

Approaching the MIMO Capacity with a Low-Rate Feedback Channel in V-BLAST

Seong Taek Chung

STAR Laboratory, Stanford University, Stanford, CA 94305-9515, USA
Email: stchung@dsl.stanford.edu

Angel Lozano

Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: aloz@lucent.com

Howard C. Huang

Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: hchuang@lucent.com

Arak Sutivong

Information Systems Laboratory, Stanford University, Stanford, CA 94305-9510, USA
Email: arak@stanfordalumni.org

John M. Cioffi

STAR Laboratory, Stanford University, Stanford, CA 94305-9515, USA
Email: cioffi@stanford.edu

Received 8 December 2002; Revised 30 October 2003

This paper presents an extension of the vertical Bell Laboratories Layered Space-Time (V-BLAST) architecture in which the closed-loop multiple-input multiple-output (MIMO) capacity can be approached with conventional scalar coding, optimum successive decoding (OSD), and independent rate assignments for each transmit antenna. This theoretical framework is used as a basis for the proposed algorithms whereby rate and power information for each transmit antenna is acquired via a low-rate feedback channel. We propose the successive quantization with power control (SQPC) and successive rate and power quantization (SRPQ) algorithms. In SQPC, rate quantization is performed with continuous power control. This performs better than simply quantizing the rates without power control. A more practical implementation of SQPC is SRPQ, in which both rate and power levels are quantized. The performance loss due to power quantization is insignificant when 4–5 bits are used per antenna. Both SQPC and SRPQ show an average total rate close to the closed-loop MIMO capacity if a capacity-approaching scalar code is used per antenna.

Keywords and phrases: adaptive antennas, BLAST, interference cancellation, MIMO systems, space-time processing, discrete bit loading.

1. INTRODUCTION

Information theory has shown that the rich-scattering wireless channel can support enormous capacities if the multipath propagation is properly exploited, using multiple transmit and receive antennas [1, 2, 3]. In order to attain the closed-loop multiple-input multiple-output (MIMO) capacity, it is necessary to signal through the channel's eigenmodes with optimal power and rate allocation across those modes [4, 5]. Such an approach requires instantaneous chan-

nel information feedback from the receiver to the transmitter, hence a closed-loop implementation. Furthermore, a very specialized transmit structure is required to perform the eigenmode signaling. Therefore, it is challenging to incorporate the closed-loop MIMO capacity-achieving transmit-receive structures into existing systems.

Open-loop schemes that eliminate the need for instantaneous channel information feedback at the transmitter have also been proposed [6, 7, 8, 9, 10, 11]. These schemes can be divided into two categories: multidimensional coding

(e.g., space-time coding) and spatial multiplexing (e.g., vertical Bell Laboratories layered space-time (V-BLAST)). Multidimensional coding [7] requires very specialized coding structures and complicated transceiver structures. Furthermore, its complexity grows very rapidly with the number of transmit antennas. Among spatial multiplexing approaches, V-BLAST [9, 10, 11] uses simple scalar coding and a well-known transceiver structure. This paper focuses on the V-BLAST transmission scheme.

In V-BLAST, every transmit antenna radiates an independently encoded stream of data. This transmission method is much more attractive from an implementation standpoint; the transmitter uses a simple spatial demultiplexer followed by a bank of scalar encoders, one per antenna. The receiver uses a well-known successive detection technique [12]. Furthermore, this scheme is much more flexible in adapting the number of antennas actively used. This flexibility is a strong advantage for the following reasons. First, the channel estimation process requires more time as the number of transmit antennas increases; consequently, the overall spectral efficiency—including training overhead—could actually degrade with an excessive number of transmit antennas in rapidly fading channels. Hence, MIMO systems may need to adapt the number of antennas actively used depending on the environment. Second, it is expected that during initial deployment, not all base stations and terminal units may have the same number of antennas. Therefore, the number of antennas actually being used may need to be adapted, for example, during hand-off processes between different cells.

As previously mentioned, the main weakness of open-loop V-BLAST is that it attains a part of the closed-loop MIMO capacity; as the transmitter cannot adapt itself to the channel environment in an open-loop fashion, V-BLAST simply allocates equal power and rate to every transmit antenna. Consequently, the performance is limited by the antenna with the smallest capacity, as dictated by the channel. Hence, it is natural to consider per-antenna rate adaptation using a low-rate feedback channel.

Using a low-rate feedback channel, [13] introduced rate adaptation at each antenna in V-BLAST to overcome this problem. We extend their approach to both rate and power adaptations at each antenna and theoretically prove that this new scheme, denoted as V-BLAST with per-antenna rate control (PARC), achieves the performance of an open-loop scheme with multidimensional coding. A similar approach was taken at OFDM/SDMA in the downlink of wireless local networks [14]. We show that with per-antenna rate and power control, V-BLAST achieves higher performance than the other open-loop schemes. Moreover, V-BLAST with PARC attains the open-loop MIMO capacity.

In developing the optimal PARC, similarities are noted between the V-BLAST with PARC and the Gaussian multiple-access channel (GMAC) problems. Every transmit antenna within the V-BLAST can be regarded as an individual user in a GMAC. As shown in [15], with optimum successive decoding (OSD), the total sum capacity of the GMAC can be achieved at any corner point of the capacity region. As

will be shown, this result translates directly to the V-BLAST context by simply incorporating the notion of PARC.

Next, these theoretical results are applied to practical modulation scenarios. In order to apply the idealized capacity results to a real system, the following points should be considered. First, the idealized results assume an infinite-length codebook to achieve vanishingly small bit error rates (BERs), but in a real system, current coding techniques and practical system requirements allow only for a finite-length coding with nonzero error rates [16]. Second, the idealized results assume a continuous rate set, but in a real system, only rates from a discrete rate set are feasible.

The first issue can be easily solved by adopting the concept of a gap (Γ) [17]:

$$b = \log_2 \left(1 + \frac{\text{SINR}}{\Gamma} \right). \quad (1)$$

The number of bits transmitted at a specific SINR and specific coding and BER can be expressed as (1), where b is the number of bits transmitted per symbol, SINR is the signal-to-interference-and-noise ratio, and Γ is a positive number larger than 1, which is a function of the BER and specific coding method. Note that this is a capacity expression, except that the SINR is scaled by a penalty Γ , which is a function of the target BER and coding method. Γ can take various values; for uncoded M-QAM with the target BER 10^{-3} , Γ is 3.333 (5.23 dB). For a very powerful code (e.g., Turbo code), Γ is close to 1 (0 dB). When Γ equals 0 dB, the gap expression (1) equals the actual capacity [17]. Works in [13] also utilize the gap expression in considering the rate adaptation per antenna.

