Near-Optimal Sensor Placement for Linear Inverse Problems

Juri Ranieri, *Student Member, IEEE*, Amina Chebira, *Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—A classic problem is the estimation of a set of parameters from measurements collected by only a few sensors. The number of sensors is often limited by physical or economical constraints and their placement is of fundamental importance to obtain accurate estimates. Unfortunately, the selection of the optimal sensor locations is intrinsically combinatorial and the available approximation algorithms are not guaranteed to generate good solutions in all cases of interest. We propose FrameSense, a greedy algorithm for the selection of optimal sensor locations. The core cost function of the algorithm is the *frame potential*, a scalar property of matrices that measures the orthogonality of its rows. Notably, FrameSense is the first algorithm that is near-optimal in terms of mean square error, meaning that its solution is always guaranteed to be close to the optimal one. Moreover, we show with an extensive set of numerical experiments that FrameSense achieves state-of-the-art performance while having the lowest computational cost, when compared to other greedy methods.**

***Index Terms*—Frame potential, greedy algorithm, inverse problem, sensor placement.**

I. Introduction

In many contexts, it is of interest to measure physical phenomena that vary in space and time. Common examples are temperature, sound, and pollution. Modern approaches tackling this problem are often based on wireless sensor networks (WSN), namely systems composed of many sensing nodes, each capable of measuring, processing and communicating information about the surrounding environment.

Challenges and trade-offs characterize the design of a WSN. One of the key aspects to design a successful WSN is the optimization of the spatial locations of the sensors nodes, given the location's impact on many relevant indicators, such as coverage, energy consumption and connectivity. When the data collected by the WSN is used to solve inverse problems, the optimization of the sensor locations becomes even more critical. In fact, the location of the sensor nodes determines the error of the solution of the inverse problem and its optimization represents the difference between being able to obtain a reasonable solution or not. In this work we consider linear inverse problems defined as

$$\boldsymbol{f} = \boldsymbol{\Psi}\boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{f} \in \mathbb{R}^N$ is the measured physical field, $\boldsymbol{\alpha} \in \mathbb{R}^K$ are the parameters to be estimated and $\boldsymbol{\Psi} \in \mathbb{R}^{N \times K}$ is the known linear model representing the relationship between the measurements and the parameters. Note that this simple model can be easily adapted to more complicated scenarios. For example, if the collected measurements are linear combinations of the physical field, as in the presence of a sampling kernel, we simply consider $\boldsymbol{\Psi} = \boldsymbol{\Phi}\boldsymbol{\Theta}$, where $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ represent the sampling kernel and the physical phenomenon, respectively.

The role of $\boldsymbol{\alpha}$ depends on the specific inverse problem. For example, if the WSN is designed for *source localization*, $\boldsymbol{\alpha}$ represents the location and the intensity of the field sources. On the other hand, if we are planning to *interpolate* the measured samples to recover the entire field, we may think of $\boldsymbol{\alpha}$ as its low-dimensional representation. In other scientific applications, for example [1]–[3], the solution of a linear inverse problem is a step within a complex procedure and $\boldsymbol{\alpha}$ may not have a direct interpretation. Nonetheless, the accurate estimation of $\boldsymbol{\alpha}$ is of fundamental importance.

It is generally too expensive or even impossible to sense the physical field $\boldsymbol{f}$ with N sensor nodes, where N is determined by the resolution of the discrete physical field. Assume we have only $L < N$ sensors, then we need to analyze how to choose the L sampling locations such that the solution of the linear inverse problem (1) has the least amount of error. Namely, we would like to choose the *most informative* L rows of $\boldsymbol{\Psi}$ out of the N available ones. One could simply adopt a brute force approach and inspect all the possible combinations for the L sensor locations. In this case, the operation count is exponential due to its combinatorial nature, making the approach unfeasible even for modest values of N.

It is possible to significantly reduce the computational cost by accepting a sub-optimal sensor placement produced by an approximation algorithm. In this case, near-optimal algorithms are desired since they *always* produce a solution of *guaranteed quality*. We measure this quality as the ratio between the value of the approximated solution and the value of the optimal one, and we call it the *approximation factor*.

A. Problem Statement and Prior Art

We consider the linear model introduced in (1) and a WSN measuring the field at only $L < N$ locations. We denote the sets of measured locations and of available locations as $\mathcal{L} =$

Manuscript received June 17, 2013; revised September 26, 2013; accepted December 04, 2013. Date of publication January 10, 2014; date of current version February 07, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Eduard Axel Jorswieck. This research was supported by an ERC Advanced Grant—Support for Frontier Research—SPARSAM Nr: 247006.

J. Ranieri and M. Vetterli are with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail juri.ranieri@epfl.ch).

A. Chebira is with Swiss Center for Electronics and Microtechnology (CSEM), CH-2002 Neuchâtel, Switzerland.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2299518

$\{i_1, \ldots, i_L\}$ and $\mathcal{N} = \{1, \ldots, N\}$, respectively. Note that $\mathcal{L} \subseteq \mathcal{N}$ and $|\mathcal{L}| = L$.

The measured field is denoted as $\boldsymbol{f}_{\mathcal{L}} \in \mathbb{R}^L$, where the subscript represents the selection of the elements of $\boldsymbol{f}$ indexed by $\mathcal{L}$. Consequently, we define a pruned matrix $\boldsymbol{\Psi}_{\mathcal{L}} \in \mathbb{R}^{L \times K}$, where we kept only the rows of $\boldsymbol{\Psi}$ indexed by $\mathcal{L}$. We obtain a smaller linear system of equations,

$$\boldsymbol{f}_{\mathcal{L}} = \boldsymbol{\Psi}_{\mathcal{L}} \boldsymbol{\alpha}, \tag{2}$$

where we still recover $\boldsymbol{\alpha}$, but with a reduced set of measurements, $L \geq K$. Note that we have by definition $\boldsymbol{\Psi}_{\mathcal{N}} = \boldsymbol{\Psi}$ and $\boldsymbol{f}_{\mathcal{N}} = \boldsymbol{f}$.

Given the set of measurements $\boldsymbol{f}_{\mathcal{L}}$, there may not exist an $\widehat{\boldsymbol{\alpha}}$ that solves (2). If it exists, the solution may not be unique. To overcome this problem, we usually look for the least squares solution, defined as $\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\Psi}_{\mathcal{L}} \boldsymbol{\alpha} - \boldsymbol{f}_{\mathcal{L}}\|_2^2$. Assume that $\boldsymbol{\Psi}_{\mathcal{L}}$ has rank K, then this solution is found using the Moore-Penrose pseudoinverse,

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Psi}_{\mathcal{L}}^{+} \boldsymbol{f}_{\mathcal{L}},$$

where $\boldsymbol{\Psi}_{\mathcal{L}}^{+} = \left(\boldsymbol{\Psi}_{\mathcal{L}}^{*} \boldsymbol{\Psi}_{\mathcal{L}}\right)^{-1} \boldsymbol{\Psi}_{\mathcal{L}}^{*}$. The pseudoinverse generalizes the concept of inverse matrix to non-square matrices and is also known as the *canonical dual frame* in frame theory. For simplicity of notation, we introduce $\mathbf{T}_{\mathcal{L}} = \boldsymbol{\Psi}_{\mathcal{L}}^{*} \boldsymbol{\Psi}_{\mathcal{L}} \in \mathbb{R}^{K \times K}$, a Hermitian-symmetric matrix that strongly influences the reconstruction performance. More precisely, the error of the least squares solution depends on the spectrum of $\boldsymbol{T}_{\mathcal{L}}$. That is, when the measurements $\boldsymbol{f}_{\mathcal{L}}$ are perturbed by a zero-mean i.i.d. Gaussian noise with variance σ^2, the mean square error (MSE) of the least squares solution [4] is

$$\mathrm{MSE}(\widehat{\boldsymbol{\alpha}}) = \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2^2 = \sigma^2 \sum_{k=1}^{K} \frac{1}{\lambda_k}, \tag{3}$$

where λ_k is the k-th eigenvalue of the matrix $\boldsymbol{T}_{\mathcal{L}}$. We thus state the sensor placement problem as follows.

Problem 1: Given a matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N \times K}$ and a number of sensors L, find the sensor placement $\mathcal{L}$ such that

$$\arg\min_{\mathcal{L}} \sum_{k=1}^{K} \frac{1}{\lambda_k} \qquad \textit{subject to} \qquad |\mathcal{L}| = L. \tag{4}$$

Note that if $\boldsymbol{T}_{\mathcal{L}}$ is rank deficient, that is $\mathrm{rank}(\boldsymbol{T}_{\mathcal{L}}) < K$, then the MSE is not bounded.

A trivial choice would be to design algorithms minimizing directly the MSE with some approximation procedure, such as greedy ones. In practice, the MSE is not used because it has many unfavorable local minima. Therefore, the research effort is focused in finding tight proxies of the MSE that can be efficiently optimized. In what follows, we survey different approximation strategies and proxies from the literature.

B. Prior Work

Classic solutions to the sensor placement problem can be classified in three categories: convex optimization, greedy methods and heuristics.

Convex optimization methods [5], [6] are based on the relaxation of the Boolean constraints $\{0, 1\}^N$ representing the sensor placement to the convex set $[0, 1]^N$. This relaxation is usually not tight as heuristics are needed to choose the sensor locations and there is no a-priori guarantee on the distance from the optimal solution. The authors in [6] define an online bound for the quality of the obtained solution by looking at the gap between the primal and the dual problem.

Heuristic methods [7]–[12] are valid options to reduce the cost of the exhaustive search, which has a prohibitive cost. Again, even if the methods work in practice, little can be said about the quality or the optimality of the solution.

Greedy algorithms leveraging the submodularity of the cost function [13] are a class of algorithms having polynomial complexity and guaranteed performance with regards to the chosen cost function [5], [14]–[17]. Since the MSE is not submodular in general [15], [17], alternative cost functions have been considered [5], [14]–[17]. The proposed methods are theoretically near-optimal with regards to the chosen cost function, but little can be said about the achieved MSE. Moreover, the local optimization of the proposed cost functions are computationally demanding, often requiring the inversion of large matrices [14]. Therefore, approximations of the cost functions have been proposed [14], offering a significant speedup for an acceptable reduction of the solution's quality.

