

Geometrical bounds on the efficiency of wireless network coding

Petteri Mannersalo*, Georgios S. Paschos^{†‡} and Lazaros Gkatzikis[§]

*VTT Technical Research Centre of Finland. Email: petteri.mannersalo@vtt.fi

[†]Informatics & Telematics Institute, CERTH-ITI, Greece. E-mail: gpasxos@iti.gr

[‡]Laboratory for Information and Decision Systems, LIDS-MIT, USA. E-mail: gpasxos@mit.edu

[§] Dept. of Computer & Communication Engineering, University of Thessaly, Greece. E-mail: lagatzik@uth.gr

Abstract—This paper explores wireless network coding both in case of deterministic and random point patterns. Using the Boolean connectivity model we provide upper bounds for the maximum encoding number, i.e., the number of packets that can be combined such that the corresponding receivers are able to decode. For the models studied, this upper bound is of order \sqrt{N} , where N denotes the (mean) number of neighbours. Our simulations show that the \sqrt{N} law is applicable to small-sized networks as well. Moreover, achievable encoding numbers are provided for grid-like networks where we obtain the multiplicative constants analytically. Building on the above results, we provide an analytic expression for the upper bound of the efficiency of wireless network coding. The conveyed message is that it is favourable to reduce computational complexity by relying only on small encoding numbers, for example, XORing only pairs, as the resulting throughput loss is typically small.

Index Terms—encoding number, network coding, random networks, wireless

I. INTRODUCTION

Network coding is an exciting new technique promising improvements to the current limits of data transfer in wireless networks. For the case of multiple unicast sessions (or flows), the basic idea of inter-session network coding is to combine packets belonging to different sessions within the network and thereby to reduce the number of transmissions or equivalently increase the throughput region of these flows. The scheme that achieves the theoretical maximal throughput region for multiple unicasts is yet unknown, but there are practical approaches providing increased throughput. The network coding scheme called *local network coding* was one of the first practical implementations able to demonstrate throughput benefits; see COPE in [1], [2]. The idea of local network coding is to encode packets belonging to different unicast flows whenever it is possible for these packets to be decoded at the next hop. The simplicity of this scheme is promising when considering actual implementations in real-world wireless routers. For example, in the well-known Alice-relay-Bob scenario, the relay XORs outgoing packets while Alice and Bob use their own packets as keys for decoding, proving throughput improvement up to factor 4/3 by eliminating one unnecessary transmission.

Local network coding has been enhanced with the functionality of *opportunistic listen*. The wireless terminals are exposed to information traversing the channel, and [1] proposed a

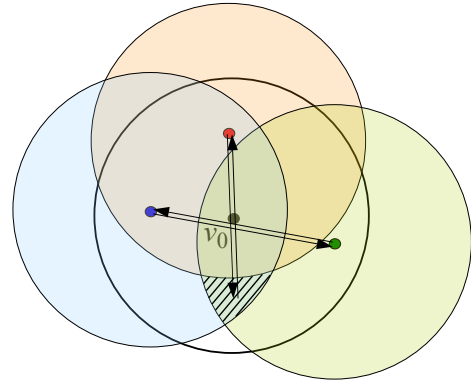


Fig. 1. Local network coding. Node v_o is a relay and the disks indicate the connectivity ranges of v_o and three other nodes. The shaded area shows the possible positioning of a node resulting in a valid combination for the four flows denoted by arrows.

smart way to make the best of this inherent broadcast property of the wireless channel. Particularly, each terminal operates in always-on mode, overhearing constantly the channel and storing all overheard packets. The reception of these packets is explicitly announced to an intermediate node, called the *relay*, which makes the encoding decisions. Finally, the relay can arbitrarily combine packets of different flows as long as the recipients have overheard the necessary packets for decoding.

The so-called *infinite wheel topology* is an unrealistic example where the benefits of opportunistic listen are maximal. In such topology, a relay locates in the middle and all nodes are connected to everyone else except the intended receiver. Thus each flow needs to be relayed, but on the other hand, the transmissions are overheard by all the others. If all possible flows are conveniently assumed to exist and opportunistic listen is utilised, as the number of nodes tends to infinity network coding can improve aggregate throughput by an order of 2 by diminishing the downlink into a single transmission; see [3]. However, such a symmetric topology is expected to appear rarely in real settings. Besides, the above calculations assume that all links have the same transmission rates; thus it takes the same amount of time to deliver a native (non-coded) packet or an encoded one.

As a more realistic scenario, consider the wireless network of Figure 1. Each disk indicates the connectivity range of a node and the shaded area shows the locations where the

positioning of an additional node would form a small wheel topology with one relay (node v_0) and four nodes. In practice, existence of large combinations is rare: the bigger the combination, the more restrictive geometric constraints. Thus, a natural question arises: *What is the expected throughput gain of local network coding with opportunistic listening in a wireless ad hoc network with arbitrarily positioned nodes?*

Deciding which packets to group together in an encoded packet is not a trivial matter as explained in [3], [4], [5]. Specifically in ER [4] and CLONE [5], the medium is assumed to be lossy, and the goal is to find the optimal pattern of retransmissions in order to maximise throughput. Finding the optimal encoding scheme can be reduced to finding a minimum clique partition of the coding graph. Work related to index coding has shown that this problem can be mapped to the Boolean satisfiability problem (SAT problem) [6]. If the coding techniques are restricted to be linear, the selection of the minimum time transmission policy is equivalent to filling a sampled matrix in order to minimise its rank [7]; a problem known to be NP-hard. Moreover, the same complexity appears when the relay node makes scheduling decisions, i.e., selects which packets to serve and what combinations to use; see e.g. [8]. Evidently, restricting the search on a space of small combination sizes reduces the complexity and hence is very desirable from the point of practical system algorithms. Thus, the second natural question arises: *What is the loss in throughput gain if instead of searching over all possible encoded packet combinations, we restrict our search in combinations of size at most m ?*

In this paper, we show that there exist inherent geometric properties bounding the maximum encoding number below a number relative to the population or density of nodes. Applying the Boolean connectivity model, it is demonstrated that opportunities for large encoding combinations rarely appear in ad hoc wireless network. We study the maximum encoding number in different scenarios including both regular and random topologies. To capture the behaviour of large (or dense) networks, the asymptotic scaling laws of the maximum encoding number are derived. In support of our findings, recent work [9] found evidence through simulations that in randomly positioned wireless networks and k -tuple coding, which is a generalisation of local coding to the multi-hop case, in most cases mainly pair combinations appear, i.e., $k = 2$.