The second issue is investigated using ad hoc methods since the optimal solution for discrete rates is difficult to obtain analytically. Successive quantization with power control (SQPC) is first proposed. Here, the rate is quantized efficiently with continuous power control. However, a continuously variable transmit power level can be impractical since the feedback channel data rate is limited. Therefore, SQPC is extended to successive rate and power quantization (SRPQ) by considering power level quantization as well.

The organization of this paper is as follows. The system model is introduced in Section 2. V-BLAST is specifically described in Section 3, with optimal PARC, when the transmit antenna powers are given. The antenna power allocation that maximizes the capacity is derived in Section 4. Section 5 shows that the open-loop capacity can be approached using V-BLAST with equal power allocation; additional power control only leads to a slight increase in capacity. Section 6 first suggests a simple discrete bit loading algorithm based on rounding off the rate from a continuous set with equal power allocation. Then, a new discrete bit loading is presented along with continuous power control, SQPC, in Section 7. In Section 8, a discrete bit loading with quantized power levels, SRPQ is suggested. Results are shown in Section 9. Conclusions follow in Section 10.

2. SYSTEM MODEL

We assume a general architecture with M transmit and N receive antennas and perfect channel estimation at the receiver. Rate and/or power information can be fed back to the transmitter. The $M \times 1$ transmit signal vector is \mathbf{x} ; the $N \times 1$ received signal vector is \mathbf{y} . The $N \times M$ channel matrix \mathbf{H} can take any value; however, for a rich scattering environment, we assume that \mathbf{H} is composed of independent zero-mean complex Gaussian random variables. The zero-mean additive white Gaussian noise (AWGN) vector at the receiver, denoted by \mathbf{n} , has a covariance matrix equal to the identity matrix scaled by σ^2 . For simplicity, we assume $\sigma^2 = 1$ and scale the channel appropriately. The average power of each component of the \mathbf{H} matrix is indicated by g , while the total power available to the transmitter is denoted by P_T . An average SNR ρ is defined as $P_T g$.

This model can be expressed mathematically as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (2)$$

where $E[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_N$ and $E[H(n_1, m_1)^* H(n_2, m_2)] = g\delta(n_1 - n_2, m_1 - m_2)$ for all n_1, n_2, m_1 , and m_2 . \mathbf{I}_N denotes the identity matrix of size $N \times N$, $\delta(m, n)$ denotes the 2-dimensional Kronecker delta function, and $H(n, m)$ indicates the n th row and m th column element of the \mathbf{H} matrix. Consistent with the open-loop V-BLAST concept, the signals radiated from different antennas are independent. Hence, the covariance matrix of \mathbf{x} can be expressed as follows when the power allocated to antenna m is equal to P_m :

$$E[\mathbf{x}\mathbf{x}^H] = \begin{bmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & P_{M-1} & 0 \\ 0 & 0 & \cdots & 0 & P_M \end{bmatrix}, \quad (3)$$

where $\sum_{m=1}^M P_m = P_T$. When we simply allocate equal power to all the transmit branches, we assign $P_m = P_T/M$. We use $(\cdot)^T$ and $(\cdot)^H$ to denote transposition and Hermitian transposition, respectively. For scalars, $(\cdot)^*$ denotes complex conjugate.

3. V-BLAST WITH PARC

With respect to minimum mean square error (MMSE) V-BLAST, the natural extension is PARC, which is explained in detail below.

The capacity of the m th transmit antenna C_m can be expressed in terms of the channel matrix and the transmit power of each antenna. We define \mathbf{h}_m as the m th column of \mathbf{H} and $\mathbf{H}(m)$ ($m = 1, \dots, M$) as the $N \times (M - m + 1)$ matrix $[\mathbf{h}_m \mathbf{h}_{m+1} \cdots \mathbf{h}_{M-1} \mathbf{h}_M]$. We also define $\mathbf{P}(m)$ as an $(M - m + 1) \times (M - m + 1)$ diagonal matrix with $(P_m, P_{m+1}, \dots, P_{M-1}, P_M)$ along the diagonal.

According to the OSD procedure described in [15], the signals radiating from the M transmit antennas are decoded

in any agreed-upon arbitrary order. In the remainder, it is assumed, without loss of generality, that they are decoded according to their index order. It is interesting to note that, unlike the open-loop V-BLAST, the ordering has no impact on the capacity attained by the sum of all M antennas.¹ It does, however, impact the fraction of that capacity that is allocated through rate adaptation to each individual antenna. It also affects the total rate when both rate and power are quantized.

The process is parameterized by a set of projection vectors \mathbf{F}_m ($m = 1, \dots, M$) and cancellation vectors $\mathbf{B}_{m1}, \mathbf{B}_{m2}, \dots, \mathbf{B}_{mm}$ ($m = 1, \dots, M - 1$), all with a dimension of $N \times 1$. In decoding the m th transmit antenna signal, interference from the $(m - 1)$ already decoded signals is subtracted from \mathbf{y} by applying the proper cancellation vectors to reencoded versions of their decoded symbols. An inner product of that cancellation process result and the projection vector corresponding to the m th antenna is fed into the m th antenna decoder.

The first antenna, in particular, is decoded based on Z_1 , which is obtained as the inner product of \mathbf{F}_1 and the receive vector $\mathbf{Y}_1 = \mathbf{y}$ expressed as $Z_1 = \langle \mathbf{F}_1, \mathbf{Y}_1 \rangle = \mathbf{F}_1^H \mathbf{Y}_1$. The decoded bits are reencoded to produce \hat{x}_1 . The second antenna is similarly decoded based on Z_2 , where Z_2 is now the inner product of \mathbf{F}_2 and a vector \mathbf{Y}_2 obtained by subtracting the vector $\mathbf{B}_{11}\hat{x}_1$ from \mathbf{y} . Therefore, $\mathbf{Y}_2 = \mathbf{y} - \mathbf{B}_{11}\hat{x}_1$ and $Z_2 = \langle \mathbf{F}_2, \mathbf{Y}_2 \rangle$. In general, the m th antenna is decoded based on $Z_m = \langle \mathbf{F}_m, \mathbf{Y}_m \rangle = \mathbf{F}_m^H (\mathbf{y} - \sum_{j=1}^{m-1} \mathbf{B}_{(m-1)j}\hat{x}_j)$. Here, it is assumed that all decoded bits are error-free, which is legitimate in the analysis of capacity [16].

