Network-Coded Diversity Protocol for Collision Recovery in Slotted ALOHA Networks

G. Cocco*, N. Alagha¶, C. Ibars* and S. Cioni¶

*Centre Tecnològic de Telecomunicacions de Catalunya – CTTC
Parc Mediterrani de la Tecnologia, Av. Carl Friedrich Gauss 7 08860, Castelldefels – Spain

¶European Space Agency - ESTEC, Noordwijk – The Netherlands

giuseppe.cocco@cttc.es, nader.alagha@esa.int, christian.ibars@cttc.es,

stefano.cioni@esa.int

Abstract

We propose a collision recovery scheme for symbol-synchronous slotted ALOHA (SA) based on physical layer network coding over extended Galois fields. Information is extracted from colliding bursts allowing to achieve higher maximum throughput with respect to previously proposed collision recovery schemes. An energy efficiency analysis is also performed and it is shown that, by adjusting the transmission probability, high energy efficiency can be achieved. A performance evaluation is carried out using the proposed algorithms, revealing remarkable performance in terms of normalized throughput.

Index Terms

Multiple access, MAC, physical-layer network coding, PNC, slotted ALOHA, collision resolution

I. INTRODUCTION

The throughput of Slotted ALOHA (SA) systems is limited by the collisions that take place when more than one node accesses the channel in the same time slot. This limitation is particularly problematic in satellite networks with random access, where the long round-trip time (RTT) greatly limits the use of feedback from the receiver. Techniques like Diversity Slotted
ALOHA (DSA) [1], in which each packet is transmitted more than once, have been proposed in order to increase the probability of successful detection. In [2] a novel scheme called network-assisted diversity multiple access (NDMA), inspired by signal separation principles borrowed from signal processing, has been presented. In the scheme the collisions are recovered through successive retransmissions assuming feedback from the receiver. This interprets the received signals across successive transmission slots as a matrix which is processed in the analog domain so that the single bursts are recovered if the matrix is full rank. In Contention Resolution Diversity Slotted ALOHA (CRDSA) [3] the transmissions are organized in frames. The collided signals are exploited using a successive interference cancelation (SIC) process. In CRDSA each packet is transmitted twice and uncollided packets are subtracted from slots in which their replicas are present. An enhanced version of CRDSA, called CRDSA++, has been presented in [4]. In CRDSA++ more than two copies of each burst (3 to 5) can be transmitted and an iterative interference cancelation is applied, if needed, within a slot. In [5] a packet-level forward error correction (FEC) code has been applied to CRDSA, while in [6] a convergence analysis and optimization of CRDSA has been proposed. Another technique that allows to extract information from colliding signals is physical layer network coding (PNC). PNC was originally proposed to increase spectral efficiency in two-way relay communication [7] by having the relay decoding the collision of two signals under the hypothesis of symbol, frequency and phase synchronism. In [8] a cooperative relaying protocol that leverages on PNC and SIC has been proposed, while in [9] PNC has been applied in the satellite context for pairwise node communication. In [10] and [11] it has been proposed to apply PNC to determine the identity of transmitting nodes in case of ACK collision in multicast networks by using energy detection and ad-hoc coding schemes. In [12] an overview of the state of the art on PNC has been presented from an information theoretical point of view. In [13] and [14] PNC has been applied for collision resolution in multiple access systems with feedback from the receiver, under the assumption of frequency synchronous transmitters.

In this paper we present a new scheme named Network-Coded Diversity Protocol (NCDP), that leverages on PNC over extended Galois fields for recovering collisions in symbol-synchronous SA systems. Once the PNC is applied to decode the collided bursts, the receiver uses common matrix manipulation techniques over finite fields to recover the original messages, which results in a high-throughput scheme. The proposed scheme and analysis differ from previous works on
collision resolutions at both system level and physical level:

- Unlike in [2] and [13] we assume that transmissions are organized in frames. We consider two different setups, one in which feedback from the receiver is not allowed and another in which feedback is allowed.
- We take into account the energy consumption in the design of our solution and evaluate jointly the spectral and the energy efficiency of the proposed scheme, comparing it with other collision resolution schemes previously proposed in the literature.
- We use extended Galois fields, i.e., $GF(2^n)$ with $n > 2$, instead of $GF(2)$, which is generally used in PNC. This allows to better exploit the diversity of the system, leading to an increased spectral efficiency and, depending on the system load, to an increased energy efficiency. Unlike in [2], in our scheme most of the processing is done using finite field arithmetics, which reduces the complexity of the system.
- We present results relative to implementation issues such as decoding in the presence of frequency offsets, channel estimation and imperfect symbol synchronism for a generic number of colliding signals. Up to our knowledge such issues have been previously addressed only for the case of two transmitters [15], [16], [17].

The rest of the paper is organized as follows. In Section II we present the system model. In Section III the proposed scheme is described, while a theoretical analysis of its performance is carried out in Section IV. Section V deals with practical issues such as decoding in presence of frequency offsets, channel estimation and imperfect symbol synchronization. In Section VI we present the numerical results, while Section VII contains the conclusions.

II. SYSTEM MODEL

Let us consider the return link (i.e., the link from a user terminal to a satellite or a base station) of a multiple access system with $M$ transmitting terminals, $T_1, ..., T_M$, and one receiver $R$. Packet arrivals at each transmitter are modeled as a Poisson process with rate $\lambda_i^M$, which is independent from one transmitter to the other. Each packet $u_i = [u_i(1), ..., u_i(K)]$ consists of $K$ binary symbols of information $u_i(\xi) \in \{0, 1\}$, for $\xi = 1, \ldots, K$. We assume that, upon receiving a message, each terminal $T_i$ uses the same linear channel code of fixed rate $r_{cc} = \frac{K}{N}$ to protect its message $u_i$, obtaining the codeword $x_i = [x_i(1), ..., x_i(N)]$, where $x_i(l) \in \{0, 1\}$
for \( l = 1, \ldots, N \). Each codeword \( \mathbf{x}_i \) is BPSK modulated (using the mapping \( 0 \rightarrow -1, 1 \rightarrow +1 \)), thus obtaining the transmitted signal