Scaling laws are of extreme interest for the network community in general because they provide valuable insights to the system designers. In this direction, the authors of [10] study the scaling capacity of wireless networks in a Gupta-Kumar way taking into account complex field network coding. [11] also consider the scaling capacity and find that network coding cannot improve the order of throughput, i.e., the $O\left(\frac{1}{\sqrt{n}}\right)$ law prevails. [12] discusses the issue of scaling network coding gain in terms of delays while [13] identifies the energy benefits of network coding both for single multicast session as well as for multiple unicast sessions. In [14], network coding is used instead of power control and the benefits are characterised. In

a similar spirit, [15] investigates the use of rate adaptation for wireless networks with intersession network coding. By utilising rate adaptation, it is possible to change the connectivity and increase or decrease the number of neighbours per node. They identify domains of throughput benefits for such a case.

Our work is in line with [16], [17], where the authors also analyse the maximum encoding number, i.e., the maximum number of packets that can be encoded together such that the corresponding receivers are able to decode. They show that this number scales with the ratio $\frac{R}{\delta}$ where δ is a region outside the communication region and inside the interference region. Note however, that this work relates the scaling law of the maximum encoding number to specific protocol properties. In networks with small δ , e.g., where a hard decoding rule is applied, the bound for the maximum encoding number becomes very loose.

The main contributions of this paper are the following. First, we prove that, assuming a Boolean connectivity model, the nodes belonging to a valid combination can be considered as the vertices of a convex polygon. Then, we continue by considering fixed separation distance networks, like a square grid, and show that in such networks, the maximum encoding number is $O(\sqrt{N})$ and $\Omega(\sqrt[3]{N})$,¹ where N is the number of nodes in transmission range of the relay. Next, we study a random network where the locations of the nodes follow a Poisson point process on the plane. In this case, the maximum encoding number is found to be asymptotically bounded by $\lambda^{\frac{1}{2}+\epsilon}$, where λ is the node density and $\epsilon > 0$ arbitrary. In other words, the law $\sqrt{E[N]}$ is valid again. Finally, we consider the case where the encoder searches for combinations of at most size $m < N$. We show that the throughput efficiency loss in this case depends on the size of the network, and for small networks the loss can be negligible. Through simulations we show that all the derived results hold quite well even for small networks.

The paper is organised as follows. In Section II, the model is described and some basic properties are given. In Section III, the main results for the grid-like networks are derived. Then in Section IV random networks are considered. A rate analysis is provided in Section V and simulation results are shown in Section VI. The paper is concluded in Section VII.

II. COMMUNICATION MODEL

We assume a set of nodes \mathcal{V} , located on the plane. Communications between these nodes are established via the Boolean connectivity model with constant communication radius (see, e.g., [18]). In this model, a link between two nodes $\{v_i, v_j\}$ exists if and only if $|\mathbf{X}(v_i) - \mathbf{X}(v_j)| \leq R$, where $\mathbf{X}(v_i)$ is the location of node v_i and R is the communication radius. In this case, we say that v_i is connected with v_j and vice versa. This corresponds to an undirected graph in the sense that only bi-directional links appear.

¹The symbol $O()$ denotes that the function is bounded above by some linear function of the expression in the brackets whereas $\Omega()$ denotes that the expression is bounded from below.

A. Information flows

Each node v_i having at least two neighbours can relay information. By employing local network coding, unnecessary transmissions can be avoided if decoding of packets at the receivers can be guaranteed. To simplify the analysis, we consider only one cell, i.e., we focus on a given node v_0 and all its neighbours, see Figure 2. Thus, we restrict \mathcal{V}_0 containing all neighbours of v_0 , with $\mathcal{V}_0 = \{v_1, v_2, \dots, v_N\}$ and $N \doteq |\mathcal{V}_0|$ the number of nodes under consideration. For a network determined by a Poisson point process with density λ , we use correspondingly the mean number of points which is given by $E[N] = \lambda\pi R^2$.

Apart from the number of neighbours, the gain analysis depends also on the activated unicast flows. In the simple Alice–relay–Bob topology, it is possible that only the flow going from Alice to Bob is activated, in which case the gain is zero. In this paper we are interested in determining an upper bound for the efficiency loss when the relay is constrained on combinations of size $m < N$ (e.g., if $m = 2$ the system is constrained to pairwise XORing). For this reason, we consider the maximum gain scenario, i.e., for each node designated as a relay, we assume that all possible two-hop flows traversing this relay are activated. This means that the node designated as a relay, has all possible different packets from which to select an XOR combination to be sent to the neighbours. Moreover, all the neighbours have opportunistically listened the uplink transmissions and stored the overheard packets. Again, the Boolean connectivity model determines which uplink transmissions can be overheard, i.e., interference is neglected. Since not all encoding combinations are possible, finding the maximum valid combination that corresponds to the maximum encoding number is a non-trivial task and will be the goal of this paper. The resulting bound will help to characterise the efficiency loss due to resorting to m -wise encoding. In real systems, some flows might not be active and interference decreases the possibilities of opportunistic listening. Thus the resulting efficiency loss from m -wise encoding will be even smaller.

