Wireless Netw (2009) 15:1126–1139
DOI 10.1007/s11276-008-0107-4

Maximizing delay-constrained throughput in multi-channel DS-CDMA ALOHA networks through power diversity and successive decoding

Yitzhak Birk · Uri Tal

Published online: 27 February 2008

Abstract We study multi-channel ALOHA networks (e.g., satellite-based networks) for online transaction processing, striving to maximize attainable throughput while meeting a deadline with near certainty. This captures the service provider's fixed costs and per-transaction revenue, the user's delay consciousness and ALOHA's probabilistic nature. Specifically, we consider CDMA channels and successive-decoding receivers. Interestingly, judicious use of power diversity is shown to be extremely effective: with a single transmission, capacity is doubled relative to that with power equalization. With the deadline permitting as few as one or two retransmission attempts upon failure, the probability of not meeting it can be virtually diminished (10^{-5} and 10^{-8}, respectively) while approaching the throughput attainable without delay constraints. This also holds for limited mean transmission power. Thus, the effect of power diversity in conjunction with CDMA depends strongly on the type of receiver and on the exact performance measure, and the proposed approach is worth considering for next-generation systems.

Keywords Multi-channel ALOHA · Deadline-constrained throughput · Successive interference cancellation (SIC) · Wireless transaction processing · ALOHA · Satellite communication · Successive decoding · Successive cancellation

Y. Birk (✉) · U. Tal
Electrical Engineering Department, Technion, Haifa 32000, Israel
e-mail: birk@ee.technion.ac.il

U. Tal
e-mail: urital12@gmail.com

1 Introduction

1.1 Multichannel slotted ALOHA

The ALOHA random access scheme was introduced by Abramson in the 1970s [1], followed by a slotted-time version [16]. It has since been studied extensively. Many studies focused on maximum throughput or the throughput-delay tradeoff. Others considered stability problems (e.g. [9]).

Presently, ALOHA is used almost exclusively for transmitting short messages in networks that utilize a shared channel when propagation delay is larger than message transmission time. (The messages may carry user information, serve for network control, or serve for reservation requests.) In such an environment, short messages render reservation schemes ineffective, and the long round-trip delay precludes the effective use of channel-sensing access schemes.

An important use of slotted ALOHA nowadays is in satellite-based networks used for on-line transactions [8]. These typically comprise many thousands of terminals (VSATs), a central hub, and a satellite that serves as a reflector. Traffic from the terminals to the hub uses multi-channel slotted ALOHA, whereby each transmission takes place over a randomly chosen channel (there are typically many tens of channels). If more than one transmission takes place on a given channel in a given time slot, a collision occurs and none are received. The hub transmits acknowledgments and other information over a contention-free broadcast channel. In multi-channel slotted ALOHA, randomization of retransmission delay is replaced with randomized channel selection.

1.2 Design goal and prior art

1.2.1 Design goal

For the aforementioned uses, Birk and Keren [5] introduced a performance measure that reflects both the user's requirements for delay that does not exceed a certain limit, the network owner's desire to maximize the attainable throughput with given channel resources, and the probabilistic nature of ALOHA: *maximizing attainable throughput ("capacity") while adhering to a maximum permissible delay (deadline) constraint with (at least) a required probability*. The permissible probability of not meeting the deadline, $P_e^{\max}$, is set to be negligible relative to probabilities of error in other links of the application chain.

Let D_r be the maximum number of permissible rounds (transmission attempts). D_r is determined by the deadline constraint and by the round-trip delay. P_e denotes the probability that a packet fails to succeed in D_r or fewer rounds. Since D_r is the maximum permissible number of rounds, failing D_r times results in the loss of the packet (although a higher level protocol may resubmit it). Our goal is thus to maximize the attainable throughput S subject to $P_e < P_e^{\max}$ and D_r.

1.2.2 Prior art

The straightforward use of multi-channel Slotted ALOHA for this purpose entails transmission of a packet over a randomly chosen channel. If an acknowledgment is not received from the hub, the packet is retransmitted over a randomly chosen channel. These *transmission rounds* are repeated until receipt of an ACK or expiration of the deadline. We refer to this as the *Baseline scheme* or as "standard multi-channel slotted ALOHA".

A general alternative approach for maximizing delay-constrained capacity was introduced in [5]: transmission policies whereby the maximum channel-resource expenditure per message is high, while the mean is kept low. The rationale is as follows:

- Spending a *large maximum* effort on a message before giving up on it reduces the probability of its failure to meet the deadline;
- The *low mean* resource expenditure minimizes "pollution," thereby allowing more active users (and thus higher throughput) at any given "working point" (offered load).

A class of policies that apply this approach entails spending a small amount of network resources on a message in the first transmission round, and increasing this amount in late rounds. Late rounds are unlikely to take place (retransmission ceases upon successful reception), so a large amount of network resources can be spent on a message before it is abandoned with little impact on the mean per-message resource expenditure [5]. This and other works, listed below, explored this idea for "narrowband" channels (at most one successful reception over any given channel in a given time slot).

In [5], an optimal *multi-copy* policy was developed: an increasing number of copies of the message are transmitted (over different, randomly chosen channels) in successive transmission rounds until success or deadline expiration. (The number of copies in each round is optimized.)

In [3], a "*multiple working point*" scheme was proposed: disjoint subsets of channels are allocated to different transmission rounds, with less loaded channels allocated to the later rounds. (Channel partition among the groups is optimized.) The "*multi-copy*" and the "*multiple working points*" schemes achieve dramatic improvements relative to the Baseline scheme. It was also discovered that multi-copy is superior, and an optimal combination of the two approaches only slightly outperforms it [3].

In [2], the multi-copy approach of [5] was generalized and extended to multi-slot packets: erasure-correcting codes computed over multiple same-packet fragments replaced replication. This allows any given "extra" (redundant) transmitted fragment to be substituted for *any* single fragment of the original packet fragments that was not received. The successful reception of any message fragment moreover made the hub aware of the need to send the remaining ones, and the hub allocated contention-free slots for them. This "*coding-reservation*" scheme can break the $1/e$ capacity barrier of Slotted ALOHA even with a delay constraint. (Note that collision resolution algorithms require too many rounds!)

1.3 Multiple power levels

1.3.1 Narrowband channels (at most one success per slot)

The use of multiple power levels, selected randomly, has been shown to increase the capacity of various multiple-access schemes, in particular ALOHA [10–12, 19]. The improvement is due to the power capture effect which, in certain cases, permits successful reception of one packet despite the concurrent transmission of others on the same channel. Capacity (no delay constraint) increases of 43% and 70% were achieved with two and three power levels, respectively, and perfect capture [12].

In [6] the judicious use of power diversity for the maximization of *delay-constrained* throughput was studied. Throughput was increased by hundreds of percents, much more than can be achieved for the unconstrained throughput. This gain was achieved by using non-stationary policies of power level selection, whereby higher power levels are used in later rounds to decrease P_e. Optimally combining multiple power levels with multiple copies further increased delay-constrained throughput.

Remark: While increasing the probability of success of a given message, multiple copies "pollute" the channels. Multiple power levels, in contrast, do not pollute the channel and merely serve as a priority mechanism. (This assumes no adjacent-channel interference, a reasonable assumption with the modulations used in actual satellite systems.)

1.3.2 CDMA channels

With the use of CDMA channels, several messages can succeed on the same channel in any given time slot. Unlike the use of power diversity in narrowband channels, which can only improve performance, its use with CDMA and "conventional" (e.g., matched-filter) receivers presents a trade-off: transmitting a message with higher power increases its probability of success; however, the use of multiple power levels reduces the maximum possible number of concurrent successful receptions.

To illustrate this, consider a situation wherein the signal-to-interference ratio must be at least 1:3 for successful reception. If a single power level is used, four concurrent successful receptions are possible. However, if one message is transmitted with power that is five times higher than that of the other messages, only it will succeed, as any other message would see a signal-to-interference ratio of at most 1:5.

In [15] it was shown that for CDMA with a "single-user" matched-filter receiver, power diversity gives little benefit if any at all, depending on system parameters and load. These results are not surprising when considering the poor near-far resistance of the matched filter receiver.