The optimal cancellation vectors are given by $\mathbf{B}_{(m-1)j} = \mathbf{h}_j$, and the optimal projection vectors are $\mathbf{F}_m = (\mathbf{H}(m + 1)\mathbf{P}(m + 1)\mathbf{H}(m + 1)^H + \mathbf{I}_N)^{-1}\mathbf{h}_m$ [15].

Furthermore, the capacity of the m th antenna can be expressed as

$$C_m = \log_2 \left(1 + P_m \mathbf{h}_m^H (\mathbf{H}(m + 1)\mathbf{P}(m + 1) \times \mathbf{H}(m + 1)^H + \mathbf{I}_N)^{-1} \mathbf{h}_m \right) \quad (m = 1, \dots, M), \quad (4)$$

and it was proved in [15] that

$$\sum_{m=1}^M C_m = \log_2 \det (\mathbf{I}_N + \mathbf{H}\mathbf{E}[\mathbf{x}\mathbf{x}^H]\mathbf{H}^H), \quad (5)$$

which, with equal power per antenna, is precisely the open-loop MIMO capacity attainable with multidimensional coding [1]. Hence, the same capacity can be achieved using scalar coding, but at the expense of rate adaptation using a low-rate feedback channel. For a practical coding scheme with a nonzero BER, the rate R_m is expressed as follows, using (1)

¹It should be emphasized that this is true only in a capacity sense. In practice, due to error propagation, error rate performances can differ depending on the ordering.

and (4):

$$R_m = \log_2 \left(1 + \frac{P_m \mathbf{h}_m^H (\mathbf{H}(m+1) \mathbf{P}(m+1) \mathbf{H}(m+1)^H + \mathbf{I}_N)^{-1} \mathbf{h}_m}{\Gamma} \right) \quad (m = 1, \dots, M) \quad (6)$$

It is interesting to note that as the number of antennas grows large, the capacities C_m become increasingly predictable from the statistics of the channel, and hence the feedback need for each transmit antenna actually vanishes progressively [18].

4. POWER CONTROL IN V-BLAST WITH PARC

In this section, the power P_m ($m = 1, \dots, M$) allocation methods are considered under the total power constraint. For any set of powers P_m ($m = 1, \dots, M$), the optimal capacity and rate are those given by (4) and (6). The optimal power allocation scheme here is different from the waterfilling solution in [4].

4.1. Optimal scheme for $N = 1$ or $N = 2$

The optimal power control was found only when the number of receive antennas is 1 or 2. The optimal power allocation for more extensive cases was independently derived in [19].

When $N=1$, the open-loop MIMO capacity can be expressed as

$$C = \log_2 \left(\sum_{m=1}^M P_m |\mathbf{h}_m|^2 + 1 \right), \quad (7)$$

where \mathbf{h}_m is a scalar. Under the total power constraint, the optimal power allocation corresponds to assigning the entire power budget to the transmit antenna with the largest $|\mathbf{h}_m|$.

When $N=2$, following (5), the open-loop MIMO capacity can be expressed as

$$C = \log_2 \left[\left(\sum_{m=1}^M P_m |H(1, m)|^2 + 1 \right) \left(\sum_{m=1}^M P_m |H(2, m)|^2 + 1 \right) - \left(\sum_{m=1}^M P_m H(1, m)^* H(2, m) + 1 \right) \times \left(\sum_{m=1}^M P_m H(2, m)^* H(1, m) + 1 \right) \right]. \quad (8)$$

Under the total power constraint, the optimal power allocation can be found using a Lagrangian method:

$$J(P_1, \dots, P_M) = \left(\sum_{m=1}^M P_m |H(1, m)|^2 + 1 \right) \left(\sum_{m=1}^M P_m |H(2, m)|^2 + 1 \right) - \left(\sum_{m=1}^M P_m H(1, m)^* H(2, m) + 1 \right) \times \left(\sum_{m=1}^M P_m H(2, m)^* H(1, m) + 1 \right) + \lambda \left(\sum_{m=1}^M P_m - P_T \right), \quad (9)$$

where $J(P_1, \dots, P_M)$ is convex with respect to P_m . The optimal power allocation should satisfy the Karush-Kuhn-Tucker condition [20]; if the optimal power allocation P_m is positive for all $m = 1, \dots, M$, then the optimal power assignment policy is found from $\partial J / \partial P_l = 0$ ($l = 1, \dots, M$) and the total power constraint. $\partial J / \partial P_l = 0$ becomes

$$\sum_{m=1}^M P_m |H(1, l)H(2, m) - H(1, m)H(2, l)|^2 = \lambda - |H(1, l)|^2 - |H(2, l)|^2 \quad (l = 1, \dots, M). \quad (10)$$

If some P_m 's are zero in the optimal power allocation, then $\partial J / \partial P_l$ should be zero only for the nonzero P_l 's and the total power constraint should be satisfied. By checking this condition numerically, the optimal power allocation can be found. Simulation results are shown in Section 5.

4.2. Suboptimal scheme for $N > 2$

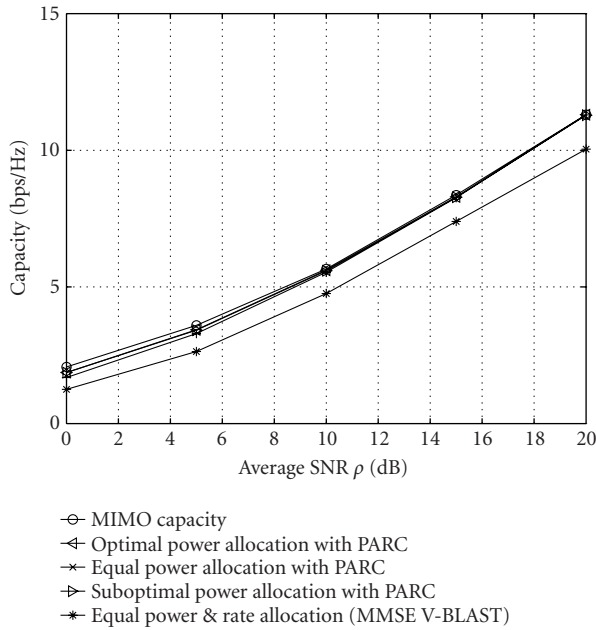
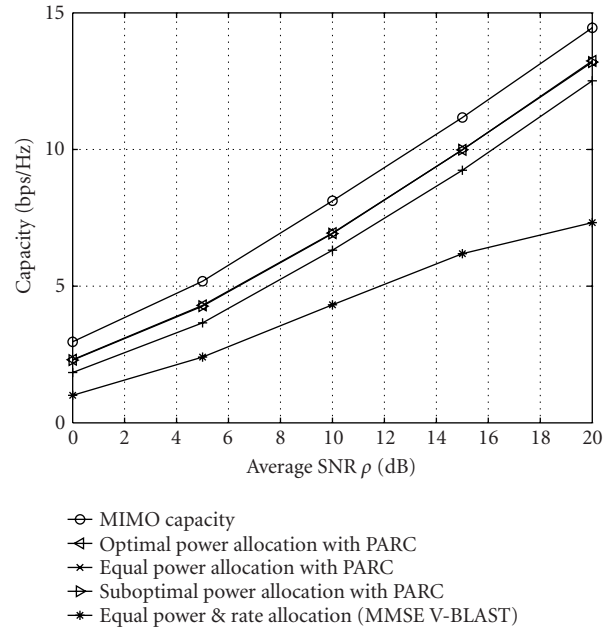
We were not able to find the optimal power and rate allocations when the number of receive antennas is more than 2. By solving the n th-order linear equations, we can get the optimal power solution, but obtaining a closed form, even for $N = 3$, is extremely complicated. However, from the optimal solution for $N = 1$ and $N = 2$, we observe the following:

