Linear Physical-Layer Network Coding for 5G Radio Access Networks

Invited Paper

Alister Burr and Dong Fang

Dept. of Electronics,
University of York
York YO10 5DD, UK
alister.burr@york.ac.uk

Abstract—**We consider the application of linear physical layer network coding to next generation radio access networks (LPNC-RAN). We introduce a linear physical layer network coding scheme based on binary matrices, and illustrate its application to a simple example network topology. We compare this to benchmark schemes based on coordinated multipoint (CoMP), and show that while its performance is poorer than ideal CoMP with unlimited backhaul, it is significantly better than a practical CoMP scheme, while also giving rise to much lower backhaul load.**

Keywords-physical-layer network coding; PNC; COMP; Network MIMO; linear physical-layer network coding

I.   INTRODUCTION

Fifth generation wireless networks will undoubtedly be called upon to provide unprecedented capacity-density: figures up to 25 Gbit/s/km^2 have been suggested [1], and even higher densities may be required in some applications. This inevitably means that cell sizes will have to further reduce, requiring vastly larger numbers of base stations. To make this feasible in terms both of cost and energy efficiency clearly requires revolutionary change in access networks: the cost of providing both backhaul capacity and power would otherwise be unsustainable.

Fortunately in the past decade it has been realized that the current cellular paradigm is highly inefficient in its use both of bandwidth and of power, and that dramatic gains are in principle possible through the use of the *network MIMO* concept, also known as *coordinated multipoint* (CoMP) – although the latter incorporates a range of options, of which what we may call "full" network MIMO is only one [2].

The essential concept here is that instead of competing with one another for resources to serve user terminals in their own cells, base stations instead cooperate to jointly serve all users across all cells. The competitive approach limits the bandwidth available to each user, and inevitably results in interference between users in neighbouring cells. In network MIMO, on the other hand, interference is effectively eliminated, because neighbouring base stations now cooperate to transmit signals to a given user. Such a system can in principle be regarded as a single multiuser MIMO system (hence the term "network MIMO"), in which the cooperating base stations across the whole network act as a single, distributed base station with a much larger number of antennas, and therefore a much larger potential MIMO gain.

The obvious disadvantage of this approach, however, is its complexity. In particular it gives rise to a very large backhaul load, since in principle the signals at each base station antenna have to be sampled and transmitted to some central point where the signal processing will take place. Moreover channel state information (CSI) is required for channels between all base stations and all user terminals, which again must be transferred via the backhaul.

The objective of this paper is to describe an alternative approach which is capable of achieving most of the benefit of network MIMO with (in principle) no increase in backhaul load compared to a non-cooperative network. This is based on another new technology developed in the course of the past decade or so: *network coding* [3], and in particular *physical-layer network coding* (PNC) [4, 5]. We will restrict our attention here to forms of PNC which may be described in a sense we will discuss below as *linear*, and will discuss a very simple example topology in which they can achieve performance very close to full network MIMO, and much better than network MIMO with realistic backhaul constraints.

In the next section we will outlined the network model assumed in this paper, including a simplified topology which we will use later to illustrate the approach. Then in section III we describe linear physical-layer network coding (LPNC), and in particular we will introduce a form of LPNC based on multiplication by binary matrices. We will then discuss the application of LPNC to radio access networks (LPNC-RAN), especially with reference to our simple example topology, on both uplink and downlink, and finally we will give simulation results comparing LPNC-RAN to both ideal and more practical CoMP, to show the potential benefit of the scheme, and conclude the paper.

