

A Rate-Splitting Approach to the Gaussian Multiple-Access Channel

Bixio Rimoldi, *Member, IEEE*, and Rüdiger Urbanke

Abstract—It is shown that any point in the capacity region of a Gaussian multiple-access channel is achievable by single-user coding without requiring synchronization among users, provided that each user “splits” data and signal into two parts. Based on this result, a new multiple-access technique called rate-splitting multiple accessing (RSMA) is proposed. RSMA is a code-division multiple-access scheme for the M -user Gaussian multiple-access channel for which the effort of finding the codes for the M users, of encoding, and of decoding is that of at most $2M-1$ independent point-to-point Gaussian channels. The effects of bursty sources, multipath fading, and inter-cell interference are discussed and directions for further research are indicated.

Index Terms—Asynchronous Gaussian multiple-access channel, rate-splitting multiple accessing, successive cancellation, stripping.

I. INTRODUCTION

THE purpose of this paper is to shed new light on the structure of the capacity regions of Gaussian multiple-access channels and to suggest a new and promising strategy for implementing practical systems.

We suppose at first that the channel is discrete-time and frame-synchronous, with power constraint $P = (P_1, \dots, P_M)$, where P_i is the power constraint for user i , and noise variance σ^2 (see Fig. 1). Its capacity region [1]–[8] is the subset of \mathbf{R}^M containing the rate M -tuples (R_1, \dots, R_M) with nonnegative components satisfying

$$\sum_{i \in S} R_i \leq \frac{1}{2} \log_2 \left(1 + \frac{\sum_{i \in S} P_i}{\sigma^2} \right) \quad \forall S \subseteq \{1, \dots, M\}. \quad (1)$$

There are points¹ in this capacity region that are known to be achievable with an implementation complexity substantially less than a general point. These are the points which, after a

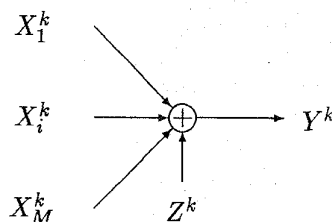


Fig. 1. Gaussian multiple-access channel: The received signal at time k is $Y^k = \sum_{i=1}^M X_i^k + Z^k$, where the noise samples Z^k are i.i.d. zero-mean Gaussian random variables with variance σ^2 , and $X_i^k \in \mathbf{R}$ is the symbol transmitted by sender i .

possible re-indexing, satisfy

$$R_j \leq \frac{1}{2} \log_2 \left(1 + \frac{P_j}{\sigma^2 + \sum_{i < j} P_i} \right), \quad j = 1, \dots, M. \quad (2)$$

For these points, one can design the multiple-access code by constructing one single-user code for each user j , $1 \leq j \leq M$, assuming that user j has power constraint P_j and noise variance $\sigma^2 + \sum_{i < j} P_i$ on his single-user channel. Unless stated otherwise, we will use the term “code” to mean an ideal *random code* so that the code for user i is obtained by selecting its $\lceil 2^{nR_i} \rceil$ codewords by choosing each codeword independently as an i.i.d. sequence of Gaussian random variables with zero mean and variance P_i . Thus the codewords of other users look like Gaussian noise to any given user. Hence, the decoder of user M can decode considering the codewords of user $1, \dots, M-1$ as noise. The contribution of user M can then be removed from the received word, and the procedure can be repeated until each codeword has been decoded. This idea is due to Bergmans and Cover [9] and to Wyner [6], who observed that the vertices of the capacity region satisfy (2) with equality (see the Appendix for a proof of this fact). Another account of this idea was given in [4]. The decoding procedure described above is known variously as *onion peeling*, *stripping*, *successive cancellation*, and *superposition coding*. In the communication-engineering community, the term *interference cancellation* has been used to describe the same decoding process, but the motivation there has been to alleviate the near/far effect in spread-spectrum multiple accessing rather than trying to achieve capacity at lower complexity. In the sequel, we will use the term *single-user coding* to describe the procedure of constructing and assigning a single-user code to each user combined with a

Manuscript received September 9, 1993; revised October 10, 1995. This work was supported by the National Science Foundation under Grants NCR-9357689 and NCR-9304763. The material in this paper was presented in part at the International Symposium on Information Theory, Trondheim, Norway, June 27–July 1, 1994.

B. E. Rimoldi is with the Department of Electrical Engineering, Electronics Systems and Signals Research Laboratory, Washington University, St. Louis, MO 63130 USA.

R. Urbanke was with the Department of Electrical Engineering, Electronics Systems and Signals Research Laboratory, Washington University, St. Louis, MO 63130 USA. He is now with AT&T Bell Laboratories, Murray Hill, NJ 07974 USA.

Publisher Item Identifier S 0018-9448(96)01349-1.

¹Throughout this paper, the terms “points” and “rate tuples” are used interchangeably.

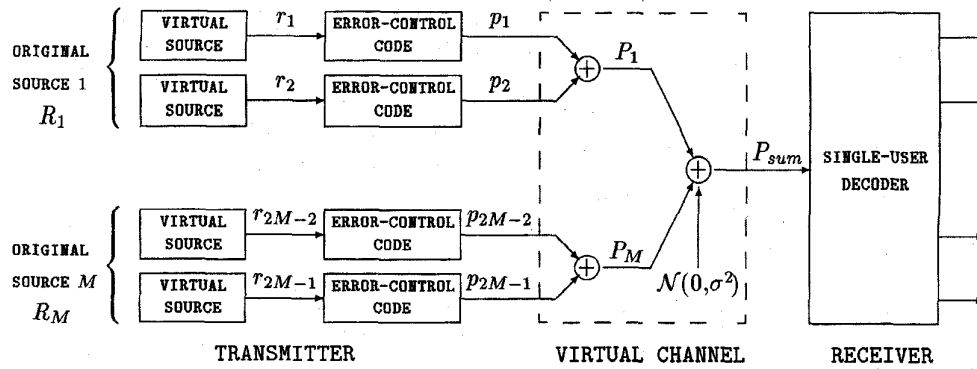


Fig. 2. A multiple-access system based on rate-splitting multiple accessing.

successive cancellation decoder at the receiver. When focusing on the actual decoding procedure, we will use the term *single-user decoding*.

The rate tuples contained in the capacity region but not of the type described above were previously known to be attainable only by one of two methods. These “difficult” rate tuples include the important case where all users have the same power constraint and the same rate. The first method to achieve these difficult rate tuples is joint encoding/decoding of all users. This is very difficult to implement in practice (even for the synchronous case) since random codes have a decoding complexity of the order of $2^{nR_{\text{sum}}}$, where $R_{\text{sum}} = R_1 + \dots + R_M$ is the *sum rate* and n is the block length, and since the construction of (joint) codes in such a way as to approach the achievable rate region seems to be a formidable task. To our knowledge, the only channel model with a nontrivial capacity region for which explicit codes have been constructed to achieve all rate points in the capacity region is the collision channel without feedback studied by Massey and Mathys [10]. For the M -user synchronous binary-input real-adder channel (which is noiseless), a family of asymptotically optimal codes with two codewords per user has been constructed [11]. For a survey on codes for specific multiple-access channels see the introduction in [12]. The second approach to achieving the difficult rate tuples is time-sharing between vertices. This approach can require as many as M multiple-access codes, each multiple-access code requiring M individual codes for each user. Thus this scheme requires on the order of M^2 individual codes. Moreover, time sharing requires synchronization among users, which can be difficult to achieve in practice.²

Quite possibly as a result of the difficulties in achieving general points in the capacity region, the throughput of practical systems is presently far from the ultimate limits. What is generally viewed as the practical approach to CDMA is decoding each user while treating all other users as noise. This decreases the achievable rates considerably, especially when bandwidth is limited. Because a Fourier bandwidth of B

hertz accommodates $2B$ uses per second of the corresponding discrete-time channel, a sum rate R_{sum} in bits per use of the discrete-time channel is equivalent to a bandwidth efficiency of $2R_{\text{sum}}$ bits per second per hertz (b/s/Hz). Thus the bandwidth efficiency of such a practical system with M users, each having the same power P , satisfies

$$2R_{\text{sum}} \leq \lim_{M \rightarrow \infty} M \log_2 \left(1 + \frac{P}{\sigma^2 + (M-1)P} \right) = 1/\log_e 2 \text{ [b/s/Hz]}.$$

There is no such upper bound for an ideal system. Indeed, the bandwidth efficiency may be as high as

$$2R_{\text{sum}} = \log_2 \left(1 + \frac{MP}{\sigma^2} \right) \text{ [b/s/Hz]}$$

which tends to infinity as the number of users increases, even if each user has very limited power. Thus there is a considerable gap between current practice and the theoretical limits.