Beside the approximation strategy, approximation algorithms are differentiated by the chosen cost function. Under restrictive assumptions, the MSE can be chosen as a cost function, see [15], [18]. In [17], the authors bounded the performance of greedy algorithms optimizing R^2, a measure of goodness of fit based on the MSE, using the concept of *submodularity ratio*. However, such algorithm is generally less performant than the greedy algorithms optimizing proxies of the MSE. Common proxies of the MSE are inspired by *information theoretic* measures such as entropy [12], cross-entropy [16], [19] and mutual information [14]. A popular choice is the maximization of the log determinant of $\boldsymbol{T}_{\mathcal{L}}$, being the volume of the confidence ellipsoid given by the measurements. This proxy has been historically introduced in D-Optimal experiment design [20], but has also been successfully proposed as a cost function for a convex relaxed method [6] and greedy algorithms [5]. Other proxies have also been introduced in optimal experiment design, such as maximization of the smallest eigenvalue λ_K (E-Optimal design) or the maximization of the trace of $\boldsymbol{T}_{\mathcal{L}}$ (T-Optimal design). A detailed description of the different choices available for experiment design can be found in [20].

Note that there exists optimal strategies with a reasonable computational cost for some specific scenarios. This is the case when we have the freedom of completely designing the matrix $\boldsymbol{\Psi}_{\mathcal{L}}$ given the dimensions L and K. More precisely, if $L = K$, the optimal basis corresponds to an orthonormal basis, while if $L > K$, then we are looking for a unit-norm tight frame [21], [22]. Benedetto *et al.* showed that each tight frame is a non-unique global minimizer of the *frame potential* (FP), that is a scalar property of the frame defined as

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) = \sum_{i,j \in \mathcal{L}} \left|\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle\right|^2,$$

where $\boldsymbol{\psi}_i$ is the ith row of $\boldsymbol{\Psi}_{\mathcal{L}}$. One of the reasons of the popularity of the FP in the frame theory community is its interesting

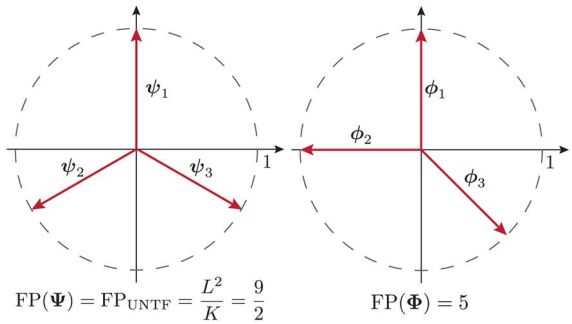


Fig. 1. A graphical representation of the unit-norm rows of two matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. On the left, the rows $\boldsymbol{\Psi}$ are optimally spread on $\mathbb{R}^2$ minimizing the FP and the MSE. This is also know as the Mercedes-Benz frame and is a typical example of a unit-norm tight frame. On the right, $\boldsymbol{\Psi}$ is the frame built by adding a vector to the orthonormal basis of $\mathbb{R}^2$ $[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2]$. The three vectors are not minimizing the FP and therefore they are not in equilibrium with regards to the frame force. You can envision a parallel example with three electrons on a unit circle under the Coulomb force. More precisely, if the three electrons are free to move, they would reach an equilibrium when located as the vectors of $\boldsymbol{\Psi}$, up to a rotation factor.

physical interpretation [23]. Namely, it is the potential energy of the so-called *frame force*, a force between vectors inspired by the Coulomb force. The frame force and its potential energy have been introduced for their *orthogonality encouraging* property: the force is repulsive when the angle between the two vectors is acute, null when they are orthogonal and attractive when the angle is obtuse. A graphical explanation of this physical interpretation is given in Fig. 1, where the unit-norm rows of two matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ belonging to $\mathbb{R}^{3\times 2}$ are represented. While $\boldsymbol{\Psi}$, that is the unit-norm tight frame minimizing the FP, has vectors as close to orthogonality as possible and therefore in equilibrium with regards to the frame force, the vectors of $\boldsymbol{\Phi}$ are not optimally close to orthogonality and the FP is thus not minimized. Note that according to frame theory [22], $\boldsymbol{\Psi}$ is the matrix that also achieves the minimum MSE (per component).

Note that the FP is also known as the *total summed correlation* in communication theory [24]. It is used to optimize the signatures of CDMA systems to achieve the Welch's lower bound [25] and maximize the capacity of the channel.

Given its interpretation and its role in defining the existence of tight frames—the optimal frames in terms of MSE—we hypothesize that the FP is an interesting cost function for an approximation algorithm.

C. Our Contributions

We propose FrameSense, a greedy sensor placement method that minimizes the frame potential to choose the sensing locations $\mathcal{L}$. We briefly summarize the innovative aspects of the proposed algorithm:

- Under some stability conditions regarding the spectrum of $\boldsymbol{\Psi}$, FrameSense is the only known algorithm, to the best of our knowledge, that is near-optimal with regards to MSE.
- FrameSense outperforms other greedy algorithms in terms of MSE.
- FrameSense is on par with the method based on convex relaxation [6], which uses heuristics to improve the local solution and has a significantly higher complexity.
- The computational cost of FrameSense is significantly lower with regards to the other considered algorithms.

The remainder of the paper uses the following notations: calligraphic letters as $\mathcal{C}$ indicate sets of elements, while bold letters as $\boldsymbol{\Psi}$ and $\boldsymbol{\psi}$ indicates matrices and vectors, respectively. The k-th largest eigenvalue of $\mathbf{T}_{\mathcal{L}} = \boldsymbol{\Psi}_{\mathcal{L}}^* \boldsymbol{\Psi}_{\mathcal{L}}$ is denoted as λ_k. Moreover, we always consider a real physical field $\boldsymbol{f}$ and real matrices $\boldsymbol{\Psi}$ for simplicity but the extension to the complex domain does not require major adjustments.

The content is organized as follows. In Section II we introduce some frame theory concepts focusing on the role of the FP. We describe FrameSense and the analysis of its near-optimality in Section III and we numerically compare its performance with various other algorithms in Section IV.

II. The Frame Potential in Frame Theory

This section briefly introduces some of the basic concepts of frame theory that are useful to understand and analyze the proposed algorithm. Frame theory studies and designs families of matrices $\boldsymbol{\Psi}_{\mathcal{L}}$ such that $\boldsymbol{T}_{\mathcal{L}}$ is well-conditioned. More precisely, $\boldsymbol{\Psi}_{\mathcal{L}}$ is a frame for a Hilbert space $\mathbb{H}$ if there exists two scalars A and B such that $0 < A \leq B < \infty$ so that for every $\mathbf{x} \in \mathbb{H}$ we have

$$A\|\mathbf{x}\|_2^2 \leq \|\boldsymbol{\Psi}_{\mathcal{L}}\mathbf{x}\|_2^2 \leq B\|\mathbf{x}\|_2^2,$$

where A and B are called frame bounds. $\boldsymbol{\Psi}_{\mathcal{L}}$ is a tight frame when $A = B$ and its columns are orthogonal by construction. Of particular interest is the case of unit norm tight frames (UNTF); these are tight frames whose frame elements—the rows $\boldsymbol{\psi}_i$—have unit norm. These provide Parseval-like relationships, despite the non-orthogonality of the frame elements of $\boldsymbol{\Psi}_{\mathcal{L}}$. In addition to (3), there are other interesting relationships between the characteristics of $\boldsymbol{\Psi}_{\mathcal{L}}$ and the spectrum of $\boldsymbol{T}_{\mathcal{L}}$. For example, we can express the FP as

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) = \mathrm{Trace}\left(\boldsymbol{T}_{\mathcal{L}}^* \boldsymbol{T}_{\mathcal{L}}\right) = \sum_{k=1}^{K} |\lambda_k|^2.$$

Moreover, the sum of the eigenvalues of $\boldsymbol{T}_{\mathcal{L}}$ is equal to the sum of the norm of the rows, $\sum_{i=1}^{L} \|\boldsymbol{\psi}_i\|^2 = \sum_{k=1}^{K} \lambda_k$.

These quantities of interest take a simplified analytical form for UNTFs. In this scenario, we know [22] that the FP is minimum with regards to all other matrices of the same size with unit-norm rows, and it is equal to $\mathrm{FP}_{\mathrm{UNTF}} = \frac{L^2}{K}$. According to [22], the optimal MSE is also achieved when the FP is minimized and it is equal to $\mathrm{MSE}_{\mathrm{UNTF}} = \frac{K^2}{L}$. Note that in this case all the eigenvalues are equal, $\lambda_{\mathrm{UNTF}} = \lambda_i = \frac{L}{K} \forall i$.

Next, we would like to intuitively explain why the FP is a good candidate to be a proxy for the MSE. Consider the distance between the FP of a matrix with unit-norm rows $\boldsymbol{\Psi}_{\mathcal{L}} \in \mathbb{R}^{L\times K}$ and the FP of a UNTF. Then, it is possible to show that this distance is always positive and equal to

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) - \mathrm{FP}_{\mathrm{UNTF}} = \sum_{k=1}^{K} \left(\lambda_k - \frac{L}{K}\right)^2,$$

where $\frac{L}{K}$ is the value of the eigenvalues λ_{UNTF}. Note that if we minimize the FP of $\boldsymbol{\Psi}_{\mathcal{L}}$, then each λ_k converges to $\frac{L}{K}$.

At the same time, the distance between the MSE of $\mathbf{\Psi}_{\mathcal{L}}$ and the MSE of the UNTF can be expressed as

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{L}}) - \mathrm{MSE}_{\mathrm{UNTF}} = \sum_{k=1}^{K} \left(\frac{1}{\lambda_k} - \frac{K}{L} \right).$$

Now, it is easy to see that if the eigenvalues converge to $\frac{L}{K}$, then $\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{L}})$ converges to the MSE of a UNTF, being also the optimal one [22].

III. FrameSense: A Near-Optimal Sensor Placement

Even if the intuition given in Section II is clear, it does not directly explain why an algorithm placing the sensor according to the FP would perform well in terms of MSE. Indeed, we need to address some complications such as matrices $\mathbf{\Psi}$ having rows with different norms and the non-uniform convergence of the eigenvalues. In what follows, we first describe the details of FrameSense and then analyze its near-optimality in terms of both the FP and the MSE.

