

## Research Article

# On the Problem of Bandwidth Partitioning in FDD Block-Fading Single-User MISO/SIMO Systems

Michel T. Ivrlač and Josef A. Nossek

*Lehrstuhl für Netzwerktheorie und Signalverarbeitung, Technische Universität München, 80290 München, Germany*

Correspondence should be addressed to Michel T. Ivrlač, ivrlac@tum.de

Received 6 November 2007; Revised 2 April 2008; Accepted 26 June 2008

Recommended by Sven Erik Nordholm

We report on our research activity on the problem of how to optimally partition the available bandwidth of frequency division duplex, multi-input single-output communication systems, into subbands for the uplink, the downlink, and the feedback. In the downlink, the transmitter applies coherent beamforming based on quantized channel information which is obtained by feedback from the receiver. As feedback takes away resources from the uplink, which could otherwise be used to transfer payload data, it is highly desirable to reserve the “right” amount of uplink resources for the feedback. Under the assumption of random vector quantization, and a frequency flat, independent and identically distributed block-fading channel, we derive closed-form expressions for both the feedback quantization and bandwidth partitioning which jointly maximize the sum of the average payload data rates of the downlink and the uplink. While we do introduce some approximations to facilitate mathematical tractability, the analytical solution is asymptotically exact as the number of antennas approaches infinity, while for systems with few antennas, it turns out to be a fairly accurate approximation. In this way, the obtained results are meaningful for practical communication systems, which usually can only employ a few antennas.

Copyright © 2008 M. T. Ivrlač and J. A. Nossek. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

In this work, we consider a single-user, frequency division duplex (FDD) wireless communication system which can be modeled as a frequency flat fading multi-input single-output (MISO) system in the *downlink*, and as a frequency flat fading single-input multi-output (SIMO) system in the *uplink*. In order to achieve the maximum possible channel capacity of such a communication system, perfect knowledge about the normalized channel vector has to be present at the receiver in the uplink, and at the transmitter in the downlink.

In the uplink, the channel between the single transmit and the multiple receive antennas (the SIMO case) can be estimated by the receiver by evaluating a received pilot sequence prior to applying coherent receive beamforming based on the estimated channel vector, so-called maximum ratio combining [1]. In the downlink, the situation is more complicated. Because of the frequency gap between the uplink and the downlink band, the channel which was estimated by the receiver in the uplink cannot be used by the transmitter in the downlink. The channel between the multiple transmit antennas and the single receive antenna

(the MISO case) has to be estimated by the receiver, and then transferred back in a quantized form to the transmitter, such that coherent transmit beamforming can be applied, so-called maximum ratio transmission [2].

The more bits are used for the quantized feedback, the higher is the obtainable beamforming gain, and hence, the downlink throughput. However, feedback is taking away resources from the uplink, which could otherwise be used to transfer payload data. It is highly desirable to reserve the “correct” amount of uplink resources for the feedback such that the overall performance of the downlink *and* the uplink is maximized. Moreover, the division of the available bandwidth into the uplink band and the downlink band should also be optimized. In this report, we will propose a way on obtaining an optimized partition of the total bandwidth into subbands for the uplink, the downlink, and the feedback.

### 1.1. Related work

Coherent transmit beamforming for MISO systems based on quantized feedback was proposed in [3]. The beamforming

vector is thereby chosen from a finite set, the so-called *codebook*, that is known to both the transmitter and the receiver. After having estimated the channel, the receiver chooses that vector from the codebook which maximizes signal-to-noise ratio (SNR). The index of the chosen vector is then fed back to the transmitter. There are different ways of designing codebooks for vector quantization [4]. By extending the work in [5], a design method for orthogonal codebooks is proposed in [6] which can achieve full transmit diversity order using quantized equal gain transmission. In [7], nonorthogonal codebooks are designed based on Grassmannian line packing [8]. Analytical results for the performance of optimally quantized beamformers are developed in [9], where a universal lower bound on the outage probability for any finite set of beamformers with quantized feedback is derived. The authors of [10] propose to maximize the mean-squared weighted inner product between the channel vector and the quantized vector, which is shown to lead to a closed form design algorithm that produces codebooks which reportedly behave well also for correlated channel vectors. Nondeterministic approaches using so-called random vector quantization (RVQ) are proposed in [11–13], where a codebook composed of vectors which are uniformly distributed on the unit sphere is randomly generated each time there is a significant change of the channel. It is shown in [11] that RVQ is optimal in terms of capacity in the large system limit in which both the number of transmit antennas and the bandwidth tend to infinity with a fixed ratio. For low number of antennas, numerical results [14] indicate that RVQ still continues to perform reasonably well.

The important aspect that feedback occupies resources that could otherwise be used for payload data, is investigated in [15, 16]. The cost for channel estimation and feedback is taken into account in [15] by scaling the mutual information that is used as a vehicle to compute the block fading outage probability. In [16], the optimum number of pilot bits and feedback bits in relation to the size of a radio frame is analyzed. In particular, for an i.i.d. block fading channel, upper and lower bounds on the channel capacity with random vector quantization and limited-rate feedback are derived, which are functions of the number of pilot symbols and feedback bits. The optimal amount of pilot symbols and feedback bits as a fraction of the size of the radio frame is derived under the assumption of a constant transmit power and large number of transmit antennas. (It is shown in [16] that for a constant transmit power, as the size of the radio frame approaches infinity forming a fixed ratio with the number  $N$  of transmit antennas, the optimal pilot size and the optimum number of feedback bits normalized to the antenna number tend to zero at rate  $(\log N)^{-1}$ .)

### 1.2. Our approach: optimum resource sharing

While [15, 16] do consider that feedback and pilot symbols occupy system resources, they treat the flow of payload data

as unidirectional, namely, flowing in the downlink from the multiantenna transmitter to the single-antenna receiver (the MISO-case). Furthermore, the asymptotic analysis in [16] for large antenna numbers keeps the transmit power constant, which leads to a receiver SNR that increases with the number of transmit antennas.

In our approach, we propose to *share* the totally available resources between *downlink*, *uplink*, and *feedback* such that the overall system performance in terms of the *sum of the throughputs* of the downlink and the uplink is maximized. In this way, we can also maintain a *given and finite SNR at the receivers with lowest amount of transmit power*. Keeping the receiver SNR constant, instead of the transmit power, has the advantage that any desired trade-off between bandwidth efficiency and transmit power efficiency can be implemented [17]. (We will see in Section 4.7 that a receive SNR of about 6 dB is optimum in the sense that it maximizes the product of bandwidth efficiency and transmit power efficiency.) To be more specific, we are interested in the following situation.

- (1) We consider an FDD system which has  $N$  transmit antennas and a single receive antenna in the downlink, and  $N$  receive antennas and a single transmit antenna in the uplink.
- (2) The system has available a total usable bandwidth  $B$ . (The term “usable” refers to the fact that the communication system may need additional bandwidth resources, e.g., for channel estimation, synchronization, traffic control channels, and guard bands. The total “usable” bandwidth is the bandwidth which the system has available for transporting downlink payload data, uplink payload data, and feedback information.)
- (3) The bandwidth  $B$  has to be partitioned into a bandwidth  $B_{DL}$  for the downlink band, and into a bandwidth  $B_{UL}$  for the uplink band. Furthermore, a part of the uplink band, with bandwidth  $B_{FB}$ , has to be reserved for feedback rather than for carrying the uplink payload data. This bandwidth partitioning is shown in Figure 1.
- (4) The uplink and the downlink bands are separated by a frequency gap, such that instantaneous channel state information obtained from the uplink *cannot* be used in the downlink, hence making feedback of instantaneous channel state information necessary. Notice that such a gap in frequency between the uplink band and the downlink band is necessary in any FDD system due to implementation issues. (The huge imbalance in receive and transmit power (usually more than 100 dB) at the basestation necessitates a significant gap in frequency in order to insure that the order of the required filters does not become too large to be implementable.)
- (5) Both the uplink band and the downlink band can be modeled as frequency flat fading.

(6) The proposed bandwidth partitioning takes place according to

$$\begin{aligned}
 & (B_{UL}^{opt}, B_{DL}^{opt}, B_{FB}^{opt}) \\
 & = \arg \max_{(B_{UL}, B_{DL}, B_{FB})} (\bar{R}_{DL}(B_{DL}, B_{FB}) + \bar{R}_{UL}(B_{UL}, B_{FB})), \\
 & \text{such that } \begin{cases} B_{DL} > 0, \\ B_{UL} > 0, \\ 0 < B_{FB} \leq B_{UL}, \\ B_{UL} + B_{DL} = B, \\ \bar{R}_{UL} = \mu \bar{R}_{DL}, \end{cases} \tag{1}
 \end{aligned}$$

where  $\bar{R}_{DL}$  and  $\bar{R}_{UL}$  denote the average payload data rates in the downlink and the uplink, respectively, and  $\mu \geq 0$  is a symmetry factor which accounts for different requirements on payload data rate in the two different directions. For  $\mu = 0$ , the communication becomes unidirectional (downlink only), that is, the whole uplink band can be used for feedback. Of course, (1) can be restated as maximization of  $\bar{R}_{DL}$  with the same constraints, since  $\bar{R}_{UL}$  is kept in a fixed ratio with  $\bar{R}_{DL}$ . However, the formulation (1) has a convenient structure which can be used to arrive at an elegant solution.