\[
s_i(t) = \sum_{l=1}^{N} b_i(l) g(t - lT_s),
\]

where \( T_s \) is the symbol period, \( b_i(l) \) is the BPSK mapping of \( x_i(l) \) and \( g(t) \) is the square root raised cosine (SRRC) pulse. The signal \( s_i(t) \) is called burst. Higher order modulations such as QPSK may also be used, but some preliminary results we obtained showed a consistent increase of the FER with respect to BPSK in case random phase and frequency offsets are present. In particular we observed that with QPSK modulation some relative phase rotations introduce an ambiguity in the received signal that impedes correct decoding even in the absence of noise. Further investigation is needed to study this issue and the potential countermeasures, such as the adoption of different quaternary modulations. In the present paper we focus on BPSK which is more robust to random phase and frequency offsets. As we will show in Section VI, even with a real-valued modulation our method outperforms other collision resolution schemes that use complex-valued modulations for certain code rates and SNR values.

In the following we will refer to a time division multiple access (TDMA) scheme. However, the techniques proposed hereafter can be also applied to other access schemes, such as multi-frequency-TDMA (MF-TDMA), in which a frame may include several carriers, or code division multiple access (CDMA), where NCDP can be used to recover collisions in each of the code sub-channels. The proposed technique still relies on single carrier transmission by each user terminal. From the user terminal perspective no significant change is required.

Transmissions are organized in frames. Each frame is divided into \( S \) time slots. The number \( S \) of time slots that compose a frame is fixed, i.e., it does not change from one frame to the other. The duration of each slot is equal to about \( N \) burst symbols. When more than one terminal transmits its burst in the same slot a collision occurs at the receiver. A collision involving \( k \) transmitters is said to have size \( k \). We assume symbol-synchronous transmissions, i.e., in case of a collision, the signals from the transmitters add up with symbol synchronism at the receiver \( R \). The received signal before matched filtering and sampling at \( R \), in case of a collision of size \( k \) (assuming, without loss of generality, the first \( k \) terminals collide), is:

\[
y(t) = h_1(t)s_1(t) + \ldots + h_k(t)s_k(t) + w(t),
\]

September 23, 2013 DRAFT
where \( s_i(t) \) is the burst transmitted by user \( i \), \( w(t) \) is a complex additive white Gaussian noise (AWGN) process while \( h_i(t) \) takes into account the channel from terminal \( i \) to the receiver. \( h_i(t) \) can be expressed as:

\[
h_i(t) = A_i e^{j(2\pi \Delta \nu_i t + \varphi_i)},
\]

(3)

where \((A_i)^2 = |h_i(t)|^2\) is a lognormally distributed random variable modeling the channel power of transmitter \( i \), while \( \Delta \nu_i \) and \( \varphi_i \) are the frequency and phase offsets with respect to the local oscillator in \( R \), respectively. The log-normality of the satellite channel power has been assumed in several previous works such as [3], [4] and [18] although, as indicated in [4], this is a pessimistic assumption. As we will refer to the schemes presented in these papers, we keep the assumption of log-normal power distribution in the following. In general a distribution characterized by large fluctuations of the channel power are likely to affect the proposed scheme because it is conditioned by the channel with less power. We will come back on this issue in the following sections. We assume that the amplitude \( A_i \) and the frequency offset \( \Delta \nu_i \) remain constant within one frame while \( \varphi_i \) is a random variable uniformly distributed in \([-\pi, +\pi]\) that changes independently from one slot to the other due to the phase noise at the transmitting terminals as assumed in [3]. The assumption of constant phase within a burst is more accurate for shorter burst lengths and assuming high class transmitting terminals with stable oscillators or a high symbol rate (typically above 2 Mbaud). We assume this model for ease of exposition. Further studies are needed to characterize the sensitivity of PNC to phase noise for a generic number of colliding signals and especially for a large burst size, but this is out of the scope of the present paper. Assuming that the frequency offset is small compared to the symbol rate \( 1/T_s \) (i.e., \( \Delta \nu T_s \ll 1 \)), the sample taken at time \( t_l \) after matched filtering signal \( y(t) \) is:

\[
r(t_l) = h_1(t_l)q_1(t_l) + \ldots + h_k(t_l)q_k(t_l) + n(t_l),
\]

(4)

where \( q(t) = s(t) \otimes g(-t) \), \( \otimes \) being the convolution operator, while \( n(t_l) \)'s are i.i.d. zero mean complex Gaussian random variables with variance \( N_0 \) in each component. Note that even in case a BPSK modulation is used, as we are assuming in this paper, both the I and Q components of the received signal are considered by the receiver. This is because the phases of the users have random relative offsets and thus both components carry information relative to the useful signal. The frequency and phase relative offsets must be taken into account by the decoder, as they can not be eliminated by the demodulator. We consider this more in detail in Section III.
We assume that the receiver has knowledge of the nodes that are transmitting, as well as the full channel state information at each time slot. As we are considering a random access scheme, the knowledge about nodes identity cannot be available \textit{a priori} at the receiver. Thus, nodes identity must be determined by $R$ starting from the received signal, even in case of collision. This can be achieved by having the transmitting nodes add a pseudonoise preamble in each transmitted burst. For a length-$N^\text{pre}$ preamble there are $2^{N^\text{pre}}$ different sequences. Using maximal length pseudonoise signature sequences the crosscorrelation between any two different sequences is $-1/N^\text{pre}$, which translates in a $10$ $\text{dB}$ gain for preambles that are just $10$ symbols long. We discuss the issue of node identification and channel estimation more in detail in Section V.

III. NETWORK CODED DIVERSITY PROTOCOL

In this section we present our network-coded diversity protocol (NCDP) which aims at increasing the throughput and reducing packet losses in Slotted ALOHA multiple access systems. In the first part of the section we describe the way the received signal is processed by the receiver in case of collision, while in the rest of the section we describe the NCDP at the transmitter and at the receiver sides.