A. The Algorithm

FrameSense finds the sensor locations $\mathcal{L}$ given the known model $\mathbf{\Psi}$ and the number of available sensors nodes L with a greedy minimization of the FP. It is a greedy "worst-out" algorithm: at each iteration it removes the row of $\mathbf{\Psi}$ that maximally increases the FP. In other words, we define a set of locations $\mathcal{S}$ that are not suitable for sensing and at each iteration we add to $\mathcal{S}$ the row that maximizes the following cost function:

$$F(\mathcal{S}) = \mathrm{FP}(\mathbf{\Psi}) - \mathrm{FP}(\mathbf{\Psi}_{\mathcal{N}\setminus\mathcal{S}}). \tag{5}$$

The pseudo-code for FrameSense is given in Algorithm 1.

Algorithm 1: FrameSense

Require: Linear Model $\mathbf{\Psi}$, Number of sensors L

Ensure: Sensor locations $\mathcal{L}$

1) Initialize the set of locations, $\mathcal{L} = \emptyset$.
2) Initialize the set of available locations, $\mathcal{N} = \{1, \ldots, N\}$.
3) Find the first two rows to eliminate, $\mathcal{S} = \arg\max_{i,j\in\mathcal{N}} |\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle|^2$.
4) Update the available locations, $\mathcal{L} = \mathcal{N} \setminus \mathcal{S}$.
5) **Repeat until L locations are found**
   a) If $|\mathcal{S}| = N - L$, stop.
   b) Find the optimal row, $i^* = \arg\min_{i\in\mathcal{L}} F(\mathcal{S} \cup i)$.
   c) Update the set of removed locations, $\mathcal{S} = \mathcal{S} \cup i^*$.
   d) Update the available locations, $\mathcal{L} = \mathcal{L} \setminus i^*$.

One may ask why we do not optimize directly the MSE, instead of minimizing the FP, which indirectly optimizes the MSE. As we have already indicated, a greedy algorithm optimizing a general function, like the MSE, converges to a local stationary point of the cost function and we have no guarantee on the distance from the global optimum. On the other hand, we can prove that FrameSense is near-optimal with regards to the FP by exploiting the submodularity of the cost function. In addition, we also guarantee the performance of FrameSense in terms of the MSE, exploiting a link between FP and MSE.

B. Near-Optimality of FrameSense With Regards to FP

We define the performance of FrameSense with regards to FP using the theory of submodular functions. We start by defining the concept of submodularity that relates to the concept of diminishing returns: if we add an element to a set $\mathcal{Y}$, the benefit is smaller or equal than adding the same element to one of the subsets of $\mathcal{Y}$. Then, we introduce a theorem by Nemhauser *et al.* [13] that defines the approximation factor of greedy algorithms maximizing a submodular function. We continue by showing that FrameSense satisfies the conditions of Nemhauser's theorem and we derive its approximation factor in terms of FP.

Definition 1 (Submodular Function): Given two sets $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathcal{X} \subset \mathcal{Y} \subset \mathcal{N}$ and given an element $i \in \mathcal{N} \setminus \mathcal{Y}$, a function G is submodular if it satisfies

$$G(\mathcal{X} \cup i) - G(\mathcal{X}) \geq G(\mathcal{Y} \cup i) - G(\mathcal{Y}). \tag{6}$$

Submodular functions are useful in combinatorial optimization because greedy algorithms have favorable properties when optimizing a function with such a property. More precisely, it has been proved that the greedy maximization of submodular functions is near-optimal [13].

Theorem 1 (Near-Optimal Maximization of Submodular Function [13]): Let G be a normalized, monotone, submodular set function over a finite set $\mathcal{N}$. Let $\mathcal{L}$ be the set of L elements chosen by the greedy algorithm, and let $\mathrm{OPT} = \max_{\mathcal{A}\subset\mathcal{N},|\mathcal{A}|=L} G(\mathcal{A})$ be the optimal set of elements. Then

$$G(\mathcal{L}) \geq \left(1 - \frac{1}{e}\right) \mathrm{G(OPT)},$$

where e is Euler's number.

Namely, if G satisfies the conditions of Theorem 1, then the solution of the greedy algorithm is always close to the optimal one. These conditions are satisfied by the cost function F in (5), as shown in the following lemma.

Lemma 1 (Submodularity of the Cost Function): The set function maximized in Algorithm 1,

$$F(\mathcal{S}) = \mathrm{FP}(\mathbf{\Psi}) - \mathrm{FP}(\mathbf{\Psi}_{\mathcal{N}\setminus\mathcal{S}}), \tag{7}$$

is a normalized, monotone, submodular function.

Proof: The set function F is normalized if $F(\emptyset) = 0$. Here, normalization is trivially shown since $\mathbf{\Psi} = \mathbf{\Psi}_{\mathcal{N}}$ by definition. To show monotonicity, we pick a generic matrix $\mathbf{\Psi}$ of N rows, a set $\mathcal{X}$ and an index $i \notin \mathcal{X}$. Then, we compute the increment of F due to i with regards to the set $\mathcal{X}$, showing that it is always positive.

$$\begin{aligned} F(\mathcal{X} \cup i) - F(\mathcal{X}) &= \mathrm{FP}(\mathbf{\Psi}_{\mathcal{N}\setminus\mathcal{X}}) - \mathrm{FP}(\mathbf{\Psi}_{\mathcal{N}\setminus\mathcal{X}\cup i}) \\ &\overset{(a)}{=} \sum_{n,m\in\mathcal{A}\cup i} |\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_m \rangle|^2 - \sum_{n,m\in\mathcal{A}} |\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_m \rangle|^2 \\ &= 2 \sum_{n\in\mathcal{A}} |\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_i \rangle|^2 + |\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_i \rangle|^2 \geq 0, \end{aligned}$$

where (a) is due to a change of variable $\mathcal{N} \setminus \mathcal{X} = \mathcal{A}$. Assuming without loss of generality that $\mathcal{Y} = \mathcal{X} \cup j$, we check the submodularity according to Definition 1.

$$\begin{aligned} &F(\mathcal{X} \cup i) - F(\mathcal{X}) - F(\mathcal{Y} \cup i) + F(\mathcal{Y}) \\ &\quad = F(\mathcal{X} \cup i) - F(\mathcal{X}) - F\left(\mathcal{X} \cup \{i,j\}\right) + F(\mathcal{X} \cup j) \\ &\quad = \mathrm{FP}\left(\boldsymbol{\Psi}_{\mathcal{A}\cup\{i,j\}}\right) - \mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{A}\cup j}) - \mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{A}\cup i}) + \mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{A}}) \\ &\quad = 2\sum_{n\in\mathcal{A}\cup j} \left|\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_i\rangle\right|^2 - 2\sum_{n\in\mathcal{A}} \left|\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_i\rangle\right|^2 \\ &\quad = 2\left|\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j\rangle\right|^2 \\ &\quad \geq 0. \end{aligned}$$

■

Now, we use Theorem 1 to derive the approximation factor of FrameSense with regards to the FP.

Theorem 2 (FP Approximation Factor): Consider a matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N\times K}$ and a given number of sensors L, such that $K \leq L < N$. Denote the optimal set of locations as $\mathrm{OPT} = \arg\max_{\mathcal{A}\subset\mathcal{N},|\mathcal{A}|=L} \mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{A}})$ and the greedy solution found by FrameSense as $\mathcal{L}$. Then, $\mathcal{L}$ is near-optimal with regards to the FP,

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) \leq \gamma \mathrm{FP}(\boldsymbol{\Psi}_{OPT}), \tag{8}$$

where $\gamma = \left(1 + \frac{1}{e}\left(\mathrm{FP}(\boldsymbol{\Psi})\frac{K}{L_{MIN}^2} - 1\right)\right)$ is the approximation factor and $L_{MIN} = \min_{|\mathcal{L}|=L} \sum_{i\in\mathcal{L}} \|\boldsymbol{\psi}_i\|^2$ is the sum of the norms of the L rows with the smallest norm.

Proof: According to Lemma 1, the cost function used in FrameSense satisfies the conditions of Theorem 1. Therefore,

$$F(\overline{\mathrm{OPT}}) - F(\mathcal{S}) \leq \frac{1}{e}\left(F(\overline{\mathrm{OPT}})\right),$$

where $F(\mathcal{S}) = (\boldsymbol{\Psi}) - \mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{N}\setminus\mathcal{S}})$ is the considered cost function, $\mathcal{S}$ is the set of rows eliminated by FrameSense and $\overline{\mathrm{OPT}} = \mathcal{N} \setminus \mathrm{OPT}$. If we consider the cost function, we obtain

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{N}\setminus\mathcal{S}}) \leq \left(1 - \frac{1}{e}\right)\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{N}\setminus\overline{\mathrm{OPT}}}) + \frac{1}{e}\mathrm{FP}(\boldsymbol{\Psi}). \tag{9}$$

Then, we note that the following minimization problem,

$$\mathrm{minimize}_{\mathcal{L}}\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) \qquad \text{subject to} \qquad |\mathcal{L}| = L,$$

is equivalent, under the change of variable $\mathcal{L} = \mathcal{N} \setminus \mathcal{S}$, to

$$\mathrm{minimize}_{\mathcal{S}}\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{N}\setminus\mathcal{S}}) \qquad \text{subject to} \qquad |\mathcal{S}| = N - L.$$

Using the equivalence in (9), we obtain

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) \leq \left(1 + \frac{1}{e}\left(\frac{\mathrm{FP}(\boldsymbol{\Psi})}{\mathrm{FP}(\boldsymbol{\Psi}_{\mathrm{OPT}})} - 1\right)\right)\mathrm{FP}(\boldsymbol{\Psi}_{\mathrm{OPT}}),$$

To conclude the proof, we bound from above the term $\frac{1}{e}\frac{\mathrm{FP}(\boldsymbol{\Psi})}{(\boldsymbol{\Psi}_{\mathrm{OPT}})}$. First, we consider the optimal solution OPT to select a tight frame whose rows have a summed norm of $L_{\mathrm{OPT}} = \sum_{i\in\mathrm{OPT}} \|\boldsymbol{\psi}_i\|^2$,

$$\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{L}}) \leq \left(1 + \frac{1}{e}\left(\mathrm{FP}(\boldsymbol{\Psi})\frac{K}{{L_{\mathrm{OPT}}}^2} - 1\right)\right)\mathrm{FP}(\boldsymbol{\Psi}_{\mathrm{OPT}}).$$

Then, we assume that OPT selects the rows having the smallest norm, $L_{\mathrm{OPT}} \geq L_{\mathrm{MIN}} = \min_{|\mathcal{L}|=L} \sum_{i\in\mathcal{L}} \|\boldsymbol{\psi}_i\|^2$. ■

Note that the FP of the original matrix influences significantly the final result: the lower the FP of $\boldsymbol{\Psi}$, the tighter the approximation obtained by the greedy algorithm. Therefore, FrameSense performs better when the original matrix $\boldsymbol{\Psi}$ is *closer* to a tight frame. In fact, the FP is bounded as follows,

$$\frac{\left(\sum_{i=1}^{N}\|\boldsymbol{\psi}_i\|^2\right)^2}{K} \leq \mathrm{FP}(\boldsymbol{\Psi}) \leq \left(\sum_{i=1}^{N}\|\boldsymbol{\psi}_i\|^2\right)^2,$$

where the lower bound is reached by tight frames and the upper bound by rank 1 matrices.

