# MULTIPLE WAVELENGTH SENSING ARRAY DESIGN

*Gal Shulkind, Stefanie Jegelka* and Gregory W. Wornell†*

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA, USA
{shulkind,stefje,gww}@mit.edu

## ABSTRACT

We design finite antenna arrays for far-field sensing at multiple wavelengths, under two design paradigms. The first design paradigm is optimized for collection of measurements at multiple wavelengths, fusing these together for joint inference over an underlying scene. The second design paradigm is robust, in a sense that it is guaranteed to allow good inference over the scene at any one single wavelength at a time. We quantify inference quality via the D-Bayes optimality criterion and limit the design space by restricting the number of allowed sensors and the positions where these can be placed.
We show that the resulting combinatorial optimization problems are instances of problems in a class known to have efficient guaranteed approximation algorithms, namely submodular optimization problems, and showcase the design of arrays under both paradigms utilizing simple greedy selection algorithms, and state-of-the-art robust submodular maximization algorithms.

***Index Terms—*** Multiple Wavelength, Far Field, Array Design, Submodular Optimization, Robust Optimization

## 1. INTRODUCTION

Sensor arrays for spatial sensing are deployed in a wide range of applications including radar, sonar, medical imaging and radio astronomy and there is a vast literature on the topics of array design and array processing from the last century [1, 2]. In this paper we focus on designing sensor arrays for multiple wavelength sensing applications. Arrays operating at multiple wavelengths have been studied in various contexts such as wideband direction of arrival estimation problems [3, 4], multi-frequency synthesis in Astronomy [5], and designing wideband array patterns [6].
A major goal in designing arrays is efficiently meeting required specifications with a limited budget of sensing elements, which are often a main determinant of system cost. However, even in the single wavelength case, the design of array geometries is a notoriously hard task, and many applications simply utilize a uniform truncated half-wavelength design, or restrictions thereof. Indeed, the problem of designing non-uniform arrays hints at combinatorial optimization and is computationally hard as we discuss later.
Various studies tackle the problem of non-unifom array design directly. In beamforming arrays methods like array thinning [7] have been put to the test. Other approaches consider methods such as swarm optimization [8], dynamic programming [9], genetic algorithms [10], inversion [11] and Bayesian compressive sampling [12].
In applications of estimating direction of arrival, other specialized techniques have emerged for finding efficient array designs such as optimizing corresponding error bounds [13], or designing nested topologies [14].
In this paper we study the related problem of inference on a scene of interest via measurements collected at a sensor array. The approach we take for designing array geometries is novel in that we consider settings where some prior on the scene is available and propose adapting the geometry accordingly to achieve efficient inference exploiting this belief. In this Bayesian setting, sensing the scene is just performing inference in the model, and the problem of array geometry design asks to select a geometry that optimizes the quality of inference.
We quantify the quality of inference through the D-Bayes optimality criterion [15, 16] and consider two design paradigms. The first design paradigm is optimized for collection of measurements at multiple wavelengths, fusing these together for joint inference over an underlying scene. The second design paradigm is robust, in a sense that it is guaranteed to allow good inference over the scene at any one single wavelength at a time.
We show that the resulting combinatorial optimization problems possess the property of submodularity [17]. Recently, there has been significant progress on the theory of optimizing submodular functions [18, 19, 20, 21, 22]. In particular, these results state that, while in general NP-hard, these problems admit variants of greedy algorithms that are guaranteed to achieve near-optimal solutions, i.e., within a constant factor. Our connections and formulations open avenues for leveraging those results for efficient array design with strong guarantees. We demonstrate this by showcasing the design of array geometries under both paradigms in settings with arbitrary apertures. Together with our new formulations, the exploitation of prior knowledge leads to higher quality inference at lower cost in terms of the number of sensing elements.

## 2. PROBLEM FORMULATION

In this section we formulate the multiple wavelength array design problem for far field sensing applications, and frame sensing as Bayesian inference. To simplify the exposition, we focus on the one-dimensional case.

