# Coding Mask Design for Single Sensor Ultrasound Imaging

van der Meulen, Pim; Kruizinga, Pieter; Bosch, Johannes G.; Leus, Geert

**DOI**
10.1109/TCI.2019.2948729

**Publication date**
2020

**Document Version**
Final published version

**Published in**
IEEE Transactions on Computational Imaging

# Coding Mask Design for Single Sensor Ultrasound Imaging

Pim van der Meulen, Pieter Kruizinga, Johannes G. Bosch, and Geert Leus

***Abstract*—We study the design of a coding mask for pulse-echo ultrasound imaging. We are interested in the scenario of a single receiving transducer with an aberrating layer, or 'mask,' in front of the transducer's receive surface, with a separate co-located transmit transducer. The mask encodes spatial measurements into a single output signal, containing more information about a reflector's position than a transducer without a mask. The amount of information in such measurements is dependent on the mask geometry, which we propose to optimize using an image reconstruction mean square error (MSE) criterion. We approximate the physics involved to define a linear measurement model, which we use to find an expression for the image error covariance matrix. By discretizing the mask surface and defining a discrete number of mask thickness levels per point on its surface, we show how finding the best mask can be posed as a variation of a sensor selection problem. We propose a convex relaxation in combination with randomized rounding, as well as a greedy optimization algorithm to solve this problem. We show empirically that both algorithms come close to the global optimum. Our simulations further show that the optimized masks have better a MSE than nearly all randomly shaped masks. We observe that an optimized mask amplifies echoes coming from within the region of interest (ROI), and strongly reduces the correlation between echoes of pixels within the ROI.**

***Index Terms*—Coded aperture, compressed sensing, experimental design, sparse sensing, ultrasound imaging.**

## I. Introduction

In RECENT years there has been a considerable interest in reducing the amount of measured sensor data in the ultrasound imaging field. Data compression has great potential, for example, in minimally invasive surgery, where the number of data cables is limited by the available space inside the catheter. This prohibits transferring all data from a dense imaging array through the catheter, making it very challenging to design high frame-rate, high quality imaging devices for this scenario. One possible approach to avoid using a Nyquist-sampled sensor array (either in time or space), is to have the data somehow compressed during sensing. That is, to reduce the amount of sampled data without any complicated electronic hardware actively compressing the measured array signals.

Recently, the field of compressed sensing (CS) showed that such compression is possible if the signal to be estimated has a sparse representation in a known domain [1]–[4]. This led to a large number of studies in ultrasound using CS in an attempt to get better image reconstructions using less data (see, for example [5]–[11], amongst others). In compressive sensing, it is assumed that the compressed data is a linear combination of the original, uncompressed, signals. Moreover, there are probabilistic performance guarantees if the compression is implemented by a *random* linear combination of the original signals. However, implementing such a random linear combination in hardware is not necessarily easy, since these compressors might have to combine samples across both space and time, resulting in large sample buffers. Another popular compression technique is randomly subsampling in time and/or space. However, this means that a large amount of information is not used at all (the ignored samples are not in any way present in the compressed measurement), and additionally results in a degradation of SNR. More importantly, we are currently not aware of a domain where ultrasound images are very sparse.

Some of the most striking examples of CS for imaging have been demonstrated in the field of optical imaging, the most well-known being the single-pixel camera [12]. There, random linear combinations of the image are obtained by spatial light modulators (SLM), which basically integrate a different subselection of the image pixels per measurement. This has led to a great number of works using SLM setups for compressed imaging for a single sensor ([13]–[15], to name a few). Instead of subsampling or implementing a (random) linear compressor in hardware, one could also place a contrasting medium in front of the sensor array. In optics, for example, researchers have placed a heterogeneous medium between the sensor array and the imaging scene, causing multiple scatterings inside this heterogeneous medium, after which the scrambled light field is measured [16]–[19]. This scrambled light field can be seen as a random linear combination of all image pixels, and CS can be used for image reconstruction. The random compression thus occurs naturally, and is achieved without any electronic hardware. Similar techniques (not employing CS) have also been used in the medical ultrasound community, by transmitting

Manuscript received February 17, 2019; revised July 3, 2019; accepted August 26, 2019. Date of publication October 21, 2019; date of current version January 13, 2020. This work is part of the OTP research programme ASPIRE with project number 14926 which is financed by the Netherlands Organisation for Scientific Research (NWO). The associate editor coordinating the review of this manuscript and approving it for publication was J. Portilla. *(Corresponding author: Pim van der Meulen.)*

P. van der Meulen and G. Leus are with the Circuits and Systems Group, Delft University of Technology, Delft 2628, The Netherlands (e-mail: p.q.vandermeulen@tudelft.nl; g.j.t.leus@tudelft.nl).

P. Kruizinga is with the Department of Neuroscience, Erasmus Medical Center, Rotterdam 3015, The Netherlands (e-mail: p.kruizinga@erasmusmc.nl).

J. G. Bosch is with the Department of Biomedical Engineering, Thorax Center, Erasmus Medical Center, Rotterdam 3015, The Netherlands (e-mail: j.bosch@erasmusmc.nl).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TCI.2019.2948729

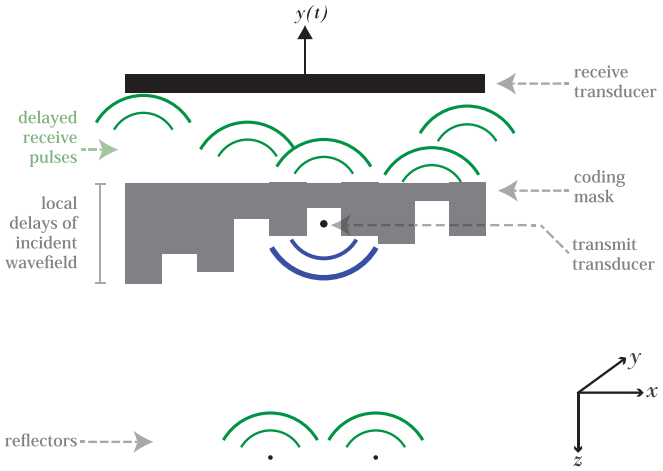


Fig. 1. Experimental setup considered in this article. A point source emits a spherical undistorted wave into the ROI. Reflected echoes propagate through the coding mask before being measured by a single transducer. The coding mask has a different speed of sound than its surrounding medium, and distorts the received wavefield in a DOA-dependent manner. As a result, scatterers can be resolved based on the information residing in the temporal waveforms of the pulse-echo signal.

randomized waves and measuring on a single sensor [20], or by placing a heterogeneous cavity in front of a single sensor [21].

We recently demonstrated a similar approach for 3D ultrasound imaging [22], [23], where we have shown how a single sensor with a 'coding mask' can successfully reconstruct 3D objects from multiple pulse-echo measurements, facilitating the design of new imaging devices with reduced output data. Instead of placing a heterogeneous scattering medium in front of the single sensor, we used a homogenous, but irregularly shaped piece of plastic, to which we refer as a 'coding mask' (Fig. 1). In contrast to CS, we have shown that reconstruction is possible without sparse regularization techniques, removing the necessity of an a priori known sparse domain for the ultrasound images.

The main idea of this imaging approach is the following. Consider the experimental setup shown in Fig. 1, but without a coding mask. Due to the lack of spatial sampling, one has to rely on the information in the temporal dimension of the measured pulse-echo signal. Only the distance to a reflector can be determined (based on an echo's time-of-arrival (TOA)), but nothing can be said about the direction-of-arrival (DOA), since the pulse-echo signals will look roughly the same for all directions.[1] However, once we place a small layer of irregularly shaped plastic in front of the receiving sensor, in which ultrasound waves have a different speed of sound than the imaging medium, the received wavefields are significantly distorted. If the mask shape is designed correctly, this distortion is direction-dependent, and we are now able to infer the DOA in addition to the distance to the object by exploiting the signal structure in the time-domain waveforms. Through collecting additional measurements by rotating or translating the sensor with the coding mask, the image object is probed from multiple points of view.

In our experiments so far, we used randomly shaped masks to remove ambiguities between pixel signals. Naturally, we would like to know if random masks are a good design choice, and if not, how the coding mask should be shaped to get the best possible imaging performance. In this paper, we investigate coding mask design algorithms using sensor selection techniques, and propose a convex and a greedy optimization program. As such, this study falls within the larger framework of (sparse) ultrasound array design [24], [25]. Using various simulations, we demonstrate that the masks obtained using our techniques exhibit better imaging mean square error (MSE) compared to a random mask design.

To design a mask, we will first introduce an approximate model, where the mask is discretized in all dimensions $(x, y, z)$. Based on this model, we want to minimize the imaging MSE, which is a function of the error covariance matrix, and consequently, of the measurement matrix mapping the discretized image domain to the sampled pulse-echo measurement domain. We will show how optimization problems of this form relate to sensor selection problems. These problems are typically solved in literature using convex relaxations employing the $\ell_1$-norm [26], [27], or using more efficient greedy methods by formulating a sub-modular cost-function [28]–[32]. These are relatively well-understood techniques, and various functions of the reconstruction error covariance matrix can be optimized near-optimally using such convex and greedy optimization algorithms. There are some fundamental differences however, between our problem and the typical sensor selection problem.