### 1.3. Major assumptions

In order to solve (1), we make the following assumptions.

- (1) In the downlink, the  $N$  transmit antennas are used for maximum ratio transmission based on quantized channel feedback.
- (2) An i.i.d. frequency-flat block-fading channel is assumed for the uplink and the downlink. That is, the channel is assumed to remain constant within the time  $T_{dec}$ , and then to abruptly change to a new, independent realization.
- (3) The channel coefficients between any receive and transmit antenna are uncorrelated.
- (4) Channel estimation errors at the receivers are negligible.
- (5) The bandwidth  $B$  is completely usable for payload and feedback. There are additional resources needed for channel estimation, however, those have to be present with or without the feedback scheme, so we do not consider those resources as part of the optimization.
- (6) The quantization of the normalized channel vector is performed by RVQ using  $b$  bits per antenna. The codebook, therefore, consists of  $2^{Nb}$  (pseudo)-random vectors which are chosen uniformly from the unit sphere. Each time the channel changes, a new realization of the codebook is generated. In this way, the performance of the RVQ is averaged over all random codebooks (uniform on the unit sphere).

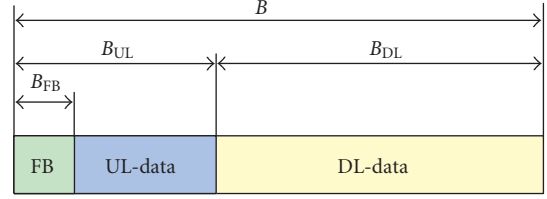


FIGURE 1: Partitioning of the available bandwidth into a downlink band and an uplink band, where the latter accommodates also a band reserved for feedback. Note that the gap in frequency between the uplink and the downlink band is *not* shown in this figure.

- (7) The quantized feedback bits are protected by capacity approaching error control coding.
- (8) Capacity approaching error control coding is also used for the payload data both in the uplink and the downlink.
- (9) The feedback bits can be decoded correctly with negligible outage.
- (10) Feedback has to be received within the time  $T$ , where  $T \ll T_{dec}$ .

## 2. GENERIC SOLUTION

From the assumptions in Section 1.3, we can write with the help of the newly introduced parameter  $\eta$  (which is used as a nice mathematical way to obtain the notion of *outage* capacity while, in effect, only *ergodic* capacity has to be computed):

$$\frac{N \cdot b}{T} = \eta \cdot B_{FB} \cdot E[\log_2(1 + \text{SNR}_{UL})], \tag{2}$$

since  $Nb$  bits of feedback have to be *reliably* transferred within  $T$  seconds, requiring an information rate of  $Nb/T$  bits per second. It is important to note that the instantaneous receive SNR in the uplink ( $\text{SNR}_{UL}$ ) and hence, the instantaneous uplink channel capacity, fluctuates randomly because of the block fading channel. Nevertheless, it is highly important that the feedback information can be decoded correctly in most cases. In order to ensure correct decoding with a given probability, we include the factor  $\eta$ , with  $0 < \eta \leq 1$ . Therefore, in (2), we equate the information rate  $Nb/T$  with  $\eta$  times the ergodic uplink channel capacity. The smaller the value of  $\eta$ , the higher is the probability that the instantaneous channel capacity is above  $\eta$  times its mean value, and hence, the smaller is the probability of a channel outage. For instance, with  $N = 4$  and i.i.d. Rayleigh fading with an average uplink SNR of 6 dB, it turns out, that correct decoding is possible with 99% probability when we set  $\eta = 0.4$ . Therefore, assuming these parameters, (2) equates the feedback information rate  $Nb/T$ , with the 1%-outage capacity of the feedback channel. In the following, we consider  $\eta$  as a given system parameter. Note that for large number of antennas, the fluctuation of  $\text{SNR}_{UL}$  around its mean value becomes small. Hence,  $\eta$  can be chosen close to unity:

$$\lim_{N \rightarrow \infty} \eta = 1. \tag{3}$$

Furthermore,

$$\bar{R}_{UL}(b) = \underbrace{B_{UL} \cdot E[\log_2(1 + \text{SNR}_{UL})]}_{\bar{R}_{UL} \text{ without feedback}} - \frac{Nb}{T\eta}, \quad (4)$$

and finally,

$$\bar{R}_{DL}(b) = B_{DL} \cdot E[\log_2(1 + \text{SNR}_{DL})], \quad (5)$$

where  $\text{SNR}_{DL}$  is the receive SNR in the downlink. Since the obtainable beamforming gain depends on the quantization resolution,  $\bar{R}_{DL}$  is a function of the number  $b$  of feedback bits per antenna. The original optimization problem (1) can now be solved in *three* steps.

(1) Assuming that we know  $B_{DL}^{\text{opt}}$ , find the optimum quantization resolution.

$$\begin{aligned} b_{\text{opt}}(B_{DL}^{\text{opt}}) &= \arg \max_b (\bar{R}_{DL}(b) + \bar{R}_{UL}(b)), \\ &= \arg \max_b \left( B_{DL}^{\text{opt}} \cdot E[\log_2(1 + \text{SNR}_{DL})] - \frac{Nb}{T\eta} \right), \end{aligned} \quad (6)$$

since  $\text{SNR}_{UL}$  does *not* depend on  $b$ . Note that this  $b_{\text{opt}}$  will depend on  $B_{DL}^{\text{opt}}$ , whose value is unknown at this moment, but will be computed in the following step.

(2) Find the optimum bandwidth partition.

From (2) immediately follows that

$$B_{FB}^{\text{opt}}(B_{DL}^{\text{opt}}) = \frac{N \cdot b_{\text{opt}}(B_{DL}^{\text{opt}})}{\eta \cdot T \cdot E[\log_2(1 + \text{SNR}_{UL})]}. \quad (7)$$

Using the last constraint in (1), it follows from (4) and (5) that

$$B_{UL}^{\text{opt}}(B_{DL}^{\text{opt}}) = \underbrace{\mu \frac{E[\log_2(1 + \text{SNR}_{DL})]}{E[\log_2(1 + \text{SNR}_{UL})]}}_{\geq 0} B_{DL}^{\text{opt}}(B_{DL}^{\text{opt}}) + B_{FB}^{\text{opt}}(B_{DL}^{\text{opt}}). \quad (8)$$

With the second to last constraint in (1), it follows from (8) that

$$B_{DL}^{\text{opt}} = \frac{B - B_{FB}^{\text{opt}}(B_{DL}^{\text{opt}})}{1 + \mu(E[\log_2(1 + \text{SNR}_{DL})]/E[\log_2(1 + \text{SNR}_{UL})])}. \quad (9)$$

Note that (9) is an *implicit* solution, since it contains the desired  $B_{DL}^{\text{opt}}$  both on its left-hand side and on its right-hand side. However, we will see in Section 4.5 that (9) can be transformed into an *explicit* form, where  $B_{DL}^{\text{opt}}$  is given as an explicit function of known system parameters.

(3) Obey the remaining constraints.

As long as

$$\begin{aligned} b_{\text{opt}} &> 0, \\ B_{FB}^{\text{opt}} &< B, \end{aligned} \quad (10)$$

we can see from (7)–(9) that the remaining first three constraints of (1) are fulfilled. As a consequence, (10) is necessary and sufficient for the existence of the solution.

The original constraint optimization problem (1) is, therefore, essentially reduced to the unconstrained problem (6) of finding  $b_{\text{opt}}$ .