A. Multi-User Physical Layer Network Coding

When a collision of size $k$ occurs, i.e., $k$ bursts collide in the same slot, the receiver tries to decode the bit-wise XOR of the $k$ transmitted messages. This can be done by feeding the decoder with the appropriate log-likelihood ratios (LLR). The calculation of the LLRs for a collision of generic size $k$ in case of BPSK modulation was presented in [13]. In the following we include the effect of frequency offset in the calculation of the LLR’s, which was not taken into account in [13].

When signals from $k$ transmitters collide, the received signal at $R$ is given by (2). Each codeword $x_i$ is calculated from $u_i$ as $x_i = C(u_i)$, where $C(\cdot)$ is the channel encoder operator. All nodes use the same linear code $C(\cdot)$. Starting from $r(t)$, the receiver $R$ wants to decode codeword $x_s \triangleq x_1 \oplus x_2 \oplus \ldots \oplus x_k$, where $\oplus$ denotes the bit-wise XOR. In order to do this the decoder of $R$ is fed with vector $L^\oplus = [L^\oplus(1), ..., L^\oplus(N)]$ of LLRs for $x_s$, where:

$$
L^\oplus(t) = \ln \left\{ e^{ \sum_{j=1}^{k+1} \sum_{m=1}^{\binom{k}{j-1}} e^{- \frac{\|r(t_j) - d^e(2i-1,m)Tr(t_j)\|_2^2}{2N_0}} } \right\}, \quad \text{ (5)}
$$
being a column vector containing the channel coefficients of the $k$ transmitters at time $t_l$ (which change at each sample due to frequency offsets), while $\mathbf{d}^e(2i - 1, m)$ and $\mathbf{d}^e(2i, m)$ are column vectors containing one (the $m$-th) of the $\binom{k}{2i-1}$ or $\binom{k}{2i}$ possible permutations over $k$ symbols (without repetitions) of an odd or even number of symbols with value “+1”, respectively. Equation (5) is derived considering that an even or an odd number of symbols with value +1 adding up at $R$ must be interpreted by the decoder as a 0 or a 1, respectively (see [19] and [20] for an extension to higher order modulations). If the decoding process is successful, $R$ obtains the message $\mathbf{u}_r \triangleq \mathbf{u}_1 \oplus \ldots \oplus \mathbf{u}_k$. In Section V the frame error rate (FER) curves for different collision sizes obtained using these LLR values are shown.

B. NCDP: Transmitter Side

We call active terminals the nodes that have packets to transmit in a given frame. Each message is transmitted more than once within a frame, i.e., several replicas of the same message are transmitted. We will give details about the number of replicas transmitted within a frame in the next section. Assume that node $i$ has a message $\mathbf{u}_i$ to deliver to $R$ during a given frame, i.e., node $T_i$ is an active terminal. Before each transmission, node $i$ pre-encodes $\mathbf{u}_i$ as depicted in Fig. 1. The pre-coding process works as follows. $\mathbf{u}_i$ is divided into $L = \frac{K}{n}$ blocks of $n$ bits each.

At each slot a different coefficient $\alpha_{ij}, j \in \{1, \ldots, S\}$, is drawn randomly according to a uniform distribution in $GF(2^n)$. If $\alpha_{ij} = 0$, terminal $T_i$ does not transmit in slot $j$. Each of the $L$ blocks $\mathbf{u}_{ij}^r, r \in \{1, \ldots, L\}$, is interpreted as an element in $GF(2^n)$ and multiplied by $\alpha_{ij}$. We call $\mathbf{u}_{ij}'$ the message $\mathbf{u}_i$ after the multiplication by $\alpha_{ij}$. $\mathbf{u}_{ij}'$ is then channel encoded, generating the codeword

\[
\mathbf{s}_{ij} = M(\mathbf{x}_{ij}')
\]

\[
\mathbf{x}_{ij} = C(\mathbf{u}_{ij}')
\]
$x_{ij} = C(u'_{ij})$. After channel coding, a header $p_i$ is added to $x_{ij}$. Such header is generated using a pseudonoise sequence generator such as the ones used in CDMA. The increase in complexity at the transmitter side with respect to the case in which all nodes use the same preamble is not large if maximal length pseudonoise sequences are used, as each node just needs to choose a random seed and feed it to a shift register which is the same for all nodes and known at the receiver. On the other hand, at the receiver side the complexity associated to the detection phase increases in a quasi-linear fashion with the number of correlators used. However, such increase in complexity may not be an issue if the receiver is located at the gateway station, as it is likely to have good computational capabilities. The same header $p_i$ is used for all transmissions of node $T_i$ within frame $f$. Once the header is attached, $x_{ij}$ is BPSK modulated and transmitted. Note that the multiplication of $u_i$ by $\alpha_{ij}$ is needed to introduce randomness in the MAC mechanism and does not modify the number of information bits transmitted.

The choice of the coefficients and of the header is done as follows. Node $T_i$ draws a random number $\mu$. $\mu$ is fed to a pseudo-random number generator in $GF(2^n)$, which is the same for all terminals and is known at $R$. The first $S$ outputs of the generator are used as coefficients. The header is uniquely determined by $\mu$, i.e, there is a one-to-one correspondence between the set of values that can be assumed by $\mu$ and the set of available pseudo-noise sequences. The cross-correlation properties of the preambles allow the receiver to know which of the active terminals in frame $f$ is transmitting in each time slot. Moreover, as the header univocally determines $\mu$ and thus the set of coefficients used by each node, $R$ is able to know which coefficient is used by each transmitter in a given slot. As we will see in Section III-C, this is of fundamental importance for the decoding process. As said before, the set of headers is a set of pseudo-noise sequences, such as those usually adopted in CDMA. The fundamental difference with respect to a CDMA system is that in the latter the (quasi-)orthogonality of the codes is used to (quasi-)orthogonalize the channels and expand the spectrum, while in NCDP the low cross-correlation of the preambles is used only for determining the identity of the transmitting node, which is obtained without any spectral expansion, as the symbol rate $1/T_s$ is equal to the chip rate (i.e., the rate at which the modulated symbols are transmitted over the channel).

The discrimination capability of pseudo-noise preambles may suffer a consistent degradation in the presence of strong Doppler shifts which are typical of Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) satellite systems. For this reason such node identification method is mostly
suited for geostationary satellite networks as well as terrestrial multiple access systems.