II. Network Model

Figure 1 shows the general network we consider in this paper. We assume that K user terminals (T1 – TK) are served by L base stations (acting as relays, R1 – RL), which in turn are connected via a backhaul network to a hub, which is further connected to the core network. Here we assume that the backhaul is perfect in the sense that the data is transferred without error, but that it may have limited capacity, and also that the data is transferred in digital rather than analogue form.
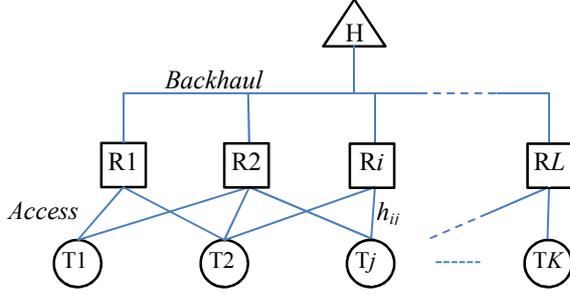
Figure 1 Network Model

The wireless links between terminals and base stations are referred to as the access network: they are assumed to be subject to uncorrelated flat fading, with the fade coefficient between the ith relay and the jth terminal being denoted as h_{ij}. In this paper for simplicity we assume that both terminals and base stations are equipped with only one antenna each. We also assume that all terminals and base stations use the same resources: that is, the network employs 100% frequency reuse. Terminal Tj transmits symbol s_j, modulated as signal x_j, and the corresponding received signal at relay Ri is denoted y_i. Then:

$$y_i = \sum_{j=1}^{K} h_{ij} x_j + n_i$$

where n_i denotes the noise at the ith relay.

The benchmark scheme for comparison here is network MIMO, as mentioned in section I. Here we assume that the base stations collaborate perfectly to serve the terminals: that is, the analogue signals received by the base stations (on the uplink) are available at full precision at the hub for joint processing, and conversely are perfectly transferred from the hub to base stations on the downlink. In addition perfect CSI for all access links is available at the hub. It is clear that in principle this requires infinite backhaul capacity. We will also consider, however, the effect of limited backhaul on this scheme. This benchmark scheme acts as a multiuser MIMO system with (on the uplink) K single-antenna users and L receive antennas.

For purposes of illustration we will analyse a much simpler network, including only two terminals and two base stations, as illustrated in Figure 2. This will suffice to demonstrate the principle and the potential benefits available.
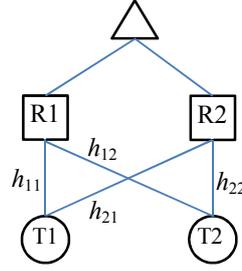
Figure 2 Simplified topology

III. Linear Physical-Layer Network Coding for Radio Access Networks (LPNC-RAN)

In *network coding* a node in a network which receives more than one stream of data simultaneously forwards some function of both streams, rather than selecting only one to forward. However this requires the two streams to be received on separate channels, which in a wireless network requires additional time/spectrum resources. *Physical-layer network coding*, on the other hand, allows the two streams to be received simultaneously at the node on the same channel. The node then decodes the required function directly from the received signal, even if it is not possible to decode the two data streams separately.

A. Example 1: BPSK on Defined Channel

We use a simple example based on the topology of Figure 2 to demonstrate this. Assume for the purpose of illustration that in Figure 2 $h_{11} = h_{21} = h_{22} = 1$, and $h_{12} = 0$, and also that both terminals transmit using BPSK. The scenario is illustrated in Figure 3, which shows also the constellations transmitted from the terminals and received at the relays.
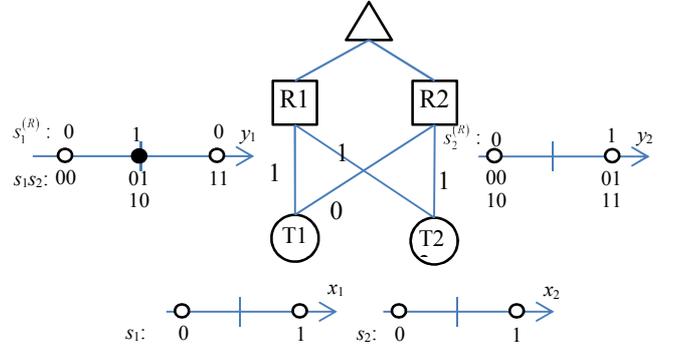
Figure 3 Illustration of PNC in simplified topology