## 2.1. Simplifications

For the sake of mathematical tractability, we will use the approximation:

$$E[\log_2(1 + \text{SNR}_{DL})] \approx \log_2(1 + E[\text{SNR}_{DL}]). \quad (11)$$

Note that

$$\begin{aligned} E[\log_2(1 + \text{SNR}_{DL})] &\rightarrow \log_2(1 + E[\text{SNR}_{DL}]) \\ &\text{for } \begin{cases} \text{SNR}_{DL} \rightarrow 0, \\ N \rightarrow \infty. \end{cases} \end{aligned} \quad (12)$$

That is, the approximation (11) becomes an almost exact equality either in the low SNR regime, or for large number of antennas. The latter is due to the fact that with increasing  $N$  the diversity order increases, such that the SNR varies less and less around its mean value. Using this approximation in (6), we obtain the optimization problem:

$$\tilde{b}_{\text{opt}} = \arg \max_b \left( B_{DL} \cdot \log_2 \left( 1 + \underbrace{E[\text{SNR}_{DL}]}_{\text{function of } b} \right) - \frac{Nb}{T\eta} \right), \quad (13)$$

which is much easier to solve than (6). Because of (12), it follows that

$$\tilde{b}_{\text{opt}} \rightarrow b_{\text{opt}} \quad \text{for } \begin{cases} \text{SNR}_{DL} \rightarrow 0, \\ N \rightarrow \infty. \end{cases} \quad (14)$$

## 2.2. Preview of key results

In the following sections, we present a detailed derivation of the solution to the problem (13) and the associated optimum bandwidth partitioning problem in closed form. More precisely, for a given system bandwidth  $B$  and a symmetry factor  $\mu$ , we obtain analytical expressions for the optimum quantization resolution and the optimum bandwidth that should be allocated for the downlink, the uplink, and the feedback.

While the solution is asymptotically exact as the number of antennas approaches infinity, we will see that it is also fairly accurate for low antenna numbers. In this way, the obtained solution is not only attractive from a theoretical point of view, but also applicable for practical communication systems. For instance, in the process of standardization of future wireless communication systems, the proposed solution may provide valuable input for the discussion about how fine to quantize channel information and how much resources to reserve for its feedback.

In order to gain a better feeling about what can be done with the solution developed in this manuscript, we would like to present some of the obtained results. For the sake of clarity, let us look at a concrete example system, where a totally usable bandwidth  $B$  has to be partitioned. Let the time  $T$  during which the feedback has to arrive be given by  $T = 100/B$ . The considered system should be a symmetrical one, where the average

payload data rates are the same in uplink and downlink (symmetry factor  $\mu = 1$ ). Moreover, let us assume that the encoded feedback can be decoded correctly with high probability, say 99%. This can be accomplished by setting the factor  $\eta$  (see (2) and the discussion in Section 2) properly. (The actual value for the factor  $\eta$  depends on the fading distribution in the uplink, which also depends on the number  $N$  of receive antennas. In the case of i.i.d. Rayleigh fading it turns out that  $\eta = (0.175, 0.4, 0.57, 0.7, 0.79, 1)$  in conjunction with  $N = (2, 4, 8, 16, 32, \infty)$  guarantees decoding errors below 1%.) In both the uplink and the downlink, the average SNR is set to 6 dB, which is the optimum value for a single-stream system that attempts to be both bandwidth-efficient *and* power-efficient at the same time (see the discussion in Section 4.7 for more details). Using the results derived in this manuscript, we obtain the optimum bandwidth partition for the described example system for different number of antennas  $N \in \{2, 4, 8, 16, 32, \infty\}$ , as shown in Figure 2. Note that starting from about 5.4% of the total bandwidth for  $N = 2$  antennas, the optimum amount of feedback bandwidth increases strictly monotonic with increasing antenna number, reaching almost 10% for  $N = 8$ . In case that  $N \rightarrow \infty$ , it turns out that it is optimum to reserve exactly 20% of the totally available bandwidth for feedback. It is interesting to note that this last asymptotic result essentially only depends on the symmetry factor  $\mu$ , but *not* on system parameters like bandwidth  $B$ , or time  $T$ . By setting the symmetry factor  $\mu = 0$ , we obtain a pure downlink system, which makes use of the whole uplink band for feedback. As we will see in Section 4.6, this system is most happy with a feedback bandwidth of exactly 1/3 of the available bandwidth, as the number of antennas approaches infinity.

### 3. RANDOM VECTOR QUANTIZATION

As described in Section 2, the optimum bandwidth partitioning problem can essentially be reformulated in the unconstrained optimization problem (13). As a prerequisite for its solution, we need to know the functional relationship:

$$b \mapsto E[\text{SNR}_{\text{DL}}], \quad (15)$$

that is, in what way the average SNR in the downlink is influenced by the resolution with which the channel information is quantized. In this section, the function (15) is derived, assuming random vector quantization (RVQ). The motivation for RVQ is both mathematical tractability [13], and the fact that it can indeed be optimal for large number of antennas [11].

#### 3.1. Transmit beamforming

In the downlink, the frequency flat i.i.d. block fading channel between the  $N$  transmit antennas and the single receive antenna is described by the channel vector  $\mathbf{h} \in \mathbb{C}^{N \times 1}$ . The transmitter applies beamforming with a beamforming vector

$\mathbf{u} \in \mathbb{C}^{N \times 1}$  such that the signal,

$$r = \sqrt{\frac{P_{\text{T}}}{\|\mathbf{u}\|_2^2 \cdot E[|s|^2]}} \cdot \mathbf{h}^T \mathbf{u} \cdot s + \nu, \quad (16)$$

is received, in case that the signal  $s \in \mathbb{C}$  is transmitted with power  $P_{\text{T}}$ . Herein, the term  $\nu \in \mathbb{C}$  denotes receiver noise with power  $\sigma_{\nu}^2$ . The receive SNR in the downlink, therefore, becomes

$$\begin{aligned} \text{SNR}_{\text{DL}} &= \frac{E[|r - \nu|^2 | \mathbf{h}, \mathbf{u}]}{E[|\nu|^2]}, \\ &= \underbrace{\frac{P_{\text{T}} \cdot \|\mathbf{h}\|_2^2}{\sigma_{\nu}^2}}_{\text{SNR}_{\text{DL}}^{\text{max}}} \cdot \underbrace{\frac{|\mathbf{h}^T \mathbf{u}|^2}{\|\mathbf{h}\|_2^2 \cdot \|\mathbf{u}\|_2^2}}_{\gamma}, \end{aligned} \quad (17)$$

where  $\text{SNR}_{\text{DL}}^{\text{max}}$  is the maximum obtainable downlink SNR, while  $0 \leq \gamma \leq 1$  is the *relative* SNR, which is maximum for coherent beamforming, that is, if  $\mathbf{u} = \text{const} \cdot \mathbf{h}^*$ .

#### 3.2. Quantization and feedback procedure

The receiver generates quantized feedback in the following way.

- (1) The channel vector  $\mathbf{h}$  is estimated (with negligible error).
- (2) A sequence of  $2^{Nb}$  i.i.d. pseudorandom vectors  $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{2^{Nb}})$  is generated such that

$$\mathbf{u}_i \propto \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{I}_N). \quad (18)$$

- (3) The transmitter generates the same sequence of pseudorandom vectors.
- (4) In case that  $\mathbf{u}_i$  is chosen as the beamforming vector, the resulting relative SNR will be

$$\gamma_i = \frac{|\mathbf{h}^T \mathbf{u}_i|^2}{\|\mathbf{h}\|_2^2 \cdot \|\mathbf{u}_i\|_2^2}. \quad (19)$$

- (5) The vector  $\mathbf{u}_{i_*}$  is selected as the beamforming vector according to

$$i_* = \arg \max_{i \in \{1, 2, \dots, 2^{Nb}\}} \gamma_i. \quad (20)$$

- (6) The  $Nb$  bit long binary representation of the index  $i_*$  is protected by capacity approaching error control coding and fed back to the transmitter.

- (7) Upon successful decoding of the encoded feedback data, the transmitter begins to use the beamforming vector  $\mathbf{u}_{i_*}$ , which leads to an SNR:

$$\text{SNR}_{\text{DL}} = \text{SNR}_{\text{DL}}^{\text{max}} \cdot \gamma_{i_*}. \quad (21)$$

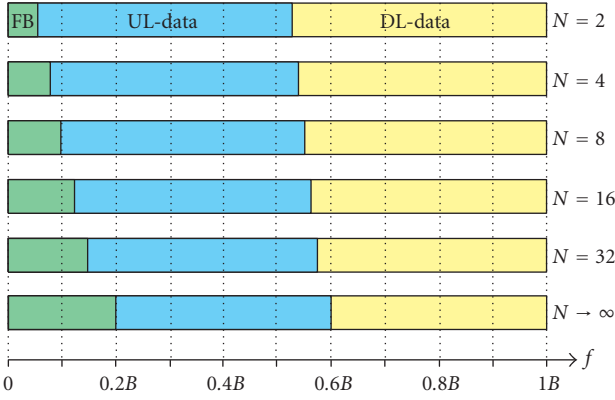


FIGURE 2: Optimum partitioning of the available bandwidth for a symmetric ( $\mu = 1$ ) system operating at average SNR of 6 dB with a bandwidth-time product of  $BT = 100$ .