The situation used to be similar for point-to-point additive white Gaussian noise (AWGN) channels, but advances in the past decade have removed much of the gap between theory and practice. At low rates (fractions of 1 b/s/Hz), “turbo codes” allow one to approach capacity closely with binary codes and antipodal modulation (see also [14] for an information-theoretic treatment of turbo codes). At high signal-to-noise ratios (SNR), the gap reduction has been obtained by a process triggered by Ungerböck [15], and refined by many researchers [15]–[26] (see also [27] and the comprehensive bibliography therein for constant-energy codes). As a tangible result of this evolution, there are commercially available modems that operate close to capacity on telephone lines.

The practical implication of this paper is that the tasks of constructing good codes for an M -user Gaussian multiple-access channel, of encoding, and of decoding can be reduced to those corresponding to a set of up to $2M-1$ independent (and asynchronous) point-to-point AWGN channels. Hence, all of the above mentioned breakthroughs for point-to-point channels can be used to achieve significant performance improvements on multiple-access channels. More precisely, we prove that any point in the capacity region of a Gaussian multiple-access channel can be achieved in the manner shown by the general block diagram in Fig. 2. From the original M -input

²We have omitted frequency-division multiple accessing (FDMA) from this discussion since the corresponding discrete-time channel model does not fall under the general model shown in Fig. 1. FDMA will be discussed in Section V.

channel we create a "virtual channel" with up to $2M - 1$ inputs as shown. Power constraints are assigned to "virtual inputs" so as to satisfy the original constraints. Accordingly, the M original sources are split into $2M - 1$ "virtual sources." Virtual sources are interfaced to virtual inputs via usual point-to-point error-control codes for AWGN channels. The receiver performs single-user decoding. Proving that the rate tuple $R = (R_1, \dots, R_M)$ can be achieved as shown in Fig. 2 is equivalent to showing that the rate tuple $r = (r_1, \dots, r_{2M-1})$ is a vertex of the capacity region described by $p = (p_1, \dots, p_{2M-1})$ and σ^2 . A by-product of this result is that the capacity of (discrete-time or continuous-time) Gaussian multiple-access channels is not affected by the lack of synchronism among users. Time-division multiple accessing (TDMA) and frequency-division multiple accessing (FDMA) are based on breaking time and frequency, respectively, into disjoint intervals, one for each user. Our proposed multiple-access technique relies rather on splitting the rate (and power) of single users. For this reason, we will refer to it as *rate-splitting multiple accessing* (RSMA). For the Gaussian case considered in this paper, an equally good name would be *power-splitting multiple accessing*. However, we prefer the former terminology since our technique can be used on arbitrary discrete memoryless channels [28], in particular on those that do not have power constraints.

This paper is organized as follows: In Section II we prove that single-user coding can indeed achieve the vertices of the capacity region. Section III is the core of the paper. There we show that it is always possible to split as shown in Fig. 2 in such a way that $r = (r_1, \dots, r_{2M-1})$ is a vertex of the capacity region described by $p = (p_1, \dots, p_{2M-1})$ and σ^2 . In Section IV we generalize the results of the previous sections to the frame-asynchronous case as well as to continuous-time Gaussian multiple-access channels. It is shown that the capacity region of the (discrete-time or continuous-time) Gaussian multiple-access channel is unaffected by the lack of synchronization among transmitters.

II. VERTEX ACHIEVABILITY BY SINGLE-USER CODING

In this brief section we argue that vertices of the capacity region are indeed achievable via single-user coding. We do this, since the intuitive and elegant arguments used in the past to prove achievability of vertices by means of single-user coding are based on the assumption that, when previous cancellations were successful, the noise is Gaussian. This is true for the first step in the decoding process, but in the event that the first decoding step was successful, the noise must be in the successful decoding region of this decoder and, hence, will no longer be Gaussian-distributed for the following decoding steps. This creates difficulties in bounding the probability of error of the second (and all subsequent) decoders. In this section we show that it is nevertheless easy to bound the overall probability of error (as opposed to the probability of error of each decoder.)

For clarity, we first treat the case shown in Fig. 3(a) of single-user decoding for a two-user multiple-access system. In this figure, X_i and $D_i, i = 1, 2$, represent the transmitted

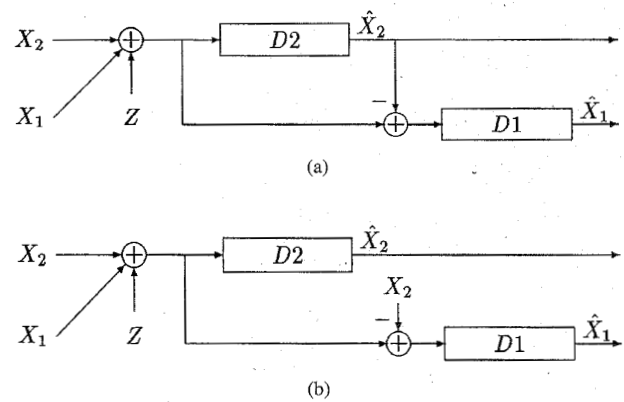


Fig. 3. (a) Single-user decoder. (b) Genie-aided decoder.

codeword and the decoder of user i , respectively, and Z represents the noise. We next consider the model shown in Fig. 3(b) for which X_2 is always removed from the input to $D1$, but everything else is identical in both models. Conceptually, we follow the lead of Wozencraft and Jacobs [29, p. 419] and postulate a genie who always knows the codeword of user 2 and who assists $D1$. In this case, the standard argument [1, p. 242] can be applied to show that $\Pr\{E_i\}$ goes to zero as the block length goes to infinity, where $E_i, i = 1, 2$, is the event that decoder i is producing a wrong estimate. By the union bound

$$\Pr\{E_1 \cup E_2\} \leq \Pr\{E_1\} + \Pr\{E_2\}. \quad (3)$$

Hence, the overall probability of error $\Pr\{E_1 \cup E_2\}$ for the model of Fig. 3(b) goes to zero. But the overall probability of error is the same for both models. This is true, since when $D2$ fails, $E_1 \cup E_2$ occurs regardless of whether E_1 occurs or not. Hence, somewhat surprisingly, the genie does not affect the overall probability of error. The argument easily generalizes to an M -user situation.

The following technical subtlety is worth pointing out. The code of a single user with small average error probability can be converted into a code with small maximum error probability (and nearly the same rate) by the standard procedure of deleting the worst half of all codewords [1, p. 202]. For the present case, we can apply the same procedure to transform the code of each user to one having small maximum error probability, where the error probability of a particular codeword is defined to be the error probability averaged over all noise samples and averaged over all combinations of yet undecoded codewords (with a uniform distribution on these combinations). Nevertheless, for a particular choice of as yet uncanceled codewords, the error probability (now averaged over the noise samples alone) might be large. We note that, for practical purposes, this is not an issue, since in a well-designed system there will be source coding which ensures an approximately uniform distribution over codewords.