### 2.1. Far-Field Sensing

The far field sensing setup is depicted in Fig. 1. A scene of interest is located far from an observation axis $x$. The scene is characterized through an illumination function $\tilde{\beta}(\theta)$, with $|\theta| \leq \frac{\pi}{2}$ the angle between the direction of observation and the broadside.

*This work was supported, in part, by NSF CAREER Award 1553284.
†This work was supported, in part, by DARPA under Contract No. HR0011-16-C-0030, and by NSF under Grant No. CCF-1319828.

We are to place $Q$ sensors along the observation axis at positions $S \equiv \{x_0, \ldots, x_{Q-1}\}$, and collect measurements at wavelengths $\Lambda \equiv \{\lambda_0, \ldots, \lambda_{L-1}\}$. We aim to choose an optimal set $S$ satisfying some constraints. We take $A$ to be a finite selection set of possible positions (e.g. a finite grid on some section of the real line), and pose the constraint $S \subseteq A$.
The noiseless reading $r(x_q;\lambda_l)$ taken at position $x{=}x_q$ and wavelength $\lambda{=}\lambda_l$ is given according to [1]:

$$r(x_q;\lambda_l) = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \tilde{\beta}(\theta) e^{j\frac{2\pi}{\lambda_l} x_q \sin\theta} \cos\theta d\theta = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \beta(\psi) e^{j\frac{4\pi}{\lambda_l} x_q \psi} d\psi \tag{1}$$

where $\psi \equiv \frac{\sin\theta}{2}$ such that $|\psi| \leq \frac{1}{2}$, $\beta(\psi) \equiv \tilde{\beta}(\sin^{-1}(2\psi))$, and we assume that the illumination function is wavelength-invariant such that the scene appears identical when probed at different wavelengths $\lambda$. Finally, we consider the effect of noise by introducing

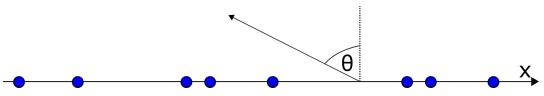


**Fig. 1**: Far-field sensing with sensing elements depicted as points.

$\tilde{\boldsymbol{f}} = [\tilde{f}_0, \ldots, \tilde{f}_{N-1}]^T$, a vector of $N \equiv QL$ noisy measurements modeled according to:

$$\tilde{f}_{q+Ql} \equiv r(x_q;\lambda_l) + w_{q+Ql} \qquad \begin{matrix} q=0,\ldots,Q-1 \\ l=0,\ldots,L-1 \end{matrix} \tag{2}$$

where $w_n$ is additive noise. Stacked in $\boldsymbol{w} = [w_0, \ldots, w_{N-1}]^T$, we assume throughout that the noise is complex, circular, Gaussian $\boldsymbol{w} \sim \mathcal{CN}(0, \boldsymbol{\Sigma}_{ww})$ [23], i.i.d. across measurements, i.e., $\boldsymbol{\Sigma}_{ww} = \sigma_w^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is an $N \times N$ identity matrix.

### 2.2. The Sensing Setup

The sensing problem we consider in this paper entails the estimation of the illumination function $\beta(\psi)$ from the set of noisy measurements $\tilde{\boldsymbol{f}}$. Even in the noiseless setting this problem is gravely ill-posed as infinitely many wildly varying scenes map to any given finite set of observations[1].
To cope with this ill-posedness, some prior belief must hence be incorporated into the model. In what follows, we take a Bayesian approach and impose a prior on the scene $\beta(\psi)$. Subsequently, sensing is equivalent to performing inference in this model. The prior may be assigned based on past observations over scenes or based on a-priori knowledge of scene properties as we discuss next.

