

# Entropy Based Sensing Schemes for Energy Efficiency in Massive MTC

Sergi Liesegang, *Student Member, IEEE*, Antonio Pascual-Iserte, *Senior Member, IEEE*,  
and Olga Muñoz, *Member, IEEE*

**Abstract**—Machine-type-communications (MTC) are crucial in the evolution of mobile communication systems. Within this context, we distinguish the so-called massive MTC (mMTC), where a large number of devices coexist in the same geographical area. In the case of sensors, a high correlation in the collected information is expected. In this letter, we evaluate the impact of correlation on the entropy of a set of quantized Gaussian sources. This model allows us to express the sensed data with the data correlation matrix. Given the nature of mMTC, these matrices may be well approximated as rank deficient. Accordingly, we exploit this singularity to design a technique for switching off several sensors that maximizes the entropy under power-related constraints. The discrete optimization problem is transformed into a convex formulation that can be solved numerically.

**Index Terms**—Machine-type-communications, data entropy, pseudo-determinant, Cholesky factorization, convex optimization.

## I. INTRODUCTION

MACHINE-type-communications (MTC) have drawn a lot of attention during the past decade. They are key for the development of a plethora of applications in future mobile generations [1]. MTC systems involve transmissions between autonomous devices without human interaction. This facilitates the rapid deployment of large networks, which leads to an exponential growth in the number of connected terminals [2]. As a result, MTC are contributing to the advance of new fields of interest such as the Internet-of-Things [3].

In this letter, a scenario with a large number of sensors is analyzed. These systems are commonly denoted as *massive* MTC (mMTC) [4] and, given the high spatial density of these devices, the sensed data can be highly correlated [5]. Thereby, a natural question that arises is how to manipulate this information to reduce the overall payload. In fact, this data redundancy entails the waste of a significant amount of power. Hence, working with only the relevant information would help increase the energy efficiency and extend the battery lifetime of the sensors, which are critical issues in MTC deployments.

To tackle this problem, we formulate strategies to decide that some devices will not sense while still ensuring that a significant amount of the information is collected [6]–[8]. The procedure is based on the entropy of quantized Gaussian sources, which allows us to represent the collected information through the (frequently ill-conditioned) data correlation matrix [9], [10]. Accordingly, since this matrix captures the correlation between the sensors data, we can decide which sensors

This work was supported in part by the Spanish Ministry of Science, Innovation and Universities through the Project 5G&B RUNNER-UPC (AEI/FEDER, UE) under Grant TEC2016-77148-C2-1-R, in part by the Fellowship FPI under Grant BES-2017-079994, and in part by the Generalitat de Catalunya–Agència de Gestió d’Ajuts Universitaris i de Recerca under Grant 2017 SGR 578.

The authors are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona 08034, Spain (e-mail: sergi.liesegang@upc.edu, antonio.pascual@upc.edu, olga.munoz@upc.edu).

will be active and which will remain silent, thus reducing the energy consumption. In particular, we will select as active devices those that maximize the information collected (i.e. entropy) under power-related constraints and silence the rest. Since devices transmit only after sensing, our approach can further reduce the power consumption. To evaluate the performance of these schemes, we study the loss in the data entropy after the device selection. To the best of our knowledge, no studies have been reported in this direction so far.

Given the discrete nature of the optimization problem (i.e. sensor selection), convex relaxation and approximations of the entropy are used to obtain sub-optimal solutions. We show the new formulated problem to be convex and, thus, numerically solvable with standard optimization methods. The performance of the proposed strategy is evaluated in an illustrative practical scenario in which the elements of the correlation matrix are generated through a simple spatial model based on the distance between sensors. Under certain conditions, this matrix can be shown to be well approximated by a singular matrix.

This letter is structured as follows. In Section II, the system model, the data entropy and the correlation are described. In Section III, the sensing scheme is formulated and the proposed solution is presented. Section IV is devoted to the numerical simulations and conclusions are shown in Section V.

## II. SYSTEM MODEL

Throughout this work, we consider a setup with  $M$  sensors deployed around one data collector node (CN), as depicted in Fig. 1. In this scenario, we focus on how much information the sensed data contains, and leave the communication aspects for future studies. Then, it is desirable to establish a metric that quantifies this information. It is natural to use the multivariate entropy as a proper indicator, since it determines the minimum number of bits needed to represent a discrete source [11].