The current work was motivated by the observation that more sophisticated receivers that employ successive decoding can mitigate the problem caused by power diversity. Indeed, the results presented in the following sections for the successive decoding technique show that power diversity, in conjunction with optimized power levels and selection probabilities, can yield a significant increase in the attainable delay-constrained throughput. In most cases, delay-constrained throughput that is very close to the maximum achievable unconstrained throughput can be attained even for extremely small $P_e^{\max}$ values using only two or three rounds.

In this work we consider DS-CDMA with SIC receivers. However, other types of multiuser detection receivers exist which also mitigate the interference caused by higher level transmissions [18]. Additionally, frequency-hopped CDMA (FH-CDMA) has been recently shown to have near-far resistance even with a single-user receiver [17].

The results presented in this paper are the output of an exhaustive search. One may argue that this is too computationally expensive and therefore cannot be used in practice. However, we have succeeded in sharply reducing the required amount of computation by using an efficient numerical approximation; also, the optimal power levels can be precomputed once per deadline value and stored in a table for subsequent use. It should moreover be noted that if the transaction rate is below the maximum attainable one, the use of the precompiled values will support that rate, albeit with suboptimal channel utilization.

The remainder of the paper is organized as follows. In Sect. 2 we present the channel and receiver models. In Sect. 3 we analyze single-round power-selection policies for several Successive-Interference-Cancellation (SIC) receiver types. In Sect. 4, we extend the single-round policies to the case of multiple rounds. The focus in the results is on the reduction in $P_e^{\max}$ under a deadline constraint without sacrificing throughput. Section 5 offers concluding remarks.

2 CDMA and successive decoding

As in [18], we embark from the basic CDMA K-user channel model consisting of the sum of antipodally modulated signature sequences embedded in additive white Gaussian noise. The analysis in this section is for a single CDMA channel, but the actual scheme entails redrawing a channel in each round.

Consider M-bit-long packets of single time-slot duration. The received signal can be expressed as

$$r(t) = \sum_{m=1}^{M}\sum_{k=1}^{K} A_k b_{km} s_{km}(t-(m-1)T-\tau_k) + \sigma n(t), \quad (1)$$

where the bit-duration T is the inverse of the data rate. $s_{km}(t)$ is the spreading sequence assigned to the kth user for the mth bit, normalized so as to have unit power, and defined over the one-bit interval $[0, T]$. A_k is the received amplitude of the kth user's signal. A_k^2 is referred to as the power of the kth user. We assume that the amplitude remains constant throughout the entire packet transmission. $b_{km} \in \{-1, +1\}$ is the mth bit in the M-bit-long packet transmitted by the kth user. $n(t)$ is white Gaussian noise with unit power spectral density. It represents thermal noise.[1]

We assume that the transmitters use random spreading sequences that are known to the receiver. However, in

[1] This paper is motivated by the long delays of geostationary satellites, and the most common setup is not mobile, and using highly directional antennas. For this, the AWGN model and the assumption of amplitude stationarity for the duration of a packet are sensible.

ALOHA, the total number of spreading sequences that the receiver needs to correlate against may be huge, while the actual number of transmissions per time-slot is much smaller. To facilitate the identification of the transmitting stations in each time-slot, one may prefix each packet with a short preamble, transmitted using a single (common) spreading sequence. If the length of the preamble is short enough relative to the length of each packet, the probability of a preamble overlap (leading to reception failure) is negligible in the relevant offered load scenarios. Further details of this issue are beyond the scope of this paper.

The transmissions are assumed to be time-slot synchronous, but they are not assumed to be synchronized at the bit level. $\tau_k \in [0, T]$ is the time-shift "phase" of the kth user relative to an arbitrary time point ($t = 0$).[2] τ_k are i.i.d. random variables, uniformly distributed over the interval [0,T].

Consider K concurrent transmissions. A successive cancellation receiver [7, 18] attempts to detect users in succession by canceling each decoded signal from the aggregate received signal $r(t)$. We assume that the order of detection is according to decreasing order of reliability. The common reliability criterion is simply the power of the output of the matched filter for each user [18]. Without loss of generality, we will henceforth use the term "user k" to refer to the kth transmission in the cancellation order.

We represent the signal used to detect data from user k as $r^{(k)}(t)$, which is the received signal, $r(t)$, after $k-1$ subtractions of previously decoded user signals:

$$r^{(k)}(t) = r(t) - \sum_{i=1}^{k-1} \hat{x}_i(t), \tag{2}$$

where $\hat{x}_i(t)$ is an estimate of the received signal from user i, defined as $x_i(t) = A_i b_i s_i(t - \tau_i)$.

Figure 1 illustrates the structure of a successive cancellation receiver. As can be seen, an estimate from the ith decoding is used for the $(i + 1)st$ decoding.

We consider three successive cancellation schemes:

2.1 Linear SIC receiver

In [7] a *linear* receiver is considered, wherein the estimate of a received signal $\hat{x}_k(t)$ is the projection of the received signal remaining after previous subtractions, $r^{(k)}$, onto the spreading code of user k:

$$\hat{x}_k(t) = z_k s_k(t), \tag{3}$$

where z_k is the matched filter output after previous subtractions, $z_k = \langle r^{(k)}(t), s_k(t)\rangle$.

[2] The actual relative time shift of different transmissions may exceed a single bit time, but is assumed to be negligible relative to packet length. For interference purposes, the "phase" (modulo T) of the time difference suffices.

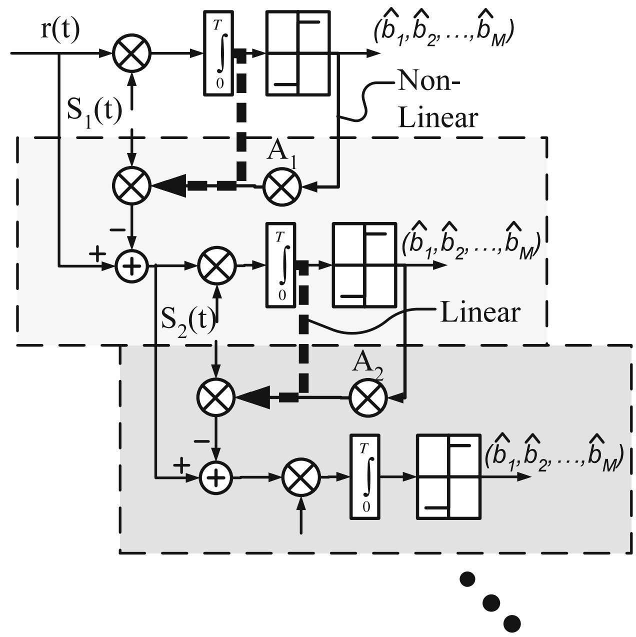

Fig. 1 Three schemes of Successive Interference Cancellation (SIC)

We define the SIR for signal k, Γ_k, as [7]

$$\Gamma_k = \frac{E^2[z_k]}{\mathrm{var}[z_k]} \tag{4}$$

Γ_k is measured in dB, and is the power of the desired signal divided by the power of thermal noise and other interferences. Γ_{thr} is the lowest SIR that still enables successful decoding of a packet by the receiver. Γ_{thr} is a function of the decoding scheme used, and its derivation is beyond the scope of this work.

A more detailed analysis of SIC receivers appears in the appendix. The output of the matched filter (for a specific bit) after subtractions in the *linear* SIC receiver follows from (2) and (3),

$$\begin{aligned} z_k^{lin} &= \left\langle r^{(k)}(t), s_k(t-\tau_k)\right\rangle = \left\langle r(t) - \sum_{i=1}^{k-1} \hat{x}_k(t), s_k(t-\tau_k)\right\rangle \\ &= \left\langle \sum_{i=1}^{K} A_i b_i s_i(t-\tau_i) - \sum_{i=1}^{k-1} z_i s_i(t-\tau_i) + \sigma n(t), s_k(t-\tau_k)\right\rangle \\ &= A_k b_k + \sum_{i=k+1}^{K} \rho_{ik}(\tau_i - \tau_k) A_i b_i \\ &\quad + \sum_{i=1}^{k-1} \rho_{ik}(\tau_i - \tau_k)(A_i b_i - z_i) + n_k. \end{aligned} \tag{5}$$

For asynchronous spreading sequences, the variance of the matched filter output, $\mathrm{var}[z_k^{lin}]$, is given by the recursive formula

$$\mathrm{var}[z_k^{lin}] = \sigma^2 + \frac{1}{3N}\sum_{i=1}^{k-1}\mathrm{var}[z_i^{lin}] + \frac{1}{3N}\sum_{i=k+1}^{K} A_i^2, \tag{6}$$

where we have used known results for the variance of the cross-correlation among random spreading sequences. (See Sect. A in the appendix.)