III. RATE-SPLITTING MULTIPLE ACCESSING ACHIEVES ANY POINT IN THE CAPACITY REGION

In the previous section, we saw that vertices of a capacity region have the desirable property of being single-user cod-

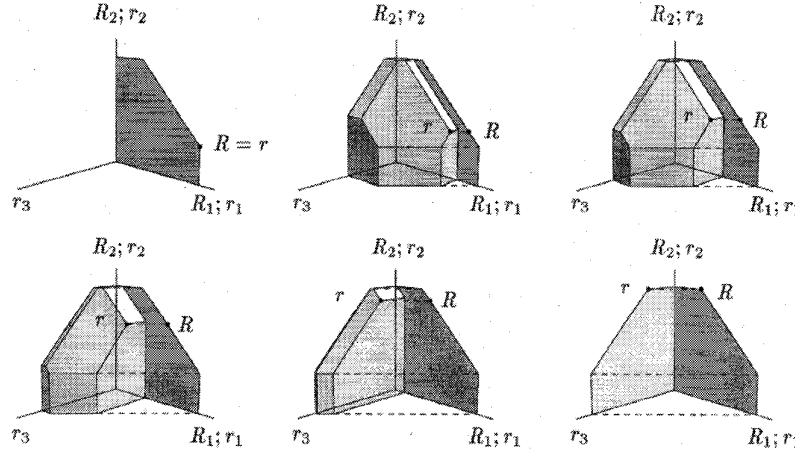


Fig. 4. Geometric relationship between $R = (R_1, R_2)$ and $r = (r_1, r_2, r_3)$ for fixed power constraint P and σ^2 and various points on the dominant face of the capacity region. Note that $r_1 + r_3 = R_1$.

able. First note that any point in a capacity region is dominated (i.e., is smaller with respect to the natural partial order on \mathbf{R}^M) by some point on the *dominant face* which is defined as the set of points that fulfill (1) with equality when $S = \{1, \dots, M\}$. Hence, we can restrict our attention to such points, and we will call them *dominant face points*. In this section, we will show that *any* point in the capacity region of a synchronous discrete-time Gaussian multiple-access channel is achievable by rate-splitting multiple accessing.

Define $C(P, \sigma^2)$ as the capacity of a white Gaussian noise channel with power P and noise σ^2 , i.e.

$$C(P, \sigma^2) = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2} \right).$$

One can easily verify that

$$C(P, \sigma^2) = C(P_1, \sigma^2) + C(P_2, \sigma^2 + P_1) \quad (\text{chain rule})$$

is valid for all nonnegative numbers P_1, P_2 , and σ^2 such that $P = P_1 + P_2$. The chain rule says that the rate R of a single user transmitting at capacity can be seen as a vertex

$$(R_1, R_2) = (C(P_1, \sigma^2), C(P_2, \sigma^2 + P_1))$$

of a two-user *virtual channel* described by (P_1, P_2) and σ^2 , where $P_1 + P_2 = P$ and $R_1 + R_2 = R$. While the usefulness of splitting the user of a point-to-point channel into two virtual users may be limited to suggesting that a higher rate code can be obtained from two lower rate codes, this observation is the seed of our main result.

More illustrative of the idea behind rate-splitting multiple accessing is the two-user case, which we now consider. Let the noise power σ^2 and power constraint (P_1, P_2) be arbitrary, and consider any rate tuple (R_1, R_2) in the dominant face, i.e., such that

$$\begin{aligned} R_1 &< C(P_1, \sigma^2) \\ R_2 &< C(P_2, \sigma^2) \\ R_1 + R_2 &= C(P_1 + P_2, \sigma^2). \end{aligned} \quad (4)$$

Let $\delta > 0$ be the unique number which satisfies³

$$R_2 = C(P_2, \sigma^2 + \delta). \quad (5)$$

Now we define a new Gaussian multiple-access channel with noise power σ^2 and three *virtual inputs*. Let the power constraint (p_1, p_2, p_3) for this channel be given by

$$\begin{aligned} p_1 &= \delta \\ p_2 &= P_2 \\ p_3 &= P_1 - \delta \end{aligned}$$

and define the rate tuple (r_1, r_2, r_3) by

$$\begin{aligned} r_1 &= C(p_1, \sigma^2) \\ r_2 &= C(p_2, \sigma^2 + p_1) \\ r_3 &= C(p_3, \sigma^2 + p_1 + p_2). \end{aligned}$$

By definition (see (2)), the above rate tuple is a vertex of the capacity region of the new channel and, therefore, is achievable via single-user coding. Also, *virtual user 2* has the same rate and power constraint as the original user 2. Moreover

$$\begin{aligned} r_1 + r_2 + r_3 &= C(p_1 + p_2 + p_3, \sigma^2) \\ &= C(P_1 + P_2, \sigma^2) = R_1 + R_2 \end{aligned} \quad (6)$$

where the first equality follows from the chain rule (used twice). Subtracting $r_2 = R_2$ from both sides, we obtain $r_1 + r_3 = R_1$. Since also $p_1 + p_3 = P_1$, we see that we may think of the original user 1, with rate R_1 and power P_1 , as being split into two independent virtual users with rates (r_1, r_3) and powers (p_1, p_3) . The geometric relationship between $R = (R_1, R_2)$ and $r = (r_1, r_2, r_3)$ is illustrated in Fig. 4 for various dominant face points R .

To consider the general case, let the quadruple (M, P, R, σ^2) denote the point $R = (R_1, \dots, R_M)$ in the capacity region of the M -user Gaussian multiple-access channel with power constraint $P = (P_1, \dots, P_M)$ and noise variance σ^2 . Given

³Alternatively, we could have chosen δ to fulfill with equality the *first* inequality in (4).

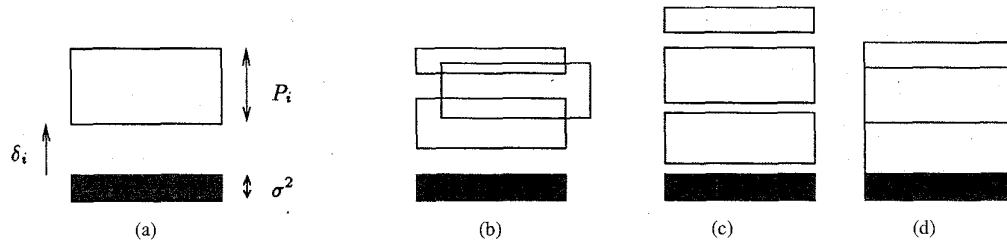


Fig. 5. Convenient representation: (a) represents user i . The bottom (bold) box represents the noise and its height is proportional to the noise power σ^2 . The user is represented by the white rectangle whose height and position are proportional to the power constraint P and to δ , respectively. The nonzero widths of rectangles and their horizontal positions (as in (b)) are chosen for convenience and bear no information; (b)–(d) represent several (three) users in a similar way.

the quadruple (M, P, R, σ^2) we define $\delta_i, i = 1, \dots, M$, as the unique nonnegative number that satisfies

$$R_i = C(P_i, \delta_i + \sigma^2).$$

In words, $\delta_i + \sigma^2$ is the maximum variance of a Gaussian noise that a user with power P_i and rate R_i can tolerate on a single-user channel. We see that for a given power P_i , the rate R_i determines δ_i and *vice versa*. To each user i we associate a representation as shown in Fig. 5(a). In general, labels can and will be omitted since they are implicitly determined by the height and vertical position of the rectangles. Fig. 5(b)–(d) shows several possible scenarios for $M = 3$. To emphasize the underlying graphic representation, we will speak of the *configuration* (M, P, R, σ^2) .

Using (2) we see that the configuration (M, P, R, σ^2) is single-user codable if, after a possible re-indexing

$$\delta_{i+1} \geq \delta_i + P_i, \quad i = 0, 1, \dots, M-1$$

where we defined $\delta_0 = 0$ and $P_0 = \sigma^2$.

As mentioned at the beginning of this section, we need only consider rate tuples on the dominant face of the capacity region. The corresponding configurations will be called *tight*. For instance, the configuration in Fig. 5(b) is not single-user-codable but could be tight (it depends on the actual values), the configuration in Fig. 5(d) is both single-user-codable and tight, whereas the configuration in Fig. 5(c) is single-user-codable but not tight. We give a formal proof of this last fact since it is a crucial point in the following theorem. What we prove is that for a tight configuration there are at least two users that are either “adjacent” (as in Fig. 5(d)) or “overlap” (as the two “top” users in Fig. 5(b)).