To describe the sensed data, we define a random vector  $\mathbf{X} = [X_1, \dots, X_M] \in \mathbb{R}^M$  with distribution  $f_{\mathbf{X}}(\mathbf{x})$ , where each element represents the information collected by each sensor. Since this information is continuous, an infinite number of bits are needed to represent  $\mathbf{X}$ . Thus, it is necessary to quantize the source to work with finite precision. We denote the quantized version as  $\mathbf{X}^\Delta$  with associated mass probability distribution  $p_{\mathbf{X}^\Delta}(\mathbf{x}^\Delta)$ . The multivariate entropy (in bits) is

$$H(\mathbf{X}^\Delta) = - \sum_{\mathbf{x}^\Delta} p_{\mathbf{X}^\Delta}(\mathbf{x}^\Delta) \log_2 p_{\mathbf{X}^\Delta}(\mathbf{x}^\Delta), \quad (1)$$

and, as shown in the Appendix, it can be lower bounded by

$$\tilde{H}(\mathbf{X}^\Delta) \triangleq h(\mathbf{X}) - M \log_2 \Delta \leq H(\mathbf{X}^\Delta), \quad (2)$$

where we have considered a uniform quantization step  $\Delta$ . The term  $h(\mathbf{X})$  is the differential entropy of the continuous source:

$$h(\mathbf{X}) = - \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \log_2 f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (3)$$

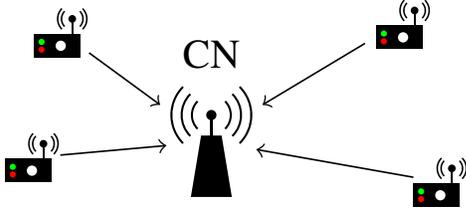


Fig. 1. System setup for  $M = 4$ .

Note that the previous lower bound  $\tilde{H}(\mathbf{X}^\Delta)$  becomes tight for  $\Delta \rightarrow 0$  [11], as shown in Section IV. This is why it will be used as an approximation for the actual entropy  $H(\mathbf{X}^\Delta)$ .

Considering the Gaussian distribution with zero mean and (full-rank) correlation matrix  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , i.e.  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ , we have [11]:

$$h(\mathbf{X}) = \frac{1}{2} (M \log_2 2\pi e + \log_2 \det(\mathbf{C})). \quad (4)$$

Note that the Gaussian distribution is taken as reference as it requires the largest number of bits for accurate representation once quantized (hence, it can be understood as a challenging case) [11]. Furthermore, the Gaussian assumption is typical in many sensing scenarios [9], [10]. Also, it allows us to measure the amount of collected information as a function of matrix  $\mathbf{C}$ , which models the correlation between sensors data.

The elements of  $\mathbf{C}$  can be constructed as follows [10]:

$$[\mathbf{C}]_{i,j} = \sigma_i \sigma_j K(d_{i,j}), \quad 1 \leq i \leq M, \quad 1 \leq j \leq M, \quad (5)$$

where  $\sigma_i$  is the standard deviation of the information signal collected by sensor  $i$ . The term  $K(d_{i,j})$  is the correlation factor between the data of sensors  $i, j$  with  $|K(d_{i,j})| \leq 1$ , and that usually depends on the distance between them  $d_{i,j}$ . Different models can be adopted for the function  $K(d_{i,j})$  [12]. Just as an illustrative example, a well-known and accepted model for sensor networks is a quadratic-distance exponential [13]:

$$K(d_{i,j}) = \exp(-(d_{i,j}/\theta)^2), \quad (6)$$

where  $\theta > 0$  controls the range of correlation. The extreme values in (6) are 0 and 1 for  $d_{i,j} = \infty$  and  $d_{i,j} = 0$ , respectively. Hence, as devices are located at different positions, the spatial dependency of the correlation is incorporated in the model.

Note that the strategy developed in this letter is based on a generic matrix  $\mathbf{C}$  and is also valid if other underlying models for matrix  $\mathbf{C}$  are adopted (see examples [5], [9], [10], [13]).

Under some assumptions, the correlation matrix  $\mathbf{C}$  can be well approximated by a singular matrix for  $M \rightarrow \infty$  [9]. Intuitively, this can be understood by the fact that its elements depend only on the distance and, given the high spatial density, their values can be very similar. From now on, and to ease notation,  $\mathbf{C}$  will represent the singular (rank deficient) matrix approximating the original correlation matrix. The differential entropy of the resulting degenerate Gaussian distribution is [9]

$$h(\mathbf{X}) = \frac{1}{2} (r(\mathbf{C}) \log_2 2\pi e + \log_2 \text{Det}(\mathbf{C})), \quad (7)$$

where  $r(\mathbf{C})$  and  $\text{Det}(\mathbf{C})$  are the rank and pseudo-determinant of  $\mathbf{C}$ , respectively [14]. The latter is computed as the product of the non-zero eigenvalues of  $\mathbf{C}$ . Finally, the entropy of the quantized vector  $\mathbf{X}^\Delta$  yields [9]

$$H(\mathbf{X}^\Delta) \geq h(\mathbf{X}) - r(\mathbf{C}) \log_2 \Delta = \tilde{H}(\mathbf{X}^\Delta). \quad (8)$$

### III. ENTROPY BASED SENSOR SELECTION

#### A. Problem Formulation

The purpose of this letter is to present a strategy that minimizes the loss of information when silencing some sensors. For this task, we take advantage of the high correlation in the sensed data, which yields an ill-conditioned correlation matrix  $\mathbf{C}$ , and select as active sensors those with the highest impact on  $\mathbf{C}$ . Note that as matrix  $\mathbf{C}$  is singular, the number of sensors to keep active does not increase substantially with  $M$  [9].