There are three main contributors to $\mathrm{var}[z_k^{lin}]$: the background noise, σ^2, the residuals from imperfect subtractions, $\mathrm{var}[z_i^{lin}]$, and user signals that have yet to be decoded, A_i^2.

2.2 Non-linear SIC receiver with stop on failure

Instead of using the direct output of the matched filter as in the linear receiver, here we take hard-decision output of the decoder and re-modulate it using an estimated amplitude and the known spreading sequence. $\hat{x}_k(t)$ thus becomes

$$\hat{x}_k(t) = \hat{A}_k \hat{b}_k s_k(t-\tau_k), \tag{7}$$

where $\hat{A}_k$ is the estimated received amplitude and $\hat{b}_k$ is the output of the decoder.

It should be emphasized that the estimate $\hat{b}_k$ is not a simple per-bit hard-decision as is schematically illustrated in Fig. 1. If the SIR is above a certain threshold, Γ_{thr}, we assume that decoding is successful, and that all the bit estimates in the packet are accurate (the amplitude estimation may nonetheless be imperfect). Furthermore, the receiver can perform a CRC to verify a successful decoding. If the SIR is below the threshold, we assume that decoding fails and that the output bits are incomprehensible. Subtracting such a signal introduces a non-uniform noise among different bits, thus significantly complicating the model. To simplify the model, we assume that a packet whose decoding fails is not subtracted. Note that in such a case, the remaining signals inevitably fail because their SIR would be even lower (the decoding order is by decreasing SIR).

The output of the matched filter after subtractions in the *non-linear* SIC receiver similarly follows from (2) and (7):

$$z_k^{non-lin} = A_k b_k + \sum_{i=k+1}^{K} \rho_{ik}(\tau_i - \tau_k) A_i b_i + \sum_{i=1}^{k-1} \rho_{ik}(\tau_i-\tau_k)(A_i b_i - \hat{A}_i \hat{b}_i) + n_k. \tag{8}$$

The use of erasure correcting codes and CRC to check for correct decoding justifies an approximation whereby $\hat{b}_i = b_i$ if $\Gamma_i \geq \Gamma_{thr}$. We use a simple model to account for inaccurate amplitude estimation of successfully decoded signals, by assuming a fixed relative error,

$$\frac{|\hat{A}_k - A_k|}{A_k} \approx \alpha. \tag{9}$$

The SIR numerator, $E^2[z_k^{non-lin}]$, is A_k^2, and the denominator is

$$\mathrm{var}[z_k^{non-lin}] = \frac{1}{3N}\sum_{i=k+1}^{K} A_i^2 + \frac{1}{3N}\sum_{\substack{i<k \\ \Gamma_i<\Gamma_{thr}}} A_i^2 + \frac{1}{3N}\sum_{\substack{i<k \\ \Gamma_i \geq \Gamma_{thr}}} \alpha^2 A_i^2 + \sigma^2. \tag{10}$$

Equation 10 can be solved iteratively, starting from var $[z_1]$.

2.3 Hybrid SIC receiver

To prevent the failure of all the remaining signals after the first decoding failure in the non-linear receiver, we consider another variant of an SIC receiver: when the decoding of a user fails, we subtract it *linearly*, as in (3). This receiver is a combination of the non-linear decision-driven receiver and linear subtractions when decoding fails. Hence we name it *Hybrid SIC receiver*. The variance of z_k becomes

$$\mathrm{var}[z_k^{hyb}] = \frac{1}{3N}\sum_{i=k+1}^{K} A_i^2 + \frac{1}{3N}\sum_{\substack{i<k \\ \Gamma_i<\Gamma_{thr}}} \mathrm{var}[z_i^{hyb}] + \frac{1}{3N}\sum_{\substack{i<k \\ \Gamma_i \geq \Gamma_{thr}}} \alpha^2 A_i^2 + \sigma^2. \tag{11}$$

The next section deals with adapting the above deterministic models to the stochastic scenario of ALOHA and to the probabilistic choice of power level. We first consider a single round, and seek transmission power policies that yield high throughput in the stochastic scenario. Then, in Sect. 4, we explore multi-round power-selection policies.

3 Single-round power selection policies

3.1 Transmission policy and channel working point

The network comprises ground stations that transmit single-slot messages over randomly chosen time-slotted channels. A hub monitors all channels and ACKs all successful receptions. ACKs are sent over separate, contention-free lossless channels. The absence of an expected ACK indicates a destructive collision. With CDMA, several transmissions may succeed and some may fail in one time-slot. By abuse of terminology we say that a transmission collided if it has failed. This is different than in the classical ALOHA model, in which a collision meant that *all* transmissions failed.

We assume that in a single transmission round, each station selects one of m possible power levels, $(W_1, \ldots, W_m)$ with probabilities $(P_1, \ldots, P_m)$, respectively. The number of transmissions with W_i is a Poisson random variable with parameter $G_{w_i} = G \cdot P_i$, with the index corresponding to the power level index. These random variables jointly constitute the *channel working point*. Without loss of generality, $W_i > W_{i+1}$, so W_1 is the highest power level.[3]

A channel working point determines the throughput in the network and the collision probabilities. Let P_{cW_i} denote the collision probability for a transmission with power W_i. P_c is the average collision probability:

$$P_c = \sum_{i=1}^{m} P_i P_{cW_i}. \tag{12}$$

Note that P_{cW_i} depends on the powers chosen by all transmitters, so (12) is not a truly simple expression.

The resulting throughput S is

$$S = G \cdot (1 - P_c). \tag{13}$$

In the following (P_c, S) graphs, each (P_c, S) point is obtained by picking an offered load G and a power-level scheme, and optimizing the powers and selection probabilities to maximize the throughput S. The result is a point on the (P_c, S) graph for that scheme.

In [7], Buehrer analyzed a linear SIC receiver for K concurrent transmissions (deterministic), and derived the power levels that attain equal SINR. The model there is of K stations, where K is a deterministic number, that can choose to be received in any SNR. Buehrer shows analytically how to choose the power levels used by each station to attain an equal SINR for all stations for the linear SIC receiver. The chosen power levels are not drawn at random by the stations, and are used without change throughout the sessions. The analysis in [7] assumes an imperfect cancellation of the linear SIC receiver. This imperfection is shown to limit the number of stations K that can be admitted into the channel without any constraint on the dynamic range of the power levels.

In our network model, the number of transmissions per time slot is a Poisson random variable, so the results in [7] cannot be used "as is". As a baseline, one can use the optimal power levels derived in [7] in a naive way. First, set K to be the mean number of transmissions, $E[G]$. Then,

[3] Throughout this work, we conveniently mention transmission power levels, W_i, but these are actually power levels of the signal arriving at the hub's receiver. With proper feedback from the hub, we assume that VSAT transmission levels can be adjusted to yield the desired relative levels at the receiver.

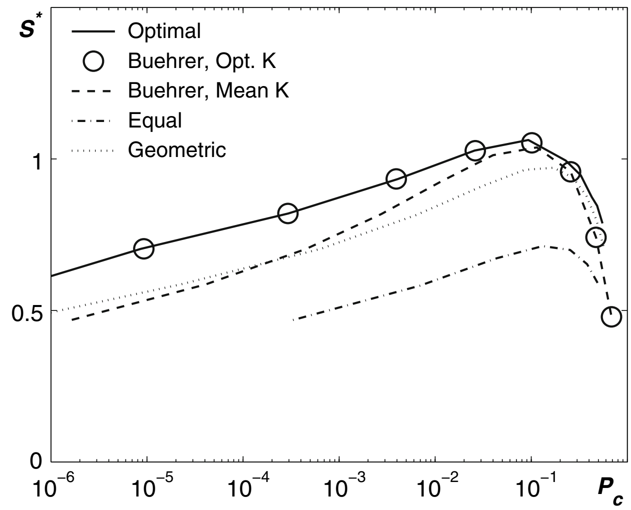

Fig. 2 Attainable throughput S vs. probability of collision, P_c. Spreading gain: $N = 64/3$. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. Single power level: *dash-dot*. Using a geometric sequence of power levels: *dots*. Using the power levels obtained in [7] for $K = E[G]$: *dashed*. Using the method in [7] but with a higher "fake" K: *circles*. Using optimal power levels obtained through exhaustive search: *solid*

use Buehrer's method to calculate the required power levels. Finally, with no reason to do otherwise, use equal selection probabilities. Setting K equal to the mean number of transmissions is only a baseline for comparison, and is not guaranteed in any way to be optimal or even near-optimal.