Lemma 1: For a tight configuration (M, P, R, σ^2) , after a possible re-indexing, there exists at least one index i for which

$$\delta_i \leq \delta_{i+1} \leq \delta_i + P_i. \quad (7)$$

Proof: We re-index so that $0 = \delta_0 \leq \delta_1 \leq \delta_2, \dots, \leq \delta_M$, and assume that the claim is false, i.e., that

$$\delta_{i+1} > \delta_i + P_i, \quad i = 0, 1, \dots, (M-1). \quad (8)$$

It follows that

$$\begin{aligned} \sum_{i=1}^M R_i &= \sum_{i=1}^M C(P_i, \sigma^2 + \delta_i) \\ &\stackrel{(a)}{<} \sum_{i=1}^M C\left(P_i, \sigma^2 + \sum_{j<i} P_j\right) \\ &\stackrel{(b)}{=} C\left(\sum_{i=1}^M P_i, \sigma^2\right) \end{aligned}$$

which contradicts the assumption that the configuration is tight, i.e., that the point is on the dominant face of the capacity region. In the above chain of equations (a) follows from (8) and the strict monotonicity of $C(\cdot, \cdot)$ with respect to the second parameter, and (b) follows from a repeated application of the chain rule. \square

The following lemma will also be useful in the proof of our main result.

Lemma 2: Consider a quadruple (M, P, R, σ^2) and assume that

$$\delta_j \geq \delta_i + P_i \quad (9)$$

for some $i, j \in \{1, 2, \dots, M\}$. Let δ be the unique nonnegative number such that $R_i + R_j = C(P_i + P_j, \delta)$. Then

$$\delta_i \leq \delta \leq \delta_j - P_i. \quad (10)$$

In words, if two “nonoverlapping” users (like the “top” and the “bottom” user in Fig. 5(b)) are combined by putting their power together, in order to keep their sum rate constant, the bottom user has to “move up” (first inequality in (10)) and the top user has to move “down” (second inequality in (10)).

Proof: The first inequality in (10) follows from

$$\begin{aligned} C(P_i + P_j, \delta) &= R_i + R_j = C(P_i, \delta_i) + C(P_j, \delta_j) \\ &\leq C(P_i, \delta_i) + C(P_j, \delta_i + P_i) \\ &= C(P_i + P_j, \delta_i). \end{aligned}$$

Note that strict inequality in (9) implies strict inequality in (10). The second inequality in (10) follows from

$$\begin{aligned} C(P_i + P_j, \delta) &= R_i + R_j = C(P_i, \delta_i) + C(P_j, \delta_j) \\ &\geq C(P_i, \delta_j - P_i) + C(P_j, \delta_j) \\ &= C(P_i + P_j, \delta_j - P_i). \end{aligned} \quad \square$$

The notion of splitting users into virtual users and channel inputs into virtual channel inputs in a meaningful way is formalized by the following definition.

Definition 1: The quadruple (m, p, r, σ^2) is a *spinoff* of (M, P, R, σ^2) if there exists a surjective mapping $\phi: \{1, \dots, m\} \rightarrow \{1, \dots, M\}$ such that for all $i \in \{1, \dots, M\}$ we have

$$P_i \geq \sum_{j \in \phi^{-1}(i)} p_j$$

and

$$R_i \leq \sum_{j \in \phi^{-1}(i)} r_j$$

where $\phi^{-1}(i)$ is the set of all $j \in \{1, \dots, m\}$ that map into i by means of ϕ .

Theorem 1: For every M -user configuration (M, P, R, σ^2) , there exists a spinoff (m, p, r, σ^2) which is single-user-codable. Moreover, one can guarantee that one user is un-split and that no user is split into more than two virtual users. Hence $m \leq 2M - 1$. How to justify it?

Proof: Without loss of generality we may assume that (M, P, R, σ^2) is tight. The proof is by induction on M . For $M = 1$ the claim is trivially true and for $M = 2$ it was proved earlier in this section. We now assume that it is true for an arbitrary positive M and consider the case of $M + 1$ users.

After a possible re-indexing, from Lemma 1 we may assume that

$$\delta_M \leq \delta_{M+1} \quad (11)$$

$$\delta_{M+1} \leq \delta_M + P_M \quad (12)$$

i.e., that users M and $(M + 1)$ overlap. We reduce the original $(M + 1)$ -user configuration $(M + 1, P, R, \sigma^2)$ to an M -user configuration $(M, \hat{P}, \hat{R}, \sigma^2)$ by defining

$$\hat{P}_i = P_i, \quad i = 1, 2, \dots, M - 1$$

$$\hat{P}_M = P_M + P_{M+1}$$

$$\hat{R}_i = R_i, \quad i = 1, 2, \dots, M - 1$$

$$\hat{R}_M = R_M + R_{M+1}.$$

By the induction hypothesis, there exists a spinoff (m, p, r, σ^2) of $(M, \hat{P}, \hat{R}, \sigma^2)$ with $m \leq 2M - 1$ and with the property that one user is un-split and no user is split into more than two virtual users. Assume first that combined user M with power \hat{P}_M now consists of two virtual users with powers $p_a \geq 0$ and $p_b \geq 0$, respectively, where $\hat{P}_M = p_a + p_b$. The situation is depicted in Fig. 6. It remains to be shown that it is possible to split these two virtual users into (at most) four parts so as to create (at most) two virtual users for each of the original users M and $M + 1$. By the chain rule for any such split, the sum power and the sum rate, taken over all four parts, are constant and equal to $P_M + P_{M+1}$ and $R_M + R_{M+1}$, respectively. Hence, it is enough to show that for one original user, say user M , the split leads to (at most) two virtual users whose sum power is P_M and whose sum rate is R_M .

Consider the sequences of splits shown in Fig. 7 for the four cases (i) $p_a \geq P_M, p_b \geq P_M$, (ii) $p_a \geq P_M, p_b \leq P_M$, (iii) $p_a \leq P_M, p_b \geq P_M$, and (iv) $p_a \leq P_M, p_b \leq P_M$, in

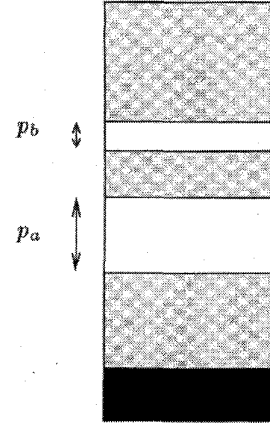


Fig. 6. Situation after the induction step. Boundaries are shown only for the users of interest.

which, for clarity, all but the two virtual users of interest have been omitted. In each frame of each sequence, gray represents the virtual user(s) associated with original user M and, for ease of reference, it (they) will be called the offspring(s) of user M . The offspring(s) of original user $(M + 1)$ is (are) shown in white. Nowhere in these sequences there are more than two offsprings for each of the two original users. In each frame the sum power over the offsprings of the original user $i, i = M, M + 1$, is constant, however the sum rate, hereafter called ρ_i , is not. For each of the four cases, when going from the constellation (a) to the constellation (b), ρ_M decreases continuously from its maximum to its minimum value.

We now argue that for the constellations (a) in Fig. 7, $\rho_M \geq R_M$. We need only distinguish the two cases $p_a \geq P_M$ and $p_a \leq P_M$. We start with the case $p_a \geq P_M$. For convenience Fig. 7ii(a) has been redrawn in Fig. 8(a). Assume that the claim is false, i.e., that for the situation in Fig. 8(a), $\rho_M < R_M$ and thus that $\rho_{M+1} > R_{M+1}$. Equivalent conditions are

$$\Delta_M > \delta_M \quad (13)$$

$$\Delta_{M+1} < \delta_{M+1} \quad (14)$$

where Δ_i is defined by $\rho_i = C(P_i, \Delta_i), i = M, M + 1$. Combining the two nonoverlapping offsprings of the original user $(M + 1)$ (see Fig. 8(a) and (b)), we conclude from the first inequality in (10) (with $\delta_i = \Delta_M + P_M$ and $\delta = \Delta_{M+1}$) that

$$\Delta_{M+1} \geq \Delta_M + P_M.$$

Thus using (13) and (14) (see Fig. 8(c))

$$\delta_{M+1} > \Delta_{M+1} \geq \Delta_M + P_M > \delta_M + P_M$$

which contradicts (12). The case $p_a \leq P_M$ can be handled in a similar manner using the second inequality in (10).

With obvious changes to the above argument, using (11) instead of (12), one shows that for the configurations (b) in Fig. 7, $\rho_M \leq R_M$.