We are interested in designing a device selection matrix  $\mathbf{B} = \text{diag}(\mathbf{b}) = \text{diag}([b_1, \dots, b_M])$ , with binary elements  $b_i \in \{0, 1\}$ , that reduces power consumption while preserving the sensed information to be transmitted. It represents an on/off activity strategy applied to the sensors. The reason behind this hard allocation scheme is the simplicity of sensors, which may be unable to handle various sensing states (they can be either active or asleep, but not in intermediate phases).

Let us denote  $\mathbf{X}_p^\Delta$  as the quantized version of the selected data  $\mathbf{X}_p = \mathbf{B}\mathbf{X}$ . Then, the goal is to find a selection matrix  $\mathbf{B}$  that maximizes  $H(\mathbf{X}_p^\Delta)$  under a norm constraint. To that end, we consider the following optimization problem:

$$\mathbf{b}^* = \arg \max_{\mathbf{b} \in \{0,1\}^M} H(\mathbf{X}_p^\Delta) \quad \text{s.t.} \quad \|f(\mathbf{b})\| \leq \delta, \quad (9)$$

where  $\|f(\mathbf{b})\|$  represents a power-related constraint. The constant  $\delta > 0$  determines the bound applied to the norm.

Furthermore, since the correlation matrix of  $\mathbf{X}_p^\Delta$ , i.e.  $\mathbf{BCB}$ , may also be rank deficient, the corresponding approximate entropy of this new quantity is

$$\tilde{H}(\mathbf{X}_p^\Delta) \triangleq h(\mathbf{X}_p) - r(\mathbf{BCB}) \log_2 \Delta \leq H(\mathbf{X}_p^\Delta), \quad (10)$$

with differential entropy

$$h(\mathbf{X}_p) = \frac{1}{2} (r(\mathbf{BCB}) \log_2 2\pi e + \log_2 \text{Det}(\mathbf{BCB})). \quad (11)$$

Given that the lower bound  $\tilde{H}(\mathbf{X}_p^\Delta)$  becomes tight for small  $\Delta$ , we can use it as the objective function in (9). The exact error of this approximation will be illustrated in Section IV. When maximizing the lower bound, the problem in (9) becomes

$$\mathbf{b}^* = \arg \max_{\mathbf{b} \in \{0,1\}^M} \tilde{H}(\mathbf{X}_p^\Delta) \quad \text{s.t.} \quad \|f(\mathbf{b})\| \leq \delta. \quad (12)$$

Regarding the constraint function  $\|f(\mathbf{b})\|$ , we can use different approaches. For instance, some reasonable options are:

- (i) Number of active sensors:  $\|\mathbf{b}\|_0$ ,
- (ii) Sum of powers:  $\|\mathbf{b}^2 \circ \boldsymbol{\gamma}\|_1$ ,
- (iii) Maximum power:  $\|\mathbf{b}^2 \circ \boldsymbol{\gamma}\|_\infty$ ,

where  $\circ$  is the element-wise or Hadamard product and  $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_M]$  is the set of individual sensor power consumptions. When applied to (12), (i) determines the sparsity of  $\mathbf{b}$  and, although it does not consider the power  $\boldsymbol{\gamma}$ , it can reduce the consumption since it limits the number of active sensing devices. On the other hand, since (ii) and (iii) consider power consumption, both can improve the system energy efficiency better than (i). However, since  $\boldsymbol{\gamma}$  is usually fixed, (iii) has no special interest as the solution would be a trivial thresholding irrespective of  $\tilde{H}(\mathbf{X}_p^\Delta)$ , i.e.  $b_i = 1$  for all sensors with power  $\gamma_i$  below  $\delta$ . This is why we will focus on (i) and (ii).

Overall, given the binary nature of  $\mathbf{b}$ , we formulate a lossy information approach that prioritizes the relevant components of the sensed data (i.e. those with more information). This way, we end up with a system that delivers the most relevant information with the benefit of a reduced consumption, i.e. the battery lifetime can be extended with similar performance.

Unfortunately, the solution to problem (12) is combinatorial and might be unfeasible to solve for large  $M$ . This is the reason why, in the upcoming section we propose some simplifications to reduce computational cost.

### B. Proposed Solution

To reduce the problem to a feasible complexity, in this letter we take the following steps: relaxation of the constraint set, approximation of the objective function, and rounding of the obtained (sub-optimal) solution. This way, we end up with a convex optimization problem, easy to solve with standard numerical methods, while still providing good performance.