First, the inverse error covariance matrix of the typical sensor selection problem is $linear$ with respect to the selection variables, whereas for our selection problem it is quadratic, making the cost-function non-convex. We are currently not aware of any sensor selection literature studying the same problem structure. Hence, we propose to linearize the problem by lifting the optimization variables to a higher dimensional space and impose a block-matrix sparsity constraint on the matrix formed by these lifted coefficients.

Second, in our optimization problem we want to select one sensor from a specific set, select another one from another set, etc., for a given number of such sets. This second obstacle is also considered in [33], but we are not aware of any similar studies. In [33], each sensor position is associated with a small pool of candidate sensors, corresponding to different sensor types. However, since the inverse covariance matrix in [33] is still a linear function of the selection variables, we cannot employ similar techniques to solve our problem.

In [34], we made a first attempt at mask optimization. There, we proposed a model similar to the one in this work, but using a different convex relaxation, utilizing $\ell_1$-reweighting [35]. In this paper, we use a different convex relaxation with a smaller feasible set, without $\ell_1$-reweighting, and we also propose a more efficient greedy coordinate-descent optimization scheme. Furthermore, we provide an analysis of the optimized masks, lacking in our previous work.

[1]To be more accurate, the pulse-echo signal for a flat circular sensor will look exactly the same for all reflectors on a circle with the same elevation angle, for all azimuth angles. The azimuth and elevation are with respect to the plane in which the sensor is positioned. However, pulse-echo signals will only change slowly with respect to the elevation angle, so there will be high ambiguity for reflectors with similar DOA.

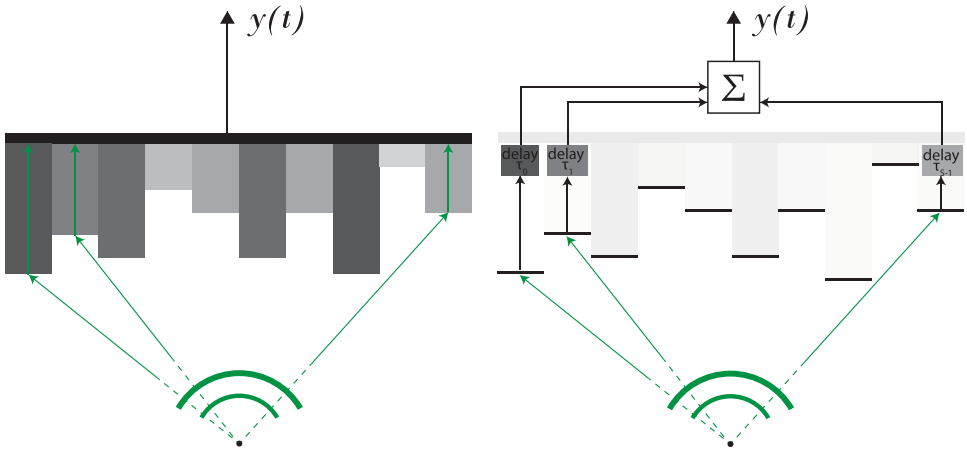


Fig. 2. Left: We approximate the physics of a transducer with a mask by discretizing the transducer surface into separate propagation channels. Right: Approximate signal model. Each channel delays the incident wavefield, and the spatial integration by the transducer surface is approximated by summing all delayed channel signals.

The remainder of this paper is organized as follows. In Section II, we describe an approximate linear measurement model for a masked transducer. In Section III, we show how this model can be parameterized using a discrete selection variable. Using this parameterization, we then formulate the mask optimization problem and propose a convex relaxation, and greedy selection algorithm to find the best mask parameters. Simulation results are discussed in Section IV, where we show various examples to evaluate the performance of the optimized masks as compared to the true optimum and a large set of randomly generated masks. In Section V the resulting masks are discussed, and the paper is concluded in Section VI.

### A. Notation

Throughout this paper we use the following notational conventions. Bold lower-case variables (e.g., $\mathbf{x}$) represent vectors, and bold upper-case variables (e.g., $\mathbf{A}$) represent matrices. The transpose and Hermitian transpose are denoted by $^{\mathsf{T}}$ and $^{\mathsf{H}}$, respectively. We use $\mathbf{W} \succeq \mathbf{0}$ to indicate that $\mathbf{W}$ is positive semi-definite. Alternatively, the set $\mathbb{S}_+$ is the set of positive semi-definite matrices. By $\text{diag}(\mathbf{W})$ we represent the vector formed by taking the coefficients on the diagonal of $\mathbf{W}$, whereas $\text{diag}(\mathbf{w})$ is the diagonal matrix with the vector $\mathbf{w}$ on its diagonal. Next, $\|\mathbf{w}\|_0$, $\|\mathbf{w}\|_1$ and $\|\mathbf{w}\|_2$ denote the cardinality, $\ell_1$-norm, and Euclidean norm of $\mathbf{w}$, respectively. To index matrices or vectors, we use $[\mathbf{W}]_{i,j}$ to represent the $(i,j)$-th entry of $\mathbf{W}$. In the case of a 3D vector representing a $(x,y,z)$-coordinate in Cartesian space, we use $[\mathbf{r}]_z$ to index the $z$-component in $\mathbf{r}$. Finally, $\mathbf{1}$ is the all-one vector, $\mathbf{I}$ denotes the identity matrix, and $\mathbb{E}(\cdot)$ represents the expectation operator. The Gaussian and complex Gaussian distribution with mean $\mu$ and covariance matrix $\mathbf{C}$ are written as $\mathcal{N}(\mu, \mathbf{C})$ and $\mathcal{CN}(\mu, \mathbf{C})$, respectively.

## II. Signal Model

We will assume a pulse-echo setup with separate co-located transmit and receive transducers (Fig. 1), and the coordinate system shown in the same figure. Throughout this paper, we assume that a point source is located at $(x,y,z) = (0,0,0)$, and the masked receive transducer in a plane at $z = 0$. As will be explained at the end of this section, we cannot easily take into account the effect of the mask on the transmitted wave, so we will assume that the transmit transducer emits a spherical wave that is unaffected by the mask.

We first discretize the mask surface in the width and length ($x$ and $y$) dimensions into many small patches, and regard each of these discrete points as an independent sensor *channel*. Each channel delays the reflected echo field according to the thickness of the mask at the channel and the speed of sound inside the mask. Put differently, we assume that every channel takes the received pulse-echo ultrasound field at the mask-medium interface as channel 'input'. The ultrasound field propagates through each channel independently to arrive at the transducer. This approximate model is visualized in Fig. 2.

The basic idea behind the approximate physical model just described is based on ray-tracing physics, and relies on the assumption that the mask is *thin enough* to assume that if a part of the echo wavefront is incident on the mask at position $(x,y)$, it will reach the transducer surface in a sufficiently close neighborhood of $(x,y)$, even with strong wave refraction. Consequently, only a phase change occurs proportional to the mask thickness around $(x,y)$, leading to the independent channels as shown in Fig. 2. Further assuming that scattering sources are not extremely close to the transducer, the incident wave angle will be limited, reducing the effect of refraction. This requires a proper discretization of the mask surface: if the mask patches (and consequently, the patch size on the transducer surface) are taken too large, they do not sufficiently sample the incident wavefield, and will neither accurately approximate the incident wave on the transducer surface, nor the output signal of the transducer. On the other hand, as patches get smaller, it becomes possible that most of the wave starts interfering into other channels, causing the model to lose its validity. To avoid both cases, we will use a patch size of around 0.3 to 0.5 times the wavelength inside the

mask material throughout this work, i.e., wavefields are sampled close to the required spatial sampling (Nyquist) rate of 0.5 $\lambda$, but not strongly exceeding it.

This approximate model has been succesfully applied in the design of holographic masks, where it was experimentally verified that these approaches can be used to design phase masks that generate a desired pressure pattern [36], [37], and we partly base the validity of our model on these studies. We have also compared this model to a true simulated wavefield obtained using the k-wave simulator [38] in the Supplementary Material accompanying this paper.

Next, we will formalize the model laid out in the previous paragraphs into a linear signal model. Assume that the output signal of the transducer, ignoring the electro-mechanical transfer function for now, is the integral of the pressure field over the transducer surface (after propagation through the mask). If we denote the spatial 3D coordinate-vectors that include all points in space on the transducer surface by $X$, then the transducer output signal $y(t)$, given the spatio-temporal pressure field $y(t, \mathbf{r})$ on the receiving transducer surface (after propagating through the mask), is:

$$y(t) = \int_{\mathbf{r} \in X} y(t, \mathbf{r}) d\mathbf{r}, \tag{1}$$

where $\mathbf{r}$ is a 3D spatial vector. For our model, if the sensor surface $X$ is discretized into $S$ channels, this can be approximated by

$$y(t) = \sum_{s=0}^{S-1} y(t, \mathbf{r}_s), \tag{2}$$

where $\{\mathbf{r}_s\}_{s=0}^{S-1}$ are the sampled positions on the transducer surface.

In the remainder of this work we will work in the frequency domain instead of the time domain. Needless to say, equations (1) and (2) are equally valid in the frequency domain if $y(t)$ and $y(t, \mathbf{r}_s)$ are replaced by their frequency domain equivalents. We will denote the sampled continuous time Fourier transform of $y(t)$ and $y(t, \mathbf{r}_s)$ by $y[\omega]$ and $y_s[\omega]$, respectively, with $\omega \in \Omega$, sampled at frequencies in the set $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{N-1}\}$. We will use equidistantly taken samples in the pulse bandwidth of the positive side of the frequency spectrum.