C. NCDP: Receiver Side

The decoding scheme at the receiver side is illustrated with an example in Fig. 2(a) and Fig. 2(b). In the example, a frame with $S = 4$ slots and $N_{tx} = 3$ active terminals is considered. In the figures only bursts with non-zero coefficients are shown. In each slot the receiver uses the pseudo-noise preamble of each burst to determine which node is transmitting and which coefficient has been used for that burst. As described in Section III-B, the coefficients used by a node in each burst are univocally determined by the preamble. The preamble can be determined at the receiver $R$ using a correlator which calculates the correlation of the received signal with the maximal length sequence for each possible shift. In order to increase the number of available sequences of coefficients and to avoid the problems due to the eventual unsuccessful decoding of some of the slots, the system can be designed so that for each preamble a different coefficient is associated to each slot. In this way the sequence of coefficients associated to the preamble changes depending on the slots where the burst that uses that preamble is transmitted. The total number of different sequences associated to a given preamble is, thus, equal to the number of possible dispositions of the $d$ repetitions over the $S$ slots of the frame, that is, $\binom{S}{d}$. Note that the use of a preamble is not a peculiarity of NCDP, as usually practical systems make use of a preamble to perform channel estimation. The preamble is also used by $R$ to estimate the channel for each of the transmitters. More details about the channel estimation are given in [21] and are recalled in Section V. Once the channel has been estimated, the receiver applies PNC decoding to calculate the bitwise XOR of the transmitted messages, as detailed in Section III-A. According to arithmetics in Galois fields and to what is stated in Section III-B, the bitwise XOR is interpreted as a sum in $GF(2^n)$. Thus the slots that have been correctly decoded are interpreted as a system of linear equations in $GF(2^n)$ with coefficients $\alpha_{ij}$, which are known to the receiver through the headers (see Fig. 2(a)). In order to simplify the notation, in the figure we indicated the vector $u_{ij}' = [\alpha_{ij}u_1', \ldots, \alpha_{ij}u_L']$, representing the network coded packet, as $\alpha_{ij}u_i$. At this point, if the coefficient matrix $A$ has full rank, $R$ can recover all the original messages using common matrix manipulation techniques in $GF(2^n)$ (see Fig. 2(b)). If $A$ is not full rank, not all the transmitted packets can be recovered. However, a part of them can still be retrieved using Gaussian elimination. The decoding process in case of rank deficient
A coefficient matrix is analyzed in Section IV. Note that, while in [2] the coefficient matrix $A$ (called mixing matrix) is a complex matrix whose elements are the terminals’ channel gains, in NCDP $A$ is a matrix in a Galois Field. In NCDP each slot is processed only once in the complex domain (PNC decoding), while all matrix manipulations are done in $GF(2^n)$. In [2], instead, the matrix $A$ is processed entirely in the complex domain. Operating in $GF(2^n)$ has an important advantage in terms of complexity, as all the processing can be done in the digital domain and avoids numerical problems that may derive from using a complex matrix, especially in case of small channel gains. If, on the one hand, using a complex coefficient matrix leads to a high probability of having full rank (which, however, also depends on the precision of the quantization in the sampling process), on the other hand in NCDP a relatively small field size (e.g., $GF(2^8)$) already achieves almost the same performance in terms of throughput as in the case of a complex matrix, as we show in Section IV.

IV. THROUGHPUT AND ENERGY EFFICIENCY ANALYSIS

A. Throughput

During each frame the terminals buffer the packets to be transmitted in the following frame. Each terminal transmits its packet more than once within a frame, randomly choosing a new coefficient in $GF(2^n)$ independently at each transmission. As described in the previous section, the coefficients can be generated using a pseudo-random number generator fed with a seed which is univocally determined by the chosen pseudo-noise preamble. Using the preamble the receiver can build up a coefficient matrix $A \in [GF(2^n)]^{S \times N_{tx}}$ for each frame, with $A_{j,i} = \alpha_{ij}$, $\alpha_{ij} \in \{1, \ldots, 2^n - 1\}$. The rows of $A$ represent the time slots while the columns represent the
active terminals, i.e., the terminals that transmit in the present frame. If \( \alpha_{ij} = 0 \), terminal \( T_i \) does not transmit in slot \( j \). During slot \( j \), \( R \) receives the sum of the bursts with \( \alpha_{ij} \neq 0 \). From the received signal, \( R \) tries to obtain the bit-wise XOR of the encoded messages as described in Section II. The XOR is interpreted by \( R \) as a linear equation in \( GF(2^n) \), the coefficients of which are derived through the pseudo-noise preamble as described in Section III. If \( N_{tx} \) is the number of active terminals in a frame and assuming that all the received signals are decoded correctly, a linear system of equations in \( GF(2^n) \) is obtained with \( S \) equations and \( N_{tx} \) variables. Each variable corresponds to a different source message. If \( A \) has rank equal to \( N_{tx} \), then all the messages can be obtained by \( R \). A necessary condition for \( A \) to be full rank is \( N_{tx} \leq S \), i.e., the number of active terminals in a frame must be lower than the number of slots in a frame.

Assuming Poisson arrivals with aggregate intensity \( G \), the probability of such event is:

\[
Pr\{N_{tx} \leq S\} = \sum_{n=0}^{S} \frac{(GS)^n e^{-GS}}{n!},
\]

that includes also the case in which there are no active terminals during a frame. For instance, in case of \( S = 100 \) slots and \( G = 0.8 \) the probability expressed by Eqn. (6) is on the order of 0.99. Even if \( N_{tx} < S \), however, it can still happen that \( A \) is not full rank, i.e., not all the messages can be recovered. The probability that \( A \) is full rank for a given \( N_{tx} < S \) depends on the MAC policy, and particularly on the probability distribution used to choose the coefficients. One possibility is to use a uniform distribution for the coefficients (i.e., each coefficient can assume any value in \( \{0, \ldots, 2^n - 1\} \) with probability \( 2^{-n} \)). In this case the number \( d \) of transmitted replicas is a random variable, and the probability that \( A \) is full rank is [22]:

\[
P(S, N_{tx}) = \prod_{k=0}^{N_{tx}-1} \left(1 - \frac{1}{2^n(S-k)}\right).
\]