We start by substituting the discrete constraint on  $\mathbf{b}$ , i.e.  $\mathbf{b} \in \{0, 1\}^M$ , with a convex relaxation. In particular, now  $\mathbf{b}$  can take continuous values between 0 and 1, i.e.  $\mathbf{b} \in [0, 1]^M$ . Thereby, we can express the new continuous problem as

$$\mathbf{b}^* = \arg \max_{\mathbf{b} \in [0, 1]^M} \tilde{H}(\mathbf{X}_p^\Delta) \quad \text{s.t.} \quad \|f(\mathbf{b})\| \leq \delta. \quad (13)$$

Note that the  $\log_2 \text{Det}$  function in  $\tilde{H}(\mathbf{X}_p^\Delta)$  is not concave in  $\mathbf{B}$  as the  $\text{Det}$  function is discontinuous [14]. This is why, in order to deal with proper determinants, we make use of the Cholesky decomposition of the correlation matrix [15]. Since  $\mathbf{C}$  is symmetric and positive semi-definite ( $\mathbf{C} \geq \mathbf{0}$ ), there exists a lower triangular matrix  $\mathbf{L} \in \mathbb{R}^{M \times M}$  such that  $\mathbf{C} = \mathbf{L}\mathbf{L}^T$ . Note that for rank-deficient  $\mathbf{C}$ , the matrix  $\mathbf{L}$  will have  $M - r(\mathbf{C})$  columns equal to zero, i.e.

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & 0 & \dots & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & & \vdots \\ l_{M1} & l_{M2} & \dots & l_{Mr(\mathbf{C})} & 0 & \dots & 0 \end{bmatrix} = [\mathbf{L}_r \mid \mathbf{0}], \quad (14)$$

where  $\mathbf{L}_r \in \mathbb{R}^{M \times r(\mathbf{C})}$  is the reduced version of  $\mathbf{L}$ . With that, the correlation matrix can be expressed as  $\mathbf{C} = \mathbf{L}_r \mathbf{L}_r^T$ . In addition, given that the pseudo-determinant is also translation invariant [14], we have that

$$\text{Det } \mathbf{C} = \text{Det}(\mathbf{L}_r \mathbf{L}_r^T) = \text{Det}(\mathbf{L}_r^T \mathbf{L}_r) = \det(\mathbf{L}_r^T \mathbf{L}_r). \quad (15)$$

Similarly, we can write  $\text{Det}(\mathbf{B}\mathbf{C}\mathbf{B}) = \text{Det}(\mathbf{L}_r^T \mathbf{B}\mathbf{B}\mathbf{L}_r)$ , which becomes a proper determinant if the selection matrix is positive definite, i.e.  $\mathbf{B} > \mathbf{0}$ . This way, the  $\log_2 \text{Det}$  in  $\tilde{H}(\mathbf{X}_p^\Delta)$  yields the (continuous)  $\log_2 \text{det}$ . However, the  $\log_2 \text{det}$  function is still not concave in  $\mathbf{B}$ . In what follows, we use instead matrix  $\mathbf{P}$ , defined as  $\mathbf{P} = \text{diag}(\mathbf{p}) = \mathbf{B}\mathbf{B} = \mathbf{B}^2$ , and reformulate the problem in (13) as follows:

$$\mathbf{p}^* = \arg \max_{\mathbf{p} \in [0, 1]^M} \tilde{H}(\mathbf{X}_p^\Delta) \quad \text{s.t.} \quad \|f(\mathbf{p})\| \leq \delta, \quad (16)$$

where now, defining  $\mathbf{\Gamma} = \text{diag}(\boldsymbol{\gamma})$  and neglecting (iii), the constraint functions in  $\mathbf{p}$  are given by:

- (i) Number of active sensors:  $\|\mathbf{p}\|_0$ ,
- (ii) Sum of powers:  $\|\mathbf{p} \circ \boldsymbol{\gamma}\|_1 = \text{tr}(\mathbf{P}\mathbf{\Gamma})$ .

As a second step, we substitute the rank in the objective function  $\tilde{H}(\mathbf{X}_p^\Delta)$  in (16) by its nuclear norm [16], producing a lower bound of the objective function that is to be maximized and that can be shown to be concave in  $\mathbf{P}$  for  $\mathbf{P} > \mathbf{0}$ . Refer to the Appendix for more details on the derivation of this lower bound, denoted by  $\tilde{H}_{lb}(\mathbf{X}_p^\Delta)$ . In addition, we also substitute the  $l_0$  quasi-norm in (i) by its closest convex approximation, namely the  $l_1$  norm  $\|\mathbf{p}\|_1 = \text{tr}(\mathbf{P})$  [17].

Note that the condition  $\mathbf{P} > \mathbf{0}$  can be ensured by imposing  $\mathbf{P} - \nu \mathbf{I} \geq \mathbf{0}$  for a sufficiently small  $\nu > 0$ , which is equivalent to  $p_i \geq \nu \forall i$ . In fact, since we eventually round the solution,  $p_i = \nu$  or  $p_i = 0$  would lead to the same binary value (i.e. the same final solution). This practice is widely used in the area of sparse sensing (see [18] and references therein).