Note that at R1 the same signal is received for $s_1 s_2$ = '10' and for $s_1 s_2$ = '01'. This means that it is not possible for R1 reliably to decode either of the symbols s_1 or s_2 from T1 or T2. At R2 on the other hand, it is clear that only s_2 can be decoded, because the link from T1 is fully faded. However R1 can decode the function $s_1^{(R)} = s_1 \oplus s_2$ of the two symbols, since '10' and '01' both correspond to $s_1^{(1)} = 1$. Then if R1 forwards $s_1^{(R)} = s_1 \oplus s_2$ and R2 forwards $s_2^{(R)} = s_2$, both symbols can be

decoded at the hub, since $s_1^{(R)} \oplus s_2^{(R)} = s_1 \oplus s_2 \oplus s_2 = s_1$. To achieve this with either conventional decode and forward relaying, or even with network coding, would require T1 and T2 to signal in separate channels, occupying twice the spectrum resource on the access network.

Note that the PNC approach maintains the same total backhaul load, 2 symbols, as would a more conventional scheme in which each relay decoded the symbol from the corresponding terminal (if that were possible). The CoMP or network MIMO solution would quantize and forward the received signal at each relay, which would increase the backhaul load several-fold.

B. LPNC over Rings and Fields

This is in fact a very simple example of LPNC, since the exclusive OR function $s_1 \oplus s_2$ is a linear function of the two symbols in the binary field GF(2) or $\mathbb{F}_2$. Here we generalize that to allow relay Ri to decode the result of any linear function of the form:

$$f_i\left(s_{j_1}, s_{j_2} \ldots\right) = a_{i1} \odot s_{j_1} \oplus a_{i2} \odot s_{j_2} \oplus \ldots \quad (1)$$

where $j_1, j_2\ldots$ denote the terminals in the connection set of relay Ri, that is those from which a significant signal is received. The a's denote coefficients of the linear function. We use $\oplus$ and $\odot$ to denote addition and multiplication since these are defined on some finite algebraic ring or field. This generalization allows us to use larger constellations.

We can allow the symbols and the coefficients to be drawn from any ring or field. Although it is conventional to use an algebraic field in these functions, we have shown elsewhere that it is also possible also to use a ring [6], provided the coefficients are not zero-divisors and have unique inverses in the ring. This provides added flexibility.

The simplest approach is to use an integer ring, in which symbols are the integers 0 to q-1, where q is the size, or *cardinality*, of the ring (and hence of the signaling constellation). Then $\oplus$ and $\odot$ denote addition and multiplication, respectively, modulo-q. It is easy to show that a coefficient a has a unique inverse, and is not a zero-divisor, if and only if it is relatively prime to q, so all other elements can be used as coefficients.

If q is prime, then the ring becomes a field, and all elements may be used as coefficients. However with the exception of $q = 2$, the field size is then not a power of 2, and hence the constellation size is inconvenient for a practical communication system.

An alternative framework is provided by the extension fields, and especially the binary extension fields, in which the field elements consist of polynomials of order p-1 in some dummy variable (here we use t) with binary coefficients. Addition $\oplus$ is then addition of coefficients modulo-2, and multiplication $\odot$ is multiplication of polynomials (using modulo-2 arithmetic for the coefficients) modulo some irreducible polynomial of order p. The cardinality of these fields is 2^p, and their elements can readily be converted to binary form.

Table 1 lists the non-zero elements of GF(4), giving them also in binary form and as powers of a primitive element $\alpha = t$.

α^0	1	10	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
α^1	t	01	$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$
α^2	$t^2 \bmod\left(1+t+t^2\right) = 1+t$	11	$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$
$\alpha^3 = \alpha^0$	$\left(t+t^2\right)\bmod\left(1+t+t^2\right) = 1$	10	

Table 1 Elements of GF(4)

This field can also be represented in the form of binary $p \times p$ matrices, whose rows are the binary representation of an element, in the form given in Table 1, followed by the next p-1 powers of α in binary form. They are given in this form in the last column of Table 1. Addition and multiplication are then modulo-2 addition and multiplication of the matrices. It will be noted that all the non-zero matrices in Table 1 are full rank, and therefore invertible, and also that none are zero divisors.