To make this more precise, we define source-destination pairs designating 2-hop flows that cross the relay, similar to [4]. Let \mathcal{F} denote all the possible packet flows on \mathcal{V} . Each flow $f \in \mathcal{F}$ has a source $S(f)$, a destination $D(f)$, a set of nodes $H(f) \subset \mathcal{V}$ having its packets (either by overhearing or ownership) and a set of nodes $N(f) \subset \mathcal{V}$ needing its packets; in our study the latter is reduced to the destination of the flow, i.e. $N(f) \equiv \{D(f)\}$. Also, note that the set $H(f)$ is the set of nodes connected to the source plus the source itself. Two flows f_i, f_j are called symmetric when they satisfy the property $S(f_i) = D(f_j)$ and $D(f_i) = S(f_j)$.

B. Constraints

Here we focus on network coding opportunities appearing in the described network model around the relay v_0 . This is done by summarising the previous subsection in the form of formal definitions and properties.

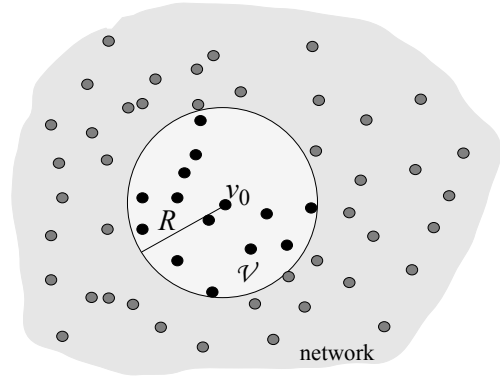


Fig. 2. The relay v_0 and the valid nodes located within distance R .

The following definitions guarantee that a valid flow cannot be 1-hop flow, i.e., the source cannot transmit the packets directly to the destination and the relay v_0 is needed. Moreover, the allowed combinations of flows and the maximum size of such sets are defined.

Definition 1. (Valid node): A node $v_i \in \mathcal{V}$ is a valid node if $|\mathbf{X}(v_i) - \mathbf{X}(v_0)| \leq R$, i.e., it lies within the communication radius of the relay v_0 .

Definition 2. (Valid flow): A flow $f \in \mathcal{F}$ is a valid flow if $S(f)$ and $D(f)$ are valid nodes satisfying $|\mathbf{X}(S(f)) - \mathbf{X}(D(f))| > R$, i.e., they are not directly connected to each other.

Definition 3. (Valid combination): A set of valid flows is a valid combination if any combined XOR-packet sent by the relay can be immediately and correctly decoded by the destination nodes.

Definition 4. (Maximum encoding number): The maximum encoding number C_{\max} is the maximum cardinality among all valid combinations $\mathcal{C} \subseteq \mathcal{F}$.

Note that if the locations of the nodes are random, then C_{\max} is evidently a random variable. We could also impose additional constraints. For example, if a flow can be routed more efficiently through a node other than v_0 , then this flow should be excluded from the set of valid flows. This would further restrict the set of valid combinations and thus, by omitting this constraint we derive an upper bound for C_{\max} .

Next we state the properties that a set of flows need to satisfy in order to be usable in local network coding.

Lemma 1. If a subset of flows $\mathcal{C} \subseteq \mathcal{F}$ is a valid combination, then every pair of flows $f_i, f_j \in \mathcal{C}, f_i \neq f_j$ satisfies

- (i) $D(f_i) \in H(f_j)$,
- (ii) $D(f_i) \neq D(f_j)$,
- (ii) $S(f_i) \neq S(f_j)$.

Proof: (i) Consider a pair of flows $f_1, f_2 \in \mathcal{C}$. Then the relay could send a packet $f_1 \oplus f_2$. However, if $D(f_1) \notin H(f_2)$ such message could not be decoded at node $D(f_1)$. This would contradict the assumption that all XOR combinations within \mathcal{C} can be immediately and correctly decoded. Cases (i) and (ii) can be proved analogously. ■

When the set of sources is identical to the set of destinations, the combination consists of symmetric flows only. Trivially, the size of such combination must be even. Next, we show that in order to calculate an upper bound of the network coding combination size, it is enough to resort to the case of symmetric flows.

Lemma 2. *For any valid combination there exists at least one combination of the same or larger size that contains only symmetric flows.*

Proof: We will show that for any flow we can add the symmetric one without invalidating the combination as long as it is not already counted.

In a bipartite graph with all the nodes \mathcal{V}_0 on one side and the destinations of \mathcal{C} on the other, consider a directional link ℓ_f , between the source of flow f and its destination, for each $f \in \mathcal{C}$. Note now that the nodes having out-degree one, i.e., the active sources in \mathcal{C} , may or may not be identical to one of the destination nodes. We can make a partition of the set of active sources by assigning those with the above property to the set $\overline{\mathcal{T}}_{\text{sym}}$ and the rest to the complementary set $\overline{\mathcal{T}}_{\text{sym}}$. If $\overline{\mathcal{T}}_{\text{sym}} = \emptyset$, then the Lemma is proved since \mathcal{C} is a valid combination with symmetric flows only. If not, then we can create a new combination \mathcal{C}' which has more flows than the original one using the following process. For each transmitter in $\overline{\mathcal{T}}_{\text{sym}}$, say $S(f_i)$ the transmitter of flow f_i , add one extra flow f'_i with $S(f'_i) = D(f_i)$ and $D(f'_i) = S(f_i)$. This flow does not belong to \mathcal{C} (because $S(f_i) \in \overline{\mathcal{T}}_{\text{sym}}$) and it does not invalidate the combination due to the bidirectional properties of the model. Note that f'_i is a valid flow because $S(f'_i)$ cannot be connected to $D(f'_i)$ due to validity of f_i . Note also that $S(f'_i)$ is connected to $D(f)$ for all $f \in \mathcal{C}$ since this is again required for the decoding of the original flows. Thus, for any flow we can add the symmetric one without invalidating the combination. ■

In graph theory terms, a valid combination with symmetric flows can be thought of as a graph created by a clique of $C+1$ nodes, minus a matching with $\frac{C}{2}$ edges, with all symmetric flows defined by this matching activated. The node with zero degree in the matching is the relay node. This graph is called in [1] *wheel topology*.