As shown in the Appendix, thanks to the previous steps, we end up with the following convex problem (maximization of a concave function subject to convex constraints), where the optimization variable is the diagonal matrix  $\mathbf{P}$ :

$$\mathbf{P}^* = \arg \max_{\mathbf{P} = \text{diag}(\mathbf{p}), \mathbf{p} \in [\nu, 1]^M} \tilde{H}_{lb}(\mathbf{X}_p^\Delta) \quad \text{s.t.} \quad \|f(\mathbf{P})\| \leq \delta. \quad (17)$$

Note that, due to the determinant, an analytic closed-form solution is hard to find, even for small  $M$ . This is the reason why, given the convexity of the problem, standard optimization methods are used to find the global solution numerically [19]. More specifically, in order to solve (17), we use the successive approximation method with the SDPT3 solver of the CVX software package, and we set the precision to 'high' [20].

Finally, once the solution is found in the continuous domain, we need to return to the discrete regime. To do so, we threshold (round) the solution such that the constraints are satisfied, i.e. activate sensors with largest  $p_i^*$  until (i) or (ii) are fulfilled.

## IV. NUMERICAL RESULTS

In this section, several simulations are presented to illustrate the performance of the sensing scheme. More specifically, the relative loss  $\varepsilon$  in information after the device selection is evaluated for different sweeps where the parameters  $\delta$  and  $M$  are changed. This metric corresponds to the difference between the entropies of the quantized data defined in (8) and (10), i.e.

$$\varepsilon = |\tilde{H}(\mathbf{X}^\Delta) - \tilde{H}(\mathbf{X}_p^\Delta)| / |\tilde{H}(\mathbf{X}^\Delta)|. \quad (18)$$

Regarding the scenario, we consider sensors to be normally distributed in a small region around the CN. This small area reduces the computational complexity of the numerical methods, as we can use smaller values of  $M$  while preserving the high spatial density of sensors. As an example of application, we consider temperature as the phenomenon to be measured, and we make use of the quadratic-distance exponential model in (6) with  $\sigma_i = \sigma = 50^\circ$  and  $\theta = 3.08$  for the correlation matrix  $\mathbf{C}$ . In addition, we also consider sensors to have different levels of power consumption, i.e.  $\gamma_i \in [0.01, 0.2]$  W [21].

The sweeps over  $\delta$  and  $M$  are depicted in Fig. 2 and Fig. 3, respectively. Note that the values of  $\delta$  represent a percentage of the different constraint functions applied directly to  $\mathbf{\Gamma}$ , i.e. (i)  $\delta = \alpha \text{tr}(\mathbf{I}) = \alpha M$  and (ii)  $\delta = \alpha \text{tr}(\mathbf{\Gamma})$ , with  $\alpha \in [0, 1]$ . Then,  $\alpha = 0.5$  is used in Fig. 3 and  $M = 50$  is set in Fig. 2. In addition, we set  $\Delta$  according to the Appendix with  $\nu = 10^{-3}$ .

In order to actually represent the consumed power after the selection, in Fig. 2 we plot the relative entropy loss  $\varepsilon$  with respect to (w.r.t.) the relative consumed power  $\mu = \text{tr}(\mathbf{P}\mathbf{\Gamma})/\text{tr}(\mathbf{\Gamma})$  instead of  $\alpha$ . In fact, note that  $\alpha$  is equal to  $\mu$  in constraint (ii) while it is indirectly related to this ratio in the case of (i).

Also, since in real scenarios the exact distances might not be known, but estimated, in Fig. 2 and Fig. 3 we show the entropy loss when the distances are perfectly known and also for an additive Gaussian estimation error with zero-mean and standard deviation of 10% w.r.t. to the actual distances.

As observed in Fig. 2, higher values of  $\mu$  ( $\alpha$ ) yield smaller entropy losses. This is not surprising as we allow more sensors to be active. For instance, for  $\mu = 40\%$  we have a relative loss of 7.30% and 4.25% for constraints (i) and (ii). This shows that it is possible to reduce significantly the consumed power while preserving most of the relevant information.

In Fig. 3 we can see that the entropy loss diminishes with the total number of sensors  $M$  because the dimension of the null-space of  $\mathbf{C}$  increases with  $M$ . For an increasing  $M$  and a constant  $\alpha$ , a significant number of sensors can be silent while reducing the entropy loss. A large error  $\varepsilon$  is then obtained for small  $M$ , when the correlation matrix is full-rank (all sensors provide relevant information). For instance, for  $M = 10$  the losses are around 45% and 25% for constraints (i) and (ii), whereas for  $M = 40$ , the entropy loss is only 4.5% and 3.3%.

Note that a better performance is attained for constraint (ii) because the optimal sensors (i.e. those with the highest impact on  $\mathbf{C}$  and smaller consumption) can also be distinguished in terms of consumed energy. Hence, the difference in power consumption can play an important role in the optimization as more relevant information is taken into account in the problem.