- (i) the optimal power allocation scheme usually corresponds to selecting 1 or 2 antennas while switching off the remaining ones completely;
- (ii) with suboptimal power allocations (e.g., equal-power allocation), the capacity loss is small.

Based on these observations, we suggest a suboptimal power allocation algorithm that works for any combination of M and N . First, divide the total power P_T by M and consider P_T/M as a power unit. There are M such power units. Then, consider every possible power unit distribution over antennas, calculate the sum capacity (5) of each distribution, and select the one that yields the largest sum capacity of all the distributions.

5. CAPACITY RESULTS

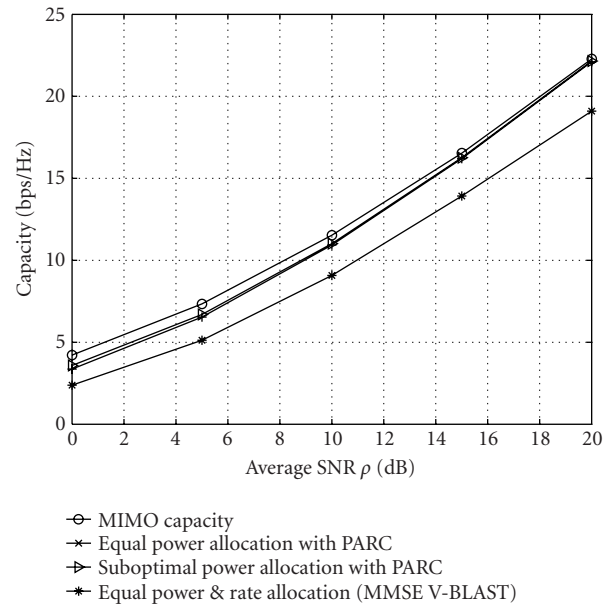
Numerical values for the capacity are shown in this section. Equation (1) is equivalent to the capacity formula for two dimensions when the gap (Γ) is 0 dB. The average (ergodic)

FIGURE 1: Average capacity when $M = 2$ and $N = 2$.FIGURE 2: Average capacity when $M = 4$ and $N = 2$.

capacity is used as a performance measure. We have also tested the outage capacity at small levels of outage, which shows a performance trend similar to that of the average capacity. Hence, the outage capacity results are not shown here.² Figures 1, 2, and 3 show such average capacity for various combinations of M and N . For each combination, the following cases are depicted: MIMO capacity, optimal power allocation with PARC, equal power allocation with PARC, suboptimal power allocation with PARC, and equal power and equal rate allocation. The MIMO capacity is the maximum rate achievable by transmitting over the channel eigenmodes when both the transmitter and the receiver know the channel matrix [4]. In other words, the MIMO capacity here is the closed-loop MIMO capacity. Furthermore, the spectral efficiency of equal power allocation with PARC is equal to the open-loop MIMO capacity.

In a moderate to high SNR regime, equal power allocation across antennas works almost as well as the optimal (or suboptimal) power allocation as long as the rate is controlled under OSD. Hence, power adaptation becomes largely irrelevant with PARC in a moderate to high SNR region. However, in a low SNR region, it is observed that power allocation improves the capacity. This is in line with conclusions drawn in other research literatures in similar cases. In a single user time-varying channel, a close-to-optimal performance is achieved by transmitting a constant power when the channel path gain is larger than a certain threshold value [21].

²In general, unless all the schemes produce the same probability density function of achievable capacity, the outage capacity does not follow the same trend as the average capacity.

FIGURE 3: Average capacity when $M = 4$ and $N = 4$.

Results also show that the capacity loss relative to the closed-loop MIMO capacity is not significant (except in Figure 2, where the gap between MIMO capacity and equal-power capacity is not reduced even though we increase the average SNR). Therefore, equal power allocation combined with PARC under OSD is a practical and efficient method to approach the MIMO capacity. All the schemes proposed in

this paper perform better than the equal power and rate allocation (MMSE) V-BLAST.³

6. SIMPLE ROUNDING-OFF

Here, a simple, discrete bit loading algorithm is proposed. Given that PARC under equal power allocation achieves the open-loop MIMO capacity as seen in (5), a natural practical extension is to simply round off each rate per antenna with equal power allocation. Here, it is assumed that all the decisions are correct during OSD process.

Given the rate R_m as described in (6), round off R_m and assign the rounded-off rate $\overline{R_m}$, where \overline{x} is the largest integer which is smaller than or equal to x . The rate set can be reduced further by considering only every q th integer. In this case, the rounded-off rate is $q\overline{R_m/q}$. This quantization method does not limit the maximum rate used, but simulation results in Section 9 show that the maximum rate per antenna calculated with this algorithm is less than or equal to 16 QAM when an average SNR is 10 dB. Hence, clipping in quantization is not considered.

As there is no power control, this is simpler than the following two schemes. However, unlike in the continuous rate case, results in Section 9 show that the spectral efficiency loss is significant when power is not adapted.