Figure 2 depicts the attainable throughput vs. the probability of collision for several power selection schemes. Note that we are assuming an amplitude estimation inaccuracy of 5%. In all cases, power levels are drawn at random by all the stations. It can be seen that by using the naive $K = E[G]$, sub-optimal results are obtained. By using a larger "fake" K, higher throughput is attained.

The difference between the naive K and the optimal K is 15–30%, and is more pronounced for lower P_c values. The reason for this difference is that most of the collisions occur in events where $K > E[G]$. Although these events are less likely to happen than ones with fewer transmissions, they still have more impact on the total collision probability.

This phenomenon is illustrated in Fig. 3. The dashed curve is the Poisson probability function, P_k, where the average G is set to $E[G] = 20$. The solid curve is $P(c|k)$, the probability of collision given that k transmissions occurred. The total collision probability P_c is

$$P_c = \sum_{k=0}^{\infty} P(c|k) P_k.$$

The dash-dot curve is the point-wise multiplication of $P(c|k)$ and P_k. Note that most of the contribution to the

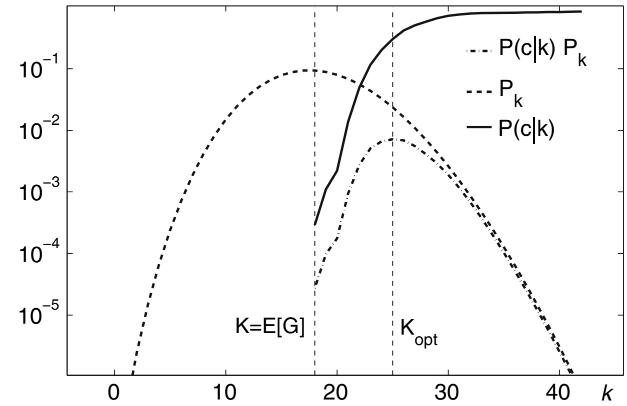

Fig. 3 The impact of rare events on the probability of collision, P_c

summation of P_c is around the value $k = 27$, and not $k = E[G] = 20$.

Consider an event of k transmissions at m different powers. For this specific event, the fate of each transmission can be calculated. We use $k_i^{(fail)}$ to denote the number of power-level-i transmissions that fail. $P_{c_{W_i}}(k_1,\ldots,k_m)$ are the fractions of transmissions at the respective power levels that collide, given that $\sum_{i=1}^{m} k_i = k$ transmissions occur:

$$P_{c_{W_i}}(k_1,\ldots,k_m) = \frac{k_i^{(fail)}}{k_i}, \tag{14}$$

If we had sufficient computing power, we could use a brute-force method to calculate $P_{c_{W_i}}$:

$$P_{c_{W_i}} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} P_{c_{W_i}}(k_1,\ldots,k_m) \cdot \Pr(k_1,\ldots,k_m). \tag{15}$$

Practically, the summation in (15) is not done over an infinite number of elements. Only the elements with a significant probability need to be included in the summation. Even so, for large values of G_{w_i} and m, the brute-force method is too demanding computationally, even as an offline scheme. For this reason we have developed an efficient numerical method for the calculation of $P_{c_{W_i}}$ which is elaborated upon in the appendix. Even by using this method, the optimization can take hours on a powerful PC.

3.2 Results

The optimal power levels $(W_1,\ldots,W_m)$ and the respective selection probabilities $(P_1,\ldots,P_m)$ are obtained through exhaustive search because the throughput in (13) depends on the collision probabilities $P_{c_{W_i}}$, for which we do not have an explicit expression.

Although our method is based upon an exhaustive search, we have succeeded in diminishing the required computation as can be seen in the appendix. Furthermore, this exhaustive search is needed only once. After performing the one-time search, the optimal system parameters are known.

Figure 4 depicts the attainable throughput versus the probability of collision for four types of receivers; the optimal set of power levels and selection probabilities is used for each set of system parameters.

The attainable throughput of the hybrid receiver is higher by 10% than that of the non-linear receiver. This is because the hybrid receiver is able to decode some transmissions when an excessively large number of transmissions occur, while the non-linear receiver simply stops once a single user fails.

3.2.1 *Transmission power considerations*

In many cases, the network provider would like to minimize the mean transmission power, thereby minimizing adjacent channel interference and satellite power consumption. In Fig. 4, for example, the mean received SNR is limited to 10 dB.

Figure 5 depicts the tradeoff between throughput and collision probability for different values of the permissible maximum allowable received SNR. Power levels and probabilities are reoptimized for each mean-SNR value.

In this section we found optimal power levels and selection probabilities for a single transmission round. We found the working points that attain the maximum throughput. However, this maximum throughput is

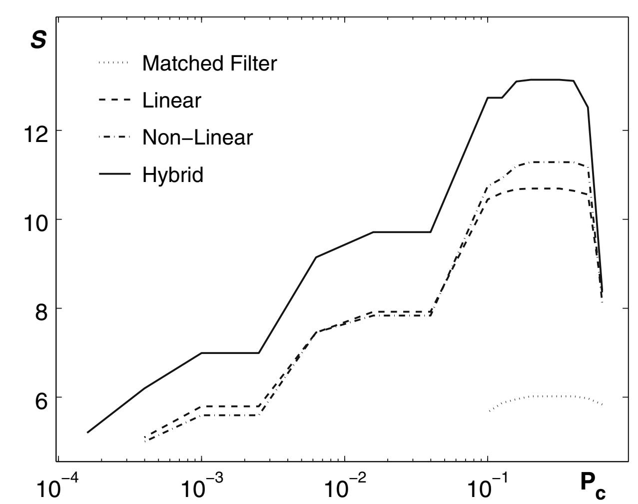

Fig. 4 Throughput for matched filter, linear, non-linear and hybrid receivers vs. the probability of collision, P_c. Spreading gain: $N = 64/3$, Mean transmission power: $E[W] = 10$ dB, Decoding threshold: $\Gamma_{thr} = 6$ dB. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. The optimal set of power levels and selection probabilities is used for each set of system parameters. The steps in the curves are due to the fact that we are performing an exhaustive search over a large, but finite set of power levels

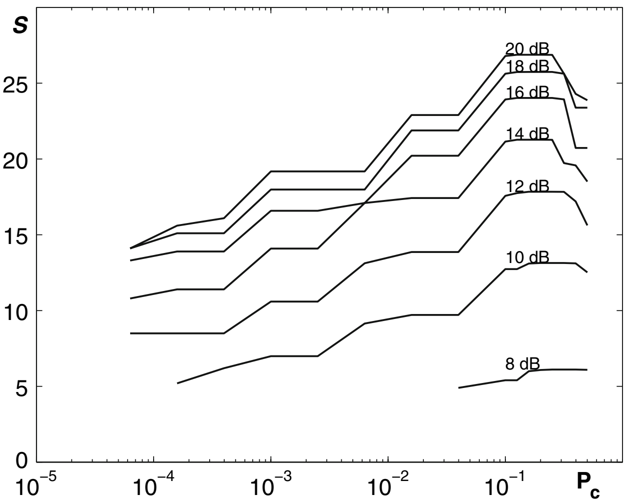

Fig. 5 Maximal attainable throughput for different values of the maximum allowable mean received SNR. Spreading gain: $N = 64/3$; Decoding threshold: $\Gamma_{thr} = 6$ dB. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. Power levels and probabilities are reoptimized for each mean-power value. The steps in the curves are due to the fact that we are performing an exhaustive search over a large, but finite set of power levels

attainable in working points at which the probability of collision is very high. In the next section we show how these working points and similar ones can be utilized in conjunction with multiple round schemes to nearly diminish the probability of packet discarding.

4 Multiple round policies

4.1 Useful notations and assumptions

Let D_r and $P_e^{\max}$ denote the number of rounds until the deadline and the permissible loss probability, respectively. Because messages can be dropped, albeit with a low probability, a distinction was made in [5] between the generation rate of new messages, S_g, and the throughput S. Specifically, $S = (1-P_e)S_g$.