To compare the performance of our system, the relative loss of a random selection is also illustrated in Fig. 2 and Fig. 3. It represents the case where sensors are randomly chosen until constraint (ii) is fulfilled. As expected, a higher loss is obtained with this naive approach, specially for small  $\alpha$  and  $M$ , where the optimal solution (i.e. device selection) is less flexible.

In contrast to the approaches described in [7], [8], where the sensor selection is based on the estimation of the parameter to be measured, we rely on the correlation between measurements (represented with the matrix  $\mathbf{C}$ ). Thus, the selection is based on this statistical information. In addition, since the optimization procedure is done at the CN and requires low signalling (i.e. binary messages to report the on/off decisions), there is no significant additional power consumption of the sensors.

To justify the approximation of the original entropy  $H(\mathbf{X}^\Delta)$  by its lower bound  $\tilde{H}(\mathbf{X}^\Delta)$ , both magnitudes, jointly with the relative error, are illustrated w.r.t.  $\Delta/\sigma$  in Fig. 4 with  $M = 1$  ( $\mathbf{X}^\Delta = X^\Delta$ ). The values of  $H(X^\Delta)$  are obtained numerically through simulations. As shown in the figure, the approximation is very tight, specially for low values of  $\Delta/\sigma$ .

Finally, some practical issues should be pointed out. First, as we observe in Fig. 2 and Fig. 3, our method is robust to the uncertainty in the distance estimation since the resulting entropy loss does not increase substantially. Second, as reducing the number of active devices reduces the number of transmissions (sensors will transmit if they collect information), our strategy can reduce the number of collisions as well.

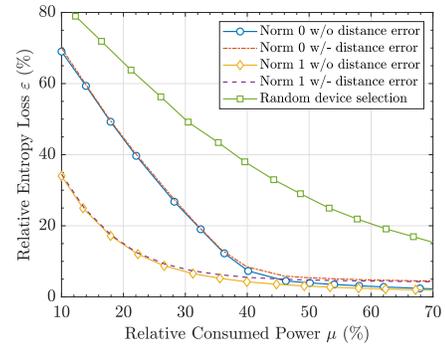


Fig. 2. Relative entropy loss  $\varepsilon$  versus consumed power  $\mu$  with  $M = 50$ .

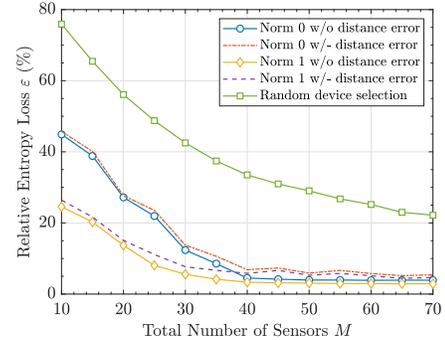


Fig. 3. Relative entropy loss  $\varepsilon$  versus  $M$  with  $\alpha = 0.5$ .

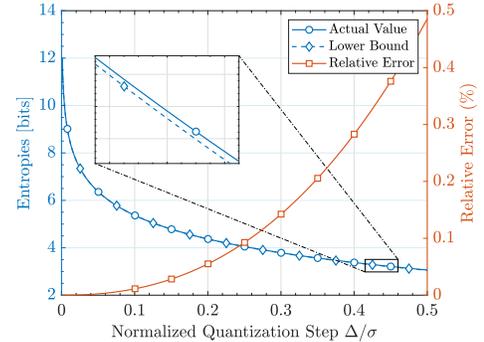


Fig. 4. Actual entropy and lower bound (left) and relative error (right) versus normalized quantization step  $\Delta/\sigma$  with  $M = 1$ .

## V. CONCLUSIONS

In this letter, we have designed a selection strategy in dense MTC networks that minimizes the loss of information to be transmitted (i.e. sensed data entropy) subject to power-related constraints, i.e. minimize the loss in information when some of the sensors do not collect or report the sensed information. A degenerate Gaussian distribution is used as a model to measure the correlated sensed information. Given the discrete nature of the resulting optimization problem, some steps are taken to find a sub-optimal solution. Simulation results prove that most of the information (i.e. entropy) can be retrieved with a smaller number of active devices. This effect is more pronounced in mMTC, where a higher correlation is experienced.

## APPENDIX

### Entropy of Quantized Source

According to [22], the entropy of a quantized source  $\mathbf{X}^\Delta$  already defined in (1) can also be written in the following way:

$$H(\mathbf{X}^\Delta) = - \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \log_2 \bar{f}(\mathbf{x}) d\mathbf{x} - M \log_2 \Delta, \quad (19)$$

where  $\bar{f}(\mathbf{x})$  is the average value of  $f_{\mathbf{X}}(\mathbf{x})$  over each quantization interval. Therefore, it can be seen that (19) can be lower bounded by (2). In fact, this bound is tight for  $\Delta \rightarrow 0$ , i.e. when  $f_{\mathbf{X}}(\mathbf{x}) \approx \bar{f}(\mathbf{x})$  is valid (high rate assumption).