Finally, we provide a result on the topology of a valid combination. Let $\mathcal{X}_{\mathcal{C}}$ represent the set of locations of all nodes being either the source or destination of a flow belonging to a valid combination \mathcal{C} .

Lemma 3. *For any valid combination \mathcal{C} of size 3 or larger, there exists a convex polygon with the set of vertices equal to $\mathcal{X}_{\mathcal{C}}$.*

Proof: Consider a valid combination defined by flows

$$\mathcal{C} = \{f_i, i = 1, \dots, C\},$$

where $C \geq 3$. Consider also the set of nodes that are sources and/or destinations in \mathcal{C}

$$\mathcal{V}_{\mathcal{C}} = \cup_i S(f_i) \cup_i D(f_i)$$

and the induced set of locations $\mathcal{X}_{\mathcal{C}}$ such that we have a bijective mapping for each element $v_j \in \mathcal{V}_{\mathcal{C}}$ with an element $\mathbf{X}(v_j) \in \mathcal{X}_{\mathcal{C}}$.

Assume that there is a node $v_j \in \mathcal{V}_{\mathcal{C}}$ which is an interior point of the convex hull² of $\mathcal{X}_{\mathcal{C}}$. Its location $\mathbf{X}(v_j)$ can be written as $\mathbf{X}(v_j) = \sum_{i \neq j} \alpha_i \mathbf{X}(v_i)$ where $\sum_{i \neq j} \alpha_i = 1$ and $\alpha_i \geq 0$ for all i .

On the other hand, there is a unique $v_{j^*} \in \mathcal{V}_{\mathcal{C}}$, which is the communicating pair (source or destination) of v_j in at least one flow, so that

$$|\mathbf{X}(v_j) - \mathbf{X}(v_{j^*})| > R. \quad (1)$$

All the other nodes (destinations or sources) in $\mathcal{V}_{\mathcal{C}}$ should be able to reach the node v_{j^*} directly. Thus,

$$|\mathbf{X}(v_j) - \mathbf{X}(v_{j^*})| \leq \sum_{i \neq j} \alpha_i |\mathbf{X}(v_j) - \mathbf{X}(v_{j^*})| \leq \sum_{i \neq j} \alpha_i R \leq R,$$

which is a contradiction to (1). Consequently the node v_j , as well as all other nodes of the combination, necessarily lie on the perimeter of the convex hull. Thus, the nodes of a valid combination are the vertices of a convex polygon. ■

III. ANALYSIS IN GRID-LIKE TOPOLOGIES

In this section we focus on networks where the nodes are deterministically located on a square lattice or a different form of a grid. Grid topologies often offer an insightful first step approach towards the random positioning behaviour. Also, the investigation of grids answers the question whether it is possible to achieve high network coding gain by arranging the locations of the nodes.

We therefore assume a network with the additional property $|\mathbf{X}(v_i) - \mathbf{X}(v_j)| \geq d$, for any pair of nodes $v_i, v_j \in \mathcal{V}$, where d is minimum Euclidean distance of any two nodes in the network. A topology satisfying this geometric property is called *fixed-separation network*. This pertains to regular grids such as the square, the triangular and the hexagonal grid as well as other grids with non-uniform geometry. We impose nevertheless the property that the node density is the same over all cells and thus the geometry should be homogeneous. The number of nodes inside a disk of radius R will be $N = O\left(\left(\frac{R}{d}\right)^2\right)$ for these networks and the corresponding node density $\lambda^{\text{grid}} = O\left(\left(\frac{1}{d}\right)^2\right)$.

Theorem 1. (Upper bound) *The maximum encoding number in fixed-separation networks is $O\left(\sqrt{N}\right)$ where N is the number of nodes or equivalently $O\left(\sqrt{\lambda^{\text{grid}}}\right)$.*

Proof: From Lemma 3 we know that the nodes belonging to the maximum combination form a convex polygon. Any such polygon fitting inside the disk of radius R must have perimeter smaller than $2\pi R$. Since the nodes on the perimeter should be at least d away from each other, we conclude that the maximum encoding number is

$$C_{\max} < \frac{2\pi R}{d}.$$

²The convex hull of points \mathcal{X} is the minimal convex set containing \mathcal{X} .

This combined with $N = O\left(\left(\frac{R}{d}\right)^2\right)$ or respectively $\lambda^{\text{grid}} = O\left(\left(\frac{1}{d}\right)^2\right)$ yields the result. ■

So far we have shown that any network with fixed separation distance d and uniform density, will have maximum encoding number upper bounded by $O(\sqrt{N})$, where N is the number of nodes connected to the relay. A particular case of the above bound is the square grid. The number of nodes inside the disk is $N = \pi\left(\frac{R}{d}\right)^2 + e\left(\frac{R}{d}\right)$ where $e\left(\frac{R}{d}\right) \leq 2\sqrt{2}\pi\frac{R}{d}$ is an error inversely proportional to d . Thus, we obtain an upper bound

$$C_{\max}^{\text{square}} < \sqrt{4\pi N}.$$

This gives us for the square grid a constant of $2\sqrt{\pi}$. however, our numerical studies show that the actual maximum encoding number is approximately half of that calculated above. The reason for that is basically that the valid polygon is always smaller than the disk of radius R and often close to the size of a disk of radius $\frac{R}{2}$. In particular, the constant can be determined for any given grid.