We denote the probability of collision in the ith round by P_{c_i}. Let G_i denote the contribution of round-i transmissions to the overall offered load $G = \sum_i G_i$. Clearly,

$$G_i = S_g \prod_{j=1}^{i-1} P_{c_j}. \tag{16}$$

4.1.1 Stability and control policy

In practice, there may be situations in which the attempted throughput exceeds capacity and the ALOHA protocol becomes unstable. (Message discarding upon deadline expiration does, nonetheless, guarantee that unlike with conventional ALOHA, the system cannot “crash” due to overload.) For those cases, which would typically be infrequent, a background process could be used by the hub to estimate the offered load and then use the contention-free outbound channel to instruct stations to back off probabilistically. (Such a scheme does not require the hub to know the identities of the contending ground stations.) Such activities are not very demanding, and are not part of the core random-access scheme. Moreover, their incorporation is not expected to substantially alter the results of this paper, which are derived assuming “good” stable operation.

4.1.2 Independence

We assume an infinite number of stations and a large number of channels. The number of transmissions over any given channel in any given time slot is modeled as a Poisson random variable whose mean is denoted G. The analysis is carried out under an independence assumption. The independence is between the fates of different transmissions of the same message. Strictly speaking, the fates of retransmissions depend on those of past transmissions, but randomized channel selection causes this dependence to diminish. This was confirmed by simulation [5].

4.2 Stationary vs. non-stationary policies

When using multiple transmission rounds, a station can use a different power-selection rule in different rounds. This is called a *non-stationary policy*. If a station uses the exact same power selection rule in all rounds, we say that it uses a *stationary policy*.

For convenience, we use the same set of power levels for all transmission rounds. Nonetheless, since a non-stationary policy can assign probability zero to any given power level in any given round, there is no loss of generality.

Let P_{ij} denote the probability that power level W_j is chosen in the ith round. With a stationary policy, P_{ij} degenerates into P_j, as in the single round case, and the selection rule $\{P_j\}$ is used in all transmission rounds.

4.3 Using single-round optimal working points

In Sect. 3 we derived channel working points that yield maximum throughput. The maximum unconstrained throughput that can be achieved in multiple rounds is the same as the maximum throughput obtained in a single round. This is because throughput is determined only by the statistics of transmission arrivals and power levels, independently of their transmission round. Consequently, the maximum delay-constrained multi-round throughput is bounded from above by the unconstrained single-round maximum throughput.

Consider a single-round working point that maximizes the throughput for a certain P_c. To get to this throughput in the multi-round case, the probabilities of transmitting at any given power level in any given round must be chosen such that the sum over rounds of the transmission rates at this level must equal the transmission rate at this level in the chosen single-round working point. Clearly, the choice of those probabilities affects the probability of collision, so this is not a simple expression. We shall revisit this shortly.

In this section, we use such working points, and select P_{ij} so as to minimize P_e. The design space P_{ij} is limited because of the artificial constraint on the channel working-point. However, having chosen high-throughput working points, throughput is guaranteed to be close to the unconstrained one, so we only need to focus on reducing P_e. Such an approach is not optimal, but it is intuitive and simple, and as will be seen, the bottom line results are not far from the maximum throughput bound. Therefore, using a more general approach cannot yield a significant further increase in throughput.

The multi-round throughput can be expressed in terms of the corresponding single-round channel working-point:

$$S = \sum_{i=1}^{D_r} G_i(1 - P_{c_i}) \tag{17}$$

$$S = \sum_{i=1}^{D_r} G_i \sum_{j=1}^{m} P_{ij}(1 - P_{cw_j}) \tag{18}$$

$$S = \sum_{j=1}^{m} \underbrace{\left(\sum_{i=1}^{D_r} G_i P_{ij} \right)}_{G_{W_j}} (1 - P_{cw_j}) \tag{19}$$

$$S = \sum_{j=1}^{m} G_{W_j}(1 - P_{cw_j}), \tag{20}$$

where G_i is the offered load of messages in the ith round, and (17) follows from summing over the throughput contributions from the D_r rounds. Equation 18 expresses the success probability $(1 - P_{c_i})$ in each round in terms of the per-power-level collision probability, P_{cw_j}. Equation 19 changes the summation order and identifies G_{W_j}, the rate of transmissions with power W_j. The design space for P_{ij} can be described by the following constraints:

$$\sum_{j=1}^{m} P_{ij} = 1, \quad 1 \le i \le D_r \tag{21}$$

$$\sum_{i=1}^{D_r} G_i \cdot P_{ij} = G_{W_j}, \quad 1 \le j \le m. \tag{22}$$

However, G_i in (22) is not a parameter of the single-round working point. We express G_i in terms of the single round parameters via (16):

$$G_i = S_g \prod_{r=1}^{i-1} P_{c_r} \tag{23}$$

$$G = \sum_{i=1}^{D_r} G_i = S_g \sum_{i=1}^{D_r} \prod_{r=1}^{i-1} P_{c_r} \tag{24}$$

P_{c_r} in (23) can be expressed in terms of P_{ij} and the single round collision probabilities P_{cw_j}:

$$P_{c_r} = \sum_{j=1}^{m} P_{rj} \cdot P_{cw_j}. \tag{25}$$

Substituting P_{c_r} from (25) in (23), and then G_i in (22), results in non-linear constraints on P_{ij}. To avoid solving a set of non-linear equations, we further limit the design space by adding an artificial constraint of picking values for $(P_{c_1}, \ldots, P_{c_{D_r}})$.[4] This results in linear constraints for P_{ij}. Furthermore, (25) becomes an additional (linear) constraint on the permissible P_{ij} values.

Using exhaustive search, we derived the maximum attainable delay-constrained throughput. As already mentioned, the use of exhaustive search to obtain the optimal system parameters does not complicate the system itself, because it is carried out once offline.

As a baseline, we first use the single-round working points that yield the maximum throughput for a given single-round collision probability P_c. The results are presented in Fig. 6 for the hybrid SIC receiver with 5% amplitude estimation error.

The reason for the fairly small difference between stationary and non-stationary policies is the lack of a priority mechanism: the single-round working points chosen thus far exhibit a negligible P_c diversity (different values of P_{c_j}) among power-level selections. In other words, choosing different power levels has no significant impact on the respective collision probabilities.

4.4 Using sub-optimal working points with higher P_c diversity

The lack of P_c diversity renders the P_{ij} degree of freedom useless. To extend the design space of P_{ij}, we now consider single-round working points that, albeit slightly sub-optimal in terms of attainable unconstrained throughput, exhibit significant P_c diversity.

For each of several mean-SNR constraints, we examined several such working points and picked the best one. The results are shown in Fig. 7. Observe that P_e is reduced by two orders of magnitude relative to that attained by using a stationary policy.

[4] We add here another loop of exhaustive search that goes over many vectors of $(P_{c_1}, \ldots, P_{c_{D_r}})$.

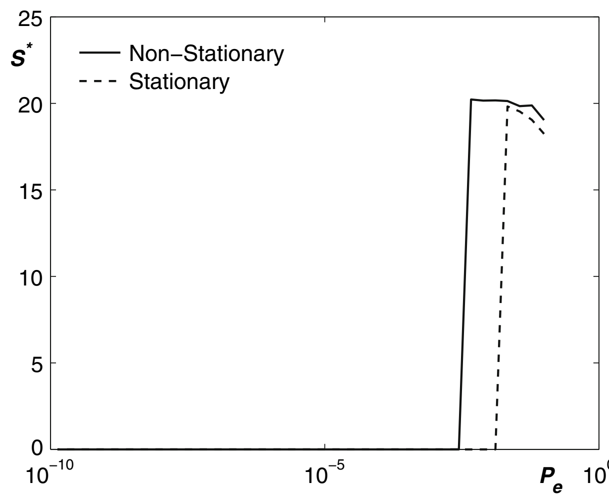

Fig. 6 Multi-round results with single-round working points. Number of rounds: $D_r = 3$, Spreading factor: $N = 64/3$; Mean SNR limit: 20 dB; Decoding threshold: $\Gamma_{thr} = 6$ dB. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. The steps in the curves are due to the fact that we are performing an exhaustive search over a large, but finite set of power levels