Next, we want to express the pressure fields $y_s[\omega]$ on the transducer surface in terms of the image we try to recover. The pixelized image consists of $M$ pixel reflection coefficients, denoted by the vector $\mathbf{x} \in \mathbb{R}^M$. We assume that the transmitted wavefield to each pixel is known. By using the Born approximation (see e.g. [39], [40]), we can define a measurement model that is linear with respect to the reflection coefficients; any measurement $y_s[\omega]$ is a superposition of the individual pulse-echo signals $\{a_{s,m}[\omega]\}_{m=0}^{M-1}$ of all scatterers, ignoring multiple reflections between scatterers:

$$y_s[\omega] = \sum_{m=0}^{M-1} x[m] a_{s,m}[\omega]. \tag{3}$$

Here, $x[m]$ is the $m$-th component of $\mathbf{x}$, corresponding to the scattering intensity of pixel $m$, and $a_{s,m}[\omega]$ is the pulse-echo signal of pixel $m$ at frequency $\omega$ measured at position $\mathbf{r}_s$.

Using the approximate model described earlier, and illustrated in Fig. 2, we have the following expression for the term $a_{s,m}[\omega]$ in (3):

$$\begin{aligned} a_{s,m}[\omega] = \frac{p[\omega]}{4\pi(\|\mathbf{r}_{px,m}\|_2 + \|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2 + \|[\mathbf{r}_{ch,s}]_z\|_2)} \\ \times \exp\left(-j\frac{\omega}{c_0}(\|\mathbf{r}_{px,m}\|_2 + \|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2) \right. \\ \left. -j\frac{\omega}{c_1}\|[\mathbf{r}_{ch,s}]_z\|_2\right). \end{aligned} \tag{4}$$

Here, $\mathbf{r}_{px,m}$ is the 3D position vector of pixel $m$, and $\mathbf{r}_{ch,s}$ is the 3D position vector of the mask-medium interface of channel $s$. The variables $c_0$ and $c_1$ are the medium and mask speed of sound, respectively. The signal $p[\omega]$ is the frequency domain representation of the excitation pulse $p(t)$. The first term accounts for the geometric spreading of the pressure field, and the exponential term is the phase shift due to the travelling distance from the transmitting transducer to the pixel and back to the masked receive sensor. In other words, it is the total delay of a single path from transmit to the 'output' of a single mask channel, as visualized in Fig. 2. This total path and its corresponding delays can be broken down into several parts, following next. The distance $\|\mathbf{r}_{px,m}\|_2$ is the transmit path length from the transmitter to pixel $m$, and the sum $\|\mathbf{r}_{px,m} - \mathbf{r}_{ch,s}\|_2 + \|[\mathbf{r}_{ch,s}]_z\|_2$ is the total receive path from pixel $m$ to the mask-medium interface of channel $s$, and the echo propagation through the channel. Similarly, the first terms in the exponent yield a phase shift due to the transmit/receive paths in the medium, whereas the last term represents the phase shift due to the propagation through mask channel $s$.

Storing all samples $\{y_s[\omega]\}_{\omega \in \Omega}$ in the vector $\mathbf{y}_s \in \mathbb{C}^N$, and all $\{a_{s,m}[\omega]\}_{\omega \in \Omega}$ in the vectors $\mathbf{a}_{s,m} \in \mathbb{C}^N$, we can use matrix-vector notation to write (3) as:

$$\begin{aligned} \mathbf{y}_s &= [\mathbf{a}_{s,0}\ \mathbf{a}_{s,1} \ldots \mathbf{a}_{s,M-1}]\,\mathbf{x} \\ &= \mathbf{A}_s \mathbf{x}. \end{aligned} \tag{5}$$

Similarly, we can rewrite (2) into its discretized, frequency domain counterpart as:

$$\mathbf{y} = \sum_{s=0}^{S-1} \mathbf{y}_s = \sum_{s=0}^{S-1} \mathbf{A}_s \mathbf{x} = \mathbf{A}\mathbf{x}, \tag{6}$$

with $\mathbf{A} = \sum_{s=0}^{S-1} \mathbf{A}_s$. Assuming that measurements are corrupted by complex Gaussian noise in the frequency domain, our final measurement model becomes:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{7}$$

where $\mathbf{n} \in \mathbb{C}^N$ represents complex Gaussian noise with covariance matrix $\mathbf{C_n}$: $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_n})$. Thus, the imaging problem consists of estimating $\mathbf{x}$ from $\mathbf{y}$. If additional measurements are acquired by, for example, translation of the masked transducer, they can be easily incorporated by adding additional rows to this system of equations (see (10) in the next section). In the next section, we parameterize $\mathbf{A}$ as a function of the mask shape in order to optimize for the mask geometry.

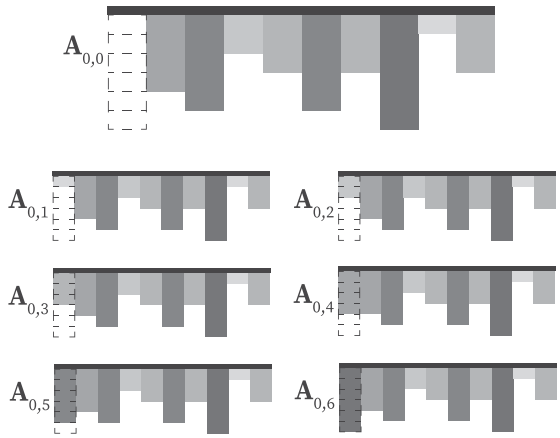


Fig. 3. Each mask channel is discretized in the mask thickness dimension, and each mask thickness level is associated with a specific measurement matrix for that channel. The goal of the mask optimization problem is to select one thickness level per channel such that the imaging MSE is minimized.

In (6), the total measurement matrix $\mathbf{A}$ is separated into independent channels, resulting in a summation of matrices. Our convex mask optimization algorithm in the next section relies on the fact that we can pre-compute each $\mathbf{A}_s$ for a pre-determined number of mask thickness levels. However, it is impossible to pre-compute $\mathbf{A}_s$ if it is not known what pressure field is emitted by each scatterer, which in turn can only be known if the transmit mask is known. Although in [22], [23] we used the same mask in transmit and receive, in this work we assume that a separate (co-located) transducer is used for transmit, and the masked transducer for receive. This makes it always possible to pre-compute any $\mathbf{A}_s$ using (4), circumventing this problem. Specifically, throughout this paper, we will assume that a point source (without a mask), located in the center of the receiver, transmits a spherically diverging wave.

Finally, we point out that the greedy algorithm proposed in this paper does not suffer from this limitation. However, for the sake of consistency and comparison to the convex algorithm, we will use the non-masked co-located transmitter setup throughout this paper for all simulations.

## III. Mask Optimization by Sensor Selection

Having formulated the linear measurement model in the previous subsection, we now want to use it to find how to optimally choose the mask thickness levels, i.e., we want to find the best channel lengths or delays. We will again use discretization (of the channel lengths), and show that channel length design can be posed as a sensor selection problem. Such problems are hard to optimize exactly, but can be either relaxed into a convex problem formulation, or approximately solved using a greedy optimization algorithm. Using a specific thickness level for a channel is equivalent to using a particular matrix $\mathbf{A}_s$ for that channel (Fig. 3). Consequently, our goal is to choose one particular $\mathbf{A}_s$ per mask channel, such that the trace of the error covariance matrix of the mask-encoded measurement system is optimal.

First, we rewrite (5) in terms of the measurement matrices for all thickness levels. Suppose there are $R$ potential thickness levels for each channel $s$. Hence, there are $R$ candidate matrices $\{\mathbf{A}_{s,0}, \mathbf{A}_{s,1}, \ldots, \mathbf{A}_{s,R-1}\}$ that can be used for $\mathbf{A}_s$ at channel $s$. Let $w[r+Rs] \in \{0,1\}$ indicate whether sensor candidate $r \in \{0,1,\ldots,R-1\}$ is used for channel $s$ or not. Then (7) can be rewritten as

$$\mathbf{y}(\mathbf{w}) = \sum_{s=0}^{S-1} \sum_{r=0}^{R-1} w[r+Rs]\mathbf{A}_{s,r}\mathbf{x} \tag{8}$$

where $\mathbf{w} = [w[0], w[1], \ldots, w[RS-1]]^{\mathsf{T}}$. To make our notation simpler, we define $\mathbf{A}[r+Rs] = \mathbf{A}_{s,r}$, and linearize $s$ and $r$ into a single index $t = r + Rs$. Consequently we can write (8) as:

$$\begin{aligned}\mathbf{y}(\mathbf{w}) &= \sum_{t=0}^{RS-1} w[t]\mathbf{A}[t]\mathbf{x} \\ &= \mathbf{A}(\mathbf{w})\mathbf{x},\end{aligned} \tag{9}$$

where $\mathbf{A}(\mathbf{w}) = \sum_{t=0}^{RS-1} w[t]\mathbf{A}[t] = \sum_{s=0}^{S-1}\sum_{r=0}^{R-1} w[r+Rs]\mathbf{A}_{s,r}$. Note that in this formulation only one sensor should be selected per channel, resulting in only $S$ non-zero values for $\mathbf{w}$, and each non-zero value has to be within a specific support range of $\mathbf{w}$. The vector $\mathbf{w}$ is the mask optimization variable, and has to lie within the constraints just described.