C. LPNC over Ring of Binary Matrices

This matrix notation provides a convenient form for LPNC: the coefficients at a relay can be represented by $p \times p$ binary matrices, and the symbols by length p binary vectors $\mathbf{s}_j$. However this representation of binary extension fields includes only a small subset of the possible $p \times p$ binary matrices, and in fact does not include all those that are full. This inspires an extension to a larger set of matrices.

In fact the set of all $p \times p$ binary matrices is a ring in which addition is again element-by-element addition modulo-2, and multiplication is matrix multiplication using modulo-2 arithmetic. The additive identity is the all zeros matrix; the multiplicative identity is the identity matrix. Not all non-zero ring elements have inverses, since some matrices are singular, and some elements are zero divisors. Hence this is not a field.

Note however, as mentioned, that there are some full rank $p \times p$ matrices which do not occur in the representation of the extension field described above. These are also not zero divisors. Hence this ring can be used provided only full rank coefficient matrices are used. It will provide a much larger range of functions which can be used at the relays, which may have advantages when used in WPNC. For example among the 2×2 matrices there are three matrices:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

which are full rank but which do not appear in Table 1.

We will use the notation $\mathbf{A}_{ij}$ to refer to the coefficient matrix of $\mathbf{s}_j$ in the function at Ri. Then we may write the function output:

$$\mathbf{s}_i^{(R)} = \mathbf{A}_{i1}\mathbf{s}_1 + \mathbf{A}_{i2}\mathbf{s}_2 = \begin{bmatrix} \mathbf{A}_{i1} & \mathbf{A}_{i2} \end{bmatrix}\begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \quad (2)$$

The combined data received at the hub is then:

$$\mathbf{s}^{(R)} = \begin{bmatrix} \mathbf{s}_1^{(R)} \\ \mathbf{s}_2^{(R)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} = \mathbf{As} \qquad (3)$$

where the coefficient matrices of all symbols at all relays are combined in a single composite binary matrix **A**. It is clear that this matrix must be invertible, and hence full rank, if the symbols from all terminals are to be recoverable at the hub.

IV. PERFORMANCE ON FADING WIRELESS CHANNELS

A. Effect of Fading

The performance of the network using LPNC is evidently significantly affected by the fading of the access links, and it is necessary for the encoding functions to adapt to the channel from the terminals to each relay. For certain values of fading coefficients, referred to as *singular fade states*, two different pairs of transmitted symbols result in the same received signal. This occurs in Example 1 in section III.A, where the signals corresponding to data '10' and '01' are the same at R1. As we noted, this implies that these pairs cannot be distinguished. However if, as in Example 1, these pairs result in the same network code function output, this is not a problem: we say the singular fade state is *resolved*.

Of course it is possible for nearly singular fade states to occur, where the signals corresponding to different symbol pairs are close, and may be confused due to noise. Hence for any fade state the optimum (or at least near optimum) approach may be to choose the network code function such that symbol pairs that result in different network coded symbols correspond to received signals which are as far apart as possible. Note, however, as mentioned in section III.C above, the resulting matrix **A** due to functions at all relays must be full rank, so that the data can be recovered at the hub. Hence we seek the full rank matrix **A** which maximizes the minimum distance between constellation points corresponding to symbol pairs yielding different network coded symbols.

B. Example 2: QPSK over Random Channels

We now consider a second example, this time using QPSK transmission and randomly-chosen fading channels. Figure 4 shows the constellation of signals received at a relay for three different pairs of complex random fade coefficients on the links between the terminals and the relay. Appropriate mapping

functions are applied at the relay which assign different received points to different network coded symbols. The points corresponding to the same network coded symbol are plotted in the same colour in the figure.