By the previously observed continuity of ρ_M there is a configuration in Fig. 7 for which $\rho_M = R_M$. It follows that also $\rho_{M+1} = R_{M+1}$.

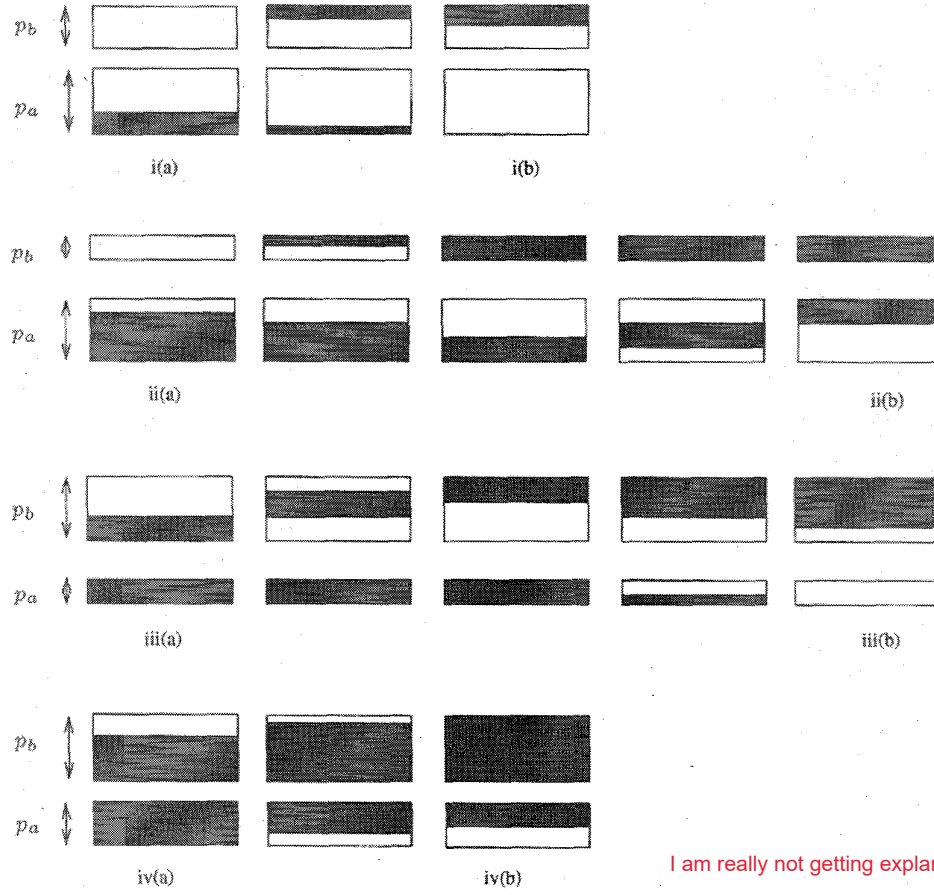


Fig. 7. How to vary the sum rate ρ_M associated to the gray offspring(s) continuously from its minimum to its maximum value without creating more than two virtual users for each of the original users M and $M+1$, for the four cases (i) $p_a \geq P_M, p_b \geq P_M$, (ii) $p_a \geq P_M, p_b \leq P_M$, (iii) $p_a \leq P_M, p_b \geq P_M$, and (iv) $p_a \leq P_M, p_b \leq P_M$.

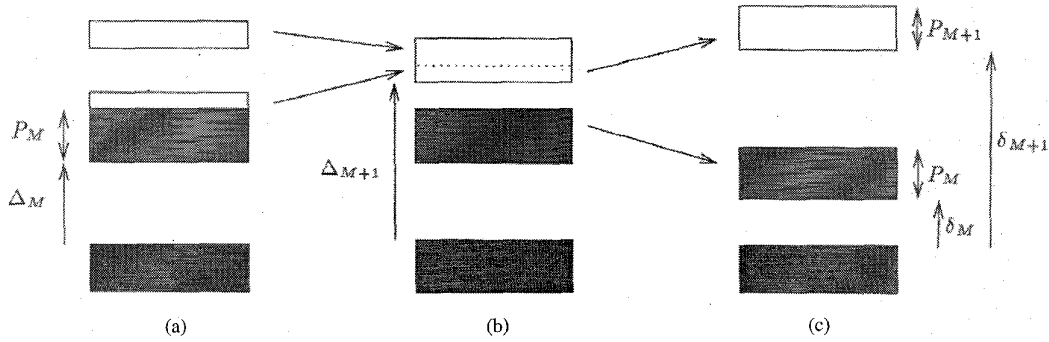


Fig. 8. Proof, by contradiction, that in Fig. 7(a) $\rho_M \geq R_M$.

If, after the induction step, combined user M is un-split, then we need to show that this user can be split into (at most) three parts so as to leave one of the original users un-split. This situation is subsumed in the last three configurations of Fig. 7ii with $p_b = 0$. \square

IV. GENERALIZATIONS

We have seen that any point in the capacity region of a discrete-time synchronous Gaussian multiple-access channel

is achievable via rate-splitting multiple accessing (RSMA) at relatively low coding complexity. A waveform channel may be reduced to a discrete-time synchronous channel if transmitters have access to a common clock. This is hard to achieve in practice. However, the fact that each single-user decoder considers yet uncanceled users as noise suggests that the existence of a common clock is not necessary, provided that each single-user decoder can synchronize with the corresponding transmitter. This is a much more realistic assumption. The purpose of

this section is to show that the weaker assumption is indeed sufficient.

For discrete (time and alphabet) memoryless channels, the lack of (frame) synchronism has been studied in [30] and [31]. Under the assumption that the signal waveforms have disjoint support (which is a suboptimal situation for band-limited channels), the continuous-time Gaussian multiple-access channel has been investigated in [32] and [33]. The effect of asynchronism on the cutoff rate was studied in [34].

We consider only the two-user case since the M -user case is a straightforward generalization. The discrete-time case is considered first. Here asynchronism means frame asynchronism. More precisely, it means that there is a fixed but arbitrary relative shift among the codewords corresponding to the two users. This shift is assumed to be known to the receiver but unknown to the transmitters. There are two technical details that need to be considered when the codewords are not aligned. First, a specific codeword of the “top” user will in general overlap with two codewords of the “bottom” user. This introduces some “memory.” Second, we will show the existence of a code which has low error probability for *all* possible shifts.

Assume that $R = (R_1, R_2)$ is a vertex of the capacity region of the two-user Gaussian multiple-access channel with power constraint $P = (P_1, P_2)$ and noise variance σ^2 . To be specific, assume

$$\begin{aligned} R_1 &= C(P_1, \sigma^2) \\ R_2 &= C(P_2, \sigma^2 + P_1). \end{aligned}$$

User 2 transmits “on top” of user 1. Let n be the codeword length and $\tau \in \{0, \dots, n-1\}$ be the relative delay between the two users.⁴ We model τ as a discrete random variable uniformly distributed on $\{0, \dots, n-1\}$. Let X_2^l, X_2^{l-1} be two consecutive codewords of user 2 and let X_1^l be the first codeword of user 1 that does not precede X_2^l (Fig. 9). Let E_i^l be the event that an error occurs in decoding X_i^l , $i = 1, 2$. We consider again the two models of Fig. 3 with the obvious modifications. In particular, we now define the overall probability of error to be $\Pr\{E_1^l \cup E_2^l \cup E_2^{l-1}\}$. This probability is the same for both models since, when $E_2^l \cup E_2^{l-1}$ occurs, then $E_1^l \cup E_2^l \cup E_2^{l-1}$ occurs and, when $(E_2^l \cup E_2^{l-1})^c$ occurs, then E_1^l occurs in either both models or in neither. For the genie-aided model we can bound the overall probability of error using the union bound, and the proof will be completed if we can show that given $\epsilon > 0$ there exists codes C_1 and C_2 such that for *any* shift $\tau \in \{0, \dots, n-1\}$ and n sufficiently large $\Pr\{E_i^l\} < \epsilon$, $i = 1, 2$, and $l \in \mathbb{Z}$. This can be shown using standard arguments which we now outline. For details see [1, ch. 10]. For every $n = 1, 2, \dots$, consider the following random coding experiment (for notational simplicity the dependence on n is suppressed):

- 1) Generate codes C_1 and C_2 of size 2^{nR_1} and 2^{nR_2} , respectively, choosing each component of each codeword independently and identically distributed (i.i.d.)