In case $K$ different pulse-echo events are taken using the same coding mask, the corresponding measurement equations can be stacked into a single system of equations, assuming $\mathbf{x}$ does not change in between measurements:

$$\mathbf{y}(\mathbf{w}) = \begin{bmatrix} \mathbf{y}_0(\mathbf{w}) \\ \mathbf{y}_1(\mathbf{w}) \\ \vdots \\ \mathbf{y}_{K-1}(\mathbf{w}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0(\mathbf{w}) \\ \mathbf{A}_1(\mathbf{w}) \\ \vdots \\ \mathbf{A}_{K-1}(\mathbf{w}) \end{bmatrix} \mathbf{x} = \mathbf{A}(\mathbf{w})\mathbf{x}. \tag{10}$$

For example, the sensor and the mask can be rotated (as in [22]), or translated (as in [23]) to obtain additional information of $\mathbf{x}$. Note however, that the same mask is used in each case, and that $\mathbf{A}(\mathbf{w})$ is still linear in $\mathbf{w}$.

This model can also be generalized to include multiple measurements involving differently shaped masks. In that case, a number of $Q$ measurements with $Q$ distinct masks are stacked vertically into a larger system of equations:

$$\begin{bmatrix} \mathbf{y}(\mathbf{w}_0) \\ \mathbf{y}(\mathbf{w}_1) \\ \vdots \\ \mathbf{y}(\mathbf{w}_{Q-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\mathbf{w}_0) \\ \mathbf{A}(\mathbf{w}_1) \\ \vdots \\ \mathbf{A}(\mathbf{w}_{Q-1}) \end{bmatrix} \mathbf{x}, \tag{11}$$

which we denote as

$$\mathbf{y}(\mathcal{W}) = \mathbf{A}(\mathcal{W})\mathbf{x},$$

where $\mathcal{W}$ represents the set $\{\mathbf{w}_q\}_{q\in\mathcal{Q}}$, $\mathcal{Q} = \{0,1,\ldots,Q-1\}$. Again, $\mathbf{A}(\mathcal{W})$ is still linear in $\mathcal{W}$. The aforementioned constraint on the number of selected channels means that if $\mathbf{w}_{q,s}$ only contains the entries of $\mathbf{w}_q$ that correspond to the $R$ selection coefficients for channel $s$, i.e., $\mathbf{w}_{q,s} =$

$[\, w_{q,s}[0],\, w_{q,s}[1],\, \ldots,\, w_{q,s}[R-1]\,]^{\mathsf{T}}$, then we want $\|\mathbf{w}_{q,s}\|_0 = 1$ for all channels $s$.

To optimally select mask thickness levels, we use $\hat{\mathbf{x}}$ to denote an estimate of $\mathbf{x}$, and define the image estimation error as $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. We will utilize the covariance matrix of this image estimation error throughout this work, defined as

$$\mathbf{C_e} = \mathbb{E}\left((\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^{\mathsf{H}}\right). \tag{12}$$

We will try to find an optimized mask by minimizing the mean squared error (MSE) under additive zero-mean i.i.d. white Gaussian measurement noise, which is a function of $\mathbf{C_e}$. Since $\mathbf{A}(\mathcal{W})$ is typically ill-conditioned, some prior information about the solution is required. We assume that the covariance matrix $\mathbf{C_x}$ of $\mathbf{x}$ is known, and that a Wiener estimate of the image is used, so that we can minimize the Bayesian MSE (the error covariance associated with the Wiener estimator). More specifically, we minimize

$$f(\mathcal{W}) = \frac{1}{M}\text{trace}(\mathbf{C_e}) \tag{13}$$

$$= \frac{1}{M}\text{trace}\left(\left(\mathbf{C_x^{-1}} + \mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{C_n^{-1}}\mathbf{A}(\mathcal{W})\right)^{-1}\right) \tag{14}$$

where $\mathbf{C_x}$ and $\mathbf{C_n}$ are the image and noise covariance matrices, respectively. Throughout this paper we assume that $\mathbf{C_x} = \sigma_x^2\mathbf{I}$ and $\mathbf{C_n} = \sigma_n^2\mathbf{I}$.

With these choices for the covariance matrices, the Wiener estimator can be rewritten as a regularized least squares estimator, i.e., $\hat{\mathbf{x}} = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y}(\mathcal{W}) - \mathbf{A}(\mathcal{W})\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$, where $\lambda$ depends on the ratio between $\sigma_x^2$ and $\sigma_n^2$. In this case, $\lambda\|\mathbf{x}\|_2^2$ only serves to make sure that a unique solution is found without large outliers in $\hat{\mathbf{x}}$, due to the ill-posedness of the inverse imaging problem, as well as to increase robustness to noise. From this perspective, minimizing $f(\mathcal{W})$ can also be seen as a way to decrease the ill-posedness of the inverse imaging problem, by finding a mask that increases the orthogonality between the columns of $\mathbf{A}(\mathcal{W})$. As a result, we expect that the resulting mask will also be beneficial for other (non-linear) imaging algorithms besides the Wiener estimator used here, since it is expected to decrease the ill-posedness of the problem.

Furthermore, it is expected that the optimal solution of $f(\mathcal{W})$ is not overly sensitive to incorrect values for $\sigma_x^2$ and $\sigma_n^2$, as long as the following two extreme cases are avoided. The first one where the ratio of these variances causes the inverse term in $f(\mathcal{W})$ to be dominated by $\mathbf{C_x^{-1}}$, and the second case where $\mathbf{C_x^{-1}}$ is negligible, so that no regularizing term is present. By choosing these values sufficiently in between these two extreme cases, we hope to find a mask that is relatively robust to the variances of $\mathbf{x}$ and $\mathbf{n}$.

Ideally, the optimization problem is then posed as

$$\begin{aligned} &\min_{\mathcal{W}} f(\mathcal{W}) \\ &\text{s.t.} \quad \|\mathbf{w}_{q,s}\|_0 = 1 \\ &\qquad \mathbf{w}_q \in \{0,1\}^{RS} \end{aligned} \tag{15}$$

for all $q \in \mathcal{Q}$ and $s \in \mathcal{S}$. However, neither the objective function nor the constraints are convex with respect to $\mathcal{W}$, and it is hard to solve this problem exactly. Hence, we will relax (15) into a problem that is convex, and also propose a more efficient coordinate descent type algorithm.

### A. Mask Optimization by Convex Relaxation

To find a convex relaxation of (15), both the cost function and the two constraints in (15) have to be relaxed. We will discuss each of these separately.

1) In the typical sensor selection problem, $\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W})$ is an affine function of the selection variables $\mathcal{W}$, and the resulting cost function is convex since the composition of a convex function with an affine function is convex. In our case, however, the input matrix is not affine with respect to the selection variables, and hence our objective function is not necessarily convex. To see that it is indeed not, note that, using (9) and (11), $\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W})$ is equal to:

$$\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W}) = \sum_{q=0}^{Q-1}\sum_{t=0}^{RS-1}\sum_{u=0}^{RS-1} w_q[t]w_q[u]\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u], \tag{16}$$

which is quadratic w.r.t. $\mathcal{W}$. As a simple counter-example of the convexity of (13), consider the simple case where $Q = 1$, so that $\mathbf{A}(\mathcal{W})$ only depends on a single selection vector $\mathbf{w}$, and consider that each $\mathbf{A}[t] \in \mathbb{C}^{1\times 1}$. Then

$$\mathbf{C_e^{-1}} = \sigma_x^{-2} + \sigma_n^{-2}\mathbf{A}(\mathbf{w})^{\mathsf{H}}\mathbf{A}(\mathbf{w}) \tag{17}$$

$$= \sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{B}\mathbf{w}, \tag{18}$$

where the coefficients of $\mathbf{B} \in \mathbb{S}_+^{RS\times RS}$ consist of the $1\times 1$ matrices $\{\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u]\}_{t,u=0}^{RS-1}$ (using (16)). Consequently, $\text{trace}(\mathbf{C_e}) = 1/(\sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{B}\mathbf{w})$, and it is straightforward to see that this function is not convex with respect to $\mathbf{w}$.