#### *2.2.1. Discrete Representation*

Assigning a prior on $\beta(\psi)$ can be simplified if $\beta(\psi)$ may be expanded in a countable basis of functions such that the prior is imposed in the discrete domain of expansion coefficients. With $\beta(\psi)$ having constrained support in $|\psi| \leq \frac{1}{2}$, we can expand it by means of Fourier basis functions $\left\{e^{j2\pi m\psi} | m \in \mathbb{Z}\right\}$ in that domain [25]:

$$\beta(\psi) = \sum_m \beta_m e^{j2\pi m\psi}, \qquad \beta_m \equiv \int_{-\frac{1}{2}}^{+\frac{1}{2}} \beta(\psi) e^{-j2\pi m\psi} d\psi \tag{3}$$

[1]The mapping between $\beta(\psi)$ and a finite set of its Fourier transform samples $r(x_q;\lambda_l)$ is not bijective [24].

where $\{\beta_m\}$ are the Fourier expansion coefficients.
In lieu of the prior on $\beta(\psi)$ we impose a prior over $\{\beta_m\}$. This description is especially suited for applications involving smooth functions $\beta(\psi)$ as suggested by the following property of the Fourier series expansion [26]:

**Lemma 2.1.** *Let $\beta(\psi) \in C^r$ where $C^r$ is the space of $r$-times continuously differentiable functions over some domain. Then $|\beta_m| \leq \frac{\alpha}{|m|^r}$ with $\alpha = \sup_\psi |\frac{\partial^r}{\partial \psi^r} \beta(\psi)|$.*

Thus, for any nicely behaved $\beta(\psi)$ the high frequency Fourier expansion coefficients diminish polynomially to zero, allowing good approximate representation through a finite subset of low frequency coefficients, which in the sequel we designate via the finite vector $\boldsymbol{\beta}$. We defer a full account of the quality of this approximation to an extended future publication.
Next, impose i.i.d. Gaussian priors $\beta_m \sim \mathcal{CN}(0, \sigma_m^2)$, where $\sigma_m^2$ is set following some initial measurements of sample functions $\beta(\psi)$ or taking into account prior knowledge. For example if we have a-priori knowledge that $\beta(\psi) \in C^r$ we may set, following Lemma 2.1:

$$\sigma_m^2 \propto m^{-2r} \tag{4}$$

which represents some least informative Gaussian prior respecting polynomial variance decay. For the sequel we adopt the prior in (4) and take $r{=}1$ to promote continuously differentiable functions.

#### *2.2.2. Observation Model*

With the prior stated in the discrete $\boldsymbol{\beta}$ domain as described above, our next goal is to circumvent $\beta(\psi)$, directly stating the problem in terms of the measurements $\tilde{f}_n$ and the coefficients $\beta_m$, replacing the continuous representation with a discrete counterpart. Substituting (3) into (1) we have (where $n = q + Ql$):

$$r(x_q;\lambda_l) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \sum_m \beta_m e^{j2\pi m\psi} e^{j\frac{4\pi}{\lambda_l} x_q \psi} d\psi = \sum_m K_{nm} \beta_m \tag{5}$$

$$K_{nm} \equiv \int_{-\frac{1}{2}}^{+\frac{1}{2}} e^{j2\pi(\frac{2}{\lambda_l} x_q + m)\psi} d\psi = \mathrm{sinc}(m + \frac{2}{\lambda_l} x_q) \tag{6}$$

Plugging this into (2) we retrieve the observation model

$$\tilde{f}_n = \sum\nolimits_m K_{nm} \beta_m + w_n \qquad n = 0, \ldots, N-1 \tag{7}$$

and the sensing problem amounts to estimating the coefficients $\boldsymbol{\beta}$ given the observation vector $\tilde{\boldsymbol{f}}$. As we have assumed Gaussian distributions throughout, the posterior $\mathbb{P}(\boldsymbol{\beta}|\tilde{\boldsymbol{f}})$ is Gaussian with a convenient analytic form.