7. SUCCESSIVE QUANTIZATION WITH POWER CONTROL

A more efficient discrete bit loading algorithm is proposed by also adapting the power levels at each transmit antenna. Obviously, the performance is maximized by using optimal power control under the assumption that discrete rates are available at each transmit antenna. However, a closed-form solution for the optimal discrete rate and continuous power control cannot be found analytically; furthermore, an exhaustive search over the set of rate and power levels is too complicated to be conducted in real time. Hence, instead of the optimal rate and power control scheme, an ad-hoc discrete bit loading method, successive quantization with power control (SQPC) (Figure 4), is suggested in the following. Here also all the decodings are assumed perfect in OSD.

The transmit antennas are labeled according to the order in which they are decoded at the receiver. The SINR of the k th transmit antenna contains interference from all the antennas decoded after it (i.e., $k + 1, \dots, M$). The available rates are assumed to be $0, q, 2q, 3q$, and so on. Therefore, q is the interval between rate quantization levels. Again, there is no clipping; from numerical calculations, the maximum rate

³Equal power and rate allocation should be interpreted carefully. This is achieved when a codebook designer knows the channel and then allocates equal power and rate across the antennas. However, in practice, MMSE V-BLAST is designed without any prior knowledge regarding the channel. Therefore, one MMSE V-BLAST can achieve one point on the curve not the entire curve.

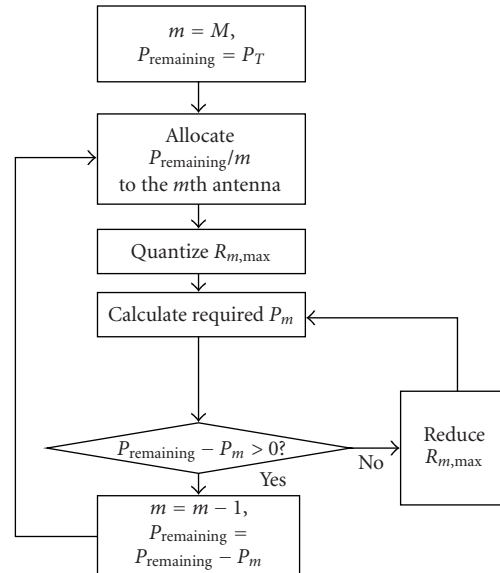


FIGURE 4: SQPC algorithm.

per antenna is less than or equal to 16 QAM when an average SNR is 10 dB.

First, the power and rate for the M th antenna are allocated. The rate of this M th antenna is independent of the power of all other antennas. P_T is divided by M and then assigned as the transmit power of the M th antenna. Then, we calculate the maximum rate $R_{M,\max}$ possible for $P_M = P_T/M$ from (6). Next, round $R_{M,\max}$ and recalculate how much P_M is needed to support rounded $R_{M,\max}$ from (6). Here “round x ” means $q\{x/q\}$, where $\{x\}$ means the integer closest to x . If that power exceeds P_T , then subtract q from $R_{M,\max}$. Then, recalculate how much power is necessary to support the reduced $R_{M,\max}$ from (6).

Second, the power and rate for the $(M - 1)$ th antenna are allocated. Given the interference due to the M th antenna from the previous stage, calculate the maximum rate for the $(M - 1)$ th antenna, assuming $(P_T - P_M)/(M - 1)$ is allocated as the transmit power of the $(M - 1)$ th antenna. Round $R_{M-1,\max}$ and recalculate how much P_{M-1} is needed to support this rounded $R_{M-1,\max}$. If $(P_M + P_{M-1})$ exceeds P_T , then subtract q from $R_{M-1,\max}$ and recalculate P_{M-1} which can support the reduced $R_{M-1,\max}$.

Iteratively, at step j ($j < M - 1$), the power and rate for the $(M - j)$ th antenna are determined. The exact amount of interference from $M, M - 1, \dots, (M - j + 1)$ th antennas is known at this stage. Calculate the maximum rate for the $(M - j)$ th antenna, $R_{M-j,\max}$, assuming $(P_T - (P_M + P_{M-1} + \dots + P_{M-j+1}))/ (M - j)$ is allocated as the transmit power of the $(M - j)$ th antenna. Round $R_{M-j,\max}$ and calculate the new P_{M-j} which can support rounded $R_{M-j,\max}$. If $(P_M + P_{M-1} + \dots + P_{M-j})$ exceeds P_T , then reduce $R_{M-j,\max}$ by q and find the new P_{M-j} which can support the reduced $R_{M-j,\max}$.

At step $M - 1$, where the power and rate for the first antenna are determined, $R_{1,\max}$ is calculated, assuming $(P_T - (P_M + P_{M-1} + \dots + P_2))$ is allocated as the transmit power of

the 1st antenna. Round off $R_{1,\max}$ and recalculate a new P_1 which can support rounded-off $R_{1,\max}$. Here, rounding up is not an option since it would violate the power budget.

SQPC will inherently leave some part of the total power P_T unused. This residual power is not sufficient to increase the rate of any antenna to the next higher quantized level.

8. SUCCESSIVE RATE AND POWER QUANTIZATION

SQPC in Section 7 can become infeasible, especially when frequent rate and power level updates are necessary. As power levels still assume infinite precision, frequent power level updates cannot be supported due to a limited data rate on the feedback channel. Here, we look into the case in which both rate and power are adapted, while limiting the number of available rate and power levels. Here also, a closed-form solution for the optimal discrete rate and discrete power control does not exist; again, an exhaustive search over the set of rates and powers is too complicated to be conducted in real time. Hence, an ad hoc suboptimal discrete bit loading, successive rate and power quantization (SRPQ) (Figure 5), is also suggested as follows. Here also, all the decoding stages are assumed perfect during OSD.

We use the same notation for the antenna labeling and the achievable rates as in Section 7. Furthermore, the available transmit power levels are $0, P_T/(N_p - 1), 2P_T/(N_p - 1), \dots, P_T$, where N_p is the number of available transmit power levels. In SQPC, only rate per antenna was quantized while the power levels could take any continuous values.

First, the power and rate for the M th antenna branch are allocated. P_T is divided by M and then assigned to the M th branch. Then, the maximum rate $R_{M,\max}$ possible is calculated for $P_M = P_T/M$ from (6). Next, round $R_{M,\max}$ and recalculate how much P_M is needed to support rounded $R_{M,\max}$ from (6). Then round up P_M considering the number of power levels available. In other words, P_M is updated as $q_p \lceil P_M/q_p \rceil$, where $q_p = P_T/(N_p - 1)$ and $\lceil x \rceil$ means the integer closest to and larger than x . Round-off is not an option since it would ruin the reliability according to (1). If that power exceeds P_T , then subtract q from $R_{M,\max}$. Recalculate how much power is required to support the reduced $R_{M,\max}$ from (6). Then round up P_M so that P_M can take one of N_p transmit power levels as before. If this P_M still violates the power budget, subtract q from $R_{M,\max}$ again and repeat the process until the power budget is satisfied.