To circumvent this problem, we absorb the quadratic terms $w_q[t]w_q[u]$ into a single variable $W_q[t,u]$:

$$\mathbf{A}(\mathcal{W})^{\mathsf{H}}\mathbf{A}(\mathcal{W}) = \sum_{q=0}^{Q-1}\sum_{t=0}^{RS-1}\sum_{u=0}^{RS-1} W_q[t,u]\mathbf{A}[t]^{\mathsf{H}}\mathbf{A}[u]. \tag{19}$$

Whereas the cost-function (13) was quadratic in the elements of $\mathbf{w}_q$, it is linear with respect to the variables $W_q[t,u]$. Expression (16) is only equivalent to (19) if the variables $W_q[t,u]$ have a certain structure. Specifically, if we store the variables $W_q[t,u]$ in a matrix $\mathbf{W}_q$ such that $[\mathbf{W}_q]_{t,u} = W_q[t,u]$, then equivalence between (16) and (19) only holds if $\mathbf{W}_q = \mathbf{w}_q\mathbf{w}_q^{\mathsf{T}}$. In other words, $\mathbf{W}_q$ has to be a matrix with $\text{rank}(\mathbf{W}_q) = 1$. Although the cost-function is now convex with respect to $\mathbf{W}_q$, the additionally required rank constraint is not. Keeping in mind that if $\mathbf{w}_q \in \{0,1\}^{RS}$, then $\text{diag}(\mathbf{w}_q\mathbf{w}_q^{\mathsf{T}}) = \mathbf{w}_q$, a common technique [41], [42] to relax the constraint

$\mathbf{W}_q - \mathbf{w}_q\mathbf{w}_q^\mathsf{T} = \mathbf{0}$ is:

$$\mathbf{W}_q - \mathbf{w}_q\mathbf{w}_q^\mathsf{T} \succeq \mathbf{0}, \tag{20}$$

$$\text{diag}(\mathbf{W}_q) = \mathbf{w}_q \tag{21}$$

for all $q \in \mathcal{Q}$.

2) For the second relaxation, we relax the $l_0$ quasi-norm to the $\ell_1$ norm:

$$\|\mathbf{w}_{q,s}\|_1 = 1. \tag{22}$$

Since each $\mathbf{W}_q \succeq \mathbf{0}$, the diagonals of each $\mathbf{W}_q$ are non-negative, and thus we can write this constraint as $\mathbf{w}_{q,s}^\mathsf{T}\mathbf{1} = 1$.

3) Finally, we have to relax the binary constraint $\mathbf{w}_q \in \{0,1\}^{RS}$. We relax it to the box constraint

$$\mathbf{w}_q \in [0,1]^{RS}. \tag{23}$$

This constraint is already implied by the previous two relaxations. Since each $\mathbf{W}_q \succeq 0$, the diagonals of each $\mathbf{W}_q$ are non-negative. Hence the entries of $\mathbf{w}_q$ are all greater or equal to zero. Together with this observation, due to the constraint $\mathbf{w}_{q,s}^\mathsf{T}\mathbf{1} = 1$, the diagonal elements of each $\mathbf{W}_q$ cannot be larger than one. Consequently, the constraint $\mathbf{w}_q \in [0,1]^{RS}$ does not have to be explicitly enforced in our optimization routine.

So far we have not taken into account the full structure that a rank-1 matrix with entries in $\{0,1\}$ and $\|\mathbf{w}_{q,s}\|_0 = 1$ has. We will thus further exploit the following two properties that $\mathbf{W}_q$ should exhibit.

1) First, if $\mathbf{w}_q$ is a solution to the original problem (15), the matrix

$$\mathbf{W}_q = \mathbf{w}_q\mathbf{w}_q^\mathsf{T} = \begin{bmatrix} \mathbf{W}_{q|0,0} & \mathbf{W}_{q|0,1} & \cdots & \mathbf{W}_{q|0,S-1} \\ \mathbf{W}_{q|1,0} & \mathbf{W}_{q|1,1} & \cdots & \mathbf{W}_{q|1,S-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{q|S-1,0} & \mathbf{W}_{q|S-1,1} & \cdots & \mathbf{W}_{q|S-1,S-1} \end{bmatrix} \tag{24}$$

will have $S$ block-matrices of size $R \times R$ on its diagonal that are themselves diagonal. In other words, each matrix $\mathbf{W}_{q|s,s}$ is in the set $\mathcal{D}$ of $R \times R$ diagonal matrices.

2) If $\|\mathbf{w}_{q,s}\|_0 = 1$ for all $s$ and $q$, then there is only one non-zero entry in each submatrix of (24), and that entry should be equal to one. This implies

$$\mathbf{1}^\mathsf{T}\mathbf{W}_{q|s,s'}\mathbf{1} = 1 \tag{25}$$

for all $s, s' \in \{0,1,\ldots,S-1\}$. Seen from a different point of view, we impose an $\ell_1$ sparsity constraint on all the $R \times R$ block-matrices $\mathbf{W}_{q|s,s'}$ instead of just on all vectors $\mathbf{w}_{q,s}$. In fact, the sparsity constraint $\mathbf{1}^\mathsf{T}\mathbf{w}_{q,s} = 1$ is implied by the two constraints we just described, and hence can be omitted now.

3) It is straightforward to see that each entry of $\mathbf{W}$ should be non-negative, so we add the constraint

$$\mathbf{W} \geq 0. \tag{26}$$

We will use the set $\mathcal{W}' = \{\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_{Q-1}\}$ to summarize the relaxed problem as follows:

$$\begin{aligned} \min_{\mathcal{W}'} \; & f(\mathcal{W}') \\ \text{s.t.} \; & \mathbf{W}_q - \text{diag}(\mathbf{W}_q)\text{diag}(\mathbf{W}_q)^\mathsf{T} \succeq 0 \\ & \mathbf{W}_{q|s,s} \in \mathcal{D} \\ & \mathbf{1}^\mathsf{T}\mathbf{W}_{q|s,s'}\mathbf{1} = 1 \\ & \mathbf{W} \geq 0 \end{aligned} \tag{27}$$

for all $s, s' \in \mathcal{S}$ and $q \in \mathcal{Q}$. Note that, as discussed before, the box and sparsity constraints are implicitly included. Since all the constraints on $\mathbf{w}_q$ are now enforced through constraints on the matrices $\mathbf{W}_q$, we replaced the constraints (20) and (21) by the first constraint in (27).

### B. Random Rounding Scheme

The solution to the convex program above is not guaranteed to have rank 1 with entries according to the original discrete problem (15). One could take the leading eigenvector of $\hat{\mathbf{W}}_q$ as the closest solution rank-1 estimate, however it will not necessarily obey the constraints in the original problem formulation (15), and one would still have to round the vector to the discrete solution set in (15). Instead, we will interpret $\hat{\mathbf{W}}_q$ as a covariance matrix to generate multiple candidate solutions, and project those solutions to the original constraint set. This approach is favored in literature since it has proven itself in many studies, and even has theoretical performance guarantees for a number of discrete optimization problems [43]–[45]. Although we are currently unable to provide a lower bound on the optimality of this approach, we have observed that this approach typically outperforms solutions obtained from the leading eigenvector. Note that, if $\hat{\mathbf{W}}_q$ turns out to be rank-1, the random vectors generated using $\hat{\mathbf{W}}_q$ as covariance matrix will anyway be equal to the leading eigenvector of $\hat{\mathbf{W}}_q$, up to a scalar value.

To describe this in more detail, we use $\hat{\mathbf{W}}_q$ as a covariance matrix, and generate many random $\tilde{\mathbf{w}}_q \in \mathbb{R}^{RS}$ from a Gaussian distribution:

$$\tilde{\mathbf{w}}_q \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{W}}_q), \tag{28}$$

and round each $\tilde{\mathbf{w}}_q$ by selecting the mask thickness level per channel with the highest absolute value:

$$[\hat{\mathbf{w}}_{q,s}]_r = \begin{cases} 1, & \text{if } r = \arg\max_v |\,[\tilde{\mathbf{w}}_{q,s}]_v\,| \\ 0, & \text{otherwise} \end{cases}, \tag{29}$$

assuming each $\hat{\mathbf{w}}_{q,s}$ has a unique maximum. If there are multiple maxima, we randomly pick one of them.

The intuition behind the random rounding method comes from the following observation (adapting the interpretation in [43] for the problem discussed here). Consider again the example in (17), where $Q = 1$, and each $\mathbf{A}[t] \in \mathbb{C}^{1\times 1}$. The original optimization problem cost function can then be rewritten as

$$\min_{\mathbf{w}} (\sigma_x^{-2} + \sigma_n^{-2}\mathbf{w}^\mathsf{T}\mathbf{B}\mathbf{w})^{-1}.$$

One could instead replace $\mathbf{w}$ by a random variable $\boldsymbol{\nu}$ which has covariance $\mathbf{W}$, $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{W})$, and try to minimize:

$$\min_{\mathbf{W} \in \mathbf{S}_+} \mathbb{E}_{\boldsymbol{\nu}} \left( (\sigma_x^{-2} + \sigma_n^{-2} \boldsymbol{\nu}^{\mathsf{T}} \mathbf{B} \boldsymbol{\nu})^{-1} \right). \tag{30}$$

That is, find a covariance matrix $\mathbf{W}$ for the random variable $\boldsymbol{\nu}$, such that the expected value of the cost function above is minimized. Since all terms are now scalar values, this can be rewritten as (maximizing the inverse of the above equation, using the properties of the trace, and using $\mathbb{E}\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}} = \mathbf{W}$):

$$\begin{aligned} &\max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{\nu}} \left( \boldsymbol{\nu}^{\mathsf{T}} \mathbf{B} \boldsymbol{\nu} \right)) \\ &= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{\nu}} \left( \text{trace}(\boldsymbol{\nu}^{\mathsf{T}} \mathbf{B} \boldsymbol{\nu}) \right)) \\ &= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \mathbb{E}_{\boldsymbol{\nu}} \left( \text{trace}(\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}} \mathbf{B}) \right)) \\ &= \max_{\mathbf{W} \in \mathbf{S}_+} (\sigma_x^{-2} + \sigma_n^{-2} \text{trace}(\mathbf{W}\mathbf{B})). \end{aligned} \tag{31}$$

This is similar in form to a lifted problem, since we have replaced $\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}$ by a matrix $\mathbf{W}$. This result suggests that random solutions generated using $\mathbf{W}$ will give good solutions on average, when $\mathbf{W}$ is a higher-dimensional, lifted, version of $\mathbf{w}$.