### 3.3. Average receive SNR in the downlink

The average receive SNR in the downlink can now be written as [13]

$$\begin{aligned} E[\text{SNR}_{\text{DL}}] &= \frac{P_{\text{T}}}{\sigma_{\text{v}}^2} \cdot E[\|\mathbf{h}\|_2^2 \cdot E[\gamma_{i_*} | \mathbf{h}]], \\ &= \frac{P_{\text{T}}}{\sigma_{\text{v}}^2} \cdot \underbrace{E[\|\mathbf{h}\|_2^2]}_{\overline{\text{SNR}}_{\text{DL}}^{\text{max}}} \cdot \left(1 - 2^{Nb} \cdot \text{B}\left(2^{Nb}, \frac{N}{N-1}\right)\right), \\ &= \overline{\text{SNR}}_{\text{DL}}^{\text{max}} \cdot \left(1 - 2^{Nb} \cdot \text{B}\left(2^{Nb}, \frac{N}{N-1}\right)\right), \end{aligned} \quad (22)$$

where  $\overline{\text{SNR}}_{\text{DL}}^{\text{max}}$  denotes the maximum possible average SNR that is obtainable in the downlink, and  $\text{B}(\cdot, \cdot)$  is the beta function [18, 19]. Notice that  $b \rightarrow \infty$  implies  $E[\text{SNR}_{\text{DL}}] \rightarrow \overline{\text{SNR}}_{\text{DL}}^{\text{max}}$ , while  $b = 0$  implies  $E[\text{SNR}_{\text{DL}}] = \overline{\text{SNR}}_{\text{DL}}^{\text{max}}/N$ .

### 3.4. Simplifications

While (22) provides an *exact* expression for the average SNR in the downlink, it does not seem particularly attractive to use it directly in the optimum quantization resolution problem given in (14) since  $b$  appears both outside and inside the beta function. We propose to apply some approximation to (22) in order to facilitate the solution of the optimum bandwidth partitioning problem. From [16, 20], an upper and lower bound on  $E[\gamma_{i_*} | \mathbf{h}]$  for  $b > 0$  can be given:

$$1 \leq \frac{E[\gamma_{i_*} | \mathbf{h}]}{1 - 2^{-b}} \leq 1 + \Psi(b, N), \quad (23)$$

where

$$\Psi(b, N) = \frac{1 + (C_{\Gamma} - 1)2^{-b} + 2^{-Nb}}{(1 - 2^{-b})(N - 1)}, \quad (24)$$

and  $C_{\Gamma} = 0.577216\dots$  is the Euler Gamma constant [18, 19]. A consequence of

$$\lim_{N \rightarrow \infty} \Psi(b, N) = 0 \quad (25)$$

is that for a constant number  $b > 0$  of bits per antenna, the upper and lower bounds in (23) converge towards each other, hence,

$$E[\gamma_{i_*} | \mathbf{h}] \rightarrow 1 - 2^{-b} \quad \text{for } N \rightarrow \infty, \quad b = \text{positive constant.} \quad (26)$$

The situation is more complicated in case that  $b$  approaches zero as  $N$  approaches  $\infty$ . Note that  $b$  should never approach zero more quickly than  $1/N$  because, otherwise, the total number of feedback bits per time  $T$  would drop below unity, which we may consider pathological for a system that attempts to use feedback. For  $b = \beta/N$ , with  $\beta \geq 1$  being a constant, we find

$$\lim_{N \rightarrow \infty} \Psi\left(\frac{\beta}{N}, N\right) = \frac{C_{\Gamma} + 2^{-\beta}}{\beta \cdot \log_e 2} < 1.56. \quad (27)$$

For large  $\beta$ , we obtain from (27)

$$\lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \Psi\left(\frac{\beta}{N}, N\right) = 0. \quad (28)$$

For  $\beta \geq 84$ , the upper bound in (23) is less than 1% ahead of the lower bound. In this way, we can use the approximation (26) even when  $b$  goes linearly down with increasing  $N$ , provided that the factor of proportionality  $\beta$  is large enough. In practice,  $\beta \geq 100$  should be sufficient. We will now make a final adjustment and propose to use the following approximation:

$$E[\gamma_{i_*} | \mathbf{h}] \approx 1 - 2^{-b} \frac{N-1}{N} \geq \frac{1}{N}. \quad (29)$$

This does not change the asymptotic behavior for large  $N$ , but makes the approximation exact for  $b = 0$  since  $E[\gamma_{i_*} | \mathbf{h}]$  is lower bounded by  $1/N$ . By substituting (29) into (22), we finally arrive at the approximation which we will make use of subsequently:

$$E[\text{SNR}_{\text{DL}}] \approx \overline{\text{SNR}}_{\text{DL}}^{\text{max}} \cdot \left(1 - 2^{-b} \frac{N-1}{N}\right). \quad (30)$$

It is interesting to note that from (30),

$$\begin{aligned} (b = 1) &\rightarrow E[\text{SNR}_{\text{DL}}] \\ &\approx \frac{\lim_{b \rightarrow \infty} E[\text{SNR}_{\text{DL}}] + \lim_{b \rightarrow 0} E[\text{SNR}_{\text{DL}}]}{2}, \end{aligned} \quad (31)$$

that is, for 1 bit quantization per antenna, one can already achieve *half* of the maximum possible gain obtainable by the feedback. For large number of transmit antennas, the loss in performance compared to ideal coherent beamforming approaches 3 dB from below, when  $b = 1$  quantization bit per antenna is used.

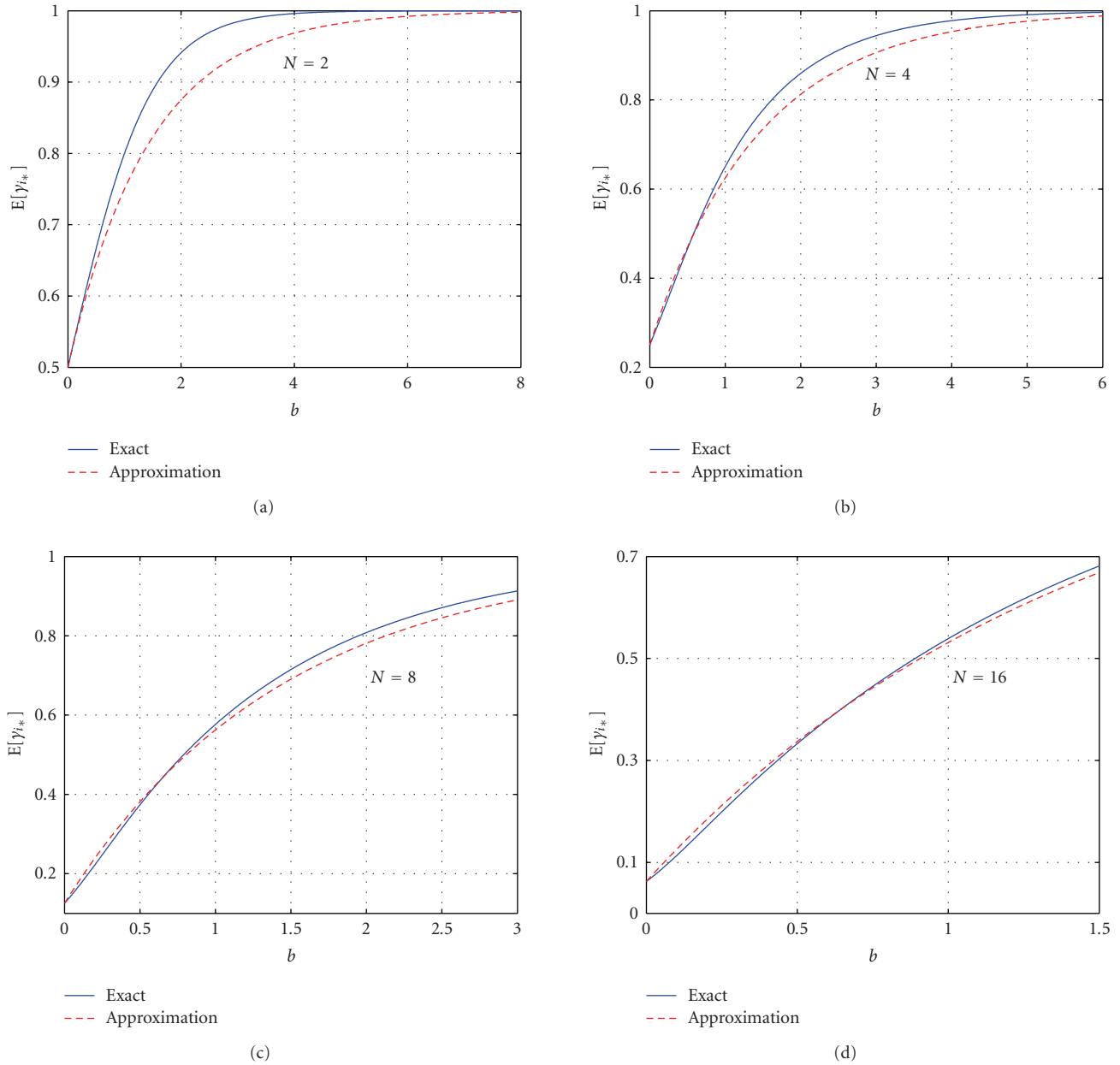


FIGURE 3: Comparison of the exact value of  $E[\gamma_{i_*}]$  from (22) and the approximation from (30).