In principle a search should be performed through all the possible pairs of mappings, corresponding to all the 2^{4p^2} binary matrices **A** to determine the full rank matrix which maximises the minimum distance over the received mapped constellations at both relays. This however would require a very large number of possible mappings to be examined. In practice we have found that several different mappings give the same minimum distance for a given set of channel fading coefficients, and in fact that a much reduced subset of the possible mappings is sufficient to ensure that an optimum mapping can always be found among them.

a	1	0	0	0
b	0	1	0	0
c	0	0	1	0
d	0	0	0	1
e	1	0	1	0
f	0	1	0	1
g	1	0	0	1
h	0	1	1	0
i	1	1	0	1

Table 2 Rows of network code matrix **A**

These matrices always have rows chosen from Table 2. Hence rows may be selected from this table to fulfil the requirements. It is interesting to consider the effect of some of these mapping matrices. For example a matrix at one relay made up of rows *a* and *b* would simply map according to the first terminal symbol, neglecting the other terminal. This would be optimum if one signal is much stronger than the other, as in the constellation illustrated in Figure 4(a). Similarly rows *c* and *d* would map according to the second terminal symbol. Rows *e* and *f* effectively form the XOR of the two bits of the constellation labels. This might be optimum in the case of the constellation of Figure 4(b); in some cases rows *g* and *h* may be optimum there (depending on the relative phase shift).
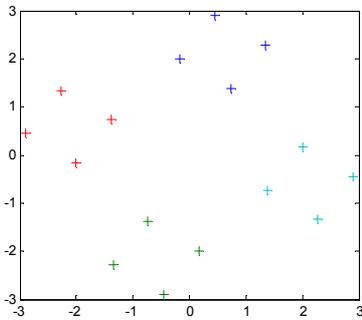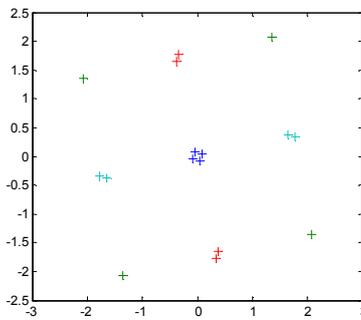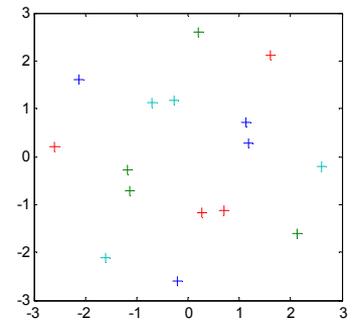
(a)   (b)   (c)

Figure 4 Signal constellations at relay for three different random channels

Note that in some cases, where symbols from both terminals are heavily attenuated at one relay, it may be preferable to ignore that relay, and instead decode both symbols at the other relay and forward that to the destination. This case can also be treated as a special case of the network coding in which the matrix at one relay is set to unity while the other is null, and selected when optimum in the same way as the others.

V. LPNC-RAN DOWNLINK TRANSMISSION

The LPNC-RAN approach can also be applied in the downlink direction. In this case the vector of data at the hub to be transmitted to the terminals is multiplied by a precoding matrix **B**, which in general is defined over the ring used for the network coding functions on the uplink, and in our case is again a binary matrix. This defines the symbols transmitted from the relays. The matrix **B** is again chosen to be full rank, and also to maximize the minimum distance in the constellations received at the terminals between signal points corresponding to different symbols at that terminal. In general, if the access link channels are reciprocal, **B** should be the inverse of the transpose of **A** – which of course is invertible since it is required to be full rank.
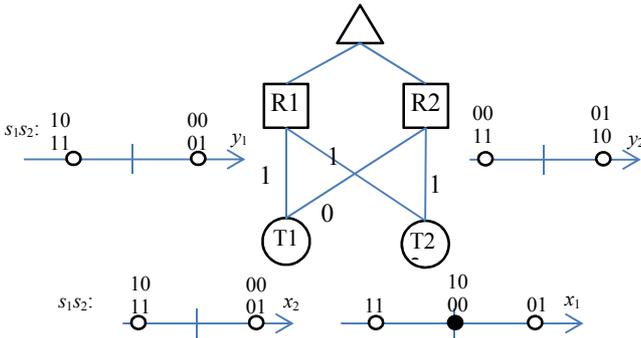
Figure 5 Simple downlink transmission example