⁴For the purpose of single-user coding, delays which are equal modulo n are equivalent.

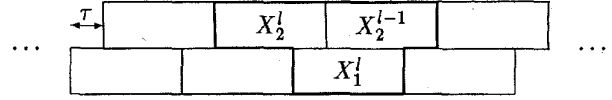


Fig. 9. Relative offset of codewords.

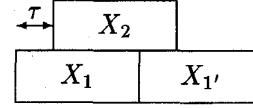


Fig. 10. Random coding experiment.

according to Gaussian distributions with variance P_1 and P_2 , respectively.

- 2) Choose a delay τ in the range $\{0, \dots, n-1\}$ according to a uniform distribution.
- 3) Pick one codeword X_2 for user 2 and two codewords X_1 and X_1' for user 1.
- 4) Transmit these codewords as shown in Fig. 10.
- 5) Assuming the genie-aided decoder, decode X_1 and X_2 by means of a typical sequence decoder, where the second user views the first user as noise.

Let $\overline{\Pr}\{E_i\}$, $i = 1, 2$, be the incurred error probability, where the double bar indicates the averaging over all codes as well as over all shifts. Observe that, although X_2 will in general overlap with X_1 as well as X_1' , and we may have $X_1 = X_1'$, the “noise” experienced by X_2 is still Gaussian and each component is independent of all others, since no component of X_1 will appear twice in the window covered by X_2 .

Since at each level the noise is Gaussian, we can invoke the single-user coding theorem, [1, ch. 10] to conclude that, by choosing the block length n large enough, the average error probability $\overline{\Pr}\{E_i\}$, $i = 1, 2$, can be made as small as desired. By the usual argument [1, ch. 10], there exists at least one code pair C_1, C_2 such that $\overline{\Pr}\{E_i\}$, $i = 1, 2$, is upper-bounded by ϵ , where the bar indicates the average over all shifts.

Instead of the average error probability over all shifts being small we would like each code to have small error probability for *any* shift. For any particular shift (for the codes C_1, C_2 at hand) we must have

$$\Pr\{E_i\} \leq n \overline{\Pr}\{E_i\}.$$

But $\overline{\Pr}\{E_i\}$ and $\overline{\Pr}\{E_i\}$ decrease exponentially with the length n and, hence, so does $\Pr\{E_i\}$ (with the same exponent to the first order).

We will now generalize this idea to continuous-time band-limited (low- or bandpass) asynchronous Gaussian multiple-access channels. We concentrate on the two-user case. Fig. 11 shows the situation for low-pass channels. Let X_i^l , $i = 1, 2$, be the transmitted codeword of user i . The channel is band-limited to $|\omega| < W$. Hence, the impulse response of the ideal transmit filter is given by $p(t) = \sqrt{W/\pi}(\sin(Wt)/Wt)$ and the corresponding symbol interval is $T = \pi/W$. The channel introduces some relative shift τ between the codewords of user 1 and 2 as well as additive white Gaussian noise $n(t)$

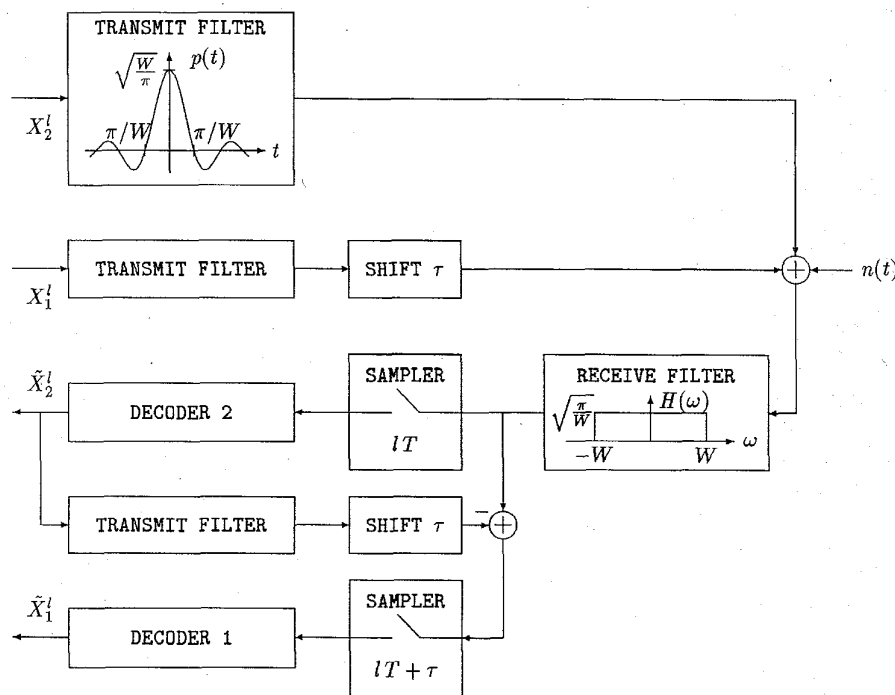


Fig. 11. The continuous-time band-limited asynchronous two-user Gaussian multiple-access channel.

with two-sided power spectral density $N_0/2$. The receiver for user 2 first band-limits the received signal to $\omega \in [-W, W]$ and then applies Nyquist sampling, taking one sample every T seconds. By assumption, the receiver of each user knows of the corresponding time shift and, hence, can choose the sampling times accordingly. Therefore, we can assume that the receiver of user 2 samples at times $t = lT, l = \dots, -1, 0, 1, \dots$. This yields $X_1^{l,\tau} + X_2^l + Z^l$, where X_2^l is the transmitted codeword of user 2, Z^l is an n -tuple of i.i.d. Gaussian random variables with variance $N_0/2$, and $X_1^{l,\tau}$ is the n -tuple stemming from the codeword of user 1. Let \hat{X}_2^l be the estimate of decoder 2. Then the estimated contribution of user 2 can be subtracted from the received signal and the result is fed into decoder 1. Note that now each component of $X_1^{l,\tau}$ depends on all symbols transmitted by user 1. From this we conclude that selecting a code randomly and then using the same code repeatedly (so that each codeword is used repeatedly) will result in (slightly) dependent components of $X_1^{l,\tau}$. The mathematical solution to this problem is to select a new code randomly for each new transmission. In this case the "noise" caused by $X_1^{l,\tau}$ will be i.i.d. according to $\sim \mathcal{N}(0, P_1)$. Arguments similar to the ones outlined for the discrete-time case can be applied to prove that for any given positive δ and ϵ there exists a (time-varying) code which results in an error probability upper-bounded by ϵ for all but a fraction δ of all shifts in the range $[0, nT)$. From an engineering point of view the above mentioned lack of independence, when using time-invariant codes, is negligible. Also, with real rather than ideal filters and with fixed rather than random codes, one has to deal with engineering judgement anyway.

The same idea applies to the bandpass case. Here each transmitter has an in-phase as well as a quadrature component which are modulated and added. The two transmitters are neither time- nor phase-synchronized. Nevertheless, the same analysis applies if we assume that each receiver can achieve phase synchronization with its corresponding transmitter.

V. CONCLUDING REMARKS

We have suggested a new multiple-access technique for M -user Gaussian multiple-access channels. This technique, called rate-splitting multiple accessing (RSMA), allows one to achieve any rate tuple in the capacity region at comparatively low implementation complexity and without the need for synchronization among users.

The key idea behind RSMA is to split some of the original channel inputs so as to create up to $2M - 1$ virtual inputs, as shown in Fig. 2. Accordingly, sources and their rates are also split to form virtual sources. Hence the name rate-splitting multiple accessing. It turns out that the degree of freedom gained by so doing allows one to reduce the tasks of constructing good codes for the given M -user channel, of encoding, and of decoding to essentially the same tasks but for a set of $2M - 1$ independent (and asynchronous) point-to-point AWGN channels. While the former task seemed to be formidable, very satisfactory solutions exist for the latter.