Before we end this section, let us briefly have a look at the difference between the approximation (30) and the exact solution (22) for the average downlink SNR. We can see in Figure 3 the average relative SNR, that is,  $E[\gamma_{i_*}]$  as a function of the number  $b$  of quantization bits per antenna number for different antenna numbers  $N$ . For small values of  $b$ , particularly for  $b \leq 1$ , the approximation does a fairly good job, even for very small (e.g.,  $N = 2$ ) antenna numbers. For larger values of  $b$ , the approximation requires higher antenna numbers to be reasonably accurate. In practice,  $N \geq 8$  might be sufficient. Note that in the limit  $N \rightarrow \infty$ , the approximation becomes exact for constant  $b$ , and for  $b = \beta/N$ , it becomes exact as also  $\beta \rightarrow \infty$ . We will make use of this property in the next section.

#### 4. OPTIMUM BANDWIDTH PARTITIONING

The results of Section 3.4 on the obtainable average receive SNR in the downlink for a given resolution of random vector quantization will be used now to solve the bandwidth partitioning problem. As our first task, we will compute the optimum quantization resolution, which maximizes the sum throughput of the uplink and the downlink. Second, we show that the product  $B_{DL}T$  has to be above a certain threshold, such that feedback can be used in a beneficial manner. We then proceed to a closed-form solution of the optimum bandwidth partitioning problem. We elaborate on the asymptotic behavior of large antenna numbers, where we also discuss the special cases of symmetrical uplink and

downlink, and a pure downlink system (which uses the whole uplink band for feedback). Finally, we treat the question of optimum SNR and its relationship with the bandwidth partitioning problem.

#### 4.1. Quantization resolution

When we substitute (30) into (13), we find

$$\tilde{b}_{\text{opt}} = \arg \max_b B_{\text{DL}} \cdot \log_2 \left( 1 + \overline{\text{SNR}}_{\text{DL}}^{\text{max}} \cdot \left( 1 - 2^{-b} \frac{N-1}{N} \right) - \frac{Nb}{T\eta} \right). \quad (32)$$

Because the second derivative of the cost function in (32) is negative for  $N > 1$  and all  $b > 0$ , the optimization problem (32) has a unique solution. It can easily be found by solving for the root of the first partial derivative of the cost function with respect to  $b$ , for which we find

$$\tilde{b}_{\text{opt}} = \log_2 \left( \left( 1 + \frac{B_{\text{DL}} T \eta}{N} \right) \cdot \frac{N-1}{N} \cdot \frac{\overline{\text{SNR}}_{\text{DL}}^{\text{max}}}{1 + \overline{\text{SNR}}_{\text{DL}}^{\text{max}}} \right). \quad (33)$$

In order to make this expression better suited to our problem, let us express  $\overline{\text{SNR}}_{\text{DL}}^{\text{max}}$  in terms of the *actual* average downlink SNR that is present for a quantization resolution of  $b = \tilde{b}_{\text{opt}}$ . Using our approximation from (30), we have

$$\overline{\text{SNR}}_{\text{DL}} = \overline{\text{SNR}}_{\text{DL}}^{\text{max}} \cdot \left( 1 - 2^{-\tilde{b}_{\text{opt}}} \frac{N-1}{N} \right), \quad (34)$$

where  $\overline{\text{SNR}}_{\text{DL}}$  is the average SNR in the downlink that we obtain in the optimum  $b = \tilde{b}_{\text{opt}}$ . By substituting (33) into (34), we obtain—after small rearrangements—the following relationship:

$$\overline{\text{SNR}}_{\text{DL}} = \frac{\overline{\text{SNR}}_{\text{DL}}^{\text{max}} (B_{\text{DL}} T \eta / N) - 1}{1 + (B_{\text{DL}} T \eta / N)}, \quad (35)$$

which we can also write in its inverse form:

$$\overline{\text{SNR}}_{\text{DL}}^{\text{max}} = \overline{\text{SNR}}_{\text{DL}} + (1 + \overline{\text{SNR}}_{\text{DL}}) \frac{N}{B_{\text{DL}} T \eta}. \quad (36)$$

By substituting (36) into (33), we obtain for the optimum quantization resolution

$$\tilde{b}_{\text{opt}} = \log_2 \left( \frac{N-1}{N} \left( 1 + \frac{B_{\text{DL}} T \eta}{N} \cdot \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}} \right) \right). \quad (37)$$

The optimum feedback information rate can be written as

$$R_{\text{FB}}^{\text{opt}} = \frac{N \cdot \tilde{b}_{\text{opt}}}{T}. \quad (38)$$

*Example 1.* The following parameters,  $B_{\text{DL}} = 20$  kHz,  $T = 50$  ms,  $N = 4$ , and  $\eta = 0.4$ , yield an optimum resolution of  $\tilde{b}_{\text{opt}} \approx 5.93$  for  $\overline{\text{SNR}}_{\text{DL}} = 4$ . This translates into a feedback information rate of about 474 bps, which is a fraction of

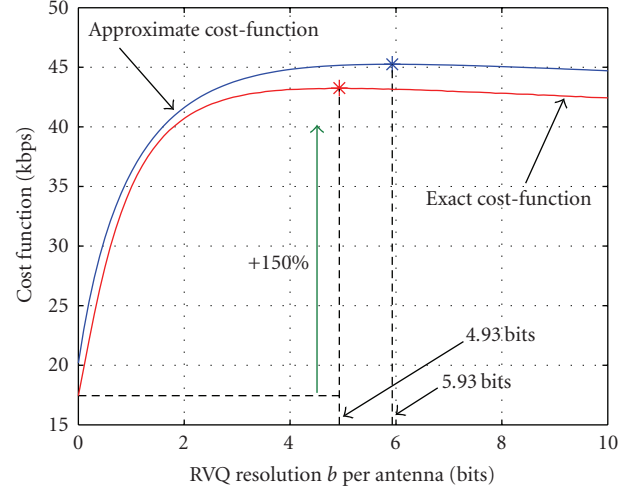


FIGURE 4: Comparison of the exact cost function (no approximations used) from (6), and the approximate cost function from (32). The former is computed numerically. The average SNR in the optimal points (star-shaped markers) is set to  $\overline{\text{SNR}}_{\text{DL}} = 4$  in both cases.

about 1.0% of the *downlink* throughput. (More precisely, this is the fraction of the feedback rate with respect to the downlink throughput of the average channel. Because the latter is an upper bound for the true average throughput, the ratio is (slightly) larger. For i.i.d. Rayleigh distributed fading, the exact ratio turns out to be about 1.07%.)

#### 4.2. Accuracy of the analytical solution

In obtaining the analytical solution (37) for the optimum resolution of the RVQ, we have made use of the two approximations from (11), and (30). While the approximation error can be made arbitrarily small by increasing the number of antennas, there remains, of course, an approximation error for finite—especially low—number of antennas. In order to check how much the proposed solution in (37) deviates from the exact one (which has to be computed numerically), we analyze the example scenario from above. We use the parameters:  $B_{\text{DL}} = 20$  kHz,  $T = 50$  ms,  $N = 4$ ,  $\eta = 0.4$ , and  $\overline{\text{SNR}}_{\text{DL}} = 4$ , when measured at the optimal value of  $b$ . Additionally, we assume i.i.d. Rayleigh fading, in which case we obtain the results displayed in Figure 4. Two curves are shown there as functions of the resolution  $b$  per antenna of the RVQ. The top-most curve corresponds to the cost function from (32) which incorporates the two approximations in (11) and (30). The lower curve shows the cost function from (6), where we use *no* approximations. The latter is computed numerically. The star-shaped markers indicate the optimum resolutions. As can be seen from Figure 4, the analytical solution from (37) slightly overestimates the true optimum resolution (in this case 5.93 bits, instead of 4.93 bits). However, since the maximum of both cost functions is rather flat for values of  $b$  which are *larger* than the respective optimum value, the results obtained from (37) represent a conservative approximation of the true optimum



resolution. From careful observation of the two curves shown in Figure 4, it turns out that the exact cost function, evaluated at the resolution  $b = 5.93$  bits, has dropped by less than 0.2% compared to its maximum value. We conclude that the proposed solution (37) is usable in practice even for as low number of antennas as  $N = 4$ .