Moreover, Theorem 2 suggests to remove from $\boldsymbol{\Psi}$ the rows whose norm is significantly smaller with regards to the others to improve the performance of FrameSense. This suggestion is intuitive, since such rows are also the least *informative*.

C. Near-Optimality of FrameSense With Regards to MSE

Having a near-optimal FP does not necessarily mean that the obtained MSE is also near-optimal. Here, we show that, under some assumptions on the spectrum of $\boldsymbol{\Psi}$, FrameSense is near-optimal with regards to the MSE.

Before going to the technical details, we generalize the concept of number of sensors to account for the norms of the rows of $\boldsymbol{\Psi}$. More precisely, we keep L as the number of rows and we define $L_{\mathcal{A}} = \sum_{i\in\mathcal{A}} \|\boldsymbol{\psi}_i\|^2$ as the sum of the norms of the rows of $\boldsymbol{\Psi}_{\mathcal{A}}$ for a generic set $\mathcal{A}$. We also define the two extremal values of $L_{\mathcal{A}}$,

$$L_{\mathrm{MIN}} = \min_{\mathcal{A}\in\mathcal{N},|\mathcal{A}|=L} \sum_{i\in\mathcal{A}} \|\boldsymbol{\psi}_i\|^2, \tag{10}$$

$$L_{\mathrm{MAX}} = \max_{\mathcal{A}\in\mathcal{N},|\mathcal{A}|=L} \sum_{i\in\mathcal{A}} \|\boldsymbol{\psi}_i\|^2, \tag{11}$$

indicating respectively the minimum and the maximum value of $L_{\mathcal{A}}$ among all possible selections of L out of N rows of $\boldsymbol{\Psi}$. $L_{\mathcal{A}}$ is also connected to the spectrum of $\boldsymbol{T}_{\mathcal{A}}$. Indeed, $L_{\mathcal{A}}$ is the trace of $\boldsymbol{T}_{\mathcal{A}}$ and thus it is also the sum of its eigenvalues,

$$L_{\mathcal{A}} = \sum_{i\in\mathcal{A}} \|\boldsymbol{\psi}_i\|^2 = \mathrm{Trace}(\boldsymbol{T}_{\mathcal{A}}) = \sum_{i\in\mathcal{A}} \lambda_i.$$

If $\boldsymbol{\Psi}$ has rows with unit-norm, then $L_{\mathcal{A}} = L_{\mathrm{MIN}} = L_{\mathrm{MAX}} = L$.

As a first step to prove the near-optimality with regards to MSE, we consider a possible placement $\mathcal{A}$ and we bound the MSE of the matrix $\boldsymbol{\Psi}_{\mathcal{A}}$ using its FP and the spectrum of $\boldsymbol{T}_{\mathcal{A}}$. To obtain such a bound, we use a known inequality [26] involving variance, arithmetic mean and harmonic mean of a set of positive bounded numbers, in this case the eigenvalues of $\boldsymbol{T}_{\mathcal{A}}$. The following lemma describes the bound, while its proof is given in Appendix A.

Lemma 2 (MSE Bound): Consider any $\boldsymbol{\Psi}_{\mathcal{A}} \in \mathbb{R}^{L\times K}$ with $|\mathcal{A}| = L \leq K$ and denote the spectrum of $\boldsymbol{T}_{\mathcal{A}}$ as $\lambda_1 \geq \ldots \geq \lambda_K$. Then the MSE is bounded according to the FP as,

$$\mathrm{MSE}(\boldsymbol{\Psi}_{\mathcal{A}}) \leq \frac{K}{L_{MIN}}\frac{\mathrm{FP}(\boldsymbol{\Psi}_{\mathcal{A}})}{{\lambda_K}^2}, \tag{12}$$

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{A}}) \geq \frac{K}{L_{MAX}} \frac{\mathrm{FP}(\mathbf{\Psi}_{\mathcal{A}})}{{\lambda_1}^2}, \tag{13}$$

where L_{MIN} and L_{MAX} are defined as in (10) and (11).

Lemma 2 is key to study the approximation factor with regards to the MSE. Specifically, it allows to analyze the two extremal cases:

- Given the optimal FP, what is the lowest MSE we can achieve?
- Given the worst case FP according to Theorem 2, what is the largest MSE we may encounter?

Lemma 2 implies the necessity to properly bound the spectrum of any $\mathbf{\Psi}_{\mathcal{A}}$ with a given FP obtained from $\mathbf{\Psi}$. While it is possible to bound λ_1 with the FP, it is also easy to build matrices with $\lambda_K = 0$, compromising the bound given in (12). Therefore, we introduce the following property to control the eigenvalues of any $\boldsymbol{T}_{\mathcal{A}}$.

Definition 2 ((δ, L)-Bounded Frame): Consider a matrix $\mathbf{\Psi} \in \mathbb{R}^{N \times K}$ where $N \geq L$ and $N > K$. Then, we say that $\mathbf{\Psi}$ is (δ, L)-bounded if, for every $\mathcal{A} \subseteq \mathcal{N}$ such that $|\mathcal{A}| = L$, $\boldsymbol{T}_{\mathcal{A}}$ has a bounded spectrum

$$\frac{L_{MEAN}}{K} - \delta \leq \lambda_i \leq \frac{L_{MEAN}}{K} + \delta,$$

where $1 \leq i \leq K$, $\delta \geq 0$ and $L_{MEAN} = \frac{L}{N} \sum_{i \in \mathcal{N}} \|\boldsymbol{\psi}_i\|^2$ is average value of $L_{\mathcal{A}}$.

The concept of (δ, L)-bounded frames is similar to the notion of matrices satisfying the restricted isometry property, that are used in compressive sensing to guarantee the reconstruction of a sparse vector from a limited number of linear measurements [27]. Moreover, it allows us to define an approximation factor for the MSE that does not depend on the FP, the cost-function we minimize.

Theorem 3 (MSE Approximation Factor for (δ, L)-Bounded Frames): Consider a matrix $\mathbf{\Psi} \in \mathbb{R}^{N \times K}$ and $L \geq K$ sensors. Assume $\mathbf{\Psi}$ to be a (δ, L)-bounded frame, let d be the ratio L_{MEAN}/K and define the optimal placement in terms of MSE as $OPT = \arg\min_{\mathcal{A} \in \mathcal{N}, |\mathcal{A}| = L} \mathrm{MSE}(\mathbf{\Psi}_{\mathcal{A}})$. Then the solution $\mathcal{L}$ of FrameSense is near-optimal with regards to the MSE,

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{L}}) \leq \eta \mathrm{MSE}(\mathbf{\Psi}_{OPT}) \quad with\ \eta = \gamma \frac{(d+\delta)^2}{(d-\delta)^2} \frac{L_{MAX}}{L_{MIN}},$$

where η is the approximation factor of the MSE and γ is the approximation factor of the FP.

Proof: First, we compute the worst case MSE when FrameSense yields the worst FP, that is for $\mathrm{FP}(\mathbf{\Psi}_{\mathcal{L}}) = \gamma \mathrm{FP}(\mathbf{\Psi}_{\mathrm{OPT}})$. Using the upper bound (12) and the bounds on the spectrum for (δ, L)-bounded frames, we have

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{L}}) \leq \frac{K}{L_{\mathrm{MIN}}} \frac{\gamma \mathrm{FP}(\mathbf{\Psi}_{\mathrm{OPT}})}{(d-\delta)^2}. \tag{14}$$

Then, we compute the best case MSE when the FP is optimal. We note that the lower bound (13) of the MSE is monotonically decreasing with regards to the FP. Therefore, we use the same strategy considered for the lower bound, and obtain

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathrm{OPT}}) \leq \frac{K}{L_{\mathrm{MAX}}} \frac{\mathrm{FP}(\mathbf{\Psi}_{\mathrm{OPT}})}{(d+\delta)^2}. \tag{15}$$

Note that we consider that the optimal MSE is achieved for the optimal FP because the lower bound of the MSE is monotonically decreasing with regards to FP. Finally, we compute the MSE approximation ratio as the ratio between (14) and (15), obtaining the desired result. ■

The definition of (δ, L)-bounded frames is key to the proof. It turns out that many families of adequately normalized random matrices satisfy Definition 2, but it is hard to build deterministic matrices with such property. Nonetheless, FrameSense works well even for $\mathbf{\Psi}$ that are not provably (δ, L)-bounded. The same happens for compressed sensing and RIP matrices [28].

