadow prices, proportional fairness and stability," *Journal of Operations Research Society*, vol. 49, pp. 237–252, Nov. 1998.

[7] S. Low. (2002) Duality model of tcp and queue management algorithms. [Online]. Available: http://netlab.caltech.edu

[8] A. Berman and R. Plemmons, *Nonnegative Matricies in the Mathematical Sciences*. Philadelphia, PA: SIAM, 1994.

[9] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[10] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Vehicular Technology*, vol. 42(4), pp. 641–646, Nov. 1993.