Using (6) and (7) we find the expression for the normalized throughput:

\[
\Phi = \frac{1}{S} \sum_{m=1}^{S} \frac{(GS)^m e^{-GS}}{m!} P(S, m) = \frac{1}{S} \sum_{m=1}^{S} \frac{(GS)^m e^{-GS}}{(m-1)!} \prod_{k=0}^{m-1} \left(1 - \frac{1}{2^n(S-k)}\right)
\]

\[
= \frac{1}{S} \sum_{m=0}^{S-1} \frac{(GS)^{m+1} e^{-GS}}{(m)!} \prod_{k=0}^{m} \left(1 - \frac{1}{2^n(S-k)}\right) = G \sum_{m=0}^{S-1} \frac{(GS)^m e^{-GS}}{m!} \prod_{k=0}^{m} \left(1 - \frac{1}{2^n(S-k)}\right).
\]

From Eqn. (8) we can see that \( \Phi \) grows with \( n \), which means that the system throughput increases with the size of the considered finite field. The throughput achievable in case of an asymptotically
large field size $n$ is:

$$
\lim_{n \to \infty} \Phi = \lim_{n \to \infty} \left[ G \sum_{m=0}^{S-1} \frac{(GS)^m e^{-GS}}{m!} \prod_{k=0}^{m} \left( 1 - \frac{1}{2^n(S-k)} \right) \right] = G \sum_{m=0}^{S-1} \frac{(GS)^m e^{-GS}}{m!}.
$$

(9)

Thus, the normalized throughput $\Phi$ tends to the probability of having less than $S$ transmitters in a frame as $n \to \infty$. Note that this is the same performance that would be achieved by a scheme that uses coefficient matrix in the complex domain, as in [2]. Further in this section we show that almost the same performance can be achieved by NCDP using a finite and relatively small field size.

B. Energy

The MAC scheme we just analyzed presents one main drawback in terms of energy efficiency. As a matter of fact, given the frame length $S$, a node transmits each message on average $E[d] = S \times p$ times, $p = (1 - 2^{-n})$ being the probability to choose a non-zero coefficient, i.e., the average number of transmissions grows linearly with $S$. In order to decrease the energy consumption, the probability of choosing the zero coefficient may be increased. However, a reduction in the transmission probability $p$ may affect the system throughput. In order to understand the relationship between the probability $p$ and the throughput $\Phi$, we refer to some results in random matrix theory. The problem can be formulated as follows: consider an $S \times N^{tx}$, $S \geq N^{tx}$, random matrix $A$ over $GF(2^n)$ with i.i.d. entries, each of which assumes value 0 with probability $1 - p$ while with probability $p$ it assumes values in $\{1, \ldots, 2^n - 1\}$. We are interested in the relationship between $p$ and the probability that $A$ is full rank. In [23] the authors show that, in order to achieve a rank $N^{tx} - O(1)$ with high probability, then, for $N^{tx}$ large, $p$ cannot be lower than the threshold probability $\frac{\log(N^{tx})}{N^{tx}}$. At high loads (i.e., $G \simeq 1$), on average $N^{tx} \simeq S$, which means that, setting $p = \frac{\log(S)}{S}$, the average number of transmissions (and so the energy consumption) for each node is $E[d] = \log(S)$, i.e., it grows logarithmically with the number of slots in a frame. On the other side, $S$ must be kept large enough, as this increases the decoding probability (see Eqn. (9)). With reference to the example considered earlier in this section the average number of transmissions corresponding to the minimum required $p$ for $S = 100$ is equal to about 4.6. We evaluated numerically the effect a reduction of $p$ has on $\Phi$ for the case $S = 100$ and $q = 2^8$. We considered three cases. In the first one the transmission probability in each slot has been set to $p = 1 - 2^{-n} = 0.9961$, which corresponds to the case studied in the first part of this
section and for which the throughput is given by Eqn. (8). In the second case we set \( p \) just above the threshold, i.e., \( p = 0.0625 > \frac{\log(S)}{S} = 0.0461 \), while in the last case \( p \) has been set exactly equal to the threshold probability. Fig. 2 shows the results together with the numerical validation of Eqn. (8). It is interesting to note how passing from \( p = 0.9961 \) to \( p = 0.0628 \), with a reduction in transmission probability (or, equivalently, in average energy per message) of about 93.7\%, leaves the throughput unchanged, while a further decrease of \( p \) of just another 1.5\% leads to a 10\% reduction in the maximum throughput with respect to the case \( p = 0.9961 \). The asymptotic analytical curve described by Eqn. (9) is also plotted in Fig. 2. Such curve represents the throughput of a system where coefficients are chosen in a finite field with asymptotically large field size. It also represents the throughput of a system derived from the NDMA scheme proposed in [2], i.e., the coefficient matrix is complex and is processed in the complex domain. It can be seen that the performance in terms of throughput is almost the same for NDMA and for NCDP with \( p = 0.0625 \), i.e., NCDP does not lose significantly with respect to NDMA, while saves in complexity by performing all the processing in a finite field instead of the complex
domain.

To further lower the energy consumption and control the number of repetitions \( d \) (which, being a Bernoulli random variable, can theoretically assume values as large as \( S \)), an alternative is to fix the number of transmitted replicas \textit{a priori}. Although this solution may decrease the probability of decoding all the transmitted messages (because the resulting \( A \) matrix would be a subset of all possible matrix of the same size), it may still be possible to recover part of them by using Gaussian elimination.

V. IMPLEMENTATION ASPECTS

For each frame the receiver \( R \) needs to know which of the active terminals are transmitting in each slot and must have CSI for each of the users. Both needs are addressed including a pseudonoise preamble, such as the spreading codes used in CDMA, at the beginning of each transmitter's burst. In [3] frequency offset and channel amplitude are derived from the clean bursts (i.e., bursts that did not experience collision) and assumed to remain constant over the whole frame. The need for an orthogonal preamble has been removed in CRDSA++ [4]. Unlike in [3], the method we propose can not rely only on clean bursts. Thus the frequency offset and the amplitude of each transmitter must be estimated using the collided bursts.