It is interesting to bound the achievable maximum encoding number from below as well. To obtain intuition about this bound we start with a non-homogeneous topology, the *cyclic grid*. We construct concentric cyclic groups of radius $R_i = id$, $i = 0, 1, \dots, \lfloor \frac{R}{d} \rfloor$ that fall inside the disk of radius R . Each cyclic group has as many nodes as possible such that the fixed separation distance condition is not violated. Such a topology exhibits different behaviour depending on the selected origin (it is not homogeneous), nevertheless it helps identify a particular behaviour of the achievable maximum encoding number. The cyclic group at R_i has $\left\lfloor \frac{2\pi}{\arccos\left(1 - \frac{1}{2i^2}\right)} \right\rfloor$ nodes. Thus, the grid of radius R will have

$$N = 1 + \sum_{i=1}^{\lfloor \frac{R}{d} \rfloor} \left\lfloor \frac{2\pi}{\arccos\left(1 - \frac{1}{2i^2}\right)} \right\rfloor.$$

A very good approximation is

$$N \approx 1 + 6 \sum_{i=1}^{\lfloor \frac{R}{d} \rfloor} i \approx 3 \frac{R^2}{d^2}.$$

Theorem 2. (Lower bound) *In networks with nodes d away from each other and cell radius R , an achievable maximum encoding number is*

- 1) $C_{\max}^{\text{cyclic}} = \Omega(\sqrt{N})$ for cyclic grid with $\text{rem}(R, 2d) \rightarrow 0$,
- 2) $C_{\max}^{\text{cyclic}} = \Omega(\sqrt[4]{N})$ for cyclic grid with $\text{rem}(R, 2d) \rightarrow d$ and
- 3) $C_{\max}^{\text{square}} = \Omega(\sqrt[4]{N})$ for the square grid of d ,

where $\text{rem}(x, y)$ is the remainder of the division x/y .

Proof:

For the cyclic grid when $\text{rem}(R, 2d) \rightarrow 0$: By focusing on the cyclic group with $\inf\{i : R_i > \frac{R}{2}\}$, note that each node is $\frac{R}{2} + \epsilon$ from the center and thus the desired connectivity properties are satisfied for all nodes on the cyclic group. In this case, we can calculate the number of nodes in the group

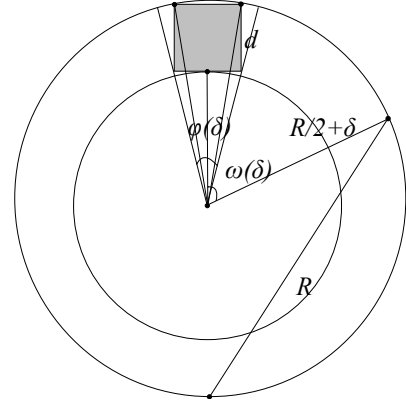


Fig. 3. Sketch for the proof of Theorem 2.

as

$$C_{\max}^{\text{cyclic}} = \left\lfloor \frac{2\pi}{\arccos\left(1 - \frac{2d^2}{R^2}\right)} \right\rfloor,$$

which for large N is bounded from below by some linear function of \sqrt{N} .

For the cyclic grid when $\text{rem}(R, 2d) \rightarrow d$: Now each node is $\frac{R}{2} + d - \epsilon$ away from the center and thus we need to select those nodes satisfying the property of valid combination. For this it is enough that we leave an empty angle ϕ such that if AOB is a diameter and AOC this angle, then $CB \leq R$. By solving this for the maximum number of points satisfying this property we get

$$C_{\max}^{\text{cyclic}} = \left\lfloor \frac{2\pi}{\arccos\left(\frac{R^2}{2\left(\frac{R}{2} + d\right)^2} - 1\right)} \right\rfloor,$$

which for large N is bounded from below by some linear function of $\sqrt[4]{N}$.

For the square grid: we construct a ring around the circle of $\frac{R}{2}$ radius. The width of the ring is δ wide enough to fit a whole square of dimensions $d \times d$. Such a square is bound to contain exactly one node of the grid. Using Figure 3, and the triangles relative to the small square, we calculate δ as

$$\delta = \frac{\sqrt{R^2 + d(5d + 4R)} - R}{2}.$$

Thus, we can show that $d \leq \delta \leq \frac{3d}{2}$. If we use the largest possible value that guarantees that the ring contains one node at each step, namely $\delta = 1.5d$, we can compute the angle that contains at least one node, which is of the order of d :

$$\phi(\delta) = \arcsin\left(\frac{\sqrt{2}d}{\sqrt{d^2 + R^2}}\right).$$

Then, we compute the angle which should be left empty in the valid combination such that any node outside this angle is reachable by the most distant node (the one at the bottom).

$$\omega(\delta) = \arcsin\left(\frac{R}{\sqrt{2}(R + 1.5d)}\right).$$

This angle is of the order of \sqrt{d} . Finally, an achievable combination is obtained if we alternate ϕ and ω until we fill the circle.

$$C_{\max}^{\text{square}} = \left\lfloor \frac{2\pi}{\phi(\delta) + \omega(\delta)} \right\rfloor,$$

which is bounded from below by a linear function of $\sqrt[4]{N}$. Note that the sparseness of the combination is due to $\omega(\delta)$ and a possible reasoning is that the bound is constructed to cover all the cases, thus also the case that the uncomfortable positioning of nodes matches the second case of the cyclic grid above. ■

In [19], relative results on convex polygons in constrained sets guarantee the existence of convex polygons of size $\Omega(\sqrt[4]{N})$ when the N thrown nodes are kept separated by some distance.

IV. STOCHASTIC ANALYSIS

Building on the insight acquired from the deterministic grid settings, we move to stochastic topologies. Assume that the locations of the nodes are determined by a Poisson point process with density λ . The connectivity properties of this model are well studied in the literature (see e.g. [20], [21]). As in the deterministic case, we assume that a relay is located at the origin. For a Poisson point process this assumption does not change the distribution of the other points. Our main result is an asymptotic upper bound in probability for the maximum encoding number.

Theorem 3. *In a random network determined by a Poisson point process with density λ , the maximum encoding number corresponding to combinations having the relay at the origin satisfies*

$$\lim_{\lambda \rightarrow \infty} \mathbb{P} \left(\frac{C_{\max}(\lambda)}{\lambda^{1/2+\epsilon}} < \delta \right) = 1,$$

for any $\epsilon > 0$ and $\delta > 0$.