Second, the power and rate for the $(M - 1)$ th antenna are allocated. Given the interference due to the M th antenna from the previous stage, calculate the maximum rate for the $(M - 1)$ th antenna while assuming that $(P_T - P_M)/(M - 1)$ is allocated as the transmit power of the $(M - 1)$ th antenna. Round $R_{M-1,\max}$ and recalculate how much P_{M-1} we need to support this rounded $R_{M-1,\max}$. Then round up P_{M-1} so that P_{M-1} can take one of N_p transmit power levels. If $(P_M + P_{M-1})$ exceeds P_T , then subtract q from $R_{M-1,\max}$ and recalculate the smallest P_{M-1} which is among the available N_p power levels and can support reduced $R_{M-1,\max}$. If the power budget cannot be satisfied, keep reducing $R_{M-1,\max}$ by q until the power budget is satisfied.

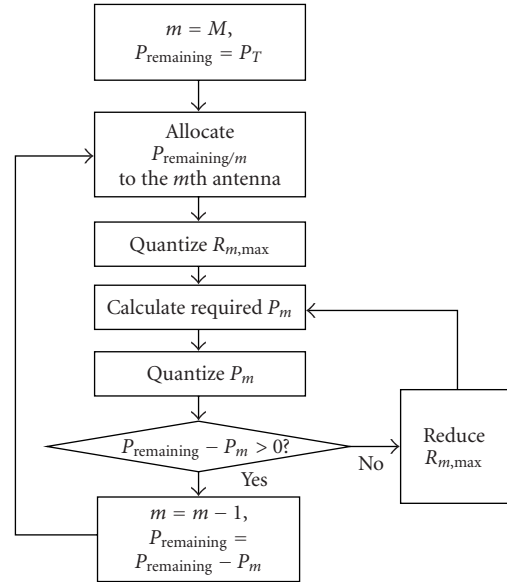


FIGURE 5: SRPQ algorithm.

Iteratively, at step j ($j < M - 1$), the power and rate for the $(M - j)$ th antenna branch are allocated. The exact amount of interference from $M, M - 1, \dots, (M - j + 1)$ th antenna branches is known. Calculate the maximum rate for the $(M - j)$ th antenna branch, $R_{M-j,\max}$, assuming $(P_T - (P_M + P_{M-1} \dots + P_{M-j+1}))/ (M - j)$ is allocated as the transmit power of the $(M - j)$ th branch. Round $R_{M-j,\max}$ and calculate new P_{M-j} which is one of the available N_p power levels and can support rounded $R_{M-j,\max}$. If $(P_M + P_{M-1} \dots + P_{M-j})$ exceeds P_T , then reduce $R_{M-j,\max}$ by q and find a new P_{M-j} which is one of the available N_p power levels and can support reduced $R_{M-j,\max}$. If the power budget is not satisfied, keep reducing $R_{M-j,\max}$ and calculate appropriate P_{M-j} .

At step $M - 1$, where the power and rate for the first antenna are decided, the maximum rate $R_{1,\max}$ is calculated assuming that $(P_T - (P_M + P_{M-1} \dots + P_2))$ is allocated as the transmit power of first branch. Round off $R_{1,\max}$ and recalculate a new P_1 , which is one of the available N_p power levels and can support rounded $R_{1,\max}$. If the power budget is not satisfied, keep reducing $R_{1,\max}$ and calculate appropriate P_1 . Here, rounding up is not an option since it would definitely violate the power budget.

Several variations are shown in the following subsections. The first one is a variation in which residual power is used efficiently to reduce error propagation, while the second one is a variation in which an efficient decoding order is found.

8.1. SRPQ1: efficient use of residual power

SRPQ inherently leaves some part of the total power P_T unused. This residual power is not sufficient to increase the rate of any antenna to the next higher quantized level. However, this residual power can be used efficiently to reduce the error rate. Therefore, by pouring residual power into the first antenna, which is decoded first, its BER performance can be improved. This reduction in BER, in turn, helps improve the

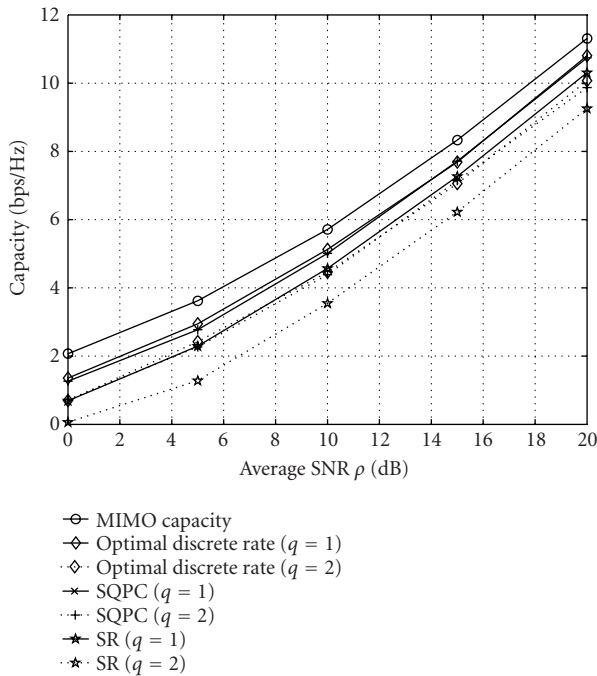


FIGURE 6: Effect of rate quantization when $M = N = 2$.

decoding reliability at later stages. Pouring all the residual power into the first antenna does not increase the feedback channel rate, even though P_1 is not within the N_p possible power levels since P_1 equals $(P_T - \sum_{m=2}^M P_m)$, which can be calculated at the transmitter once P_m ($2 \leq m \leq M$) are fed back. This variation of the SRPQ scheme is called SRPQ1.

8.2. SRPQ2: efficient decoding order

So far, the decoding order has been chosen arbitrarily. In a capacity sense, it was proved that the same total rate is achieved regardless of the decoding order. However, for the quantized rate power case, it is unclear whether the optimization of decoding order is helpful or not. Here, a decoding order is optimized by doing a full search over all possible decoding orders. This variation of SRPQ scheme is called SRPQ2.