### C. Greedy Mask Optimization

In this subsection we further describe a much more efficient greedy selection algorithm, since the convex program (27) scales polynomially with increasing problem size (more channels, thickness levels, etc.). It is based on the observation that minimizing over only a single channel has a low computational cost, since we would only have to solve the problem

$$\min_{\mathbf{w}_{q,s}} f(\mathcal{W}), \text{ s.t. } \|\mathbf{w}_{q,s}\|_0 = 1, \tag{32}$$

which is easily solved by trying all $R$ possible solutions for the current channel $s$ and measurement $q$. The algorithm operates by iterating through all channels, selecting the thickness level that most decreases the adopted cost function per channel. This is repeated, until either a local minimum or a maximum number of iterations $K_{\max}$ is reached. The algorithm is described in Algorithm1. In Algorithm1, we use $f(\mathbf{w}_{q,s})$ to indicate that $f$ is evaluated with $\mathbf{w}_{q,s}$ as a variable, but keeping all other $\mathbf{w}_{q',s'}$ fixed.

Referring back to the description of the approximate mask model, we point out that one could also directly define each $\{\mathbf{A}_s\}_{s \in \mathcal{S}}$ as a function of the mask thickness parameter for each channel. We were able to cast it as a discrete selection problem by sampling the continuous variable for the mask thickness, say $\mathbf{z} \in \mathbb{R}^S$, inside the valid mask thickness range. Keeping this in mind, Algorithm1 acts as a coordinate descent algorithm on $\mathbf{z}$, by optimizing only one element (channel) in $\mathbf{z}$ at a time. Instead of doing a line search over the allowed mask thickness range, it uses a one-dimensional grid search over $R$ pre-defined values. This line search is relatively cheap, since the maximum mask thickness is typically confined to a small number. Using a grid search has major advantages, since the proposed algorithm now easily avoids local minima in the current $[\mathbf{z}]_s$ it is optimizing (if the grid on $[\mathbf{z}]_s$ is sufficiently dense). In this way, many local minima are easily avoided, nor does the algorithm need to compute the cost-function gradient or higher order derivatives.

**Algorithm 1:** Greedy Optimization Scheme.

```
Input: W = {w_q}_{q=0}^{Q-1}, I_max
Output: Ŵ = {ŵ_q}_{q=0}^{Q-1}
ŵ_q = w_q, q = 0, 1, ..., Q − 1
i = 0
do
    f_min = f(Ŵ)
    for q = 0, 1, ..., Q − 1 do
        for s = 0, 1, ..., S − 1 do
            ŵ_{q,s} = arg min_{w_{q,s}} f(w_{q,s}), s.t. ||w_{q,s}||_0 = 1
            i = i + 1
        end for
    end for
while f(Ŵ) < f_min, and i ≤ I_max
```

The greedy routine can either be used to refine the rounded solution (29) or it can be initiated using an appropriate starting mask. In the first case, the convex solution forms a good starting point, and Algorithm1 will improve it slightly using few iterations. In the latter case, running the convex program (27) is avoided at the risk of arriving at a local minimum. In the following numerical experiments, we will use a flat starting mask as an initial guess for the greedy algorithm.

Finally, we point out that this algorithm is not dependent on pre-computing all $\mathbf{A}[t]$, and can be used for configurations where the transmitted pulse passes through the mask both on transmit and receive, using the same transducer for both events. This is due to the fact that $\mathbf{A}$ can be re-computed for the current selection of mask-channels per inner iteration of Algorithm 1. For the sake of comparison to the convex program, we have not further investigated this situation.

## IV. Simulation Results

To evaluate the proposed mask design algorithms, we consider various imaging scenarios. We first consider two scenarios with a small problem size, where we show that our methods perform well, compared to the true optimum, which we find using an exhaustive search. After these example problems, we consider some larger problems where we design masks for more realistic imaging scenarios. In all simulations, except the first one, we use a Gaussian-modulated excitation pulse $e(t) = w(t)\cos(2\pi f_0 t)$, with $f_0 = 4$ MHz. The function $w(t) = \exp\left(-(t/f_s)^2/(2\sigma_w^2)\right)$ is a smooth Gaussian window, with $f_s$ the sampling frequency, and $\sigma_w = 3.74$. This pulse is shown in Fig. 4. For the greedy algorithm, we set $I_{\max} = 5QS$ in all simulations. For the medium we use a speed of sound of $c_0 = 1491$ m/s, and we assume that the plastic mask has a speed of sound $c_1 = 2730$ m/s. We consider the following simulation scenarios throughout this section:

Scenario 1: Rotation of a masked circular transducer in the $(x, y)$ plane.

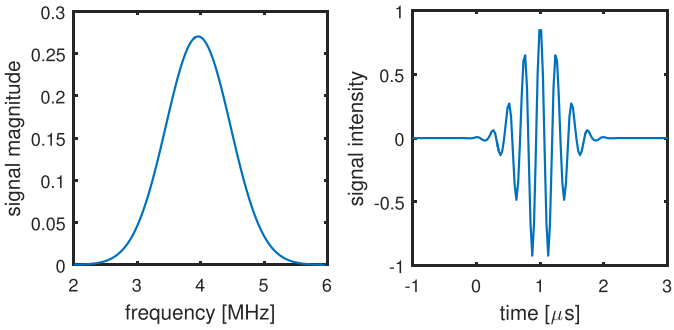


Fig. 4. Frequency spectrum and time-domain signal of the transmit pulse used in our simulated experiments.

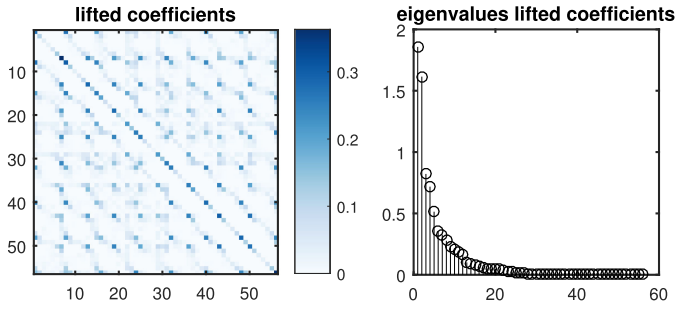


Fig. 5. Left: Optimal solution $\hat{\mathbf{W}}$ to problem (27) for the simulation described in Section IV-A. Right: Sorted eigenvalue spectrum of the matrix to the left.

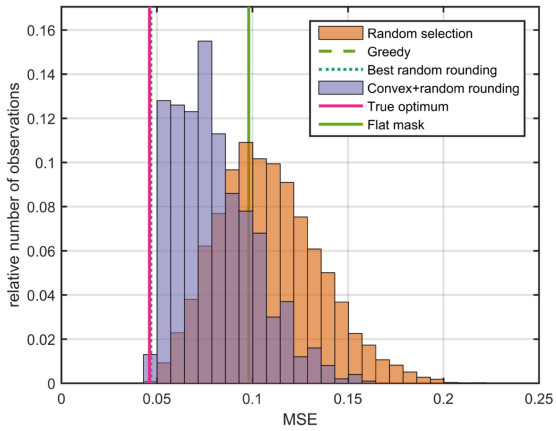


Fig. 6. MSE distribution for the problem described in Section IV-A. We show the MSE distribution for a set of 10 000 randomly generated masks, and 10 000 masks generated using the proposed random rounding procedure, the MSEs of the proposed methods, the MSE of the best mask, as well as a flat mask.

Scenario 2: Linear translation of a masked linear transducer in the $x$ direction, moving in the $x$ dimension.

Scenario 3: Using a different mask per measurement, with fixed transducer position.

### A. *Small Example 1*

In this example (using *scenario 2*), we choose all parameters such that we can find the true optimal mask of (15) by an exhaustive search within reasonable time. Although the resulting problem parameters do not correspond to a physically meaningful scenario (due to a very small ROI, and using only a few temporal frequencies), it is the only way to compare how our algorithms perform compared to the true optimum (and hence mostly tests the mathematical validity of our methods). To do this, we investigate a 2D problem, meaning that all pixel and channel positions are within the $(x, z)$ plane with $y = 0$. We use a Cartesian pixel grid of $8 \times 2$ ($M = 16$), with pixel coordinates equidistantly placed between $-2$ and 2 mm in the $x$-dimension, and at 0.5 and 1.0 cm in the $z$-dimension, with $y = 0$ for all pixel positions. The non-zero measurement frequencies are equidistantly spaced between 5 and 6 MHz ($N = 11$), which corresponds to taking 120 samples in the time-domain for a sampling frequency of 12 MHz. The mask channels ($S = 8, R = 7$) are equidistantly spaced between $-45$ and. 45 mm in the $x$-dimension, at $y = 0$ and $z = 0$, with a maximum mask thickness of 0.39 mm. This corresponds to discretization steps 2.8 times smaller than the centre wavelength in water (.37 mm) in the $x$-dimension (0.13 mm channel width), and 5.7 times smaller in the mask thickness dimension (0.065 mm per channel height segment). The noise variance was chosen such that the output SNR

$$\text{SNR} = \text{trace}\{\mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathsf{H}}\}\}/\text{trace}\{\mathbb{E}\{\mathbf{n}\mathbf{n}^{\mathsf{H}}\}\} \tag{33}$$

$$= \text{trace}\{\sigma_x^2 \mathbf{A}\mathbf{A}^{\mathsf{H}}\}/\text{trace}\{\sigma_n^2 \mathbf{I}\} \tag{34}$$

of a random mask is approximately[2] 0 dB. To add more structure to the problem, we add more measurements using the same mask ($Q = 1, K = M = 16$), by spatially shifting the masked transducer to the same $x$-coordinates as the pixels in the $x$-dimension (*scenario 2*).