Proof: First, consider a fixed λ and cover the area $[-R, R] \times [-R, R]$ by boxes of size $\lambda^{-1/2} \times \lambda^{-1/2}$. The number of nodes inside the disjoint boxes are denoted by $N_i(\lambda)$, $i = 1, \dots, n(\lambda)$, where $n(\lambda) = 4\lceil R\sqrt{\lambda} \rceil^2$. By Lemma 3, the nodes belonging to the valid combination form a convex polygon. Since the convex polygon is located inside the disk of radius R , its perimeter is covered by at most $4 \times \lceil 2\pi R\sqrt{\lambda} \rceil$ boxes. Thus

$$C_{\max}(\lambda) \leq 4 \times \lceil 2\pi R\sqrt{\lambda} \rceil \max_{i=1, \dots, n(\lambda)} N_i(\lambda) \quad \text{a.s.}$$

Next consider an increasing sequence of λ . Since for any given λ , the $N_i(\lambda)$ are identically and independently Poisson(1) distributed, there is a sequence $I_{n(\lambda)}$ such that

$$\lim_{\lambda \rightarrow \infty} \mathbb{P} \left(\max_{1 \leq i \leq n(\lambda)} N_i(\lambda) = I_{n(\lambda)} \text{ or } I_{n(\lambda)} + 1 \right) = 1, \quad (2)$$

where $I_n = O\left(\frac{\log n}{\log \log n}\right)$ (see [22], [23]). Then, for any $\delta > 0$ and $\epsilon > 0$, there exists $\tilde{\lambda} > 0$ such that

$$\frac{4 \times \lceil 2\pi R\sqrt{\lambda} \rceil}{\lambda^{1/2+\epsilon}} (I_{n(\lambda)} + 1) < \delta \quad \forall \lambda > \tilde{\lambda}.$$

Consequently,

$$\frac{C_{\max}(\lambda)}{\lambda^{1/2+\epsilon}} < \delta \frac{\max_i N_i(\lambda)}{I_{n(\lambda)} + 1} \quad \forall \lambda > \tilde{\lambda} \text{ a.s.}$$

and the proof is completed by applying equation (2). ■

V. RATE EFFICIENCY OF A NETWORK CODING COMBINATION

Given the characterisation of the maximum encoding number, we now consider to the performance loss induced by restricting to small combinations. We focus on the downlink of a valid combination of size C . Without loss of generality, assume that the rate vector $\mathbf{r} = \{r_i\}_{i=1,2,\dots,C}$ is ordered, i.e., $r_1 < r_2 < \dots < r_C$, and that the flow set is permuted accordingly so that over the link $(v_0, D(f_i))$ packets are transferred at a rate r_i .

The data rate is computed as the number of packets of size P transmitted in a virtual frame over the time needed for these transmissions. Since an encoded packet is always transmitted at the lowest rate decodable by all receivers and assuming max-min fair allocation³, we can deduce the maximum throughput rate with network coding as

$$r_{\text{NC}}(C) = \frac{CP}{\frac{P}{\min\{\mathbf{r}\}}} = Cr_1.$$

The rate without network coding would be

$$r_w(C) = \frac{CP}{\frac{P}{r_1} + \frac{P}{r_2} + \dots + \frac{P}{r_C}} = C \left(\sum_{i=1}^C \frac{1}{r_i} \right)^{-1} = r_h,$$

where r_h is the harmonic mean of \mathbf{r} . It is easy to see that if the criteria for valid combinations are fulfilled for a combination of size C then they are fulfilled for all its subsets. Choosing any $m \leq C$ and allowing for combinations of size m at most, gives an achievable rate of

$$\begin{aligned} r_m(C) &= \frac{CP}{\sum_{i=1}^{\lfloor \frac{C}{m} \rfloor} \frac{P}{r_{m(i-1)+1}} + 1_{\{\text{rem}(C,m)>0\}} \frac{P}{r_{C-\text{rem}(C,m)+1}}} \\ &= C \left(\sum_{i=1}^{\lfloor \frac{C}{m} \rfloor + 1_{\{\text{rem}(C,m)>0\}}} \frac{1}{r_{m(i-1)+1}} \right)^{-1}, \end{aligned}$$

Next, we derive the network coding gain for the maximum combination (C) and for the constrained group ($m \leq C$).

$$g(C) \doteq \frac{r_{\text{NC}}(C)}{r_w(C)} = C \frac{r_1}{r_h}.$$

Note that the gain is a linear function of C and depends on the particularities of the rate vector. Also, for the restricted

³This condition of fairness provides the best network coding opportunities and it is usually the balance point where network coding gain is computed in multiclass networks.

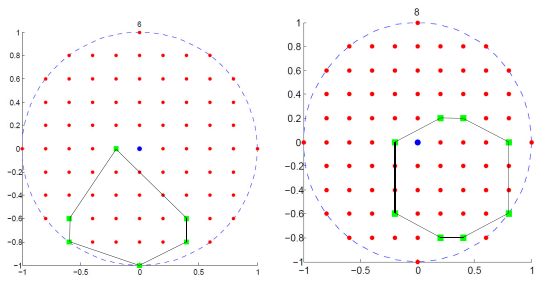


Fig. 4. Examples of valid combinations for $C = 6$ and $C = 8$ in a square grid with $N = 81$.