For simplicity, we continue with the one-dimensional case, although the same discussion holds for higher dimensions. In particular, for  $M = 1$  the difference between both entropies, now  $H(\mathbf{X}^\Delta)$  and  $\tilde{H}(\mathbf{X}^\Delta)$ , is always non-positive. The proof follows from Gibbs' inequality  $\ln x \leq x - 1$ :

$$\begin{aligned} \tilde{H}(\mathbf{X}^\Delta) - H(\mathbf{X}^\Delta) &= \int_{\mathbf{x}} f_{\mathbf{X}}(x) \log_2 \bar{f}(x) / f_{\mathbf{X}}(x) dx \\ &\leq \left( \int_{\mathbf{x}} f_{\mathbf{X}}(x) (\bar{f}(x) / f_{\mathbf{X}}(x) - 1) dx \right) \log_2 e \\ &= \left( \int_{\mathbf{x}} \bar{f}(x) dx - \int_{\mathbf{x}} f_{\mathbf{X}}(x) dx \right) \log_2 e \\ &= (1 - 1) \log_2 e = 0. \end{aligned} \quad (20)$$

### Convexity of Optimization Problem

In order to derive the convex problem in  $\mathbf{P}$ , i.e. (17), starting from (16), we need to find a new concave objective function and all constraints must be convex. Note that as we are dealing with norm-constraints, they are convex. For the  $l_0$  quasi-norm, convexity is preserved through the  $l_1$  norm relaxation. In addition, the constraint  $\mathbf{p} \in [0, 1]^M$  defines a convex set. The same holds for  $\mathbf{p} \in [\nu, 1]^M$ , which is forced with  $\mathbf{P} - \nu \mathbf{I} \geq \mathbf{0}$ . Recall that the latter condition implies  $\mathbf{P} > \mathbf{0}$ , which does not define a convex cone as the origin is not included [19].

The objective function  $\tilde{H}(\mathbf{X}_p^\Delta)$  in (16) can be approximated by a lower bound. To that end, we bound the rank function  $r(\mathbf{X})$  of a symmetric positive semi-definite matrix  $\mathbf{X}$  by its convex envelope, i.e. the nuclear norm  $\|\mathbf{X}\|_* = \text{tr}(\sqrt{\mathbf{X}^T \mathbf{X}})^1 = \text{tr}(\mathbf{X})$ , as discussed in [16]. This lower bound is valid if the spectral norm (maximum eigenvalue) of  $\mathbf{X}$  is smaller than 1, i.e.  $\|\mathbf{X}\|_2 = \lambda_{\max}(\mathbf{X}) \leq 1$ . Note that  $\|\mathbf{X}\|_*$  and  $\|\mathbf{X}\|_2$  are two different norms and can take different values.

By defining the normalized matrix  $\bar{\mathbf{C}}$  as  $\bar{\mathbf{C}} = \mathbf{C} / \|\mathbf{C}\|_2$ , we ensure that  $\|\sqrt{\bar{\mathbf{P}}} \bar{\mathbf{C}} \sqrt{\bar{\mathbf{P}}}\|_2 \leq 1$ . This comes from the fact that  $\|\mathbf{P}\|_2 \leq 1$  and that for two symmetric positive semi-definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ , it can be proved that  $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$  [23]. Therefore, given that  $r(\sqrt{\bar{\mathbf{P}}} \bar{\mathbf{C}} \sqrt{\bar{\mathbf{P}}}) = r(\sqrt{\mathbf{P}} \bar{\mathbf{C}} \sqrt{\mathbf{P}})$ , the objective function  $\tilde{H}(\mathbf{X}_p^\Delta)$  in (16) can be lower bounded as:

$$\begin{aligned} \tilde{H}(\mathbf{X}_p^\Delta) &= \frac{1}{2} \left( r(\sqrt{\bar{\mathbf{P}}} \bar{\mathbf{C}} \sqrt{\bar{\mathbf{P}}}) \log_2 \left( \frac{2\pi e}{\Delta^2} \right) + \log_2 \det(\mathbf{L}_r^T \mathbf{P} \mathbf{L}_r) \right) \\ &= \frac{1}{2} \left( r(\sqrt{\mathbf{P}} \bar{\mathbf{C}} \sqrt{\mathbf{P}}) \log_2 \left( \frac{2\pi e}{\Delta^2} \right) + \log_2 \det(\mathbf{L}_r^T \mathbf{P} \mathbf{L}_r) \right) \\ &\geq \frac{1}{2} \left( \|\sqrt{\bar{\mathbf{P}}} \bar{\mathbf{C}} \sqrt{\bar{\mathbf{P}}}\|_* \log_2 \left( \frac{2\pi e}{\Delta^2} \right) + \log_2 \det(\mathbf{L}_r^T \mathbf{P} \mathbf{L}_r) \right) \\ &\stackrel{(a)}{=} \frac{1}{2} \left( \text{tr}(\bar{\mathbf{P}} \bar{\mathbf{C}}) \log_2 \left( \frac{2\pi e}{\Delta^2} \right) + \log_2 \det(\mathbf{L}_r^T \mathbf{P} \mathbf{L}_r) \right) \\ &\triangleq \tilde{H}_{lb}(\mathbf{X}_p^\Delta), \end{aligned} \quad (21)$$