### 2.3. Cost Function

Our next step is to specify a cost function $G(S)$, used for comparisons between different designs, and for choosing the best one. Notice through (6) that the array design determines the coefficients $K_{nm}$ through the set of positions $S$. With $S$ fixed, the sensing problem amounts to performing inference leading to the posterior $\mathbb{P}(\boldsymbol{\beta}|\tilde{\boldsymbol{f}})$. The problem of determining the quality of inference has been extensively studied in the statistical literature in the context of experimental design [15, 16]. Here we adopt the Bayes D-optimality criterion whereby the quality of inference between random measurements and a hidden random variables is given by the mutual information between the two. In our setting this amounts to

$$G(S) \equiv I(\tilde{\boldsymbol{f}}_S; \boldsymbol{\beta}) \equiv H(\boldsymbol{\beta}) - H(\boldsymbol{\beta}|\tilde{\boldsymbol{f}}_S) \tag{8}$$

Where $I(\cdot;\cdot)$ is the mutual information and $H(\cdot)$ the Shannon entropy. The subscript $S$ explicitly emphasizes the dependence of the measurements on the set of sensor positions $S$. Notice that maximizing $G(S)$ as a function of $S$ can be equivalently viewed as maximizing the mutual information between $\tilde{\boldsymbol{f}}_S$ and $\boldsymbol{\beta}$, or alternatively as minimizing the uncertainty (entropy) in $\boldsymbol{\beta}$ given $\tilde{\boldsymbol{f}}_S$. Further notice that evaluation of $G(S)$ for our model (7) is easy. As we have noted all the relevant distributions are Gaussian such that evaluation of (8) may be accomplished by using

$$X \in \mathbb{R}^k, X \sim \mathcal{N}(0, \boldsymbol{\Sigma}) \quad \Rightarrow \quad H(X) = \log((\pi e)^k \det\boldsymbol{\Sigma}) \quad (9)$$

## 3. ARRAY DESIGN PARADIGMS

In this section we define two multiple wavelength array design paradigms, concisely formulated as combinatorial optimization problems.

The **fusion** problem entails designing an array $S$ for collection of measurements at a set of fixed wavelengths $\Lambda$. The full set of measurements (taken at each location in all wavelengths in $\Lambda$) is jointly used to infer $\boldsymbol{\beta}$. The set $S$ is constrained to be in $A$ and to be of size no more than $Q$:

$$S^\star = \text{argmax}_{S:S\subseteq A,|S|\leq Q}\, I(\tilde{\boldsymbol{f}}_S; \boldsymbol{\beta}). \quad (10)$$

The **robust** problem entails designing an array $S$ for collection of measurements at a single wavelength $\lambda\in\Lambda$, which is fixed but unknown. Thus we are interested in guaranteeing good quality inference for any $\lambda\in\Lambda$ and solve the robust optimization problem of maximizing the worst-case performance achieved when operating the array at any single wavelength. Concretely, let $\tilde{\boldsymbol{f}}_S^\lambda$ be the set of samples collected at a single wavelength $\lambda$ at the set of positions $S$, which satisfies the same constraints as before. The design criterion is:

$$S^\star = \text{argmax}_{S:S\subseteq A,|S|\leq Q}\, \min_{\lambda\in\Lambda} I(\tilde{\boldsymbol{f}}_S^\lambda; \boldsymbol{\beta}) \quad (11)$$

## 4. SUBMODULAR OPTIMIZATION

In this section we prescribe efficient algorithms for the solution of (10) and (11). We show that due to the structure of the cost function $G(S)$ an efficient approximation algorithm is known to exist with strong theoretical guarantees. We survey relevant results and adapt them to our needs, beginning with the following properties of set functions [17]:

**Definition 4.1.** *Let $G: 2^A \to \mathbb{R}$ be a set function.*
*(a) G is **monotonic** (increasing) if $\forall S\subseteq T\subseteq\mathcal{V}$ we have $G(S)\leq G(T)$.*
*(b) G is **submodular** if it exhibits decreasing marginals: $\forall S\subseteq T\subseteq\mathcal{V}$, and $x\in\mathcal{V}\backslash T$ it holds that $G(S\cup\{x\})-G(S) \geq G(T\cup\{x\})-G(T)$.*

As it turns out, the cost function (8) is monotonic and submodular as the next result shows (this is similar to corollary 4 in [27]):