In Fig. 5, the solution $\hat{\mathbf{W}}$ to the convex problem (27) is visualized. The constraints are able to obtain a matrix that is already quite sparse, and of low rank. Fig. 6 shows the MSE (13) distribution for a set of 10,000 randomly generated sets of masks, 10,000 masks generated using random rounding based on the solution of the convex problem (27), the MSEs of the proposed methods, the MSE of the true optimum, as well as a the MSE of a completely flat mask. The latter can still reconstruct images since multiple spatial measurements are obtained by translation of the transducer. The random rounding procedure generates solutions with a higher MSE concentration close to the true optimum than the distribution of completely random masks, and the majority of these masks also outperform the flat mask scenario. The greedy algorithm is able to find the true optimum, and provides a mask that is slightly better than the mask obtained using the convex program (27), while being much more computationally efficient.

For a complete performance description, we take the optimized and random masks from Fig. 6, and compute their MSE if the SNR changes, but the same mask is used. The resulting MSEs are shown in Fig. 7 (Fig. 6 shows the MSE distribution for one particular SNR in Fig. 7). The vertical bars in Fig. 7 indicate

[2]Note that, even if $\sigma_x^2$ is known, we can only determine $\|\mathbf{y}\|_2^2$ once the mask is known, since each mask can have a different measurement energy for the same image $\mathbf{x}$. Hence, we cannot choose $\sigma_n^2$ to give an exact output SNR. In this example, the exact output SNR of the optimized mask is 2.8 dB.

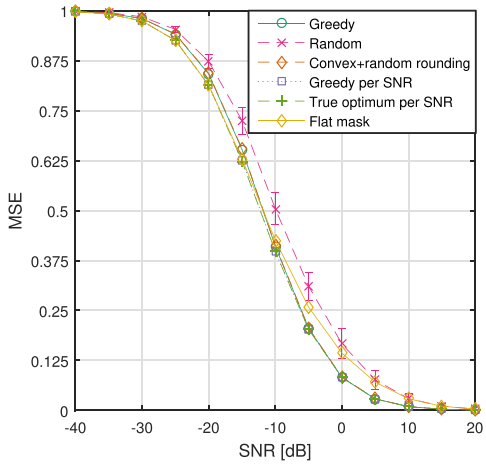


Fig. 7. Performance of various mask design strategies in terms of MSE for various output SNR scenarios, for the problem described in Section IV-A. Each curve shows the performance of a single mask that was optimized for an SNR of 0 dB. The crosses represent the mean MSE of 1 000 randomly designed masks, with the vertical bars showing the standard deviation. The MSE of the true optimal mask per SNR is shown, as well as the MSE obtained by using a greedy algorithm for each SNR.

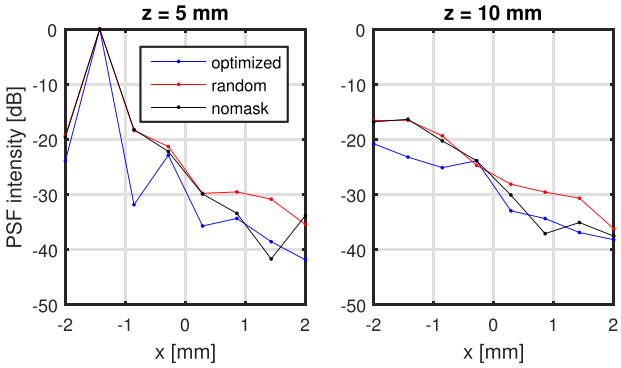


Fig. 8. PSF for a pixel at $x = -1.4$ mm, $z = 5$ mm for the scenario of Section IV-A. The subfigures show slices of the PSF at different depths.

the mean MSE plus or minus the MSE standard deviation for 1,000 randomly generated masks. Even though all methods are optimized for a particular SNR, Fig. 7 shows that all methods are able to generate a mask that performs well over the entire SNR range, although most of the MSE can be gained in the $-15$ to 10 dB range. As a benchmark, we also show the MSE of the optimal mask per SNR, as well as the MSE of the mask obtained using the greedy algorithm by optimizing per SNR (the best mask can be different for each SNR). In this example, we can see that all proposed masks, even if they are optimized for a single SNR, are able to perform close to optimal over the entire SNR range, demonstrating the robustness of the optimization method to incorrect values of the noise and image covariance matrices.

Finally, we show the PSF in dB (computed as $10\log_{10}$) for the second pixel (Fig. 8). We define the PSF for pixel $n$ as the image when the measurement is equal to: $\mathbf{y} = \mathbf{a}_n(\mathcal{W})$, the $n$-th column of $\mathbf{A}(\mathcal{W})$. For comparison, we took the PSF of a random

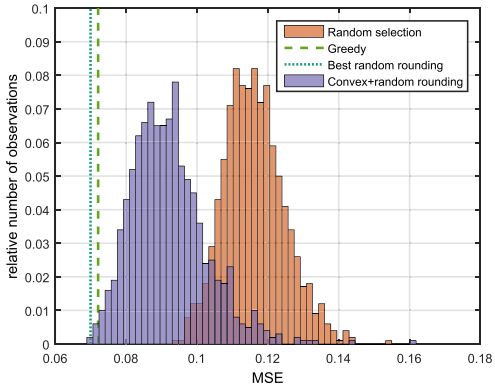


Fig. 9. MSE distribution for the problem of Section IV-B, for a set of 1 000 randomly generated masks, as well as 1 000 masks generated using the proposed random rounding procedure, and the MSEs of the proposed methods.

mask with an expected MSE of 0.106, close to the average performance of random masks according to the distribution in Fig. 6. The PSFs of the different masks re-affirm the performance as predicted by Fig. 6.

### *B. Small Example 2*

In the previous example, we constrained the problem parameters to be able to use an exhaustive search to find the global optimum. In this second small example, we make the problem slightly larger, by using more channels and multiple masks, and using *scenario 3* as described in Section II. As a result, the global optimum cannot be found within reasonable time, and is not shown in the results of this experiment. Furthermore, the distribution of random masks will be further away from the true optimum, unlike in Fig. 6. Using random rounding, however, we can still do well.

Instead of shifting the transducer, it remains in the same position, but uses $Q = 12$ different masks to obtain different measurements. We optimize the mask for 5 pixels on a line at $z = 5$ mm, located between -1 and 1 mm. The mask channels are located between -.45 and. 45 mm ($S = 8$), with a maximum mask thickness of. 26 mm ($R = 5$). The measurement frequencies are equidistantly spaced between 5 and 6 MHz ($N = 101$).

We generate 1,000 random masks and 1,000 randomly rounded masks, as well as a greedily generated mask. The MSE distributions are shown in Fig. 9. As mentioned before, the random mask design strategy does not work well, and covers a higher range of MSEs. In Fig. 10 the PSF is shown for the optimized image domain. As predicted by Fig. 9, some performance gain is possible, although not as great as in the previous simulation in Section IV-A.

### *C. Mask Design for a Single Rotating Mask*

In this subsection we optimize a single rotating mask (*scenario 1*). We assume that all the channels are located within a 5 mm disc. The only difference between the measurements is a rotation of the sensor and the mask relative to the pixels. We only consider the greedy approach so that we can use realistic problem

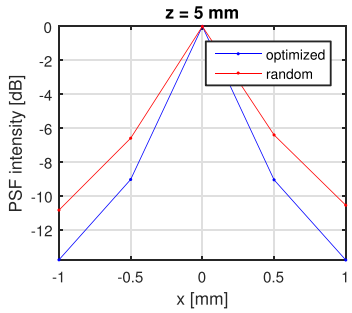


Fig. 10. PSF for the simulations of Section IV-B for a pixel at $x = 0$ mm, $z = 5$ mm.

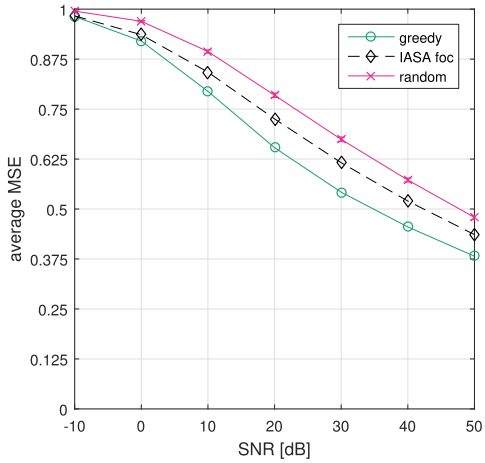


Fig. 11. MSE performance for the scenario of Section IV-C, for various SNR scenarios, optimizing for a single rotating mask. The crosses represent the mean MSE of 100 randomly designed masks, with the vertical bars showing the standard deviation. The IASA method is discussed in the Discussion section of this article.