### 4.3. Minimum required bandwidth-time product

The solution (37) is valid if and only if  $\tilde{b}_{\text{opt}} > 0$ . This sets a lower limit on the product  $B_{\text{DL}}T$ :

$$B_{\text{DL}}T > \frac{1}{\eta} \cdot \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \cdot \frac{N}{N-1}. \quad (39)$$

Because the feedback has to arrive (much) earlier than the assumed i.i.d. block fading channel changes its realization, that is,  $T \ll T_{\text{dec}}$  has to hold, it follows with (39) that

$$B_{\text{DL}}T_{\text{dec}} \gg \frac{1}{\eta} \cdot \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \cdot \frac{N}{N-1}. \quad (40)$$

### 4.4. Feedback rate for large systems

When we substitute (37) into (38), and multiply both sides by  $T$ , we obtain

$$R_{\text{FB}}^{\text{opt}}T = \log_2\left(\left(1 - \frac{1}{N}\right)^N\right) + \log_2\left(\left(1 + \frac{\alpha}{N}\right)^N\right), \quad (41)$$

where

$$\alpha = B_{\text{DL}}T\eta \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}}. \quad (42)$$

Using  $\lim_{t \rightarrow \infty} (1 + x/t)^t = e^x$ , and  $\lim_{N \rightarrow \infty} \eta = 1$ , it follows that

$$\lim_{N \rightarrow \infty} R_{\text{FB}}^{\text{opt}} = \frac{\alpha - 1}{T \log_2 2} \Big|_{\eta=1}, \quad (43)$$

$$= \log_2(e) \cdot \left( B_{\text{DL}} \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}} - \frac{1}{T} \right). \quad (44)$$

Since  $R_{\text{FB}}^{\text{opt}}$  is increasing with  $N$ , it follows that

$$R_{\text{FB}}^{\text{opt}} < B_{\text{DL}} \log_2 e. \quad (45)$$

The optimum feedback rate remains finite, even for arbitrary large number of antennas or average SNR. With (38), it follows from (44) that

$$\tilde{b}_{\text{opt}} \rightarrow \frac{\beta}{N}, \quad \text{as } N \rightarrow \infty, \quad (46)$$

where

$$\beta = \log_2(e) \cdot \left( B_{\text{DL}}T \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}} - 1 \right). \quad (47)$$

Recall from Section 3.4 that the approximation (30) that was used to arrive at the solution (37) requires  $\beta$  to have a large

value, like  $\beta > 100$ . In practice, this usually represents no problem, since at reasonably large  $\overline{\text{SNR}}_{\text{DL}}$ , say  $\overline{\text{SNR}}_{\text{DL}} = 4$ , already a relatively small bandwidth-time product of  $B_{\text{DL}}T = 88$  will guarantee  $\beta > 100$ . For large  $\beta$ , the term  $1/T$  becomes negligible in (44), so that it follows that

$$\lim_{\beta \rightarrow \infty} \left( \lim_{N \rightarrow \infty} R_{\text{FB}}^{\text{opt}} \right) = B_{\text{DL}} \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}} \log_2(e). \quad (48)$$

Because in the limit  $\beta \rightarrow \infty$  and  $N \rightarrow \infty$  the used approximations (11) and (30) become exact, the result (48) holds *exactly*.

### 4.5. Bandwidth partitioning

Recall from Section 2 that the bandwidth partitioning problem (1) is essentially solved once we know the optimum quantization resolution. By substituting (37) into (7), and applying the approximation (12) also for the uplink, we find that the bandwidth which is optimum to reserve for feedback is given by

$$B_{\text{FB}}^{\text{opt}} = \frac{N}{T\eta} \cdot \frac{\log_2\left(\left((N-1)/N\right)\left(1 + (B_{\text{DL}}T\eta/N) \cdot (\overline{\text{SNR}}_{\text{DL}}/(1 + \overline{\text{SNR}}_{\text{DL}}))\right)\right)}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})}, \quad (49)$$

where  $\overline{\text{SNR}}_{\text{UL}}$  is the average SNR in the uplink.

*Example 2.* For the case  $N = 4$ ,  $\overline{\text{SNR}}_{\text{DL}} = \overline{\text{SNR}}_{\text{UL}} = 4$ ,  $T = 50$  ms,  $B_{\text{DL}} = 20$  kHz, and  $\eta = 0.4$ , we find from (49) that  $B_{\text{FB}}^{\text{opt}} \approx 511$  Hz, or about 2.56% of the downlink bandwidth.

It is somewhat impractical that the optimum feedback bandwidth according to (49) is expressed as a function of the downlink bandwidth  $B_{\text{DL}}$  instead of the totally available bandwidth  $B$ . This problem will, however, be solved in a moment. When we substitute (49) into (9), we obtain

$$B_{\text{DL}}^{\text{opt}} \cdot \left( 1 + \mu \frac{\log_2(1 + \overline{\text{SNR}}_{\text{DL}})}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})} \right) = B - \frac{N}{T\eta} \cdot \frac{\log_2\left(\left((N-1)/N\right)\left(1 + (B_{\text{DL}}^{\text{opt}}T\eta/N) \cdot (\overline{\text{SNR}}_{\text{DL}}/(1 + \overline{\text{SNR}}_{\text{DL}}))\right)\right)}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})}. \quad (50)$$

Note that  $B_{\text{DL}}^{\text{opt}}$  appears both on the left- and the right-hand side of (50). However, it is shown in the appendix that (50) can be solved explicitly for  $B_{\text{DL}}^{\text{opt}}$ :

$$B_{\text{DL}}^{\text{opt}} = \frac{N}{T\eta} \cdot \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \cdot \left( \frac{W((N\Phi/(N-1))(1 + \overline{\text{SNR}}_{\text{UL}})^{\Phi + \text{BT}\eta/N} \log_e(1 + \overline{\text{SNR}}_{\text{UL}}))}{\Phi \log_e(1 + \overline{\text{SNR}}_{\text{UL}})} - 1 \right), \quad (51)$$

where  $W(\cdot)$  is the Lambert W-function [21, 22], and

$$\Phi \stackrel{\text{def}}{=} \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \cdot \left( 1 + \mu \frac{\log_2(1 + \overline{\text{SNR}}_{\text{DL}})}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})} \right). \quad (52)$$

Now that we know  $B_{\text{DL}}^{\text{opt}}$  explicitly as a function of the total bandwidth  $B$ , and the remaining system parameters, we can compute  $B_{\text{UL}}^{\text{opt}}$  immediately as

$$B_{\text{UL}}^{\text{opt}} = B - B_{\text{DL}}^{\text{opt}}, \quad (53)$$

while  $B_{\text{FB}}^{\text{opt}}$  can be computed from (49) by substituting  $B_{\text{DL}}$  by  $B_{\text{DL}}^{\text{opt}}$  from (51):

$$B_{\text{FB}}^{\text{opt}} = \frac{N}{T\eta} \cdot \frac{\log_e((N-1) \cdot W((N\Phi/(N-1))Z^{\Phi+BT\eta/N} \log_e Z) / (N\Phi \log_e Z))}{\log_e Z}, \quad (54)$$

where  $Z$  denotes  $(1 + \overline{\text{SNR}}_{\text{UL}})$ .

The Lambert W function has to be computed numerically. A simple but accurate approximation is given in [23] as follows:

$$W(x) \approx \begin{cases} 0.665 \cdot (1 + 0.0195 \log_e(1+x)) \cdot \log_e(1+x) + 0.04 & \text{for } 0 \leq x \leq 500, \\ \log_e(x-4) - \left(1 - \frac{1}{\log_e x}\right) \cdot \log_e(\log_e(x)) & \text{for } x > 500. \end{cases} \quad (55)$$

For  $x > 500$ , the relative error of (55) is below  $3.3 \times 10^{-4}$ .