**Theorem 4.1.** *Let $A$ be defined as before, and define the set function $G: 2^A \to \mathbb{R}$ according to $G(S) = I(\tilde{\boldsymbol{f}}_S; \boldsymbol{\beta})$. Then $G$ is submodular and monotonic (increasing).*

*Proof.* We just sketch the proof here. Monotonicity is evident from the fact that adding additional elements to $S$ cannot increase the posterior uncertainty $H(\boldsymbol{\beta}|\tilde{\boldsymbol{f}}_S)$ and thus cannot decrease $G(S)$ (per (8)). Submodularity may be proved by using generic conditioning properties of entropy. □

### 4.1. Submodular Maximization

Using Theorem 4.1 we recognize (10) as a maximization of a monotonic submodular function under a cardinality constraint. The following fundamental lemma shows that the computationally efficient greedy Algorithm 1 approximately solves it to within the best possible approximation factor:

**Algorithm 1** Greedy Submodular Maximization

```
function S=GREEDYMAX(G(·), A, Q)
    S ← ∅
    for i = 1 to Q do S ← S ∪ {argmax_{x∈A\S} G(S ∪ {x})}
```

**Lemma 4.1.** *(Nemhauser [17]) Let $G$ be a monotonic, submodular set function. Let $S^\star$=argmax$_{S\subseteq A,|S|\leq Q} G(S)$ be an optimal solutions, and $S^{gr}$ a solution retrieved by the greedy Algorithm 1. We have the following approximation guarantee: $G(S^{gr}) \geq (1-\frac{1}{e})G(S^\star)$, and no polynomial time algorithm can provide a better guarantee.*

Algorithm 1 runs in time $O(|A|Q)$, linear in the size of the set $A$ and the number of selected elements $Q$ [28]. However, more efficient variants of the algorithm have been suggested and analyzed. The 'lazy greedy' variant which was introduced in [28] was shown to offer substantial running-time improvements in practice. Our numerical experiments in Section 5 implement this more efficient variant to reduce running time.

### 4.2. Robust Submodular Maximization

Let $\{G_i(S)\}$ be as set of monotone, submodular set functions and consider the robust optimization problem

$$S^\star = \text{argmax}_{S\subseteq A,|S|\leq Q}\, \min_i G_i(S) \quad (12)$$

With Theorem 4.1 and defining $G_i(S)\equiv I(\tilde{\boldsymbol{f}}_S^{\lambda_i}; \boldsymbol{\beta})$ we immediately recognize (11) as an instance of (12).

It is known that no polynomial time algorithm approximating the solution of (12) exists [22]. However, the following lemma suggests that for integer-valued $G_i(S)$, Algorithm 3 is guaranteed to achieve approximately optimal solution:

**Lemma 4.2.** *(Krause [22]). For any integer Q, SAT (delineated in Algorithm 3) finds a solution $S^{sat}$ such that*

$$\min_i G_i(S^{sat}) \geq \min_i G_i(S^\star) \quad \textit{and} \quad |S^{sat}| \leq \alpha Q \quad (13)$$

$$\textit{where} \quad \alpha \equiv 1 + \log(\max_{s\in A} \sum\nolimits_i G_i(s)) \quad (14)$$

**Algorithm 2** Greedy submodular partial cover

```
function S=GPC(Ḡ_c(S), c)
    S ← ∅
    while Ḡ_c(S) < c do
        S ← S ∪ argmax_{s∈A} {Ḡ_c(S ∪ {s}) − Ḡ_c(S)}
```

Lemma 4.2 guarantees that SAT will find an approximate optimal set $S^{\text{sat}}$ achieving performance at least as good as the true optimal set $S^\star$, at a cost of using as many as $\alpha Q$ elements of the set $A$ in lieu of the $Q$ elements included in $S^\star$.