D. Practical Considerations on FrameSense

One point of FrameSense that needs improvement is the optimization of the sensing energy $L_{\mathcal{L}}$. In fact, the FP tends to discard the rows having a larger norm, which in theory could be more relevant to minimize the MSE. As an example, consider a matrix $\mathbf{\Psi}$ built as follows,

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_0 \\ C\mathbf{\Psi}_0 \end{bmatrix}, \tag{16}$$

where $\mathbf{\Psi}_0$ is a generic matrix and $C > 1$ is a constant. FrameSense would sub-optimally pick rows from the first matrix, discarding the ones from the second matrix and creating a sub-optimal sensor placement. In fact, the second matrix would have a lower MSE thanks to the multiplicative constant C. To limit such phenomenon, we optimize the sensor location on $\mathbf{\Psi}'$, that is a $\mathbf{\Psi}$ with unit-norm rows. This solution is not perfect: it removes the negative bias introduced by the norm of $\boldsymbol{\psi}_i$, but it does not exploit the sensing energy to improve the sensor placement $\mathcal{L}$. We leave to future work the study of a new FP-based algorithm able to exploit the information contained in the norm of the rows.

We conclude this section with a quantification of the bounds given in Theorem 2 and Theorem 3 in a simple scenario. Consider a matrix $\mathbf{\Psi} \in \mathbb{C}^{N \times K}$ filled with i.i.d. Gaussian complex random variables with zero mean and variance $\sigma^2 = 1/K$. Assume that $L = c_1 K$ and $N = c_2 K$, with $c_2 > 1 > c_1$. Then, we have the following inequalities verified in expectation,

$$\begin{aligned} \|\boldsymbol{\psi}_i\|_2 &= 1 \quad \forall i, \\ \frac{K}{{L_{\mathrm{MIN}}}^2} &= \frac{1}{{c_1}^2 K}, \\ \mathrm{FP}(\mathbf{\Psi}) &= {c_2}^2 {c_1}^2 K. \end{aligned}$$

According to Theorem 2, the approximation factor of the FP is equal to $\gamma = 1 + \frac{c_2^2 - 1}{e}$. We assume that K is sufficiently large and we consider the Marchenko-Pastur law [29] to compute the following approximated bounds for the eigenvalues of $\mathbf{\Psi}_{\mathcal{L}}$,

$$d - \delta = \sqrt{\frac{1}{c_1}}(1 - \sqrt{c_1})^2 \leq \lambda_i \leq \sqrt{\frac{1}{c_1}}(1 + \sqrt{c_1})^2 = d + \delta.$$

In this scenario, the MSE approximation factor is equal to,

$$\eta = \left(1 + \frac{{c_2}^2 - 1}{e}\right) \left(\frac{1 + \sqrt{c_1}}{1 - \sqrt{c_1}}\right)^4.$$

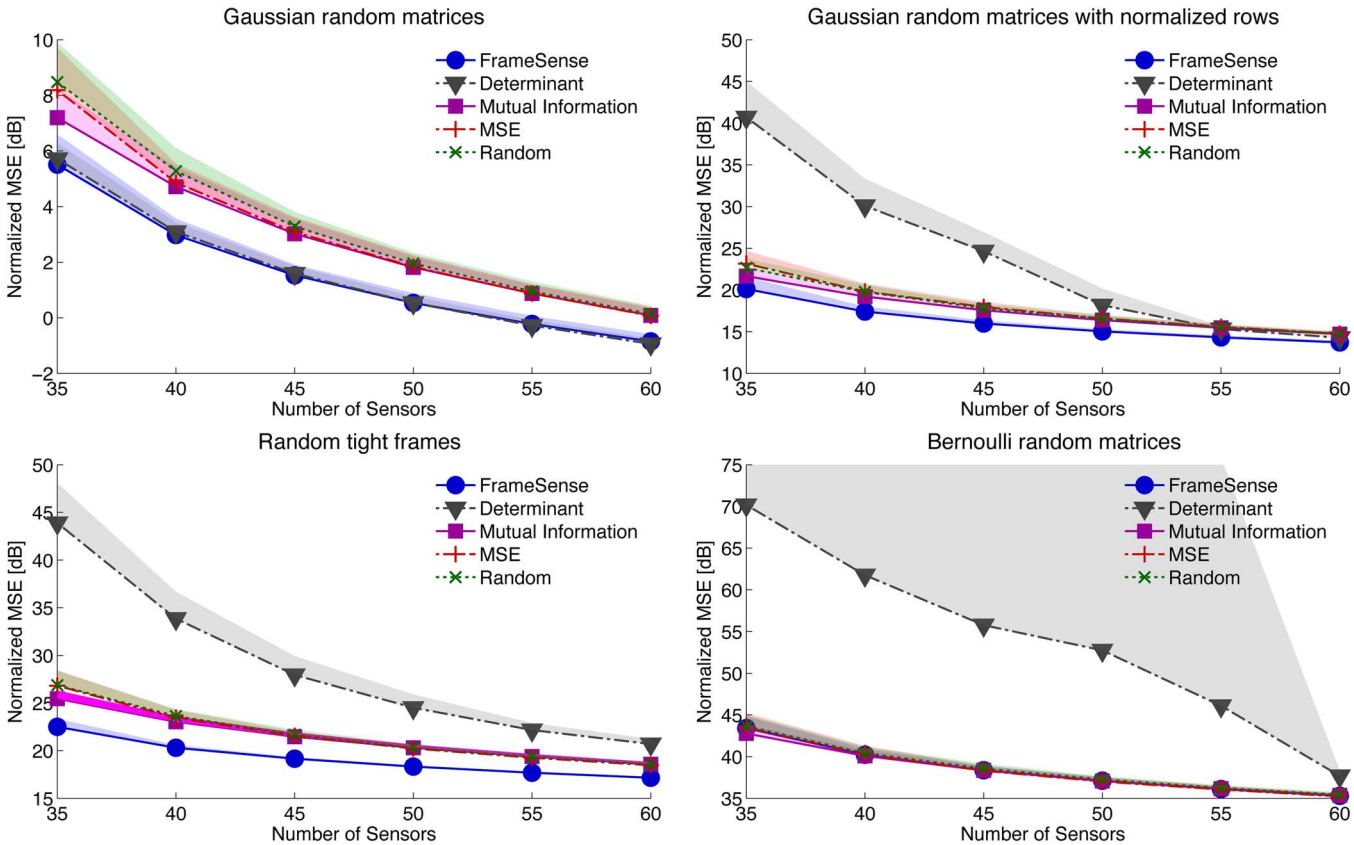


Fig. 2. Performance comparison between FrameSense and other greedy algorithms using common cost functions. We randomly generate matrices with $N = 100$ and $K = 30$ and test different greedy algorithms for a varying number of placed sensors L. The performance is measured in terms of MSE, so the lower the curve, the higher the performance. The shaded areas represent the positive side of the error bars, measured using the standard deviation over 100 realizations. We consider four different types of sensing matrices, and in all cases FrameSense outperforms the other algorithms. We underline the consistency of FrameSense over the four types of matrices.

For example, if $c_1 = 0.25$ and $c_2 = 6$, then we have $\gamma \approx 14$ and $\eta \approx 50$.

IV. Numerical Results

In this section, we analyze the performance of FrameSense and compare it with state-of-the-art algorithms for sensor placement.

A. Synthetic Data

First, we compare the FP with other cost functions when used in a naive greedy algorithm. Among the ones listed in Section I-B, we select the following three cost functions: mutual information [14], determinant of $\boldsymbol{T}_{\mathcal{L}}$ [5], and MSE [15]. We also consider an algorithm that randomly places the sensors to relate the obtained results to a random selection.

The greedy algorithms are tested on different types of sensing matrices $\boldsymbol{\Psi}$:

- random matrices with Gaussian i.i.d. entries,
- random matrices with Gaussian i.i.d. entries whose rows are normalized,
- random matrices with Gaussian i.i.d. entries with orthonormalized columns, that we call random tight frame due to the Naimark theorem [30],
- random matrices with Bernoulli i.i.d entries.

Note that the use of random matrices is sub-optimal, since we would rarely encounter such a case in a real-world scenario. However, it is the only available *dataset* that allows us to test thoroughly the different algorithms.

We consider $\boldsymbol{\Psi} \in \mathbb{R}^{100\times 30}$ and evaluate the performance in terms of MSE for $L = \{35, 35, 40, 45, 50, 55, 60\}$. We use 100 different instances for each combination, and we compute the average MSE as a function of L. Note that the MSE is always computed using (3), which assumes i.i.d. Gaussian noise perturbing the measurements $\boldsymbol{f}$ and a uniform distribution on $\boldsymbol{\alpha}$. The relatively small size of $\boldsymbol{\Psi}$ and the low number of trials are due to the lack of scalability of certain cost functions, which require the computation of large matrices. The results are given in Fig. 2. We note that FrameSense is consistently outperforming all other cost functions. In the random Gaussian matrices case, the determinant shows similar results. However, looking at the Bernoulli matrices case, we see that the determinant leads to a significantly worse MSE. Note that certain cost functions show worse performance than a random selection of the rows. While this phenomena could be partially explained by the special properties of certain families of random matrices, it indicates the importance of choosing a well-studied cost function for which we can obtain performance bounds with respect to the MSE.

According to the theory of random matrices, any selection of rows should have the same spectrum on expectation. Therefore, it should not be possible to outperform the random selection of rows. However, the theoretical analysis of random matrices is generally valid only asymptotically, while here we show results

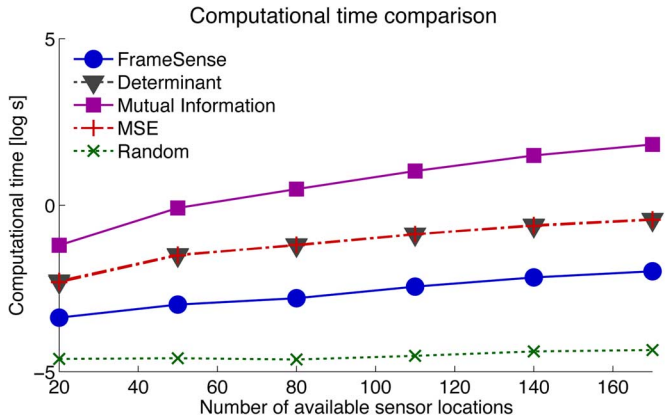


Fig. 3. Comparison in terms of computational time as a function of N between FrameSense and other greedy algorithms using commonly considered cost functions. The lower the curve the faster the algorithm. Note how FrameSense is the fastest algorithm, the exception being the random selection. Moreover, observe how the different algorithms scales equivalently with N. Errors bars are not shown because they are smaller than the markers.

for relatively small matrices. This phenomenon could also be of interest for other domains, such as compressed sensing, indicating the possibility of *optimizing* the spectrum of random matrices to improve their performance.