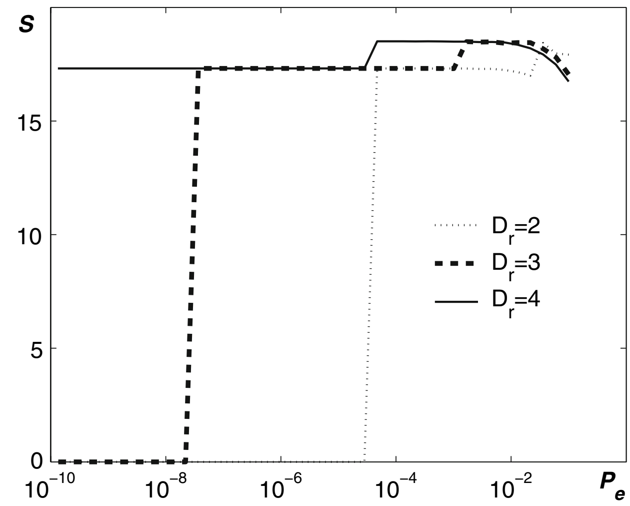

Fig. 8 Maximum delay-constrained throughput with non-stationary policies for several numbers of rounds. Spreading factor: $N = 64/3$; Maximum power limit: 14 dB; Decoding threshold: $\Gamma_{thr} = 6$ dB. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. The steps in the curves are due to the fact that we are performing an exhaustive search over a large, but finite set of power levels

A performance comparison among several values of D_r, the permissible number of transmission rounds, is shown in Fig. 8. It can be seen that the attainable delay-constrained throughput is very close to the unconstrained bound, especially for $D_r = 3$. There is no need to use $D_r > 3$, even for extremely strict requirements for meeting the deadline. For $D_r = 3$, P_e can be virtually eliminated with only a minor compromise in throughput. Also note that if the maximum permissible error probability is greater than 10^{-5}, two transmission rounds suffice.

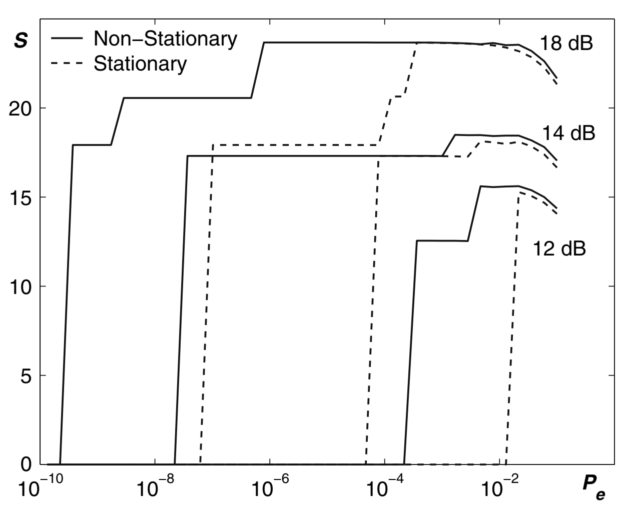

Fig. 7 Stationary vs. non-stationary power-selection-probabilities with single-round working points that exhibit high probability-of-collision diversity. Permissible number of rounds: $D_r = 3$; Spreading factor: $N = 64/3$; Decoding threshold: $\Gamma_{thr} = 6$ dB. Amplitude estimation inaccuracy: $\alpha^2 = 5\%$. Results are presented for several mean-power limitations: 12 dB, 14 dB and 18 dB. The steps in the curves are due to the fact that we are performing an exhaustive search over a large, but finite set of power levels

4.4.1 Comparison to narrow-band ALOHA

Table 1 compares the results of this work with those of [6] for $D_r = 3$ and $P_e^{max} = 10^{-5}$. It compares the benefit of using power diversity for three different models: narrow-band ALOHA with power capture, regular CDMA channel with matched-filter receivers, and CDMA with SIC receivers. The benefit of using power diversity in narrow-band is the most pronounced. In CDMA channels employing regular matched-filter receivers there is almost no use in power diversity because higher power transmissions also cause higher interferences to others. With SIC, the benefit of using power diversity is more than twice the throughput for $P_e = 10^{-5}$. For SIC power diversity in conjunction with judicious choice of power levels and the

Table 1 The benefit in using power diversity in three different models: narrowband channels (only one transmission can succeed), CDMA, and CDMA with SIC receivers

Model	Throughput at $P_e^{max} = 10^{-5}$	
	Equal powers	Optimal powers
Narrow-band, perfect capture	0.0213	0.18 (+474%)
CDMA-matched filter	0.04	0.04 (+0%)
CDMA-SIC	0.3	0.7 (+230%)

Only relative improvements should be compared since the receiver and channel models are significantly different probability of choosing them can serve as an excellent priority mechanism. This gives us the ability to allocate channel resources non-uniformly to different transmission rounds and to virtually diminish the error probability.

5 Conclusions

Multi-channel ALOHA networks are presently used primarily for the transmission of short packets when channel sensing and collision detection are impractical. This work addressed the use of power diversity and successive-decoding in ALOHA networks with CDMA channels.

We have shown that, with SIC receivers, unlike with matched filter receivers, the judicious use of power diversity with CDMA can achieve wire-like dependability with only two or three transmission attempts. The probability of failing to meet a (short) deadline due to collisions can be brought down to negligible levels relative to packet losses due to transient congestion. This enables a satellite service provider to give effectively deterministic service guarantees, even at the packet level. In fact, these guarantees are much higher than those offered by the wire-based Internet.

In this work, it was shown that by using optimized channel working points and non-stationary policies in two and three transmission rounds, an effectively deterministic service guarantee is achieved with a negligible compromise in attainable throughput. We conclude that the value of power diversity depends both on the type of receiver and on the performance measure. At the practical level, the schemes developed in this paper are worth considering for the next generation of satellite-based networks.

One could consider the combination of power diversity level and multiple copies, as was done in [2] for narrow-band channels. However, in view of the fact that the same single-round upper bound would apply and the current results are very close to it, such a combination cannot offer a significant benefit.

In this paper we explored the SIC scheme, which is a rather simple kind of multi-user detection. More complex types of multi-user detection, and other types of CDMA channels (e.g., FH-CDMA [17]) may increase the unconstrained capacity of the channel, and the question is what will the constrained capacity be relative to that. We leave this question for further research.

Appendix

A. Random spreading sequences

The performance of various demodulation strategies depends on the signal-to-noise ratio, A_k/σ, and on the similarity among the spreading sequences, quantified by their cross-correlations:

$$\rho_{ij}(\tau) = \int s_i(t)s_j(t-\tau)dt. \tag{26}$$

In the random signature model

$$s_{km}(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} d_{kmi}p_{T_c}(t - iT_c + T_c), \tag{27}$$

where N is the spreading gain. $\{d_{kmi}\}_{\substack{1\le i\le N\\ 1\le k\le K\\ 1\le m\le M}}$ are i.i.d. random variables and can be 1 or -1 with equal probabilities [18].

For further analysis of the interference among users, it is useful to know the statistical properties of the cross-correlation between the spreading sequences of each two users i, j in the mth bit: ρ_{ijm}. In [18] and [14], it is shown that

$$E\left[\rho_{ijm}^2\right] = \frac{1}{3N} i \neq j. \tag{28}$$

B. Packet collision model

Packet decoding

We assume that packets are encoded using an error-correction code. Thus, the probability that the kth user's packet is not decoded correctly, $P_{c,k}$, can be closely approximated as a unit step function of the SIR (see for example in [4]):

$$P_{c,k} = \begin{cases} 1, & \Gamma_k \geq \Gamma_{thr} \\ 0, & \Gamma_k < \Gamma_{thr} \end{cases}, \tag{29}$$

where Γ_{thr} is a parameter that depends on the specific error-correcting code used and Γ_k is the SIR of the kth user.

Let $\{b_i\}_{i=1,\dots,M}$ denote the M bits of a packet. The output of the decoder, denoted $\{\hat{b}_i\}_{i=1,\dots,M}$ is an estimate of these bits. It should be emphasized that the estimation of $\{\hat{b}_i\}_{i=1,\dots,M}$ is not a simple per-bit hard-decision but a decoding that takes into account many other neighboring bits. If decoding is successful (which can be easily determined by a simple CRC) then we assume that all of the bit estimates in the packets are accurate. (The amplitude estimation may nonetheless be imperfect). This assumption is an approximation, suitable for long coded packets (longer than 1 kb).