In the introductory section we pointed out that for the discrete-time Gaussian multiple-access channel one can achieve the "difficult" rate tuples via "time-sharing between vertices." For the continuous-time channel one can also do

“frequency-sharing between vertices.” While both techniques require in general M^2 individual codes, frequency sharing does not require synchronization among users.

It should be pointed out that there is one point in the dominant face of a Gaussian multiple-access channel that can be achieved via frequency-division multiple accessing (FDMA) at low complexity and without requiring a common clock for all transmitters [1, p. 406]. More specifically, if user i has a fraction a_i of the sum power $P_{\text{sum}} = \sum P_i$ and one assigns him a fraction a_i of the total bandwidth B , then his achievable rate is the same fraction a_i of the “sum capacity” $B \log_2 [1 + P_{\text{sum}}/(N_0 B)]$, where N_0 is the (one-sided) power spectral density of the noise. This is the only point in the dominant face of a Gaussian multiple-access channel that can be achieved via FDMA. This point corresponds to the “important” equal-rate point

$$R = (R_1, \dots, R_M)$$

$$R_i = (B/M) \log_2 [1 + P_{\text{sum}}/(N_0 B)], \quad i = 1, \dots, M$$

only for the equal-power distribution

$$P = (P_1, \dots, P_M), \quad P_i = P_{\text{sum}}/M, \quad i = 1, \dots, M.$$

The flexibility of RSMA in allowing one to approach any point on the dominant face may become even more relevant in future communication systems since the desire to integrate various services (video, voice, data, fax, etc.) and the competitiveness created by an increasing number of subscribers will require efficient use of the available resources (bandwidth and power distribution) over a wide range of rate tuples.

What is currently considered as the practical approach to code-division multiple accessing (CDMA) is to spread the spectrum of each user by means of an individual signature and to decode each user viewing all others as noise. This technique, called spread-spectrum multiple accessing (SSMA), decreases achievable rates considerably. As pointed out in the Introduction, the resulting spectral efficiency is upper-bounded by $1/\log_e 2$ (b/s/Hz). There is no such upper bound for RSMA.

A fundamental difference between SSMA and RSMA is that individual signals generated by an SSMA system have large Fourier bandwidth and relatively small Shannon bandwidth (one-half the number of dimensions of the signal space required per second), while the signal generated by a virtual user of an RSMA system has large Fourier and Shannon bandwidths. Hence, RSMA and SSMA are two instances of a CDMA system but, according to Massey’s taxonomy [35], RSMA is not a spread-spectrum system. To appreciate the fundamental difference between a spread-spectrum system and one which is not, and toward a unified and well-defined taxonomy, we recommend [35].

Implicit in the information-theoretic approach to multiple-access channels is the assumption of constant-rate sources. Moreover, the Gaussian multiple-access channel is, of course, an idealized model for real-world channels, mainly because it fails to include the effects of multipath. Finally, in many situations of practical interest one has to deal with the presence of many cells that create inter-cell interference. Now we briefly discuss the impact of these issues on RSMA.

Like SSMA, RSMA is a broadband technique in the sense that each (virtual) user’s signal occupies the entire available bandwidth. Broadband systems enjoy a variety of advantages over narrowband systems. These include: 1) Interference reduction resulting from intermittent transmitters [36]. 2) Cellular reuse factor of 1 [36]. 3) It appears that receivers for broadband systems are able to estimate the channel better than those for narrowband systems (see, e.g., [37], [38]). This is not an issue for the Gaussian multiple-access channel but estimating the channel at the receiver is a challenging problem for time-varying multipath channels. In [39] it has been shown that RSMA allows one to achieve any point in the capacity region of a time-varying multipath channel. 4) Assuming that the channel is known at the receiver (but not at the transmitters), Gallager [37] has shown that, for multipath channels, broadband systems have a capacity region which is strictly larger than that of narrowband systems. (This conclusion may reverse if inter-cell interference exceeds a given threshold and cells do not cooperate in their decoding effort [40].)

A nice property of RSMA (shared by any capacity-achieving system) is the ability of a user to use his “extra” power for the benefit of another user. This can be seen by considering a simple two-user situation. If user 1 is willing to increase his power to the point that he can achieve his rate viewing user 2 as noise, then user 2 can transmit at the same rate as if the channel were entirely dedicated to him. This property can make a difference in mobile communication where the attenuation increases with the fourth power of the distance [41] and, therefore, powers measured at the central station vary over several orders of magnitude when the transmitted powers are constant. In SSMA one controls the transmitted power in such a way that all users are received with (roughly) the same power. This procedure, which is done to cure a weakness of SSMA known as the near/far effect, reduces the capacity region of the channel model. It should be pointed out, however, that various techniques have been considered to eliminate or mitigate the near/far effect (see, e.g., [42], [43] and the references therein for recent publications on the subject.)

An additional benefit from not having to ensure that all received signals have the same strength at the central station is that it avoids a concentration of transmitted power toward the boundary of a cell where the contribution to the interference of other cells is worst.

In this paper, all single-user codes were assumed to be ideal codes in the usual information-theoretic sense (see, e.g., [1]). How RSMA behaves with practical codes is an important question. Another important question is the effect of imperfect cancellation at the receiver. This is a subject requiring further research. Promising preliminary analytic and simulation results can be found in [44].

The key ingredient to implementing practical RSMA systems is a family of good low-rate codes such as those reported in [45] and [46]. The recently discovered turbo-codes have an even greater potential. However, further research is needed to find an entire family of such codes with rates that vary over a broad range.

It is interesting to note that the approach discussed in [45, sec. IV] is a special case of RSMA in which there is no need for splitting, since the desired rate tuple is already a vertex of the capacity region. There, pseudonoise signatures are used to randomize the output of the convolutional code so that the same code can be used for all users. The bandwidth expansion is achieved through coding. We expect that the idea of randomizing practical codes by means of pseudonoise sequences can be successfully applied to general RSMA systems. This is also a subject requiring further research. To our knowledge, the first to point out that it is preferable to "spread" using low-rate codes as opposed to pseudonoise sequences was Hui [47].

While we were preparing the final version of this paper, Beat Keusch brought to our attention a related work. Carleial [48] in his unpublished thesis showed that any point in the two-user capacity region of the Gaussian multiple-access channel can be obtained by splitting either one of the two users into two virtual users. Moreover, he stated that any point in the M -user capacity region can be obtained by similar "superposition coding" but, as he gave no proof or further description of this scheme, it is unclear whether he had also observed that the required number of virtual users increases only linearly with the number of original users and, more specifically, that no user needs to be split into more than two virtual users, which is the main point in this paper.

As a final note, we point out that RSMA is a fundamental technique which is applicable also to all discrete memoryless multiple-access channels. This is shown in [28].

APPENDIX

VERTICES AND SINGLE-USER CODABLE POINTS

In this Appendix we prove that (2) is satisfied with equality if and only if R is a vertex of the capacity region, where we define a vertex of the capacity region to be a point that fulfills with equality M (out of the $2^M - 1$) inequalities in (1). Note that each inequality corresponds to one of the hyperplanes that bound the capacity region. Hence, by definition, a vertex lies in the intersection of M such hyperplanes. Let $U = \{1, 2, \dots, M\}$, $S \subseteq U$, and define

$$R(S) = \sum_{i \in S} R_i,$$

$$P(S) = \sum_{i \in S} P_i.$$

Define the sets

$$S_i = \{j \in N: 1 \leq j \leq i\}, \quad i = 0, \dots, M.$$

If (2) is satisfied with equality then

$$R_i = C(P_i, \sigma^2 + P(S_{i-1})) \quad \text{for } i = 1, \dots, M.$$

Therefore, for all $i \in \{1, \dots, M\}$ we have

$$\begin{aligned} R(S_i) &= \sum_{j=1}^i R_j \\ &= \sum_{j=1}^i C(P_j, \sigma^2 + P(S_{j-1})) \\ &= C(P(S_i), \sigma^2). \end{aligned}$$

Hence, according to our definition, R is a vertex. To prove the converse, we will need the following lemma.