Figure 5 illustrates this for the scenario of Example 1 in section III.A, where now the relays transmit BPSK signals to the terminals. In Example 1 the matrix **A** is in fact $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, denoting that R1 decodes $s_1 \oplus s_2$ and R2 decodes s_2. In the downlink scheme we use the inverse of the transpose of this matrix, which is in this case $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Thus now R1 transmits s_1 and R2 transmits $s_1 \oplus s_2$. The figure shows the constellation received at each terminal. At T1 s_1 is received: at T2 the signals corresponding to $s_2 = 0$ lie at the origin, while those corresponding to $s_2 = 1$ lie at ± 2 – hence s_2 can be decoded at T2, as required.

This example highlights that in general the constellation at the terminals may be composite: that is, in general several constellation points may correspond to one data symbol. These may appear as a cluster of points, or as several discrete points. In many cases a modulo operation can be applied to generate a constellation in which the points corresponding to one symbol always occur as a cluster.

VI. PERFORMANCE RESULTS

In Figure 6 we present performance results for the uplink of LPNC-RAN compared with three benchmark schemes based on "full" CoMP or Network MIMO. In all cases the network topology is that shown in Figure 2, and the terminals transmit QPSK. The channels between terminals and relays are subject to uncorrelated Rayleigh fading, with mean square fade coefficient $\overline{|h_{ij}|^2} = 1, \forall i, j$, which is assumed to be stationary in each frame period. We plot frame error ratio (FER) for frames of length 100 symbols against signal to noise ratio (SNR), defined as the ratio of the transmit power at each terminal to the total noise power at each relay.
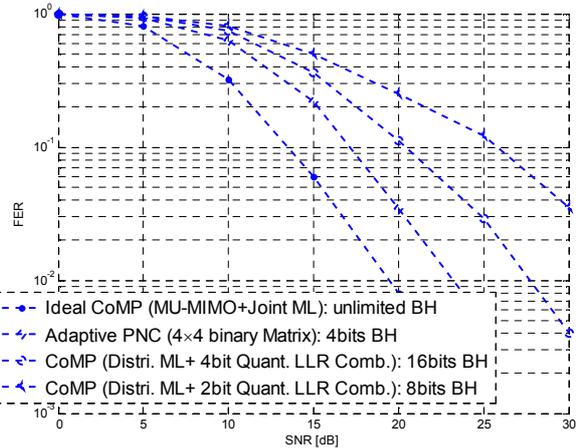
Figure 6 Simulation result for uplink LPNC compared to benchmark CoMP schemes

A. LPNC simulation

In the LPNC-RAN system (referred to as "Adaptive PNC" in Figure 6), in each frame period the random channels are determined, and the system chooses the network code matrix **A** as the full rank matrix which maximizes the minimum distance between signals at relays corresponding to different network coded symbols. Note that this assumes that full CSI for all channels is available to some central coordination function, probably located at the hub, which can then distribute the network code matrices to the relays. We then use Monte Carlo simulation to determine the probability that at least one symbol error occurs in a frame. No FEC coding is employed.

As mentioned above, in the LPNC scheme the total backhaul load is equal to the total number of bits generated: i.e. 4 bits per symbol period.