The extension of Lemma 4.2 to non integer-valued functions is discussed in [22] (section 7). One approach is porting these problems into integer-valued ones by scaling and rounding $G_i(S)$. However, this requires careful manipulations of the guarantees in Lemma

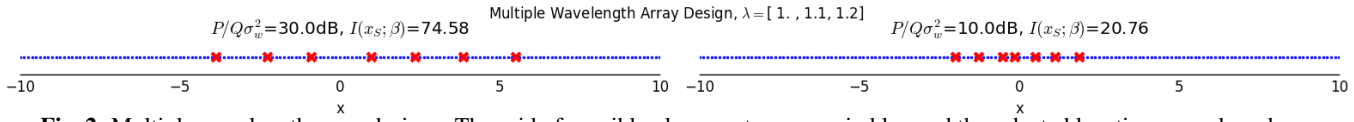


**Fig. 2**: Multiple wavelength array designs. The grid of possible placements appears in blue and the selected locations as red markers.

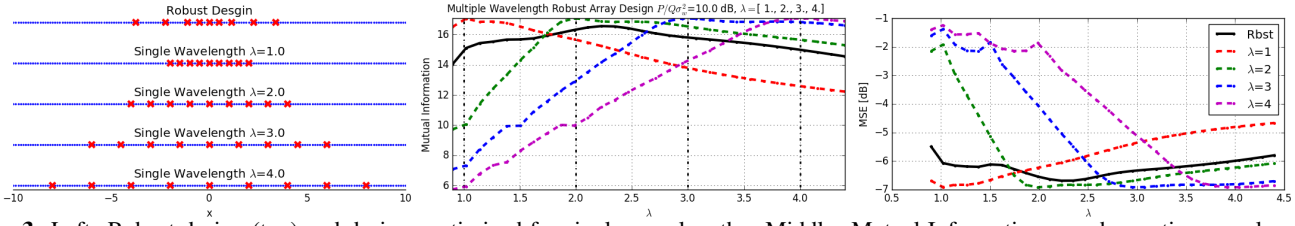


**Fig. 3**: Left: Robust design (top) and designs optimized for single wavelengths. Middle: Mutual Information vs. observation wavelength. Right: Reconstruction MSE for Monte-Carlo experiment vs. observation wavelength.

**Algorithm 3** Submodular saturation algorithm

**function** $S_{\text{BEST}}$=SAT($G_1, \ldots, G_m, A, k, \alpha$)
  $c_{\min} \leftarrow 0$ ; $c_{\max} \leftarrow \min_i G_i(A)$ ; $S_{\text{best}} \leftarrow \emptyset$
  **while** $(c_{\max} - c_{\min}) \geq \frac{1}{m}$ **do**
    $c \leftarrow (c_{\min} + c_{\max})/2$
    $\hat{S} \leftarrow GPC(\frac{1}{m}\sum_i \min\{G_i(S), c\}, c)$
    **if** $|\hat{S}| > \alpha k$ **then** $c_{\max} \leftarrow c$
    **else** $c_{\min} \leftarrow c$ ; $S_{\text{best}} \leftarrow \hat{S}$

4.2. Instead, [22] follows an empirical approach making the ad-hoc choice $\alpha$=1 for the implementation of Algorithm 3, and keeping the non integer-valued $G_i(S)$ unchanged. Based on extensive numerical experiments it is empirically shown that under these conditions SAT performs favorably. We follow this approach in our numerical experiments described in Section 5, and empirically verify the usefulness of the above choice when applied to our non integer-valued problem.

## 5. NUMERICAL EXPERIMENTS

We perform numerical experiments to exemplify array design under the paradigms of Section 3 and the algorithms in Section 4.

### 5.1. Multiple Wavelength Array Design

First, we design arrays for the fusion setting as per (10) using the lazy greedy submodular optimization algorithm of Section 4.1. We take $\Lambda=\{1, 1.1, 1.2\}$, the sensor position selection set $A$ is a uniformly spaced grid of 161 positions in $|x|\leq 10$, and we design an array consisting of $Q$=7 locations. $\boldsymbol{\beta}$ is formed by approximating $\{\beta_m\}$ via the 101 lowest frequency coefficients[2], and the prior for $\beta_m$ is set as per (4) with $r$=1. We normalize $\{\sigma_m^2\}$ for unit average scene power $P$ using Parseval:

$$P \equiv \mathbb{E}\int |\beta(\psi)|^2 d\psi = \mathbb{E}\sum\nolimits_m |\beta_m|^2 = \sum\nolimits_m \sigma_m^2 = 1 \qquad (15)$$