The spreading sequences are assumed to be independent from bit to bit, but the relative delay shifts, τ_k, do not change significantly during a given packet. This bit-to-bit error dependence was studied in [13], where the results show that the simple model that ignores inter-bit relations is quite accurate. It is also shown that resorting to a simpler model of synchronous spreading sequences ($\tau = 0$) results in a very pessimistic assessment of capacity. We therefore ignore the bit-to-bit error dependence within each packet, but do not assume bit-level synchronization among different transmissions.

C. Successive decoding SIR analysis

Let $r^{(k)}$ denote the signal used to detect data from user k, which is the received signal, $r(t)$, after $k-1$ subtractions of previously decoded user signals:

$$r^{(k)}(t) = r(t) - \sum_{i=1}^{k-1} \hat{x}_i(t), \tag{30}$$

where $\hat{x}_i(t)$ is an estimate of the received signal from user i, defined as $x_i(t) = A_i\, b_i\, s_i(t-\tau_i)$.

We also define z_k to be the matched filter output after $k-1$ previous subtractions, $z_k = \langle r^{(k)}(t), s_k(t)\rangle$.

Following a similar analysis to [7], the variance of z_k in the *Hybrid SIC receiver* model is

$$\mathrm{var}[z_k^{hyb}] = \frac{1}{3N}\sum_{i=k+1}^{K} A_i^2 + \frac{1}{3N}\sum_{\substack{i<k\\ \Gamma_i<\Gamma_{thr}}} \mathrm{var}[z_i^{hyb}] + \sigma^2. \tag{31}$$

Modeling the amplitude estimation inaccuracy

We use a simple model to account for inaccurate amplitude estimation of successfully decoded signals, by assuming a fixed relative error,

$$\frac{|\hat{A}_k - A_k|}{A_k} \approx \alpha. \tag{32}$$

Adding this to (31) yields

$$\mathrm{var}[z_k^{hyb}] = \frac{1}{3N}\sum_{i=k+1}^{K} A_i^2 + \frac{1}{3N}\sum_{\substack{i<k\\ \Gamma_i<\Gamma_{thr}}} \mathrm{var}[z_i^{hyb}] + \frac{1}{3N}\sum_{\substack{i<k\\ \Gamma_i\geq\Gamma_{thr}}} \alpha^2 A_i^2 + \sigma^2. \tag{33}$$

By iteratively calculating $E^2[z_k]$ and $var[z_k]$ for all the users, $\Gamma_k = \frac{E^2[z_k]}{var[z_k]}$ can be compared to the decoding threshold, Γ_{thr} to determine the fate of each of the concurrent transmissions.

D. A numerical method for approximating $P_{c_{W_i}}$

We assume that in all the successive cancellation schemes, transmissions are decoded in decreasing-power order. Same-power transmissions are decoded in random order.

Our method hinges on the observation that whether a "probe" transmission collides or not depends only on its SIR, and not on the numbers of concurrent transmissions K_i. The SIR of the "probe" depends on its power, W_i, and on the *aggregate* interference, $var[z]$. Consequently, we do not need to go over the many possibilities of K_i vectors as was suggested by the summation in (15); rather, we merely need to calculate the probability distribution of the SIR seen by a transmission at power level W_i.[5] $P_{c_{W_i}}$ is calculated given the power W_i, so W_i is a parameter. $var[z]$ is a random variable, because it depends on the interfering transmissions, their power levels and the cross-correlations between their spreading sequences and the probe's spreading sequence.

Let I_i denote the interference (contribution to $var[z]$) from higher- and lower-power transmissions than W_i. The interference from transmissions with the same power level W_i is excluded intentionally (The interference from these transmissions will be taken care of by the definition of the function $P_{c_{W_i}}(n,v)$).

We can simplify (15):

$$P_{c_{W_i}} = \sum_{n=0}^{\infty} \Pr(K_i = n) \int_{v=0}^{\infty} P_{c_{W_i}}(n,v)\, dP_{I_i}(v), \tag{34}$$

where $P_{I_i}(v)$ is the cumulative probability function of I_i, $P_{I_i}(v) = \Pr(I_i \leq v)$. The function $P_{c_{W_i}}(n,v)$ is the probability of collision for a W_i-power transmission, given that n transmissions with W_i occurred and the interference from other transmissions (lower + higher) equals v. This function can be pre-calculated and saved in a table for reference when calculating the integral in (34).

We separate the interference I_i into two parts: $I_i = I_i^H + I_i^L$. I_i^L results from an aggregate of lower power level transmissions that are decoded *after* the "probe" transmission is decoded. Therefore, the interference that they induce does not depend upon the success/failure of their decoding. I_i^H, in contrast, is the *residual* interference seen by the "probe" transmission, left by imperfect subtractions of previously decoded transmissions. Therefore, it depends on the success/failure of previous decodings. Furthermore, the results of previous decoding depend on the SIR levels seen by higher power level transmissions, which depend on the interference I_{i-1}^L. The analysis of I_i^H must therefore take into account the dependence on I_{i-1}^L.

Calculation of the probability distribution of I_i^L

We begin with the calculation of $P_{I_i^L}(v) = \Pr(I_i^L \leq v)$. For $i = m$, the "probe" transmission is received with the lowest of power levels, so there is no interference from lower power levels by definition. Therefore, $I_m^L = 0$ w.p. 1, and $P_{I_m^L}(v)$ is the unit step function:

$$P_{I_m^L}(v) = \begin{cases} 1, & v \geq 0 \\ 0, & v < 0 \end{cases} \tag{35}$$

$P_{I_i^L}(v)$ can be calculated iteratively using the following rule:

[5] Note that the importance of the value of W_i is not merely in that it is the numerator of the SIR expression. Rather, it also influences the decoding order. The contribution of other transmissions to the interference seen by our transmission, in turn, depends on the decoding order.

$$
\begin{aligned}
P_{I_i^L}(v) &= \Pr(I_i^L \le v) \\
&= \sum_{n=0}^{\infty} \Pr(K_{i+1}=n)\Pr(I_{i+1}^L + nW_{i+1} \le v) \\
&= \sum_{n=0}^{\infty} \Pr(K_{i+1}=n)\Pr(I_{i+1}^L \le v - nW_{i+1}) \\
&= \sum_{n=0}^{\infty} P_{I_{i+1}^L}(v - nW_{i+1}),
\end{aligned} \tag{36}
$$

Calculation of the probability distribution of $\mathrm{I_i^H}$

Because of the dependency between I_i^H and I_{i-1}^L, we calculate the conditional probability distribution function of I_i^H given I_{i-1}^L. We define $P_{I_i^H|I_{i-1}^L}(u,v)$:

$$
P_{I_i^H|I_{i-1}^L}(u,v) = \Pr(I_i^H \le u | I_{i-1}^L = v). \tag{37}
$$

To facilitate the presentation, let I_i^+ be the residual interference left by the imperfect subtractions of the W_i transmissions. The residual interference left by imperfect subtractions of $W_1, \ldots, W_{i-1}$ is I_i^H and is excluded intentionally. I_i^+ depends only on K_i, the number of transmissions with W_i, and on the total interference $I_i^H + I_i^L$. The dependency on the interference in the *linear* receiver is due to the variance of the matched filter output, which depends on the interference. In the *non-linear* receiver, the success/failure of the decoding depends on the interference, and the decoding result affects the residual interference.

We will therefore use the notation $I_i^+(K_i, I_i)$ to denote the residual interference left by K_i imperfect subtractions (of transmissions with power level W_i) given that the interference from other power levels (higher + lower) was I_i.

$I_i^+(K_i, I_i)$ and $P_{cw_i}(K_i, I_i)$ (in (34)) can be pre-calculated and saved in a table using the analysis in Sect. C. As an example, we show how this precalculation should be done for the hybrid receiver defined in Sect. 2.