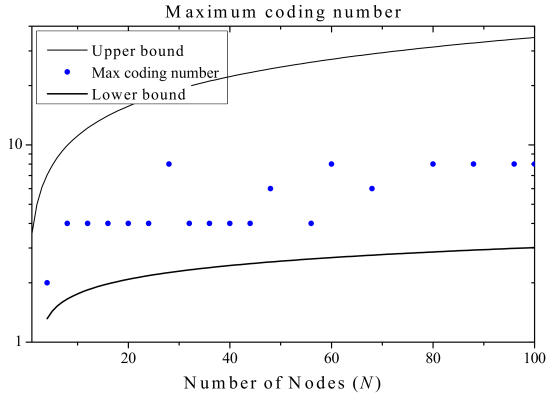


Fig. 5. Maximum encoding number of a square grid inside a disk.

case holds

$$g_m(C) \doteq \frac{r_m(C)}{r_w(C)} = \frac{C}{r_h} \left(\sum_{i=1}^{\lfloor \frac{C}{m} \rfloor + 1_{\{\text{rem}(C,m) > 0\}}} \frac{1}{r_m(i-1)+1} \right)^{-1}$$

$$\geq \frac{C}{r_h} \left(\frac{\lfloor \frac{C}{m} \rfloor}{r_1} \right)^{-1} \geq (m-1) \frac{r_1}{r_h},$$

where in the first inequality we have used that r_1 is the minimum rate, and in the second we have used $m-1 \leq \frac{C}{\lfloor \frac{C}{m} \rfloor} \leq m$.

If we choose equal rates, then we readily get $g(C) = C$ and $\max \{g_m(C)\} = m$ as the maximum gain for both. Finally, we can symbolise that $g(C) = \Theta(C)$ and moreover the difference $g(C) - g_m(C) = O(C - m)$. Therefore, the efficiency loss is at most of the order of $\sqrt{N} - m$.

VI. NUMERICAL RESULTS

In this section we present some numerical results that provide further evidence and insight for our work. For simulation purposes we consider a disk of radius $R = 1$ and a node v_0 serving as a relay situated at the center of the disk. Initially we consider a square grid of nodes over this disk and we investigate the maximum coding number, i.e., a set of nodes that satisfies the constraints of section II. Then, the scenario of uniformly random thrown nodes is considered.

A. Experiments with square grids

Figure 4 showcases examples of combination sizes $C = 6$ and $C = 8$ ($C_{\max} = 8$ in this case). During these experiments

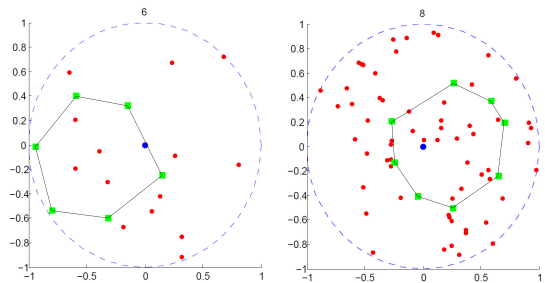


Fig. 6. Examples of valid combinations for $C = 6$ and $C = 8$ in a network with $N = 21$ and $N = 70$ nodes respectively, randomly located.

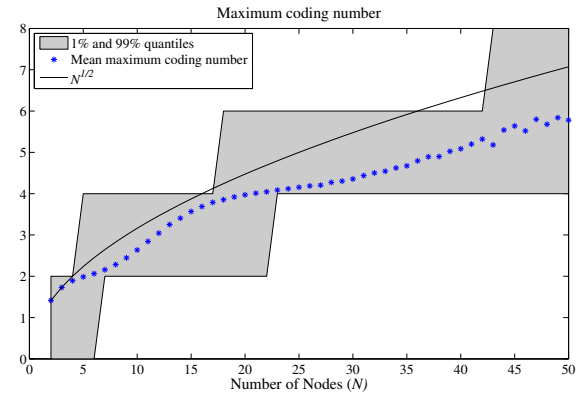


Fig. 7. Maximum encoding number for randomly positioned nodes in a disk.

we identified an interesting property. We noticed that the maximum encoding number depends only on the number of nodes inside the disk and not on the actual d used. In particular, for all $\{d : N \text{ is fixed}\}$, C_{\max} is constant. In Figure 5 we present the (deterministic) values for the maximum encoding numbers in case of the square grid. The actual values seem to be closer to the lower bound than to the upper, leading to an order closer to $\sqrt[4]{N}$ (notice the logarithmic scale of the figure). As a guideline, for grids where nodes have up to 100 connections, looking for combinations larger than 8 seems futile. Nevertheless, the oscillating effect due to the interplay between the radius and the number of nodes is evident.

B. Experiments with randomly positioned nodes

Next, we throw N uniform random nodes inside the disk of radius $R = 1$. Examples of valid encoding combinations are showcased in Figure 6. It is noted from these examples that large combinations tend to appear in a δ -ring form where the inner side of the ring is a disk of radius $\frac{R}{2}$ and the outer side is a disk of radius $\frac{R}{2} + \delta$.

The simulated values of the maximum encoding number are illustrated in Figure 7. The observed maximum encoding number was within the gray area in 98% of the experiments. The mean behaviour of the maximum encoding number is obtained by averaging over 1000 random samples. The \sqrt{N} behaviour is also depicted in this picture. Figure 8 shows the probability of existence of at least one coding combination of size C in a network of N uniformly thrown nodes. For example, the maximum component size for $20 \leq N \leq 50$ is

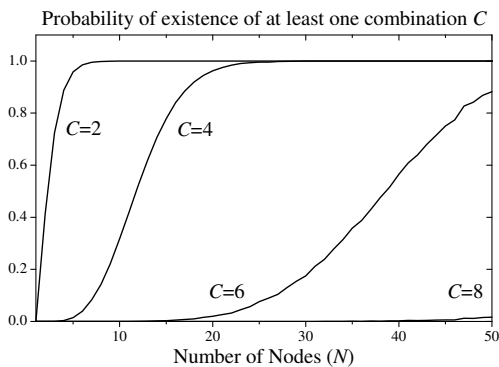


Fig. 8. Probability of existence of at least one coding combination of size C in a network of N uniformly thrown nodes.

either 4 or 6 in the majority of cases.

As a summary, these simulation results indicate that the probabilistic asymptotic bound derived in Theorem 3 predicts quite well the qualitative performance also in random networks of moderate size.