The results are summarized in Fig. 2, where blue markers denote the selection set $A$ and red markers delineate the chosen set $S$. The left subplot depicts a design for high Signal to Noise Ratio (SNR) (defined as $\frac{P}{Q\sigma_w^2}$) of 30dB. The right subplot repeats the experiment at a lower SNR of 10dB. This design tends to limit the spread of the sampling positions as samples become less reliable and there is value in limiting sampling diversity for the sake of concentrating more samples in valuable regions.

The performance in terms of mutual information $I(\tilde{\boldsymbol{f}}_S; \boldsymbol{\beta})$ for the selected locations $S$ appears in the plot title. Notice for example that for the 10dB SNR design, the achieved mutual information is 20.76. Using Lemma 4.1 we have that the optimal design cannot achieve mutual information better than $(1-\frac{1}{e})^{-1}20.76 = 32.84$.

### 5.2. Robust Single Wavelength array design

Next, we design arrays for the robust setting as per (11) using the SAT algorithm of Section 4.2. We take $A$ as before, $\Lambda=\{1, 2, 3, 4\}$, the number of elements is $Q$=9 and we assume a SNR of 10dB. In Fig. 3 (left) we plot several designs. The top configuration is the robust design generated via Algorithm 3, fixing $\alpha$=1. The bottom four configurations depict arrays each optimized for a single wavelength. These were generated by applying the greedy design scheme of Section 4.1 with a single measurement wavelength from the set $\Lambda$. The figure shows that for a single observation wavelength, we obtain configurations that are generic truncated uniform $\frac{\lambda}{2}$ arrays [1]. However, when considering observations across multiple possible wavelengths, as is done for the robust design, the resulting configuration is no longer uniform, but consists of a mixture of large and small inter-element spacings, to cater to all possibilities.

Fig. 3 (middle) plots the performance of these arrays in terms of the corresponding mutual information $I(\tilde{\boldsymbol{f}}_S; \boldsymbol{\beta})$ when the actual wavelength at which measurements are collected is swept in $0.9\leq\lambda\leq 4.4$. Each of the four single-wavelength arrays (dashed lines) maximizes the mutual information when operated at the wavelength for which it was designed, as expected. However, at mismatched wavelengths performance deteriorates. In contrast, the robust array (solid line) does not perform as well as the specialized single wavelength arrays at their target wavelengths. But, while those specialized designs are very sensitive to misspecified wavelengths, the robust design flexibly performs well across the entire range of wavelengths.

Fig. 3 (right) summarizes a Monte-Carlo experiment set up to empirically evaluate Mean Square Error (MSE) performance in scene reconstruction using our robust design. We have drawn 2000 scenes distributed as prescribed in Section 2.2.1, and collected corresponding noisy measurements at various wavelengths using the robust and the four single wavelength optimized arrays. We repeatedly performed maximum likelihood estimation of the expansion coefficients $\boldsymbol{\beta}$, and synthesized an estimated scene $\hat{\beta}(\psi)$ according to (3). The MSE discrepancy between $\hat{\beta}(\psi)$ and the true scene was recorded. Evidently, the mutual information performance of Fig. 3 (middle) is indicative of MSE performance, with the robust design exhibiting best worst-case results.

[2]We empirically find that refining the sampling grid, or including more Fourier coefficients does not significantly change the design.

## 6. REFERENCES


[1] Harry L Van Trees, *Detection, estimation, and modulation theory, optimum array processing*, John Wiley & Sons, 2004.

[2] Hamid Krim and Mats Viberg, "Two decades of array signal processing research: the parametric approach," *Signal Processing Magazine, IEEE*, vol. 13, no. 4, pp. 67–94, 1996.

[3] H Wang and M Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 4, pp. 823–831, 1985.