Example 1: Calculating $P_{cw_i}(K_i, I_i)$ and $I_i^+(K_i, I_i)$ for the hybrid receiver case.

```
Set P_cw_i(K_i, I_i) = 0 and I_i^+(K_i, I_i) = 0
Iterate on j from 1 to K_i
    Γ_j = W_i / ( (1/3N)[(K_i − j)W_i + I_i + I_i^+(K_i, I_i)] + σ^2 );
    if (Γ_j > Γ_thr) /* Successful Decoding */
        I_i^+(K_i, I_i) = I_i^+(K_i, I_i)+
            α^2 W_i; /* Power estimation error*/
    else
        P_cw_i = P_cw_i + 1/K_i;
        I_i^+(K_i, I_i) = I_i^+(K_i, I_i)
            + (1/3N)[(K_i − j)W_i + I_i + I_i^+(K_i, I_i)]
            + σ^2;
    end
end
```

Calculation of $P_{I^H|I^L}$

Now that we have $P_{cw_i}(K_i, I_i)$ and $I_i^+(K_i, I_i)$ ready, we go on to calculate the probability distribution of I_i^H given I_{i-1}^L. By definition, $I_1^H = 0$ w.p. 1. Therefore

$$
P_{I_1^H|I_0^L}(u,v) = \begin{cases} 0, & u<0 \\ 1, & u \ge 0 \end{cases} \tag{38}
$$

To calculate $P_{I_i^H|I_{i-1}^L}$ iteratively, we would like to express it in terms of $P_{I_{i-1}^H|I_{i-2}^L}$:

$$
\begin{aligned}
P_{I_i^H|I_{i-1}^L}(u,v) &= \Pr(I_i^H \le u | I_{i-1}^L = v) \\
&= \sum_{n=0}^{\infty} \Pr(K_{i-1}=n) \cdot \\
&\quad \int_{u'|I_{i-1}^+(n, v+u')+u' \le u} dP_{I_{i-1}^H|I_{i-2}^L}(u', v + nW_{i-1})
\end{aligned}
$$

The summation and integration in (39) are over all values of K_{i-1} and u', where u' is the residual interference I_{i-1}^H. Because $I_i^H = I_{i-1}^H + I_{i-1}^+$, we integrate over u' values that yield a I_i^H interference less or equal to u.

Let $u^\star$ be the solution to the equation $I_{i-1}^+(K_i, I_{i-1}^L + u^\star) + u^\star = I_i^H$. $u^\star$ can be pre-calculated and saved in a table, $u^\star(K_i, I_{i-1}^L, I_i^H)$.

$I_{i-1}^+(K_i, I_{i-1}^L + u')$ is a monotonically non-decreasing function of u', so the set of values of u' over which we integrate is simply $u' \le u^\star$. Using the table $u^\star(K_i, I_{i-1}^L, I_i^H)$, (39) can be simplified,

$$
\begin{aligned}
P_{I_i^H|I_{i-1}^L}(u,v) &= \sum_{n=0}^{\infty} \Pr(K_{i-1}=n) \cdot \\
&\quad P_{I_{i-1}^H|I_{i-2}^L}\left(u^\star(n,v,u), v + nW_{i-1}\right).
\end{aligned} \tag{39}
$$

Finally, P_{cw_i} can be calculated using the pre-calculated tables $P_{cw_i}(K_i, I_{i-1}^L)$, $P_{I_i^H|I_{i-1}^L}(u,v)$ and $P_{I_i^L}(v)$:

$$
\begin{aligned}
P_{cw_i} &= \sum_{n=0}^{\infty} \Pr(K_i = n) \int_{v=0}^{\infty} dP_{I_i^L}(v) \\
&\quad \int_{u=0}^{\infty} P_{cw_i}(n, u+v) dP_{I_i^H|I_{i-1}^L}(u, v + nW_i).
\end{aligned} \tag{40}
$$

Complexity analysis

Let m be the number of power levels and $K_{\max}$ be the maximum number of transmissions per power level.[6] The probability functions described in the analysis above are actually approximated by histograms. Let N_I be the number of bins on the interference axis of the histograms that we keep for functions like $P_{I_i^H|I_{i-1}^L}(u,v)$. The table size required for this function is N_I^2.

[6] For Poisson random variables there is not really a maximum. We pick a large enough $K_{\max}$ such that the probability of getting larger K_is is negligible for our computation purpose.

The calculations that are done according to the analysis above are in the order of $m \cdot N_I^2 \cdot K_{\max}$. If we had to do the brute-force summation described in (15), this would require calculations that are in the order of $K \cdot K_{\max}^m$. For $K_{\max} = 100$, $m = 6$ and $N_I = 1000$, the brute force method is simply infeasible while the numerical approach leads to the results shown in this paper. Furthermore, our method is more scalable in the sense that the computation needed for larger K and m grows linearly rather than exponentially.

References

1. Abramson, N. (1985). Development of the alohanet. *IEEE Transaction on Information Theory, 31*(3), 119–123.
2. Baron, D., & Birk, Y. (2002). Coding schemes for multislot messages in multichannel aloha with deadlines. *IEEE Transaction on Wireless Communications, 1*(2), 292–301.
3. Baron, D., & Birk, Y. (2002). Multiple working points in multichannel aloha with deadlines. *ACM/Baltzer Wireless Networks, 8*(1), 5–11.
4. Biglieri, E., Proakis, J., & Shamai, S. (1998). Fading channels: Information-theoretic and communication aspects. *IEEE Transactions on Information Theory, 44*(6), 2619–2692.
5. Birk, Y., & Keren, Y. (1999). Judicious use of redundant transmissions in multichannel aloha networks with deadlines. *IEEE Journal on Selected Areas in Communications, 17*(2), 257–269.
6. Birk, Y., & Revah, Y. (2005). Increasing deadling-constrained throughput in multichannel aloha networks via non-stationary multiple-power-level transmission policies. *Wireless Networks, 11*, 523–529.
7. Buehrer, R. M. (2001). Equal ber performance in linear successive interference cancellation for cdma systems. *IEEE Transactions on Communications, 49*(7), 1250–1258.
8. Heiman, R., & Kaplan, G. (1996). Issues in vsat networks, *Lecture in The 19th Convention of Electrical and Electronics Engineers in Israel*.
9. Kleinrock, L., & Lam, S. S. (1975). Packet switching in multi-access broadcast channel: Performance evaluation. *IEEE Transactions on Communications, COM-23*, 410–423.
10. Lamaire, R., Krishna, A., & Zorzi, M. (1998). On the randomization of transmitter power levels to increase throughput in multiple access radio systems. *Wireless Networks, 4*(3), 263–277.
11. Lee, C. C. (1987). Random signal levels for channel access in packet broadcast networks. *IEEE Journal on Selected Areas in Communications, SAC-5*(6), 1026–1034.
12. Metzner, J. J. (1976). On improving utilization in aloha networks. *IEEE Transactions on Communications, COM-24*, 447–448.
13. Morrow, R. K., & Lehnert, J. S. (1992). Packet throughput in slotted aloha ds/ssma radio systems with random signature sequences. *IEEE Transactions on Communications, 40*(7), 1223–1230.
14. Pursley, M. B. (1977). Performance evaluation for phase-coded spread-spectrum multiple-access communication – part i: System analysis. *IEEE Transactions on Communications, COM-25*(8), 795–799.
15. Revah, Y. (2002). *Utilization of power capture for delay-constrained throughput maximization in multi-channel aloha networks*. Master's Thesis, Electrical Engineering Department, Technion.
16. Roberts, L. G. (1975). Aloha packets, with and without slots capture. *Computer Communications Review, 5*, 28–42.
17. Sui, H., & Zeidler, J. R. (2007). A robust coded mimo fh-cdma transceiver for mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications, 25*, 1413–1423.
18. Verdú, S. (1998). *Multiuser detection*. Cambridge University Press.
19. Yue, W. (1991). The effect of capture on performance of multichannel slotted aloha systems. *IEEE Transactions on Communications, 39*(6), 818–822.

Author Biographies

Dr. Yitzhak Birk has been a faculty member in the Electrical Engineering Department at the Technion since 1991. He also heads its Parallel Systems Lab. Previously, he was a research staff member at IBM's Almaden Research Center in California. He is also involved with industry in various ways. Dr. Birk received his B.Sc. (cum laude) and M.Sc. from the Technion, and the Ph.D. from Stanford University, all in Electrical Engineering. His research interests include: computer and communication systems and subsystems, in particular parallel architectures, with focus on communication-intensive storage and information-dissemination systems. The judicious exploitation of redundancy for performance enhancement in various contexts has been the subject of much of his recent work.

Uri Tal received the B.Sc. (suma cum laude) and M.Sc. degrees in Electrical Engineering from the Technion—Israel Institute of Technology, in 1998 and 2004, respectively. His thesis was entitled "Power Diversity for Delay-Constrained Throughput Maximization in ALOHA Networks with Successive Decoding". From 1998 to 2005, he served as an Engineer in the Israel Defense Forces. His current research interests include computer and communications systems. He is currently with Broadcom.