Channel estimation can be performed using the Estimate Maximize (EM) algorithm as we have shown in [21]. We have adopted the approach described in [24], where the EM algorithm is used to estimate parameters from superimposed signals. We have applied the same approach to estimate amplitudes, phases and frequency offsets from the baseband samples of the received signal in case of a collision of size \( k \). The details on how the EM algorithm is applied can be found in [21].

In Fig. 3 the FER curves for different collision sizes obtained using the LLR values calculated in Section III-A are shown. The FER curve for the case of estimated channels using the EM algorithm are also shown. These results have the purpose of showing the feasibility of channel estimation from the collided messages. We are currently working on the enhancement of channel estimation in order to further improve the performance of PNC in terms of FER. In the simulation all channel amplitudes were assumed to be equal. We performed some preliminary simulations with unbalanced channels and observed a certain degradation in terms of FER. More specifically, the performance of PNC decoding in the case of a collision with different channel gains is
Fig. 3. FER curves for the XOR of transmitted messages for a collision of size 5. $E_b$ is the energy per information bit for each node. A tail-biting duo-binary turbo code with rate $r_{cc} = 1/2$ and codeword length 256 symbols is used by each node. Phase offsets are uniformly distributed in $(-\pi, +\pi)$, frequency offsets are uniformly distributed in $(0, \Delta \nu^{\text{max}})$ with $\Delta \nu^{\text{max}}$ equal to 1% of the symbol rate on the channel. Amplitudes are constant and equal to 1. The FER curves for the case of estimated channels using the EM algorithm are also shown [21].

slightly better than that of the weakest channel. This may require the adoption of appropriate countermeasure depending on the deployment setup. Further analysis on the effect of channel unbalance is out of the scope of the present paper and will be tackled as future work.

In order to detect eventual decoding errors, a cyclic redundancy check (CRC) can be used. The CRC operations are done in $GF(2)$ and, by the linearity of the channel encoder, the CRC field in the message obtained by decoding a collision of size $k$ is a good CRC for $u_s$, which is the bitwise XOR of the messages encoded in the $k$ collided signals. This allows to detect decoding errors, within the limits of the CRC capabilities, also in collided bursts.

Another important issue is the imperfect symbol synchronism. In [21] we have proposed several techniques, based on oversampling, aiming at reducing the impairments brought in by the lack of perfect synchronism. The different methods we proposed in [21] are all based on oversampling and show a loss of about 1 dB at $FER = 10^{-2}$ with respect to the case with ideal synchronism. We do not report here the results for a matter of space.

The lack of timing and phase synchronization determines a certain degradation in the sequence-detection performance of the code. However, such degradation can be usually tolerated in practical systems such as CDMA ones. We refer to [25] for further details on this issue.
VI. Numerical Results

Our performance metrics are the throughput $\Phi$, defined as the average number of bits per second per Hertz (bit/s/Hz) that can be correctly decoded, and the packet loss rate (PLR) $\Upsilon$, i.e., the ratio between the number of lost packets and the total number of packets that arrive at the transmitters. The relationship between these metrics is given by $\Phi = r_{\text{dec}} G (1 - \Upsilon)$, $r_{\text{dec}}$ being the rate in bits per second per Hz (bit/s/Hz) used by the transmitter and $G$ being the average rate at which the new messages (i.e., messages that are being transmitted for the first time) are injected in the network. Note that $G$ is independent from the number of times a message is repeated within a slot. Also note that $r_{\text{dec}}$ is the number of bits per second per Hertz. Thus, for instance, if a QPSK modulation is adopted, $r_{\text{dec}}$ corresponds to twice the rate of the channel encoder indicated with $r_{c}$ in Section II.

We consider two benchmarks. The first one is the CRDSA++ scheme with three repetitions. In CRDSA++ a node transmits three copies of a burst (twin bursts) in different slots randomly chosen within a frame. Each of the twin bursts contains information about the position of the other twin bursts in the frame. If one of the twin bursts does not experience a collision (i.e., it is clean) and can be correctly decoded, the positions of the other twin bursts are known. These bursts may or may not experience a collision with other bursts. If a collision occurs, these are removed through interference cancelation using the knowledge of the decoded bursts. In order to do this $R$ memorizes the whole frame, decodes the clean bursts, reconstructs the modulated signals including the effect of each user’s channel, and subtracts them from the slots in which their replicas are located. The SIC process is iterated for a number $N_{\text{iter}}$ of times, at each time decoding the bursts that appear to be “clean” after the previous SIC iteration. If at the end of the SIC process not all the bursts have been decoded, the receiver tries to decode each of the collided bursts considering the interfering bursts as noise. If a burst is successfully decoded all its replicas are subtracted from the frame and the SIC process starts back.

The second benchmark we consider is a slotted ALOHA system. In a SA system each burst is transmitted only once. Reception is successful if only one burst is transmitted within a given slot, while, in case two or more bursts collide, they are discarded. The capture effect is not considered in the SA scheme.

In both NCDP and CRDSA++ the performance at the physical layer plays an important role
in both system throughput and packet loss rate. In order to take this into account in the most
general way we adopt an information theoretical approach assuming Gaussian codebooks for
CRDSA++ and Lattice codes for NCDP. We also assume channels are symmetric. We evaluate
the performance of CRDSA++ as in [26]. In case of complex-valued channel symbols and
collision of size $k$, a burst can be correctly decoded if [27]

$$r_{dec} \leq \log_2 \left( 1 + \frac{SNR}{1 + (k - 1) \cdot SNR} \right),$$

(10)

where $r_{dec}$ is the rate in bits per second per Hertz while $SNR$ is the signal to noise ratio of the
channel (i.e., $E_s/N_0$).