In Fig. 3, we show the average computational time with regards to the number of possible sensor locations $N = \{20, 50, 80, \ldots, 200\}$. We consider 100 random Gaussian matrices $\boldsymbol{\Psi} \in \mathbb{R}^{N \times 10}$ and we place $0.5N$ sensors. We underline that FrameSense is significantly faster than any other greedy algorithm, the only exception being the random selection that has a computational time close to zero. Note that the other parameters, such as L and K, have little influence on the computational time, which strongly depends on N.

In a second experiment, we compare FrameSense with a state-of-the-art method based on convex optimization [6]. Since the algorithm proposed by Joshi *et al.* [6] is structurally different from FrameSense, we focus this analysis on two parameters: the computational time and the MSE. We fix $K = 20$ and the ratio between the number of sensors and the number of available locations as $L/N = 0.5$. Then, we vary the number of possible locations as $N = \{50, 100, 150, 200, 300, 400, 500\}$. The results for Gaussian random matrices are given in Fig. 5. First, we note that the convex method achieves a lower MSE, however the performance gap decreases when we increase the number of sensors. This is not surprising, since FrameSense is a greedy algorithm that does not require parameter fine-tuning nor heuristics, while the convex relaxation method integrates some efficient heuristics. For example, at every iteration, it refines the selection by looking at all the possible *swaps* between the chosen locations and the discarded ones. This strategy is particularly effective when $L \approx K$: in fact, swapping just one row can improve significantly the spectrum of $\boldsymbol{T}_{\mathcal{L}}$, and consequently the MSE achieved by $\boldsymbol{\Psi}_{\mathcal{L}}$. The heuristics, while effective in terms of MSE, increase also the computational cost. In fact, FrameSense is significantly faster.

The last comparison also opens an interesting direction for future work. In fact, the convex relaxed algorithm optimizes the determinant of $\boldsymbol{T}_{\mathcal{L}}$ and its near-optimality in terms of MSE has not been shown. Moreover, this cost function has been proven to be less effective compared to FP when used in a greedy algorithm (see Fig. 2). Therefore, we expect that a convex relaxed scheme based on the FP has the potential to define a new state of the art, mixing the advantages of FP and the heuristics proposed in [6].

To conclude the performance analysis, we study the trade-off between computational complexity and performance for all the considered algorithms, greedy and not. We picked 100 instances of each of the random matrices proposed in the first experiment with $N = 100$, $L = 40$ and $K = 30$. We measured the average computational time and average MSE obtained by each algorithm and the results are given in Fig. 4. We note a general trend connecting the four subfigures: FrameSense is the fastest algorithm, by at least an order of magnitude, while its performance is just second, as previously shown, to the convex relaxed method proposed by Joshi *et al.* [6].

Since the sensor placement is an off-line procedure, we may argue that the computational time is of secondary importance. While this is true in many applications, there are certain applications where it is necessary to recompute $\mathcal{L}$ regularly. This is usually the case when $\boldsymbol{\Psi}$ changes in time due to changes of the physical field and it is possible to adaptively reallocate the sensors. In other applications, such as the ones where we attempt to interpolate the entire field from L measurements, the number of possible locations N grows with the desired resolution. In this case, a lower computational time is of critical importance.

B. Temperature Estimation on Many-Core Processors

We now analyze the impact of FrameSense on a real-world problem where sensor placement is of fundamental importance. We describe the problem, followed up by a simulation showing the improvement with respect to the state of the art.

The continuous evolution of process technology increases the performance of processors by including more cores memories and complex interconnection fabrics on a single chip. Unfortunately, a higher density of components increases the power densities and amplifies the concerns for thermal issues. In particular, it is key to design many-core processors that prevent hot spots and large on-chip temperature gradients, as both conditions severely affect the overall system's characteristics. A non-exhaustive list of problems induced by thermal stress includes higher failure rate, reduced performance, increased power consumption due to current leakage and increased cooling costs. To overcome these issues, the latest designs include the thermal information into the workload allocation strategy to obtain optimal performance while avoiding critical thermal scenarios. Consequently, a few sensors are deployed on the chip to collect thermal data. However, their number is limited by area/power constraints and their optimal placement, that detects all the worst-case thermal scenarios, is still unresolved and has received significant attention [8], [32]–[35].

An improved sensor placement algorithm would lead to a reduction of the sensing cost in terms of used silicon surface. Moreover, it implies a reduction of the reconstruction error, making possible the use of more aggressive scheduling

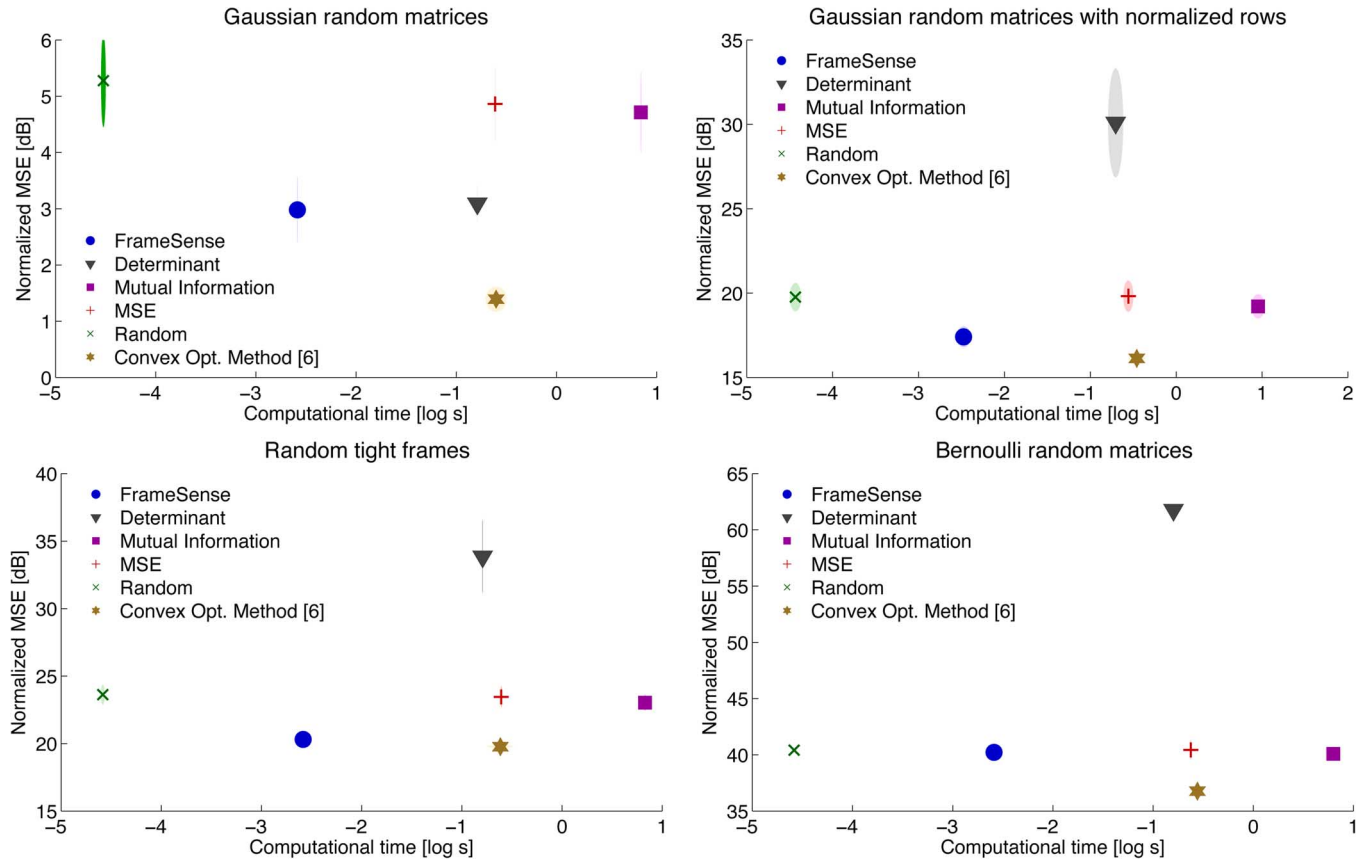


Fig. 4. Tradeoff between computational time and MSE. We randomly generate matrices—according to four different models—with $N = 100$, $K = 30$ and $L = 50$ sensors. The performance is measured in terms of MSE, therefore the lower the dot, the higher the performance. On the other hand, the computational time is measured in seconds. The error bars are represented as ellipsoids, where the length of the axes represent the standard deviation of the data point. Note that, FrameSense is the fastest algorithm by one order of magnitude and it is the second best algorithm in terms of MSE. When the legend refer to a cost function, we considered a naive greedy algorithm optimizing that cost function. On the other hand, we refer to the authors for specific algorithms implementing complicated schemes and/or heuristics.

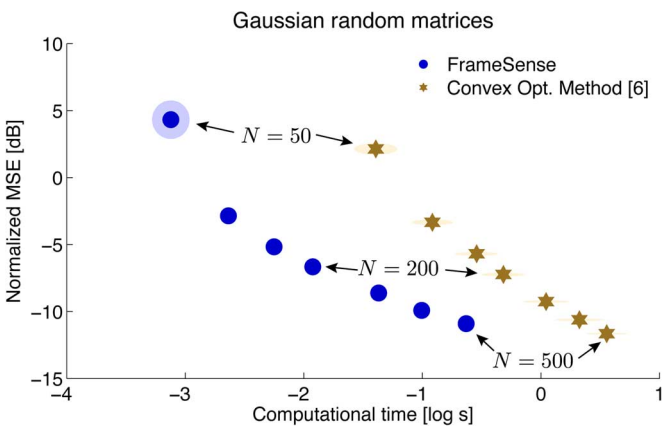


Fig. 5. Analysis of the tradeoff between computational time and MSE for FrameSense and the convex relaxed algorithm proposed by Joshi *et al.* [6]. We generate 100 Gaussian matrices with $K = 20$ and of increasing size $N = \{50, 100, 150, 200, 300, 400, 500\}$, while we place $L = 0.5N$ sensors. The error bars are represented as ellipsoids, where the length of the axes represent the standard deviation of the data point. We measure the average computational time together with the average MSE, showing that while FrameSense is significantly faster than the convex algorithm, the difference in MSE is minimal. Moreover, the gap in the quality of the solution decreases for an increasing size of the problem N.

strategies and consequently improves the processor performance. In [31], we proposed the following strategy: learn $\mathbf{\Psi}$ using a principal component analysis on an extensive set of simulated thermal maps, then place the sensors with a greedy algorithm minimizing the coherence between the rows. This work achieved a significant improvement compared to the previous state-of-the-art approach [32] that was based on specific knowledge of the spectrum of the thermal maps.