B. Benchmark schemes

We compare the results for LPNC with three benchmark schemes. The first is ideal CoMP, in which the received noisy signal from each relay is directly forwarded to the hub via the backhaul. Since we are assuming a digital implementation of the backhaul, this implicitly assumes unlimited backhaul capacity. At the hub we perform maximum likelihood (ML) detection for each symbol, that is the vector of estimates of the transmitted symbols:

$$\hat{\mathbf{s}} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \arg\min_{s_1, s_2 \to \mathbf{x}} \| \mathbf{y} - \mathbf{Hx} \|^2 \tag{4}$$

We also consider two CoMP schemes with limited backhaul load. Here we calculate the log-likelihood ratio (LLR) of each bit of each terminal's symbol at each relay, using:

$$L_{ij,k} = \log\left(\frac{\sum_{s_j : b_k(s_j) = 1} P(s_j | y_i)}{\sum_{s'_j : b_k(s_j) = 0} P(s'_j | y_i)} \right) = \log\left(\frac{\sum_{s_j : b_k(s_j) = 1} P(y_i | s_j)}{\sum_{s'_j : b_k(s_j) = 0} P(y_i | s'_j)} \right) \tag{5}$$

where $L_{ij,k}$ denotes the LLR of the kth bit of the binary representation of the symbol from the jth terminal, measured at the ith relay, and $b_k(s)$ represents the kth bit of the binary representation of symbol s. Note that this assumes that all transmitted symbols are equiprobable.

We quantize this using a scalar quantizer, and transmit to the hub, where the LLRs are combined and data decisions carried out. We consider two quantization schemes which quantize LLRs to 2 bits and 4 bits respectively. Four LLRs (two bits for each symbol at each terminal) are calculated at each relay, and hence the total backhaul load is 16 and 32 bits/symbol period for the 2 and 4 bit quantization schemes, respectively.

C. Results and Comparison

Figure 6 shows that ideal CoMP has the best FER performance, as might be expected. It achieves a diversity order of 2, again as expected given there are two receive antennas. LPNC however also achieves the same diversity order, showing that the PNC can also exploit the diversity available from the two relays. Its performance is around 3.5 dB poorer than ideal CoMP.

However restricting the backhaul capacity for CoMP has a significant impact on performance. The degradation with 4 bit quantization is 8.5 dB and with 2 bit quantization is 13.5 dB. The backhaul load in these two cases is respectively 8 and 4 times that required in LPNC. Hence LPNC has about 5 dB better performance while requiring only one eighth of the backhaul load compared with 4 bit quantized CoMP.

VII. CONCLUSIONS

We have discussed the application of linear physical layer network coding to next generation radio access networks. We have introduced a form of LPNC using network coding functions defined by binary matrices, showing how this can be obtained as a generalization of LPNC based on the binary extension fields GF(2^q). We illustrate the operation of the scheme with respect to a simple network containing two terminals communicating with a hub via two base stations acting as relays, and with give simulation results for the uplink of LPNC-RAN compared with both ideal CoMP and CoMP with restricted backhaul, While the LPNC-RAN has about 3.5 dB poorer performance than ideal CoMP, it is around 5 dB better than CoMP with a quantized backhaul, with one eighth of the backhaul load. We also discuss how the scheme can be applied in the downlink.

REFERENCES

[1] S. Liu and J. Wu "A 25 Gb/s(/km²) urban wireless network beyong IMT-Advanced" IEEE Commuications Magazine, vol. 49, no. 2, pp. 122-129, February 2011

[2] .R. Irmer, et al., "Coordinated Multipoint: Concepts, performance, and field trial results" IEEE Communications Magazine, vol. 49, no. 2, pp102-111

[3] R. Koetter and M. Medard "An algebraic approach to network coding" IEEE/ACM Trans. Networking, vol. 11, no. 5, pp. 782-795, October 2003

[4] S. Zhang, S. C. Liew, P. P. Lam "Hot topic: physical-layer network coding" Proc. Mobicom'06, pp. 358-365

[5] T. Koike Akino, P. Popovski and V. Tarokh "Denoising Maps and Constellations for Wireless Network Coding in Two-Way Relaying Systems" IEEE Globecom 2008

[6] A. Burr and D. Fang "On linear physical-layer network coding" EUSIPCO 2014, Lisbon, September 2014