As for NCDP, we refer to a result in [28] (Theorem 4) according to which the bitwise XOR
of $k$ colliding signals using the same rate $r_{dec}$ and real-valued channel symbols can be correctly
decoded if

$$r_{dec} \leq \frac{1}{2} \log_2 \left( \frac{1}{k} + SNR \right).$$

(11)

We consider two different simulation setups. In the first one the nodes do not receive any
feedback from the receiver, while in the second setup $R$ gives some feedback to the active
terminals. For this last case we consider an automatic repeat request (ARQ) scheme, in which
a node receives an acknowledgement (ACK) or a negative acknowledgement (NACK) from the
receiver in case a message is or is not correctly received, respectively. The amount of feedback
is limited to one ACK/NACK message per node and per frame\(^1\). A node that receives a NACK
enters a \textit{backlog state}. Backlogged nodes retransmit the message for which they received the
NACK in another frame, uniformly chosen at random among the next $B$ frames. We call $B$ the
maximum backlog time. The process goes on until the message is acknowledged [29]. In this
setup we also consider the average energy consumption per received message $\eta$ as performance
metric, defined as the average number of transmissions needed for a message to be correctly
received by $R$.

In both setups we assume a very large population of users and a frame size of $S = 100$.

\(^1\)An alternative to the NACK is to have the transmitters using a counter for each transmitted packet, indicating the time elapsed since it has been transmitted. If the timer exceeds a threshold value (which depends on the system’s RTT), the message is declared to be lost.
In the first setup, in which no feedback is provided by the receiver, the average amount of energy spent by a node for each message which is correctly received does not change with the system load $G$, and is equal to the average number of times a message is repeated within a frame. In Fig. 4 the normalized throughput $\Phi$ is plotted against the normalized traffic load $G$. In the figure the throughput curves of NCDP and CRDSA++ schemes for $d = 3$ replicas and rates $r_{dec} = 0.94$ and $r_{dec} = 0.5714$ (4/7) are shown. The throughput curve for NCDP in case of a constant retransmission probability $p = 0.0625$ is also shown (NCDP$_p = 0.0625$). Note that this probability is above the threshold value we mentioned in Section IV, as for $S = 100$ we have $\log(S)/S = 0.0461$. The scheme with $p = 0.0625$ outperforms all the others in terms of throughput, achieving a peak value of more than 0.7 bit/s/Hz for a rate $r_{dec} = 0.94$. The precoding coefficients of NCDP (indicated as $\alpha_{ij}$ in Section III-B) are drawn uniformly in $GF(2^8)$. The normalized throughput for NCDP with $d = 3$ and $GF(2)$ (not plotted in the figure) has a peak value of about 0.4888, with a loss of about 15% with respect to the case in which $GF(2^8)$ is used. In the figure we see how, depending on the rate and for the same number of repetitions, either NCDP or CRDSA++ achieve the highest throughput peak. This is due to the fact that the packet loss rate of CRDSA++ increases when passing from $r_{dec} = 0.5714$ bpcu.
Fig. 5. Packet loss rate $\Upsilon$ vs normalized traffic load $G$. The field size for the coefficients of NCDP is $2^8$. No feedback is assumed from the receiver. The PLR curve for SA is the same for both the considered code rates.

to $r_{dec} = 0.94$ bit/s/Hz, as confirmed by Fig. 5, where the packet loss rates of the considered schemes are shown. From the figure we also see that the PLR of NCDP is the same for both the considered rates and the difference in the peak throughput is only due to the difference in the rate at the physical level. For both the considered rates NCDP achieves a PLR as low as $10^{-3}$ for a network load of 0.655, while, for the same PLR, CRDSA++ achieves a throughput of 1.14 and 0.46 for $r_{dec} = 0.5714$ and $r_{dec} = 0.94$, respectively. Interestingly NCDP$_p = 0.0625$, although achieving the highest peak throughput, never gets to a PLR lower than $10^{-3}$, which is due partly because of the probability that a node chooses not to transmit in a given frame (that happens with probability $(1 - p)^S$).

It is interesting to note how, in NCDP, increasing the number of transmissions per message (and so the energy consumption) leads to an increase in the peak throughput of the system.

In the second setup, in which feedback is allowed, we evaluate jointly the throughput $\Phi$ and the energy consumption $\eta$ of the schemes under study. In Fig. 6, $\Phi$ is plotted against $G$ for a maximum backlog time $B = 50$ frames. The precoding coefficients of NCDP are drawn uniformly in $GF(2^8)$. The figure shows how $\Phi$ increases linearly with $G$ up to a threshold load value. Such threshold increases with the (average) number of repetitions and corresponds to the
maximum network load for which the throughput in the setup without feedback (Fig. 4) has an almost linear behavior, i.e., the PLR is low. This indicates that, if the load is such that a non negligible fraction of the messages are not decoded at the first attempt, the retransmissions saturate the channel, blocking both the iterative cancelation process of CRDSA++ and the Gaussian elimination decoding in NCDP. Note that this does not happen in the SA system, coherently to what shown in [29] for the case of large backlog time. In order to compare jointly the spectral and the energy efficiency of the different schemes, we plot the curves for the normalized throughput vs the average energy consumption per received message $\eta$, which is shown in Fig. 7. The increase in the number of repetitions corresponds to an increase in throughput but also to a higher energy consumption for a given transmitter in a given frame. However, as shown in Fig. 7, this does not necessarily imply a loss in energy efficiency. From the figure it can be seen that that there is not a scheme that outperforms the others in terms of both energy and spectral efficiency, but which scheme is best depends on the maximum target throughput. SA achieves a higher throughput with a lower energy consumption with respect to the other schemes in the region $\Phi < 0.35$, while in the region $\Phi > 0.35$ both NCDP and CRDSA++ achieve a higher throughput with lower energy consumption with respect to SA.
NCDP achieves a maximum $\Phi$ of 0.653 for $r_{dec} = 0.94$, slightly higher than CRDSA++, for which the peak value is 0.628, for $r_{dec} = 0.5714$. In the NCDP scheme with a retransmission probability of $p = 0.0625$ a peak throughput of 0.655 bit/s/Hz is achieved in correspondence of an average energy consumption of $\eta = 6.25$. The maximum average throughput that is achieved in correspondence to a packet loss rate of $10^{-3}$ is similar in the two schemes, with CRDSA++ achieving a slightly higher throughput (0.64) than NCDP (0.61).