*Example 3.* Let  $B = 20$  kHz,  $T = 50$  ms,  $N = 4$ ,  $\overline{\text{SNR}}_{\text{DL}} = 4$ ,  $\overline{\text{SNR}}_{\text{UL}} = 3$ ,  $\eta = 0.4$ , and the symmetry factor  $\mu = 1/2$ . Evaluation of (51), (53), and (54) leads to the following optimum bandwidth partition:  $B_{\text{DL}}^{\text{opt}} \approx 12.32$  kHz,  $B_{\text{UL}}^{\text{opt}} \approx 7.677$  kHz, and finally  $B_{\text{FB}}^{\text{opt}} \approx 523.7$  Hz. Therefore, the resources reserved for feedback consume about 6.8% of the uplink band, which equals about 2.6% of the total bandwidth. With (37) and (38), we can compute that the optimum RVQ should be performed with a resolution of  $\tilde{b}_{\text{opt}} \approx 5.24$  bits per antenna. In total, this amounts to about 21 bits. That means that the optimum RVQ codebook consists of some 2 million, four-dimensional, complex vectors. (If the codebook is precomputed and stored, it would require around 128 MB of memory. If it is generated on the fly, its generation would require about half a second computing time on a high-performance workstation at the time of writing. This shows that for the given example scenario, random vector quantization may not be easy-to-implement.) The optimum feedback rate equals  $R_{\text{FB}}^{\text{opt}} \approx 419$  bps, while the payload throughputs in down and uplink compute to  $\bar{R}_{\text{DL}} \approx 28.6$  kbps and  $\bar{R}_{\text{UL}} \approx 14.3$  kbps, respectively. As a

consequence, the feedback rate amounts to almost 1% of the sum-throughput of uplink and downlink, which equals 42.9 kbps. This is the highest possible sum-throughput that can be achieved with the given system parameters.

#### 4.6. Bandwidth partitioning for large systems

Recall that the approximations (12) and (30) become exact as  $N \rightarrow \infty$  and  $\beta \rightarrow \infty$ . Let us, therefore, have a look at the results for large systems, that is, systems with large number of antennas, and large bandwidth. The latter is necessary to assure that  $\beta$  as given in (47) is also large. By substituting (48) into (7), we obtain by noting that  $\lim_{N \rightarrow \infty} \eta = 1$  that

$$(N, \text{BT}) \rightarrow \infty: \frac{B_{\text{FB}}^{\text{opt}}}{B_{\text{DL}}} = \frac{\overline{\text{SNR}}_{\text{DL}}}{(1 + \overline{\text{SNR}}_{\text{DL}}) \cdot \log_e(1 + \overline{\text{SNR}}_{\text{UL}})}. \quad (56)$$

In the following, we will restrict the discussion to the important special case of

$$\overline{\text{SNR}}_{\text{UL}} = \overline{\text{SNR}}_{\text{DL}} \stackrel{\text{def}}{=} \overline{\text{SNR}}, \quad (57)$$

from which we have

$$(N, \text{BT}) \rightarrow \infty: \frac{B_{\text{FB}}^{\text{opt}}}{B_{\text{DL}}} = \frac{\overline{\text{SNR}}}{(1 + \overline{\text{SNR}}) \cdot \log_e(1 + \overline{\text{SNR}})}. \quad (58)$$

Note that

$$0 < B_{\text{FB}}^{\text{opt}} < B_{\text{DL}}, \quad (59)$$

while

$$B_{\text{FB}}^{\text{opt}} \rightarrow \begin{cases} 0 & \text{for } \overline{\text{SNR}} \rightarrow \infty, \\ B_{\text{DL}} & \text{for } \overline{\text{SNR}} \rightarrow 0. \end{cases} \quad (60)$$

In this way, the optimum amount of bandwidth that has to be reserved for feedback can be varied widely with the average SNR. While for very large SNR, this extra bandwidth becomes very small, it can raise to the size of the downlink bandwidth in case that the SNR is very small. So, what SNR should we choose? It is tempting to define the ‘‘optimum’’  $\overline{\text{SNR}}$  such that the bandwidth for feedback is neither too small nor too large, say, half-way between its minimum and maximum value. Therefore,  $\overline{\text{SNR}}_{\text{opt}}$  has to fulfill the following equation:

$$\frac{\overline{\text{SNR}}_{\text{opt}}}{(1 + \overline{\text{SNR}}_{\text{opt}}) \cdot \log_e(1 + \overline{\text{SNR}}_{\text{opt}})} = \frac{1}{2}, \quad (61)$$

from which  $\overline{\text{SNR}}_{\text{opt}}$  can be computed numerically:

$$\overline{\text{SNR}}_{\text{opt}} \approx 3.92, \quad (62)$$

which equals approximately to 6 dB. We will see in Section 4.7 that  $\overline{\text{SNR}}_{\text{opt}}$  also maximizes the product of bandwidth efficiency and transmit power efficiency, which further motivates to call this SNR the ‘‘optimum’’ SNR. In the

following, we assume  $\overline{\text{SNR}} = \overline{\text{SNR}}_{\text{opt}}$ . From (58), it follows that

$$B_{\text{FB}}^{\text{opt}} = \frac{1}{2} \cdot B_{\text{DL}}. \quad (63)$$

By substituting (63) into (9), and solving for  $B_{\text{DL}}$ , we can write for the optimum downlink bandwidth the following simple expression:

$$(N, \text{BT}) \longrightarrow \infty : \frac{B_{\text{DL}}^{\text{opt}}}{B} = \frac{2}{3 + 2\mu}. \quad (64)$$

Recall that the parameter  $\mu$  is the given ratio between the average throughput in the uplink and the average throughput in the downlink. Since  $B_{\text{DL}} + B_{\text{UL}} = B$ , we can obtain also for the optimum uplink bandwidth a simple expression:

$$(N, \text{BT}) \longrightarrow \infty : \frac{B_{\text{UL}}^{\text{opt}}}{B} = \frac{1 + 2\mu}{3 + 2\mu}. \quad (65)$$

Finally, it follows from (63) and (64) that

$$(N, \text{BT}) \longrightarrow \infty : \frac{B_{\text{FB}}^{\text{opt}}}{B} = \frac{1}{3 + 2\mu}. \quad (66)$$

Notice that for a *pure downlink* system, we have

$$\mu = 0 \longrightarrow (B_{\text{FB}}^{\text{opt}} : B_{\text{DL}}^{\text{opt}} : B_{\text{UL}}^{\text{opt}}) = (1 : 2 : 1), \quad (67)$$

that is, one third of the bandwidth is used for feedback, which occupies the whole uplink band, while the remaining bandwidth is used for the downlink. On the other hand, in a *symmetrical* system, we have

$$\mu = 1 \longrightarrow (B_{\text{FB}}^{\text{opt}} : B_{\text{DL}}^{\text{opt}} : B_{\text{UL}}^{\text{opt}}) = (1 : 2 : 3), \quad (68)$$

that is, one fifth of the total bandwidth is reserved for feedback, which occupies one third of the uplink band, while the remaining bandwidth is equally split for payload in up- and downlink.

#### 4.7. Optimum signal to noise ratio

Let us now have a second look at the “optimum” average SNR as it is implicitly defined in (61). Using the relationship between transmit power  $P_{\text{T}}$  and SNR:

$$\text{SNR} = \alpha \cdot \frac{P_{\text{T}}}{BN_0}, \quad (69)$$

where  $N_0$  is the noise power density, and  $\alpha > 0$  is a constant channel gain, the channel capacity of an additive white Gaussian noise (AWGN) channel is given by

$$C(B, P_{\text{T}}) = B \log_2 \left( 1 + \alpha \cdot \frac{P_{\text{T}}}{BN_0} \right). \quad (70)$$

The bandwidth and transmit power efficiency [17] are defined as

$$\eta_{\text{B}} \stackrel{\text{def}}{=} \frac{C(B, P_{\text{T}})}{B} = \log_2(1 + \text{SNR}), \quad (71)$$

$$\eta_{\text{P}} \stackrel{\text{def}}{=} \frac{C(B, P_{\text{T}})}{\max_B C(B, P_{\text{T}})} = \frac{\log_e(1 + \text{SNR})}{\text{SNR}}. \quad (72)$$

In this way, the bandwidth efficiency quantifies how many bits of information can be transferred per second in the given bandwidth, while the transmit power efficiency tells how much channel capacity is obtained for the given transmit power compared to what could be achieved at most with this transmit power. The equality in (72) follows from

$$\begin{aligned} \max_B C(B, P_{\text{T}}) &= \lim_{B \rightarrow \infty} C(B, P_{\text{T}}), \\ &= \alpha \cdot \frac{P_{\text{T}}}{N_0} \log_2 e, \\ &= B \cdot \text{SNR} \cdot \log_2(e). \end{aligned} \quad (73)$$

Because  $\eta_{\text{B}}$  *increases* with SNR, while  $\eta_{\text{P}}$  *decreases* with SNR, the system becomes less power efficient, when its bandwidth efficiency increases, and vice versa. Therefore, each given SNR corresponds to a specific trade-off between these two fundamental efficiencies. Since both efficiencies are important, the optimum SNR can be defined as the one which maximizes the product of bandwidth and transmit power efficiency:

$$\text{SNR}_{\text{opt}} = \arg \max_{\text{SNR}} \frac{(\log_e(1 + \text{SNR}))^2}{\log_e(2) \cdot \text{SNR}}. \quad (74)$$

By solving for the root of the derivative with respect to SNR, we find that

$$\frac{\text{SNR}_{\text{opt}}}{(1 + \text{SNR}_{\text{opt}}) \cdot \log_e(1 + \text{SNR}_{\text{opt}})} = \frac{1}{2} \quad (75)$$

must hold. Comparing (75) with (61), we can see that the SNR which maximizes the product of bandwidth and transmit power efficiency is the same  $\overline{\text{SNR}}$  which was defined optimum on the grounds of feedback bandwidth in Section 4.6.