We consider a sensing matrix $\mathbf{\Psi}$ equivalent to one defined in [31] but at a lower resolution for computational reasons. More precisely, we consider thermal maps at a resolution of 16×15 and a matrix $\mathbf{\Psi} \in \{420 \times 16\}$. Then, we compare FrameSense, our previous greedy algorithm from [31] and the other algorithms considered in Fig. 2. The results are shown in Fig. 6. We note that the performance of the placement algorithm has been further improved by FrameSense, without increasing the computational cost or changing the reconstruction strategy. Moreover, we are now able to guarantee the near-optimality of the algorithm with respect to the MSE of the estimated thermal map.

Note that we are only discussing the impact of an optimized sensor placement when we consider the $\mathbf{\Psi}$ based on the Eigenmaps described in [31]. The joint problem of sensor placement and reconstruction of thermal maps is more complex and other factors may play a fundamental role. For example, it may be more convenient to use an opportunely constructed DCT frame [32] to reduce the memory occupation at the price of a reduced reconstruction precision. The comparison of the reconstruction performance in terms of MSE for the various algorithms when

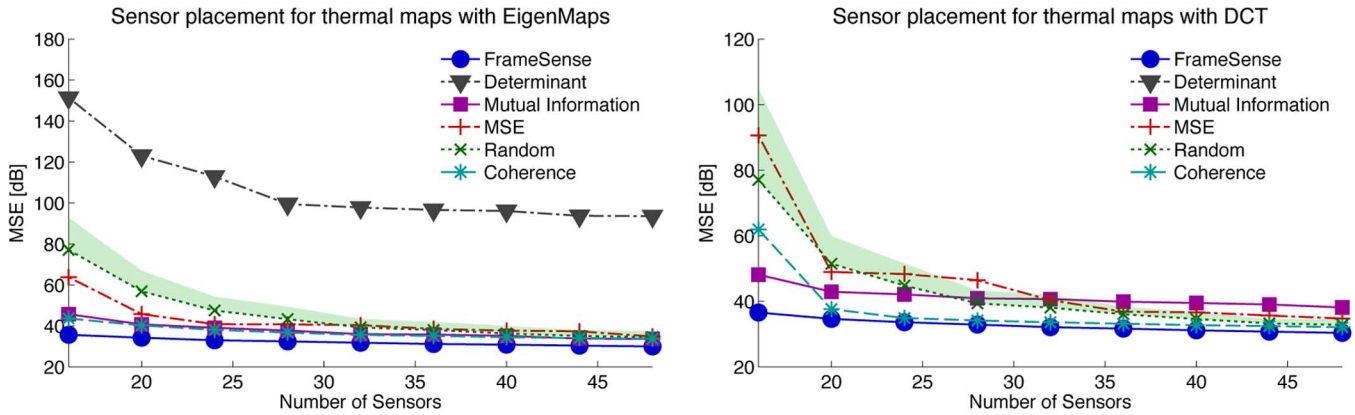


Fig. 6. Comparison between FrameSense, the other greedy algorithms and a coherence-based greedy algorithm proposed in [31]. In this experiment, we consider the sensors placement to estimate the temperature of an 8-core microprocessor using a limited number of sensors. Two different matrices are proposed: on the left, the matrix $\Psi \in \mathbb{R}^{420\times 16}$ is generated from a principal component analysis of known thermal maps as in [31]; on the right, the matrix $\Psi \in \mathbb{R}^{420\times 16}$ is the subsampled DCT matrix proposed in [32]. The shaded area represents the positive side of the error bar for the random sensor placement, measured using the standard deviation over 100 realizations. Note that FrameSense significantly outperforms the previous coherence-based method and the other greedy algorithms, in particular when the number of sensors is close to the number of estimated parameters, that is $K = 16$.

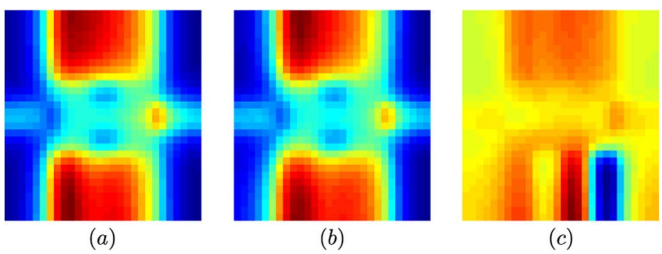


Fig. 7. Examples of reconstruction of a thermal map of resolution 56×60 using two different sensor placements. The real thermal map is depicted in (a), while the reconstructed thermal maps using FrameSense and the algorithm proposed in [31] are given in (b) and (c), respectively. $L = 16$ sensors are placed to measure $K = 16$ parameters and the measurements are corrupted by noise with an SNR of 8.5 dB. Note that the reconstruction using FrameSense is close to the real map, while the other one is extremely noisy. The details regarding the generation of the thermal maps are available in [31].

considering the $\mathbf{\Psi} \in \mathbb{R}^{420\times 16}$ based on the DCT frame are given in Fig. 6. Again, FrameSense outperforms all the other placement algorithms, in particular when the number of sensors is limited.

In Fig. 7, we present an example of a thermal map and the two reconstructions obtained by sensor placements optimized by either FrameSense or the coherence-based algorithm. Notably, the reconstruction obtained with the sensor placement proposed by FrameSense is significantly more precise.

V. Conclusions

We studied the optimization of sensor placement when the collected measurements are used to solve a linear inverse problem. The problem is equivalent to choosing L rows out of N from a matrix $\mathbf{\Psi}$ such that the resulting matrix has favorable spectral properties. The problem is intrinsically combinatorial and approximation algorithms are necessary for real-world scenarios. While many algorithms have been proposed, none has guaranteed performance in terms of the MSE of the solution of the inverse problem, which is the key merit figure.

We proposed FrameSense, a greedy worst-out algorithm minimizing the FP. Even if this chosen cost function is well-known in frame theory for its fundamental role in the construction of frames with optimal MSE, FrameSense is the first algorithm exploiting it as a cost function for the sensor placement problem. Our theoretical analysis demonstrates the following innovative aspects:

- FrameSense is near-optimal with respect to the FP, meaning that it always places the sensors such that the obtained FP is guaranteed to be close to the optimal one.
- under RIP-like assumptions for $\mathbf{\Psi}$, FrameSense is also near-optimal with respect to the MSE. Note that FrameSense is the first algorithm with this important property.

We provided extensive numerical experiments showing that FrameSense achieves the best performance in terms of MSE while having the lowest computational complexity when compared to many other greedy algorithms. FrameSense is also competitive performance-wise with a state-of-the-art algorithm based on a convex relaxation proposed in [6], while having a substantially smaller computational time.

We showed that FrameSense has appealing performance on a real-world scenario, the reconstruction of thermal maps of many-core processors. We showed a potential for reducing the number of sensors required to estimate precisely the thermal distribution on a chip, reducing the occupied area and the consumed power by the sensors.

Future work will be three-fold. First, it is foreseeable to relax the RIP-like condition on $\mathbf{\Psi}$ by considering that the characteristics of the FP are potentially sufficient to avoid $\mathbf{\Psi}_{\mathcal{L}}$ matrices with an unfavorable spectral distribution. Moreover, it would be interesting to show that there exists matrices, random or deterministic, that are (δ, L)-bounded frames. Second, we believe that a convex relaxed scheme based on the FP integrating the heuristics proposed by Joshi *et al.* [6] could improve significantly the MSE of the obtained solution, while keeping the near-optimality thanks to the FP. Third, we would like to derive a new cost function that exploits the sensing energy of the rows, as described in Section III-D.

APPENDIX

A. Proof of Lemma 2

In this section, we bound the MSE of a matrix $\mathbf{\Psi}_{\mathcal{A}} \in \mathbb{R}^{L \times K}$ as a function of its FP and the spectrum $\{\lambda_i\}_{i=1}^{K}$ of $\boldsymbol{T}_{\mathcal{L}}$.

First, consider the harmonic mean $\mathrm{H} = \frac{K}{\sum_k \frac{1}{\lambda_j}}$, the arithmetic mean $\mathrm{A} = \frac{\sum_k \lambda_k}{K}$ and the standard deviation $\mathrm{S} = \sqrt{\frac{1}{K}\sum_k (\lambda_k - A)^2}$ of the eigenvalues of $\boldsymbol{T}_{\mathcal{L}}$. All these quantities are linked to $\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{A}})$, the number of sensors L and $\mathrm{FP}(\mathbf{\Psi}_{\mathcal{A}})$. More precisely, we have

$$\mathrm{H} = \frac{K}{\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{A}})},$$
$$\mathrm{A} = \frac{L_{\mathcal{A}}}{K},$$
$$\mathrm{S} = \sqrt{\frac{1}{K}\left(\mathrm{FP}(\mathbf{\Psi}_{\mathcal{A}}) - \frac{{L_{\mathcal{A}}}^2}{K}\right)}.$$

Then, we consider the following bounds for the harmonic mean of a set of positive numbers derived by Sharma [26],

$$\frac{(M-\mathrm{S})^2}{M(M-2\mathrm{S})} \leq \frac{\mathrm{A}}{\mathrm{H}} \leq \frac{(m+\mathrm{S})^2}{m(m+2\mathrm{S})},$$

where m and M are the smallest and the largest number in the set. We use the expressions of A and H and we remove the mixed term in the denominator to obtain,

$$\frac{K^2}{L_{\mathcal{A}}}\left(1+\frac{\mathrm{S}^2}{M^2}\right) \leq \mathrm{MSE}(\mathbf{\Psi}_{\mathcal{A}}) \leq \frac{K^2}{L_{\mathcal{A}}}\left(1+\frac{\mathrm{S}^2}{m^2}\right).$$

As expected when the FP achieves its global minima, that is $\mathrm{S}=0$, we achieve the optimal MSE of a tight frame.