VII. CONCLUDING REMARKS

By considering the Boolean connectivity model, we showed that there are certain geometric constraints bounding the maximum number of packets that can be encoded together in local network coding with opportunistic listening. Particularly, due to the convexity of any valid combination, the sizes of combinations are at most of order of \sqrt{N} , N denoting the number of neighbours, for all studied network topologies. The convexity property is strongly linked to the specific connectivity model and it does not necessarily hold for other communication models. However, the underlying geometric constraints are typically quite similar to those studied in this paper, though the additional details of protocol and interference modelling may allow also non-convex valid combinations.

The derived qualitative and asymptotic behaviour of the maximum coding number is already a strong evidence that the focus should be on developing efficient algorithms that opportunistically exploit local network coding over a wide span of topologies using small XOR combinations rather than attempting to solve complex combinatorial problems in order to find the best combinations available. In order to evaluate the actual (small) quantitative performance loss due to limited combination sizes, one should extend the analysis, for example, to average encoding components with more realistic traffic models. Such study is beyond the scope of this paper and is left for further studies.

ACKNOWLEDGMENTS

The work of P. Mannersalo was partially supported by the Commune project funded by Celtic-Plus programme and the work of G.S. Paschos was supported by the WiNC project funded by National and Community Funds (European Social Fund).

REFERENCES

- [1] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," in *Proc. of ACM SIGCOMM 2006*, Pisa, Italy, Sep. 2006, pp. 243–254.
- [2] —, "XORs in the air: Practical wireless network coding," *IEEE/ACM Trans. Netw.*, vol. 16, pp. 497–510, Jun. 2008.
- [3] P. Chaporkar and A. Proutiere, "Adaptive Network Coding and Scheduling for Maximizing Throughput in Wireless Networks," in *Proceedings of MobiCom 2007*, Montréal, Québec, Canada, Sep. 2007, pp. 135–146.
- [4] E. Rozner, A. Iyer, Y. Mehta, L. Qiu, and M. Jafry, "ER: Efficient Retransmission Scheme for Wireless LANs," in *International Conference on emerging Networking EXperiments and Technologies (ACM CoNEXT)*, New York, NY, USA, Dec. 2007, pp. 1–12.
- [5] S. Rayanchu, S. Sen, J. Wu, S. Banerjee, and S. Sengupta, "Loss-Aware Network Coding for Unicast Wireless Sessions: Design, Implementation, and Performance Evaluation," in *ACM SIGMETRICS Perform. Eval. Rev.*, Annapolis, MD, USA, Jun. 2008, pp. 85–96.
- [6] M. A. R. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proceedings of IEEE INFOCOM 2008*, Phoenix, AZ, USA, Apr. 2008, pp. 1–4.
- [7] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theor.*, vol. 57, pp. 1479–1494, 2011.
- [8] G. Paschos, L. Georgiadis, and L. Tassioulas, "Optimal scheduling of pairwise XORs under statistical overhearing and feedback," in *RAWNET workshop: Workshop on Resource Allocation and Cooperation in Wireless Networks*, Princeton, NJ, USA, Apr. 2011.
- [9] N. Jones, B. Shrader, and E. Modiano, "Optimal routing and scheduling for a simple network coding scheme," in *Proceedings IEEE INFOCOM 2012*, March 2012, pp. 352–360.
- [10] T. Wang and G. B. Giannakis, "Capacity scaling of wireless networks with complex field network coding," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, ROC, Apr. 2009, pp. 2425–2428.
- [11] L. Junning, D. Goeckel, and D. Towsley, "Bounds on the gain of network coding and broadcasting in wireless networks," in *Proceedings IEEE INFOCOM 2007*, Anchorage, AK, USA, May 2007, pp. 724–732.
- [12] E. Ahmed, A. Eryilmaz, M. Medard, and A. Ozdaglar, "On the scaling law of network coding gains in wireless networks," in *Proceedings of IEEE MILCOM 2007*, Orlando, FL, USA, Oct. 2007, pp. 1–7.
- [13] A. Keshavarz-Haddad and R. Riedi, "Bounds on the benefit of network coding: Throughput and energy saving in wireless networks," in *Proceedings of IEEE INFOCOM 2008*, Phoenix, AZ, USA, Apr. 2008, pp. 376–384.
- [14] A. Goel and S. Khanna, "On the network coding advantage for wireless multicast in euclidean space," in *Proc. of International Conference on Information Processing in Sensor Networks, IPSN '08*, St. Louis, MO, USA, Apr. 2008, pp. 64–69.
- [15] Y. Kim and G. D. Veciana, "Is rate adaptation beneficial for inter-session network coding?" *IEEE J. Sel. Areas Commun.*, vol. 27(5), pp. 635–646, Jun. 2009.
- [16] J. Le, J. Lui, and C.-M. Chiu, "How many packets can we encode? - an analysis of practical wireless network coding," in *Proceedings IEEE INFOCOM 2008*, Phoenix, AZ, USA, Apr. 2008, pp. 371–375.
- [17] —, "On the performance bounds of practical wireless network coding," *IEEE Trans. Mobile Comput.*, vol. 9, pp. 1134–1146, Aug. 2010.
- [18] M. Franceschetti and R. Meester, *Random Networks for Communication*. New York, NY, USA: Cambridge University Press, 2007.
- [19] N. Alon, M. Katchalski, and W. R. Pulleyblank, "The maximum size of a convex polygon in a restricted set of points in the plane," *Discrete Comput. Geom.*, vol. 4(3), pp. 245–251, Mar. 1989.
- [20] R. Meester and R. Roy, *Continuum Percolation*. New York, NY, USA: Cambridge University Press, 1996.
- [21] O. Dousse, P. Thiran, and M. Hassler, "Connectivity in ad hoc and hybrid networks," in *Proceedings IEEE INFOCOM 2002*, New York, NY, USA, Jun. 2002, pp. 1079–1088.
- [22] C. W. Anderson, "Extreme value theory for a class of discrete distributions with applications to some stochastic processes," *Journal of Applied Probability*, vol. 7, pp. 99–113, Apr. 1970.
- [23] A. Kimber, "A note on Poisson maxima," *Probability Theory and Related Fields*, vol. 63(4), pp. 551–552, 1983.