9. RESULTS

The following schemes are considered: MIMO Capacity, SR, SQPC, SRPQ1, and SRPQ2. The MIMO capacity is the closed-loop MIMO capacity as in Section 5. For each average SNR ρ , H is generated 1000 times and the average capacity is calculated assuming that a scalar capacity-achieving code is used: $\Gamma = 1$ at (1). First, the effect of rate quantization is investigated; later, power quantization is also considered.

9.1. Effect of rate quantization levels

When q is equal to 1, both square and cross QAM (0 bits/symbol, 1 bit/symbol, 2 bits/symbol, and so on) are allowed as a signal constellation. On the other hand, when q is equal to 2, only square QAM (0 bits/symbol, 2 bits/symbol, 4 bits/symbol, and so on) is allowed. For each q , optimal

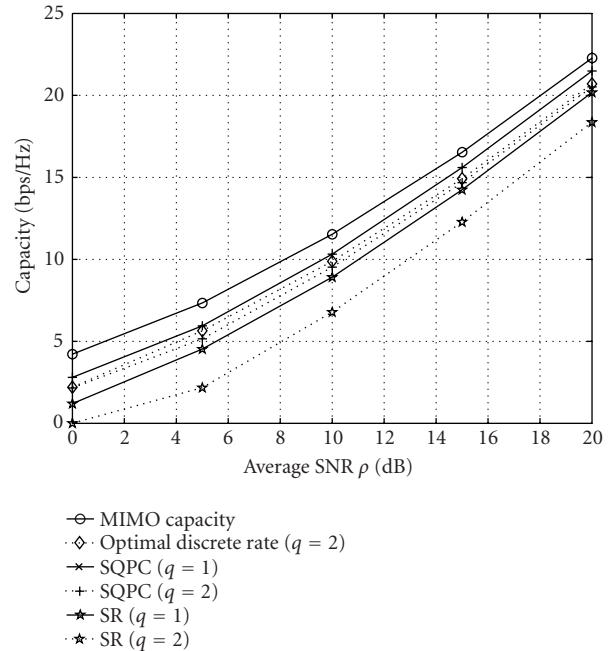


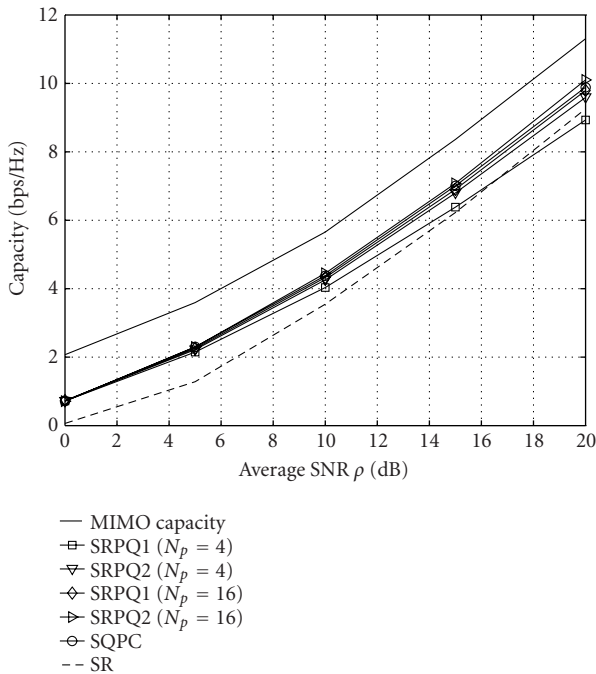
FIGURE 7: Effect of rate quantization when $M = N = 4$.

discrete rate is the case in which the spectral efficiency is maximized under a total power constraint when only discrete rates $(0, q, 2q, \dots)$ are available per antenna. In Figures 6 and 7, the average capacity is displayed as function of the quantization levels. When the power on each transmit antenna is not adapted at all (SR case), using a smaller number of discrete rate levels ($q = 2$) results in poor performance compared with using a larger number of discrete rate levels ($q = 1$). However, in other schemes (SQPC, optimal discrete rate), the performance difference is not significant between $q = 1$ and $q = 2$. The trade-off between feedback information and performance is observed; power levels at each antenna in SR do not need to be fed back. However, more rate levels (smaller q) need to be fed back for SR than for SQPC in order to achieve the same performance level. Hence, it is concluded that $q = 2$ is a reasonable quantization level choice, where power control is also available.

9.2. Effect of power quantization levels

In this section, q is assumed to be 2 and the capacities of the various schemes are compared, depending on the power quantization levels. In Figures 8 and 9, SQPC always performs better than SR for the same M, N , and q . Furthermore, the performance gap increases with M and N . Moreover, for low SNR, the capacity of SQPC falls short of the MIMO capacity by 4 dB in SNR when $q = 2$. Due to space limitations, the result for $q = 1$ cannot be presented, but in this case, the performance of SQPC is less than the MIMO capacity by 3 dB in SNR.

For a low average SNR ρ , a small number of power levels does not degrade the performance significantly from a large number of power levels. The reason is that, for a low SNR, usually only a single antenna is activated. However, for a high

FIGURE 8: Average capacity when $M = 2$ and $N = 2$ for $q = 2$.

average SNR ρ , the performance loss is considerable as the number of power levels is decreased. Indeed, when $N_p \leq 4$, the degradation caused by power control quantization becomes so great that it is better not to do power allocation at all, since SR scheme outperforms both SRPQ1 and SRPQ2.

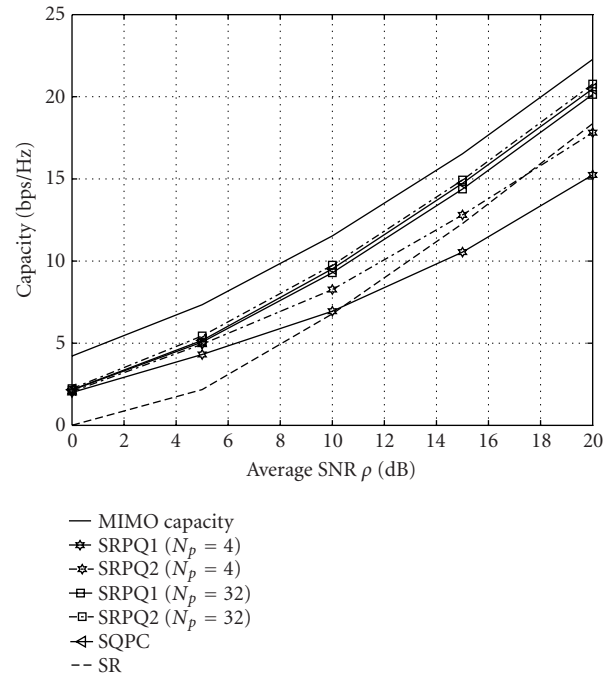
Our results suggest that $N_p = 16$ and $N_p = 32$ for $M = N = 2$ and $M = N = 4$, respectively, result in minimal degradation compared to the scheme in which continuous power is allowed. Moreover, this choice of N_p leads to only 2 dB away from the MIMO capacity if a capacity achieving scalar coding is used. Finally, as can be seen, SRPQ2 outperforms SRPQ1 in terms of spectral efficiency. This shows that the decoding order indeed matters when continuous rate and power cannot be used.