## 5. SUMMARY, CONCLUSION, AND OUTLOOK

### 5.1. Summary

An in-depth derivation of an asymptotically exact, analytical solution of the problem of optimum feedback quantization and partitioning of bandwidth in FDD-MISO/SIMO communication systems was presented in this report. While we had to introduce some approximations to facilitate mathematical tractability, the analytical solution is nevertheless asymptotically exact as the number of antennas approaches infinity. Furthermore, it turns out to be a fairly accurate approximation even for systems with only a few antennas.

### 5.2. Conclusion

From the results we may conclude the following:

- (1) The decision on the resolution of channel quantization and the amount of resources reserved for feedback should be based on the ground of a suitable optimization problem rather than done by heuristic ad hoc methods, as it possibly might have been the case in the standardization of past and current mobile communication systems.

- (2) The merits of feedback systems should always be weighted against the loss of resources that the feedback occupies.
- (3) Too less feedback can be more harmful than too much. For instance, we can observe from Figure 4 that for the example system, a resolution of about 5 bits per antenna is optimum. However, increasing the number of bits to, say 10, is much less harmful than decreasing the amount to 2 bits per antenna.
- (4) In the large-system limit, the amount of feedback is pretty large, for instance, 1/3 of the available bandwidth in a pure downlink system.
- (5) Using quantized channel feedback can boost the performance compared to a baseline system which uses no feedback, as can be observed from Figure 4.

### 5.3. Outlook

The presented results have a number of limitations and short-comings. In the following, there is a list—as brief, incomplete, and subjective as it may be—of further directions worthy to explore by the research community in the future.

- (1) The assumption of the i.i.d. block-fading channel should be given up for a more realistic, correlated block-fading channel model. This has direct impact on the quantization, since the correlations allow for predictive quantization.
- (2) While the presented results can easily be generalized to some special multiuser scenarios (like round-robin TDMA), substantial further work is required to cover multiuser systems with channel-aware scheduling.
- (3) Consider space division multiplexing (SDM) and space division multiple access (SDMA).

## APPENDIX

### DERIVATION OF EQUATION (51)

For ease of notation, let us write (50) in the following way:

$$aB_{\text{DL}}^{\text{opt}} + b\log_e(c + dB_{\text{DL}}^{\text{opt}}) + e = 0, \quad (\text{A.1})$$

where

$$a = 1 + \mu \frac{\log_2(1 + \overline{\text{SNR}}_{\text{DL}})}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})}, \quad (\text{A.2})$$

$$b = \frac{N}{T\eta \log_e(1 + \overline{\text{SNR}}_{\text{UL}})}, \quad (\text{A.3})$$

$$c = \frac{N-1}{N}, \quad (\text{A.4})$$

$$d = \frac{N-1}{N} \cdot \frac{T\eta}{N} \cdot \frac{\overline{\text{SNR}}_{\text{DL}}}{1 + \overline{\text{SNR}}_{\text{DL}}}, \quad (\text{A.5})$$

$$e = -B. \quad (\text{A.6})$$

With the substitution

$$B_{\text{DL}}^{\text{opt}} = \frac{1}{d} \left( \exp\left(-x - \frac{ed - ac}{bd}\right) - c \right), \quad (\text{A.7})$$

we can write (A.1) as

$$x \cdot \exp(x) = \frac{a}{bd} \cdot \exp\left(\frac{ac - ed}{bd}\right). \quad (\text{A.8})$$

By denoting with  $W(\cdot)$  the Lambert W-function [21, 22], which is defined by its inverse

$$W^{-1}(x) = x \cdot \exp(x), \quad (\text{A.9})$$

it follows from (A.8) that

$$x = W\left(\frac{a}{bd} \cdot \exp\left(\frac{ac - ed}{bd}\right)\right). \quad (\text{A.10})$$

When we substitute (A.10) into (A.7), we obtain

$$B_{\text{DL}}^{\text{opt}} = \frac{1}{d} \left( \exp\left(-W\left(\frac{a}{bd} \exp(-A)\right) - A\right) - c \right), \quad (\text{A.11})$$

where we have introduced

$$A = \frac{ed - ac}{bd}, \quad (\text{A.12})$$

for ease of notation. By defining

$$y = -W\left(\frac{a}{bd} \exp(-A)\right), \quad (\text{A.13})$$

it follows with (A.9) that

$$A = -\log_e\left(-\frac{bd}{a} y \cdot \exp(-y)\right). \quad (\text{A.14})$$

When we substitute (A.13) and (A.14) into (A.11), we obtain

$$B_{\text{DL}}^{\text{opt}} = \frac{b}{a} W\left(\frac{a}{bd} \exp(-A)\right) - \frac{c}{d}, \quad (\text{A.15})$$

while from (A.12), and (A.2)–(A.6) it follows that

$$A = -\log_e(1 + \overline{\text{SNR}}_{\text{UL}}) \cdot \left( \frac{\text{BT}\eta}{N} + \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \left( 1 + \mu \frac{\log_2(1 + \overline{\text{SNR}}_{\text{DL}})}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})} \right) \right). \quad (\text{A.16})$$

With

$$\Phi \stackrel{\text{def}}{=} \frac{1 + \overline{\text{SNR}}_{\text{DL}}}{\overline{\text{SNR}}_{\text{DL}}} \cdot \left( 1 + \mu \frac{\log_2(1 + \overline{\text{SNR}}_{\text{DL}})}{\log_2(1 + \overline{\text{SNR}}_{\text{UL}})} \right), \quad (\text{A.17})$$

we can write

$$\frac{a}{bd} \exp(-A) = \frac{N\Phi}{N-1} (1 + \overline{\text{SNR}}_{\text{UL}})^{\Phi + \text{BT}\eta/N} \log_e(1 + \overline{\text{SNR}}_{\text{UL}}). \quad (\text{A.18})$$

Substituting (A.18) into (A.15), we finally arrive with (A.2)–(A.5) at the explicit formula for the optimum downlink bandwidth given in (51).

## ACKNOWLEDGMENT

The authors wish to express their sincere thanks to the anonymous reviewers for their effort and comments which ultimately helped in improving the paper.

## REFERENCES

- [1] W. C. Jakes Jr., *Mobile Microwave Communication*, John Wiley & Sons, New York, NY, USA, 1974.
- [2] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, no. 10, pp. 1458–1461, 1999.
- [3] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1423–1436, 1998.
- [4] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [5] R. W. Heath Jr. and A. Paulraj, "A simple scheme for transmit diversity using partial channel feedback," in *Proceedings of the 32nd Asilomar Conference on Signals, Systems & Computers (ACSSC '98)*, vol. 2, pp. 1073–1078, Pacific Grove, Calif, USA, November 1998.
- [6] D. J. Love and R. W. Heath Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Transactions on Communications*, vol. 51, no. 7, pp. 1102–1110, 2003.
- [7] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.
- [8] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 138–159, 1996.
- [9] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.
- [10] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a VQ-based approach," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1101–1112, 2006.
- [11] W. Santipach and M. L. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '04)*, p. 290, Chicago, Ill, USA, June-July 2004.
- [12] W. Santipach and M. L. Honig, "Signature optimization for CDMA with limited feedback," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3475–3492, 2005.
- [13] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, 2007.
- [14] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, 2004.
- [15] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 785–798, 2002.
- [16] W. Santipach and M. L. Honig, "Capacity of beamforming with limited training and feedback," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 376–380, Seattle, Wash, USA, July 2006.
- [17] M. T. Ivrlač, *Wireless MIMO Systems: Models, Performance, Optimization*, Shaker, Aachen, Germany, 2005.
- [18] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and G. Mühlig, *Taschenbuch der Mathematik*, Harri Deutsch, Thun, Germany, 1995.
- [19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, NY, USA, 1972.
- [20] D. Kershaw, "Some extensions of W. Gautschi's inequalities for the gamma function," *Mathematics of Computation*, vol. 41, no. 164, pp. 607–611, 1983.
- [21] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [22] F. Chapeau-Blondeau and A. Monir, "Numerical evaluation of the Lambert W function and application to generation of generalized Gaussian noise with exponent 1/2," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2160–2165, 2002.
- [23] A. Ringwald and F. Schrempp, "QCDINS 2.0—a Monte Carlo generator for instanton-induced processes in deep-inelastic scattering," *Computer Physics Communications*, vol. 132, no. 3, pp. 267–305, 2000.