To conclude the proof, we consider the two bounds separately starting from the lower one. Let $M = \lambda_1$ and we plug in the value of S. We also consider without loss of generality $L_{\mathcal{A}} \leq L_{\mathrm{MAX}}$, since we can always improve the MSE by increasing the sensing power $L_{\mathcal{A}}$. Then,

$$\mathrm{MSE}(\mathbf{\Psi}_{\mathcal{L}}) \geq \frac{K^2}{L_{\mathrm{MAX}}}\left(1+\frac{\mathrm{FP}(\mathbf{\Psi}_{\mathcal{A}})}{K{\lambda_1}^2} - \frac{{L_{\mathrm{MAX}}}^2}{K^2{\lambda_1}^2}\right).$$

We obtain the final result by using lower bound on the largest eigenvalue: $\lambda_1 \geq \frac{L_{\mathrm{MAX}}}{K}$.

The approach to prove the upper bound is exactly symmetrical. Specifically, consider $m = \lambda_N$, $L_{\mathcal{A}} \geq L_{\mathrm{MIN}}$ and use the upper bound on the smallest eigenvalue $\lambda_N \leq \frac{L_{\mathrm{MAX}}}{K}$.

ACKNOWLEDGMENT

The authors would like to thank Ivan Dokmanić and Prof. Ola Svensson whose suggestions were fundamental to strengthen the results described in the paper.

REFERENCES

[1] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[2] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[3] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[4] M. Fickus, D. G. Mixon, and M. J. Poteet, "Frame completions for optimally robust reconstruction," *arXiv*, Jul. 2011 [Online]. Available: http://arxiv.org/pdf/1107.1912.pdf

[5] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. 49th IEEE Conf. Decision Control (CDC)*, Dec. 15–17, 2010, pp. 2572–2577.

[6] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[7] M. Al-Obaidy, A. Ayesh, and A. F. Sheta, "Optimizing the communication distance of an *ad hoc* wireless sensor networks by genetic algorithms," *Artif. Intell. Rev.*, vol. 29, no. 3–4, 2008.

[8] R. Mukherjee and S. Memik, "Systematic temperature sensor allocation and placement for microprocessors," in *Proc. Design Autom. Conf. (DAC)*, 2006.

[9] S. Lau, R. Eichardt, L. Di Rienzo, and J. Haueisen, "Tabu search optimization of magnetic sensor systems for magnetocardiography," *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 1442–1445, 2008.

[10] P. L. Chiu and F. Y. S. Lin, "A simulated annealing algorithm to support the sensor placement for target location," in *Proc. Can. Conf. Electr. Comput. Eng.*, 2004.

[11] D. J. C. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.

[12] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2004.

[13] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Math. Prog.*, vol. 14, pp. 265–294, 1978.

[14] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, 2008.

[15] A. Das and D. Kempe, "Algorithms for subset selection in linear regression," in *Proc. ACM Symp. Theory Comput. (STOC)*, 2009.

[16] M. Naeem, S. Xue, and D. C. Lee, "Cross-Entropy optimization for sensor selection problems," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, 2009.

[17] A. Das and D. Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2011.

[18] D. Golovin, M. Faulkner, and A. Krause, "Online distributed sensor selection," in *Proc. Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2010.

[19] N. Ramakrishnan, C. Bailey-Kellogg, S. Tadepalli, and V. N. Pandey, "Gaussian processes for active data mining of spatial aggregates," in *Proc. Int. Conf. Data Min. (SDM)*, 2005.

[20] D. M. Steinberg and W. G. Hunter, "Experimental design: Review and comment," *Technometrics*, vol. 26, no. 2, pp. 71–97, 1984.

[21] V. K. Goyal, M. Vetterli, and N. Thao, "Quantized overcomplete expansions in IRn: Analysis, synthesis, and algorithms," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 16–31, 1998.

[22] J. Benedetto and M. Fickus, "Finite normalized tight frames," *Adv. Comput. Math.*, vol. 18, no. 2, pp. 357–385, 2003.

[23] P. G. Casazza, M. Fickus, J. Kovačević, M. Leon, and J. Tremain, "A physical interpretation of tight frames," in *Harmonic Analysis and Applications*. New York, NY, USA: Springer, 2006, pp. 51–76.

[24] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, 1994.

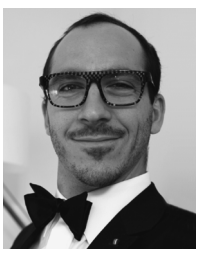
[25] J. Kovacevic and A. Chebira, "Life beyond bases: The advent of frames (Part II)," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 115–125, 2007.

[26] R. Sharma, "Some more inequalities for arithmetic mean, harmonic mean and variance," *J. Math. Inequal.*, vol. 2, pp. 109–114, 2008.

[27] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[28] H. Monajemi, S. Jafarpour, M. Gavish, D. L. Donoho, S. Ambikasaran, S. Bacallado, D. Bharadia, Y. Chen, Y. Choi, and M. Chowdhury, "Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 4, pp. 1181–1186, 2013.

[29] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24–39, 2013.
[30] J. Kovačević and A. Chebira, "Life beyond Bases: The advent of frames (Part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104, 2007.
[31] J. Ranieri, A. Vincenzi, A. Chebira, D. Atienza, and M. Vetterli, "EigenMaps: Algorithms for optimal thermal maps extraction and sensor placement on multicore processors," in *Proc. Design Autom. Conf. (DAC)*, 2012.
[32] A. N. Nowroz, R. Cochran, and S. Reda, "Thermal monitoring of real processors: Techniques for sensor allocation and full characterization.," in *Proc. Design Autom. Conf. (DAC)*, 2010.
[33] R. Cochran and S. Reda, "Spectral techniques for high-resolution thermal characterization with limited sensor data," in *Proc. Design Autom. Conf. (DAC)*, 2009.
[34] S. Reda, R. Cochran, and A. N. Nowroz, "Improved thermal tracking for processors using hard and soft sensor allocation techniques," *IEEE Trans. Comput.*, vol. 60, no. 6, pp. 841–851, 2011.
[35] S. Sharifi and T. Š. Rosing, "Accurate direct and indirect on-chip temperature sensing for efficient dynamic thermal management," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 29, no. 10, pp. 1586–1599, 2010.



Juri Ranieri (S'11) received both his M.S. and B.S. degree in Electronic Engineering in 2009 and 2007, respectively, from Universitá di Bologna, Italy. From July to December 2009, he joined as a visiting student the Audiovisual Communications Laboratory (LCAV) at EPF Lausanne, Switzerland. From January 2010 to August 2010, he was with IBM Zurich to investigate the lithographic process as a signal processing problem. From September 2010, he is in the doctoral school at EPFL where he joined LCAV under the supervision of Prof. Martin Vetterli and Dr. Amina Chebira. From April 2013 to July 2013, he was an intern at Lyric Labs of Analog Devices, Cambridge, USA. His main research interests are inverse problems of physical fields and the spectral factorization of autocorrelation functions.

Amina Chebira (M'08) is a senior research and development engineer at the Swiss Center for Electronics and Microtechnology (CSEM) in Neuchâtel, Switzerland. In 1998, she obtained a Bachelor degree in mathematics from University Paris 7 Denis Diderot. She received the B.S. and M.S. degrees in communication systems from the Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2003 and the Ph.D. degree from the Biomedical Engineering Department, Carnegie Mellon University, Pittsburgh, PA, in 2008, for which she received the biomedical engineering research award. She then held a Postdoctoral Researcher position with the Audiovisual Communications Laboratory, EPFL, from 2008 to 2012. Her research interests include frame theory and design, biomedical signal and image processing, pattern recognition, filterbanks and multiresolution theory.

Martin Vetterli (S'86–M'86–SM'90–F'95) was born in 1957 and grew up near Neuchatel. He received the Dipl. El.-Ing. degree from Eidgenossische Technische Hochschule (ETHZ), Zurich, in 1981, the Master of Science degree from Stanford University in 1982, and the Doctorat es Sciences degree from the Ecole Polytechnique Federale, Lausanne, in 1986. After his dissertation, he was an Assistant and then Associate Professor in Electrical Engineering at Columbia University in New York, and in 1993, he became an Associate and then Full Professor at the Department of Electrical Engineering and Computer Sciences at the University of California at Berkeley. In 1995, he joined the EPFL as a Full Professor. He held several positions at EPFL, including Chair of Communication Systems and founding director of the National Competence Center in Research on Mobile Information and Communication systems (NCCR-MICS). From 2004 to 2011 he was Vice President of EPFL and from March 2011 to December 2012, he was the Dean of the School of Computer and Communications Sciences. Since January 2013, he leads the Swiss National Science Foundation. He works in the areas of electrical engineering, computer sciences and applied mathematics. His work covers wavelet theory and applications, image and video compression, self-organized communications systems and sensor networks, as well as fast algorithms, and has led to about 150 journals papers. He is the co-author of three textbooks, with J. Kovacevic, *Wavelets and Subband Coding* (PrenticeHall, 1995), with P. Prandoni, *Signal Processing for Communications*, (CRC Press, 2008) and with J. Kovacevic and V. Goyal, of the forthcoming book *Fourier and Wavelet Signal Processing* (2012). His research resulted also in about two dozen patents that led to technology transfers to high-tech companies and the creation of several start-ups. His work won him numerous prizes, like best paper awards from EURASIP in 1984 and of the IEEE Signal Processing Society in 1991, 1996 and 2006, the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE Signal Processing Technical Achievement Award in 2001 and the IEEE Signal Processing Society Award in 2010. He is a Fellow of IEEE, of ACM and EURASIP, was a member of the Swiss Council on Science and Technology (2000–2004), and is an ISI highly cited researcher in engineering.