Lemma 3: Consider a point $R = (R_1, \dots, R_M)$ in the M -user capacity region given by $P = (P_1, \dots, P_M)$ and σ^2 . Let S, S' be subsets of U such that $R(S) = C(P(S), \sigma^2)$ and $R(S') = C(P(S'), \sigma^2)$. Then either $S \subseteq S'$ or $S' \subseteq S$.

Proof: (see also [49, pp. 188–192]) From $R(S) = C(P(S), \sigma^2)$ and $R(S' \cup S) \leq C(P(S' \cup S), \sigma^2)$ we get

$$\begin{aligned} R(S' - S) &= R(S' \cup S) - R(S) \\ &\leq C(P(S' - S), \sigma^2 + P(S)). \end{aligned}$$

Hence

$$\begin{aligned} C(P(S' - S), \sigma^2 + P(S)) &= C(P(S'), \sigma^2) - C(P(S' \cap S), \sigma^2) \\ &= R(S') - C(P(S' \cap S), \sigma^2) \\ &= R(S' - S) + R(S' \cap S) - C(P(S' \cap S), \sigma^2) \\ &\leq C(P(S' - S), \sigma^2 + P(S)), \end{aligned}$$

which can just be fulfilled if either $S \subseteq S'$ or $S' \subseteq S$. \square

Assume now that R is a vertex. Hence, let $S_i, i = 1, \dots, M$, be the sets which fulfill (1) with equality. Then from Lemma 3 (by finite induction) we must have (after a possible re-indexing) that $S_i \subset S_{i-1}$, $|S_1| = 1$, $S_M = U$ and $|S_i - S_{i-1}| = 1$. If we define $S_i - S_{i-1} = i$ then

$$R_i = R(S_i) - R(S_{i-1}) = C(P_i, \sigma^2 + P(S_{i-1})),$$

for $i = 1, \dots, M$

which proves the claim. \square

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor S. Verdú for helpful discussions and two anonymous reviewers for their valuable comments. The proof of Theorem 1 given in this paper is a simplification of that given in our submitted manuscript. The simpler proof was formulated by us during joint research with A. Grant and P. Whiting on a more general problem and we gratefully acknowledge the stimulation provided by them. Special thanks go to J. Massey for detailed comments on the final draft of this paper and to his doctoral student, B. Keusch, for bringing Carleial's thesis to our attention.

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

- [3] R. E. Blahut, *Principle and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.
- [4] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124-142, Mar. 1985.
- [5] T. M. Cover, "Some advances in broadcast channels," in *Advances in Communications Systems Theory and Applications* A. Balakrishnan and A. Viterbi, Eds. New York: Academic Press, 1975, pp. 229-260.
- [6] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2-10, Jan. 1974.
- [7] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. on Information Theory*, (Tsahkadsor, Armenian S.S.R., 1971). Budapest, Hungary: Hungarian Academy of Science, 1973, pp. 23-52.
- [8] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elec. Eng., Univ. of Hawaii, 1972.
- [9] P. P. Bergmans and T. M. Cover, "Cooperative broadcasting," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 317-324, May 1974.
- [10] J. L. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 192-204, Mar. 1985.
- [11] S. C. Chang and E. J. Weldon, "Coding for T -user multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 684-691, Nov. 1979.
- [12] P. Mathys, "A class of codes for a T active users out of N multiple-access communication system," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1206-1219, Nov. 1990.
- [13] C. Berrou, A. Galvieux, and P. Thitimajhima, "Near Shannon limit error correcting coding and decoding: turbo-codes," in *Proc. ICC '93* (Geneva, Switzerland, May 1993), pp. 1607-1070.
- [14] S. Shamai (Shitz) and S. Verdú, "Capacity of channels with uncoded side information," preprint.
- [15] G. Ungerböck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55-67, Jan. 1982.
- [16] J. H. Conway and N. J. A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 227-232, Mar. 1982.
- [17] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. JSAC-2, pp. 632-647, Sept. 1984.
- [18] G. D. Forney, Jr., "Coset codes—part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123-1151, Sept. 1988.
- [19] ———, "Coset codes—part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152-1187, Sept. 1988.
- [20] ———, "Multidimensional constellations II: Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 941-958, Aug. 1989.
- [21] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877-892, Aug. 1989.
- [22] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 726-740, July 1990.
- [23] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241-1260, Sept. 1991.
- [24] ———, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281-300, Mar. 1992.
- [25] T. Linder, C. Schlegel, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1735-1737, Sept. 93.
- [26] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," submitted to *IEEE Trans. Inform. Theory*.
- [27] T. Ericson and V. Zinoviev, "Spherical codes generated by binary partitions of symmetric pointsets," *IEEE Trans. Inform. Theory*, vol. 41, pp. 107-129, 1995.
- [28] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "On single-user coding for the discrete memoryless multiple-access channel," presented at the 1995 IEEE Int. Symp. on Information Theory, Whistler, B.C., Canada, Sept. 1995.
- [29] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [30] T. M. Cover, R. McEliece, and E. Posner, "Asynchronous multiple-access channel capacity," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 1409-1413, July 1981.
- [31] J. Y. N. Hui and P. A. Humblet, "The capacity region of the totally asynchronous multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 207-216, Mar. 1985.
- [32] S. Verdú, "Multiple-access channels with memory and with and without frame synchronization," *IEEE Trans. Inform. Theory*, vol. 35, pp. 605-619, May 1989.
- [33] R. S. Cheng and S. Verdú, "The effect of asynchronism on the total capacity of Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 2-13, Jan. 1992.
- [34] P. Narayan and D. L. Snyder, "The two user cutoff rate for an asynchronous and a synchronous multiple access channel are the same," Tech. Rep., Washington Univ., Biomedical Computer Lab., St. Louis, MO, Dec. 1979. Monograph 377.
- [35] J. L. Massey, "Toward and information theory of spread-spectrum systems," in *Code Division Multiple Access Communications* P. A. L. S. G. Glisic, Ed. Boston, Dordrecht, and London: Kluwer, 1995, pp. 29-46.
- [36] A. Viterbi, "Wireless digital communication: a view based on three lessons learned," *IEEE Commun. Mag.*, vol. 29, pp. 33-36, Sept. 1991.
- [37] R. G. Gallager, "An inequality on the capacity region of multiaccess multipath channels," in *Communications and Cryptography: Two Sides of One Tapestry* R. E. Blahut, D. J. Costello Jr., U. Maurer, and T. Mittelholzer, Eds. Norwell, MA: Kluwer, 1994, pp. 129-139.
- [38] A. Viterbi, "The evolution of digital wireless technology from space exploration to personal communication services," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 638-643, Aug. 1994.
- [39] B. Rimoldi, "RDMA for multipath multiple access channels: an optimum asynchronous low complexity technique," presented at the 7th Joint Swedish-Russian Int. Workshop on Information Theory, St. Petersburg, Russia, June 17-22, 1995.
- [40] S. Shamai (Shitz) and A. D. Wyner, "Information theoretic considerations for cellular multiple access channels in the presence of fading and inter-cell interference," in *Proc. 1995 IEEE IT Workshop on Information Theory, Multiple Access, and Queueing*, (St. Louis, MO, Apr. 19-21 1995), p. 21.
- [41] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. W. III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 303-312, May 1991.
- [42] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944-960, July 1995.
- [43] M. K. Varanasi, "Group detection for synchronous Gaussian code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1083-1096, July 1995.
- [44] R. Urbanke, "On multiple-access communication," Ph.D. dissertation, Washington Univ., St. Louis, MO, 1995.
- [45] A. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 641-649, May 1990.
- [46] R. McEliece, S. Dolinar, and F. Pollara, "Some easily analyzable convolutional codes," TDA Progr. Rep. 42-99, Jet Propulsion Lab., Pasadena, CA, 1989.
- [47] J. Y. N. Hui, "Throughput analysis for code division multiple accessing of the spread spectrum channel," *IEEE Trans. Commun.*, vol. COM-2, pp. 482-486, July 1984.
- [48] A. B. Carleial, "On the capacity of multiple-terminal communication networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Aug. 1975.
- [49] S. V. Hanley, "Information capacity of radio networks," Ph.D. dissertation, King's College, Univ. of Cambridge, Cambridge, UK, 1993.