The simulations show that, for the same number of repetitions, which scheme between NCDP and CRDSA++ performs better (in terms of both throughput and packet loss rate) depends on the rate $r_{dec}$. It is not straightforward to find which scheme performs better than the other for any given $(r_{dec}, SNR)$ pair. However, in the following we derive a subset of the region in the $(r_{dec}, SNR)$ plane where NCDP outperforms CRDSA++. We start by deriving an upper bound on the minimum SNR ($SNR_{min}$) required by NCDP in order to decode correctly (at the physical level, i.e., applying physical layer network coding) a collision of any size. From Eqn. (11) follows that for a collision of size $k$, the decoding at physical level of PNC with real-valued modulation is successful if:

$$SNR \geq 4^{r_{dec}} - \frac{1}{k},$$

(12)
Thus, an upper bound on the minimum SNR is:

\[ SNR_{ub}^{min} = 4^{r_{dec}}, \]  

which does not depend on the collision size. Let us now consider CRDSA++. If the iterative decoding stops, in CRDSA++ the receiver tries to decode each burst considering the others as noise in each slot. In case of symmetric channels and complex-valued modulation the maximum rate \( r_{dec} \) in bit/s/Hz such that the decoding is still possible must satisfy

\[ r_{dec} \leq \log_2 \left( 1 + \frac{SNR}{1 + SNR} \right), \]

that corresponds to the case in which there is only one interferer. The SNR below which the decoding (i.e., the collision resolution at physical level within a slot) stops is, thus:

\[ SNR \geq \frac{2^{r_{dec} - 1}}{2 - 2^{r_{dec}}}. \]

From Eqn. (13) and Eqn. (15) we find that NCDP can perform fully while CRDSA++ is limited to decoding and canceling clean bursts (as in CRDSA) if

\[ 4^{r_{dec}} \leq SNR \leq \frac{2^{r_{dec} - 1}}{2 - 2^{r_{dec}}}. \]

Now it is sufficient to note that, if PNC decoding is successful in each slot, NCDP can perform in the digital domain the equivalent of the SIC of CRDSA in case the coefficient matrix \( A \) is not full rank. This implies that, when Eqn. (16) holds, NCDP performs at least as well as CRDSA++. We plot the region defined by Eqn. (13) and Eqn. (15) in Fig. (8). In the region at the right of the picture where the required SNR of CRDSA++ is higher than that of NCDP, the latter scheme outperforms the former. This confirms the results of the simulations shown before, in which, for the same number of repetitions, when \((r_{dec} = 0.5714, SNR = 10 dB)\) CRDSA++ outperforms NCDP, while if \((r_{dec} = 0.94, SNR = 10 dB)\) NCDP performs better. Interestingly NCDP outperforms CRDSA++ at higher rates. We must also note that NCDP is limited to BPSK modulations, while CRDSA++ can be applied also with complex modulations and in principle may achieve higher rates in bit/s/Hz. However, from Fig. 8 we see that in order to work well at rates approaching 1 bit per channel symbol, CRDSA++ needs an asymptotically large \( SNR \).

Further studies are needed to address a fair comparison in case of asymmetric channels.

An important issue in the collision recovery mechanisms considered in this paper is their complexity. NCDP needs a more strict synchronization with respect to CRDSA++ (symbol level...
Fig. 8. Curves in the \((r_{\text{dec}}, \text{SNR})\) plane describing the upper bound on the minimum SNR of NCDP and the minimum SNR for CRDSA++ required to solve collisions at the physical layer. In the region on the right of the picture where the required SNR of CRDSA is higher than that of NCDP, the latter scheme outperforms the former.

versus slot level). However, NCDP may be an argument in favor of symbol synchronous multiple access systems which are currently debated in the satellite communications arena. As for the decoding complexity, although NCDP has a higher complexity in the physical layer decoder with respect to a typical point to point system (that has been studied in our previous work [21]), our scheme has the advantage that, after the PNC decoding, all the operations at the receiver are done in a finite field, which is particularly suited for implementation in the digital domain. CRDSA++, instead, applies a SIC, which requires to store and process the whole frame in the analog domain, requiring by far more memory than in case of a digital frame processing. A fair complexity comparison between the two schemes is not straightforward and can not be addressed exhaustively in the present paper for a matter of space.

VII. CONCLUSIONS

We have proposed a new collision recovery scheme for symbol-synchronous slotted ALOHA systems based on PNC over extended Galois fields. The adoption of an \(E GF\) allows to better exploit the diversity of the system, leading to an increased spectral efficiency and, depending
on the system load, to an increased energy efficiency. We have compared the proposed scheme with two benchmarks in two different setups, one without feedback and the other with feedback from the receiver. In the second setup we have evaluated jointly the spectral efficiency and the energy consumption of the proposed scheme. Once the PNC is applied to decode the collided bursts, the receiver applies common matrix manipulation techniques over finite fields, which results in a high-throughput scheme. We showed that NCDP achieves a higher or comparable spectral efficiency with respect to the considered benchmarks, while there is not a single scheme that outperforms the others in terms of both energy and spectral efficiency, but the best scheme depends on the maximum achievable throughput. For completeness, we also reported our previous results related to several physical layer issues related to multi-user PNC, namely decoding in the presence of offsets and channel estimation. As a final remark, we underline that there is room for significant improvement in the performance of the PNC decoder by using lattice codes, according to recent result in information theory [30].

ACKNOWLEDGEMENTS

This work was partially supported by the European Commission under project ICT-FP7-258512 (EXALTED), by the Spanish Government under project TEC2010-17816 (JUNTOS) and project TEC2010-21100 (SOFOCLES), and by Generalitat de Catalunya under grant 2009-SGR-940. Giuseppe Cocco was partially supported by the European Space Agency under the Networking/Partnering Initiative.

This is the pre-peer reviewed version of the following article: G. Cocco, N. Alagha, C. Ibars, S. Cioni, Network-Coded Diversity Protocol for Collision Recovery in Slotted ALOHA Networks, to appear in Wiley's International Journal of Satellite Communications and Networking, which has been published in final form at [Link to final article].

REFERENCES


September 23, 2013 DRAFT