<sup>1</sup>The matrix square root operation  $\sqrt{\mathbf{Y}} = \mathbf{X}$  implies that  $\mathbf{Y} = \mathbf{X}^2$ . Note that although this operation is different from the element-wise square root operation, both operations coincide when applied to diagonal matrices.

where (a) follows from the definition of nuclear norm and the fact that  $\text{tr}(\sqrt{\bar{\mathbf{P}}} \bar{\mathbf{C}} \sqrt{\bar{\mathbf{P}}}) = \text{tr}(\bar{\mathbf{P}} \bar{\mathbf{C}})$ , which results from the trace property:  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$  for any matrices  $\mathbf{A}$  and  $\mathbf{B}$ .

Note that the lower bound in (21) is concave in  $\mathbf{P}$  because  $\text{tr}(\bar{\mathbf{P}} \bar{\mathbf{C}})$  is linear in  $\mathbf{P}$  and  $\log_2 \det(\mathbf{L}_r^T \mathbf{P} \mathbf{L}_r)$  is concave in  $\mathbf{P}$  for  $\mathbf{P} > \mathbf{0}$  [19]. Thus, replacing the objective function in (16) by its lower bound, we obtain the convex problem (17): maximization of a concave function subject to convex constraints.

Besides, for our analysis to be consistent, the approximation in (8) and (10) must be accurate. To that end, the quantization interval  $\Delta$  must be close to the minimum eigenvalues  $\lambda_{\min}(\mathbf{C})$  and  $\lambda_{\min}(\mathbf{PC})$ . For symmetric positive semi-definite matrices, we have that  $\lambda_{\min}(\mathbf{PC}) \geq \lambda_{\min}(\mathbf{P}) \lambda_{\min}(\mathbf{C})$  [23], and, since  $\lambda(\mathbf{P}) \leq 1$ , it holds that  $\lambda_{\min}(\mathbf{P}) \lambda_{\min}(\mathbf{C}) \leq \lambda_{\min}(\mathbf{C})$ . Thus,  $\Delta \leq \lambda_{\min}(\mathbf{P}) \lambda_{\min}(\mathbf{C})$  with  $\lambda_{\min}(\mathbf{P}) = \nu$  satisfies both conditions.

### REFERENCES

- [1] H. Shariatmadari *et al.*, "Machine-type communications: current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.
- [2] Z. Dawy *et al.*, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.
- [3] M. R. Palattella *et al.*, "Internet of things in the 5G era: enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.
- [4] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
- [5] R. K. Shakya, Y. N. Singh, and N. K. Verma, "Generic correlation model for wireless sensor network applications," *IET Wireless Sensor Systems*, vol. 3, no. 4, pp. 266–276, Dec. 2013.
- [6] H. Rowaihy *et al.*, "A survey of sensor selection schemes in wireless sensor networks," in *Proc. SPIE*, vol. 6562, 2007.
- [7] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
- [8] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.
- [9] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," *Wireless Networks*, vol. 11, no. 1, pp. 149–160, Jan. 2005.
- [10] A. Dabirmoghaddam, M. Ghaderi, and C. Williamson, "Cluster-based correlated data gathering in wireless sensor networks," in *2010 IEEE MASCOTS*, Aug. 2010, pp. 163–171.
- [11] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.
- [12] J. O. Berger, V. De Oliveira, and B. Sansó, "Objective bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1361–1374, 2001.
- [13] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.
- [14] O. Knill, "Cauchy-Binet for pseudo-determinants," *Linear Algebra and its Applications*, vol. 459, pp. 522–547, 2014.
- [15] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Technical University of Denmark, Nov. 2012.
- [16] F. J. Iglesias *et al.*, "Demixing and blind deconvolution of graph-diffused sparse signals," in *2018 IEEE ICASSP*, Apr. 2018, pp. 4189–4193.
- [17] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [18] S. P. Chepuri and G. Leus, "Sparse sensing for statistical inference," *Found. Trends Signal Process.*, vol. 9, no. 34, pp. 233–368, 2016.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, 2009.
- [20] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [21] Nokia White Paper, "LTE evolution for IoT connectivity," 3rd Ed. 2016.
- [22] R. G. Gallager, *Principles of Digital Communication*. Cambridge University Press, 2008.
- [23] B. Y. Wang and M. P. Gong, "Some eigenvalue inequalities for positive semidefinite matrix power products," *Linear Algebra and Its Applications*, vol. 184, pp. 249–260, 1993.