10. CONCLUSIONS

This paper proposes an extension of V-BLAST in which the MIMO capacity is approached closely with rate and/or power control using scalar coding with successive interference cancellation. Two practical discrete bit loading algorithms are proposed: SQPC and SRPQ. Simulation results show that power control is necessary, especially in a low SNR regime. Furthermore, it is shown that 4 or 5 bits are sufficient for power quantization levels in order to sustain a similar spectral efficiency to that achieved by continuous power levels.

ACKNOWLEDGMENT

This paper was presented in part at the IEEE Vehicular Technology Conference (VTC) Fall 2001 and the 2002 IEEE Wireless Communications and Networking Conference (WCNC).

FIGURE 9: Average capacity when $M = 4$ and $N = 4$ for $q = 2$.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [2] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, 2001.
- [3] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.
- [4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Tech. Rep., AT & T Bell Laboratories Internal Technical Memorandum, 1995.
- [5] M. A. Khalighi, J. Brossier, G. Jourdain, and K. Raouf, "Water filling capacity of Rayleigh MIMO channels," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '01)*, vol. 1, pp. 155–158, San Diego, Calif, USA, September 2001.
- [6] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environment when using multiple-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [8] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal BLAST processing with multiple-access interference," in *IEEE 52nd Vehicular Technology Conference (VTC '00)*, vol. 1, pp. 87–91, Boston, Mass, USA, September 2000.
- [9] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. of URSI International Symposium on Signals, Systems, and Electronics (ISSSE '98)*, pp. 295–300, Pisa, Italy, September 1998.

- [10] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1841–1852, 1999.
- [11] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, 1999.
- [12] S. Verdú, *Multisuser Detection*, Cambridge University Press, New York, NY, USA, 1998.
- [13] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 226–235, 2002.
- [14] S. Thoen, L. Van der Perre, B. Gyselinckx, M. Engels, and H. De Man, "Adaptive loading in the downlink of OFDM/SDMA-based wireless local networks," in *IEEE 51st Vehicular Technology Conference Proceedings (VTC '00)*, vol. 1, pp. 235–239, Tokyo, Japan, May 2000.
- [15] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 1405–1409, Monterey, Calif, USA, November 1997.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.
- [17] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
- [18] A. Lozano, "Capacity-approaching rate function for layered multiantenna architectures," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 616–620, 2003.
- [19] K. J. Hwang and K. B. Lee, "Transmit power allocation with small feedback overhead for multiple antenna system," in *IEEE 56th Vehicular Technology Conference (VTC '02)*, vol. 4, pp. 2158–2162, Vancouver, Calif, USA, September 2002.
- [20] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, NY, USA, 1996.
- [21] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

Seong Taek Chung received the B.S. degree in electrical engineering from Seoul National University, Korea, in 1998. He received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Calif, in 2000 and 2004, respectively. During the summer of 2000, he was an intern with Bell Laboratories, Lucent Technologies, where he worked on multiple antenna systems. He is currently a Senior Engineer at Qualcomm Inc., San Diego, Calif. His research interest includes communication theory and signal processing.



Angel Lozano was born in Manresa, Spain, in 1968. He received the Engineer degree in telecommunications (with honors) from the Polytechnical University of Catalonia, Barcelona, Spain, in 1992 and the Master of Science and Ph.D. degrees in electrical engineering from Stanford University, Stanford, Calif, in 1994 and 1998, respectively. Between 1996 and 1998 he worked for Pacific



Communication Sciences Inc. and for January 1999 he was with Bell Laboratories (Lucent Technologies) in Holmdel, NJ. Since October 1999, he has served as an Associate Editor for IEEE Transactions on Communications. Dr. Lozano holds 6 patents.

Howard C. Huang was born in Texas in 1969. He received the B.S.E.E. degree from Rice University, Houston, TX, in 1991, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1995. He is currently a distinguished member of the technical staff in the Wireless Communications Research Department at Bell Laboratories, Lucent Technologies in Holmdel, NJ. His interests include communication theory and multiple antenna networks.



Arak Sutivong received the B.S. and M.S. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, Pa, in 1995 and 1996, respectively. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif, in 2003. From 1997 to 1998, he was a Systems Engineer at Qualcomm Inc., San Diego, Calif, developing a satellite-based CDMA system, while at Stanford University, he has served as a Technical Consultant to numerous companies. He returned to Qualcomm Inc. in October 2002, where he is currently a Staff Engineer. His research interests are in information theory and its applications, wireless communications, and signal processing.



John M. Cioffi received his B.S.E.E. degree in 1978 from the University of Illinois and his Ph.D. degree in electrical engineering from the University of Stanford in 1984. He was with Bell Laboratories from 1978 to 1984 and worked at IBM Research from 1984 to 1986. In 1986, he became a Professor of electrical engineering at the University of Stanford. Cioffi founded the Amati Communications Corporation in 1991 (purchased by Texas Instruments (TI) in 1997) and was the Officer/Director from 1991 till 1997. He is currently on the board of directors of Marvell, Teknovus, Ikanos, Clariphy, and Tranetics. He is on the advisory boards of Charter Ventures, Halisos Networks, and Portview Ventures. Cioffi's specific interest is in the area of high-performance digital transmission. Dr. Cioffi was granted the Hitachi America Professorship in electrical engineering at Stanford in 2002 and was a member of the National Academy of Engineering in 2001. He received the IEEE Kobayashi Medal in 2001 and the IEEE Millennium Medal in 2000. Moreover, he was the IEEE Fellow in 1996 and received the IEEE J. J. Thomson Medal in 2000. He is the 1999 University of Illinois Outstanding Alumnus and received 1991 IEEE Communication Magazine Best Paper Award and 1995 ANSI T1 Outstanding Achievement Award. He was the National Science Foundation (NSF) Presidential Investigator from 1987 till 1992. Cioffi has published over 200 papers and holds over 40 patents.