[4] M Wax, T.J. Shan, and T Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE transactions on acoustics, speech, and signal processing*, vol. 32, no. 4, pp. 817–827, 1984.

[5] RJ Sault and MH Wieringa, "Multi-frequency synthesis techniques in radio interferometric imaging.," *Astronomy and Astrophysics Supplement Series*, vol. 108, pp. 585–594, 1994.

[6] Dan R Scholnik and Jefrey O Coleman, "Optimal design of wideband array patterns," in *Radar Conference, 2000. The Record of the IEEE 2000 International*. IEEE, 2000, pp. 172–177.

[7] Randy L Haupt, "Thinned arrays using genetic algorithms," *Antennas and Propagation, IEEE Transactions on*, vol. 42, no. 7, pp. 993–999, 1994.

[8] Majid M Khodier and Christos G Christodoulou, "Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization," *Antennas and Propagation, IEEE Transactions on*, vol. 53, no. 8, pp. 2674–2679, 2005.

[9] M Skolnik, G Nemhauser, and J Sherman, "Dynamic programming applied to unequally spaced arrays," *IEEE Transactions on Antennas and Propagation*, vol. 12, no. 1, pp. 35–43, 1964.

[10] Randy L Haupt and Douglas H Werner, *Genetic algorithms in electromagnetics*, John Wiley & Sons, 2007.

[11] B Preetham Kumar and GR Branner, "Design of unequally spaced arrays for performance improvement," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 3, pp. 511–523, 1999.

[12] Giacomo Oliveri, Matteo Carlin, and Andrea Massa, "Complex-weight sparse linear array synthesis by bayesian compressive sampling," *Antennas and Propagation, IEEE Transactions on*, vol. 60, no. 5, pp. 2309–2326, 2012.

[13] Houcem Gazzah and Sylvie Marcos, "Cramer-rao bounds for antenna array design," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 336–345, 2006.

[14] Palghat P Vaidyanathan and Piya Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[15] Kathryn Chaloner and Isabella Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, pp. 273–304, 1995.

[16] José M Bernardo, "Expected information as expected utility," *The Annals of Statistics*, pp. 686–690, 1979.

[17] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[18] Satoru Fujishige, *Submodular functions and optimization*, vol. 58, Elsevier, 2005.

[19] Niv Buchbinder, Moran Feldman, Joseph Seffi, and Roy Schwartz, "A tight linear time (1/2)-approximation for unconstrained submodular maximization," *SIAM Journal on Computing*, vol. 44, no. 5, pp. 1384–1402, 2015.

[20] Jan Vondrák, "Symmetry and approximability of submodular maximization problems," *SIAM Journal on Computing*, vol. 42, no. 1, pp. 265–304, 2013.

[21] Baharan Mirzasoleiman, Amin Karbasi, Rik Sarkar, and Andreas Krause, "Distributed submodular maximization: Identifying representative elements in massive data," in *Advances in Neural Information Processing Systems*, 2013, pp. 2049–2057.

[22] Andreas Krause, H Brendan McMahan, Carlos Guestrin, and Anupam Gupta, "Robust submodular observation selection," *Journal of Machine Learning Research*, vol. 9, pp. 2761–2801, 2008.

[23] Fredy D Neeser and James L Massey, "Proper complex random processes with applications to information theory," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1293–1302, 1993.

[24] Jui L Yen, "On nonuniform sampling of bandwidth-limited signals," *Circuit Theory, IRE Transactions on*, vol. 3, no. 4, pp. 251–257, 1956.

[25] Alan V Oppenheim, Alan S Willsky, and Syed Hamid Nawab, *Signals and systems*, vol. 2, Prentice-Hall Englewood Cliffs, NJ, 1983.

[26] Gero Friesecke, "Lecture notes in fourier analysis [MA4064]," 2013.

[27] Andreas Krause and Carlos E Guestrin, "Near-optimal nonmyopic value of information in graphical models," *arXiv preprint arXiv:1207.1394*, 2012.

[28] Michel Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," in *Optimization Techniques*, pp. 234–243. Springer, 1978.