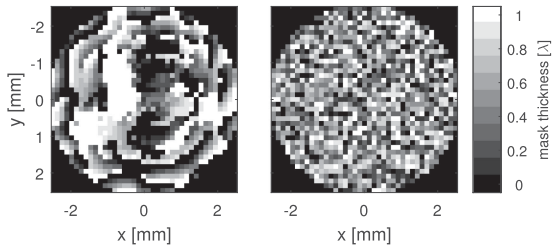


Fig. 12. Optimized and random mask used for the results in Fig. 13, using a single sensor with a rotating mask, for the simulation scenario of Section IV-C. The mask thickness values are displayed as a fraction of the wavelength inside the mask.

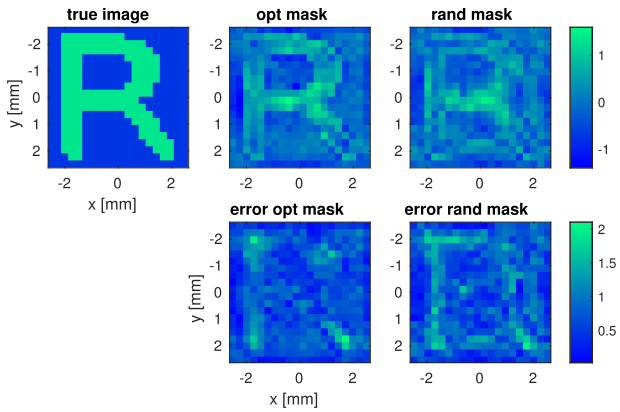


Fig. 13. Reconstruction of a letter R image for a single sensor with a rotating mask, using 40 rotations, for the simulation of scenario Section IV-C. The top figures show the true image and its reconstructions using both the optimized mask and a random mask. The bottom figures show the error maps for each reconstruction result.

sizes with acceptable computation times. We discretize the mask into 1245 square channels of equal size (0.13 mm), equivalent to approximately 2.9 channels per wavelength. Each channel was divided into 11 uniformly distributed mask thickness levels (0.13 mm per mask thickness segment), with a maximum mask thickness of 0.65 mm, equivalent to approximately 5.7 channels per wavelength. The 2D ROI of 5 by 5 mm is discretized into 20 by 20 pixels at a depth of 6 mm. The transducer is rotated 40 times, evenly divided over 360 degrees ($K = 40$). Note that, although there are multiple measurements, we optimize for a single mask ($Q = 1$). During optimization, $\sigma_n^2$ was chosen such that the expected SNR of a random mask is approximately 10 dB. We use the transmit pulse shown in Fig. 4, sampling it in the frequency domain using $\Delta f = 40$ kHz, or $N = 101$.

The imaging performance for various SNRs is shown in Fig. 11, although the mask was optimized for a single SNR. Again, it is clear that the optimized mask is able to consistently outperform all random masks. Moreover, the MSE distribution of the random masks is relatively narrow, meaning that it would take an enormous amount of tries to randomly generate a mask that performs just as well as our proposed techniques. Furthermore, the expected SNR (33) for the optimal mask is equal to 15.5 dB, whereas the SNR averaged over 100 random masks, is equal to 9.0 dB, a gain of 6.5 dB for the optimized setup.

To illustrate the imaging performance, we take a letter-shaped test image (zero-mean and unit-energy, averaged over all pixels) to generate a noisy test measurement $\mathbf{y}$. Fig. 12 shows the greedily optimized mask and a particular randomly generated mask. Fig. 13 shows some example estimated images using the masks in Fig. 12, choosing $\sigma_n^2$ such that the optimized mask has an SNR of 20 dB for the test phantom. The random mask, in contrast, causes measurements to have an SNR of 13.1 dB using the same noise variance and test phantom. We see from Fig. 13 that the optimized mask is better able to correctly estimate the phantom.

The aforesaid SNR gain is caused by a focusing effect that becomes apparent if we visualize the pulse-echo amplitude per pixel for a large region and a single frequency, shown in Fig. 14. The mask acts as a spatial filter that boosts energy of echoes originating from the ROI, whereas a random mask does not have this capability, resulting in a lower SNR.

In addition to an increased SNR, the optimized mask also causes lower similarity between pulse-echo signals of different pixels. We visualize this in Fig. 15 by plotting the normalized PSF (computed as $\mathbf{A}^{\mathsf{H}}\mathbf{a}_i$, with each column of $\mathbf{A}$ normalized so that the Euclidean norm is one) for various points in space.

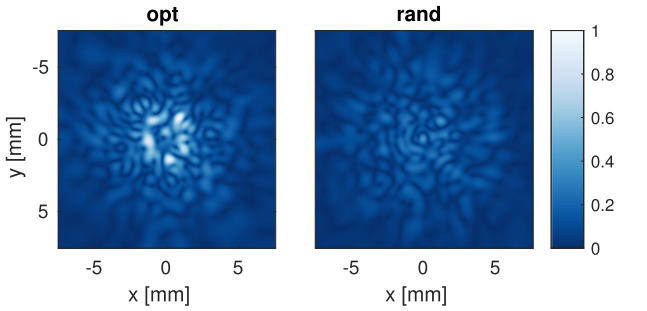


Fig. 14. Normalized amplitude per pixel at $f = 4$ MHz for an optimized rotating mask (Section IV-C) in Fig. 12, and a random mask. Note that the visualized energy extends beyond the imaging ROI, which is only 5 by 5 mm.

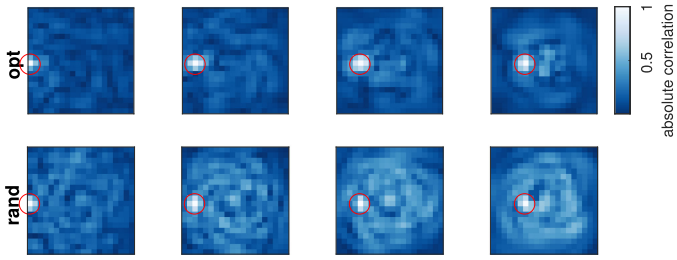


Fig. 15. Absolute PSFs for various points in space, indicated by the red circle. Top: using the greedily optimized mask. Bottom: Using a randomly selected mask. The columns in $\mathbf{A}$ are energy-normalized before computing the PSFs. Simulation scenario of Section IV-C.

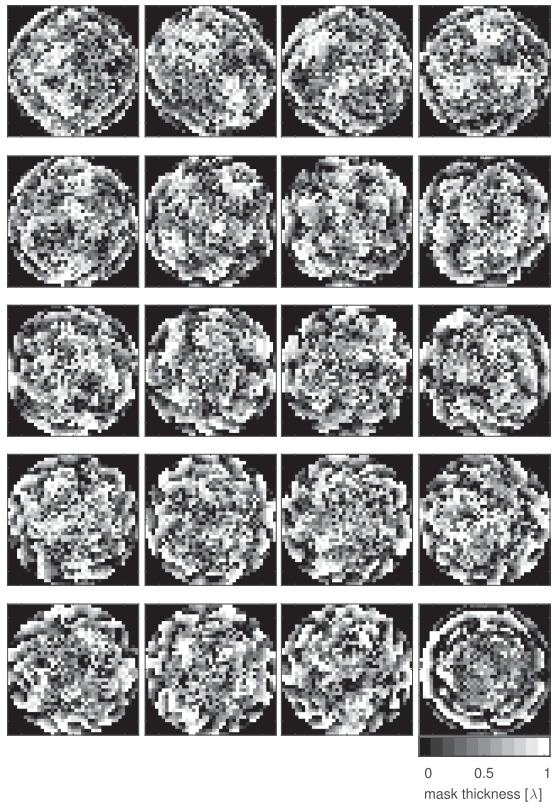


Fig. 16. A subset of the optimized masks for the problem of Section IV-D with $Q = 40$, using Algorithm1. The mask thickness values are displayed as a fraction of the wavelength inside the mask. The algorithm optimizes one mask at a time. The first mask is shown in the top left corner, while the third mask is shown to the right of the first mask, etc. (every second mask is plotted).

The optimized mask strongly decorrelates between pulse-echo signals from different pixels, whereas the random mask always has relatively high correlation across almost the entire ROI. The PSF is typically best at the edges, where the difference between different measurements obtained by rotation is the largest.

These results show that the proposed design algorithm makes columns of $\mathbf{A}(\mathcal{W})$ more orthogonal. This implies that other, possibly non-linear imaging methods will also benefit from the designed mask, since the increased orthogonality generally reduces ambiguity about $\mathbf{x}$, making the estimation of $\mathbf{x}$ easier for *any* imaging method.

### D. Mask Design for Multiple Masks

For our final simulation example, we consider an experiment where the transducer is not rotated or translated in any way, but instead the mask is changed between pulse-echo measurements (*scenario 3*). We use the greedy method to find $Q = 40$ masks, and compare the performance to 40 randomly generated masks. All other parameters are the same as in the previous simulation for a single rotating mask (Section IV-C). We choose $\sigma_n$ such that the SNR of a random mask is approximately 10 dB.

Using the greedy approach with a flat starting mask, we obtain the masks shown in Fig. 16. In Fig. 18 we show the corresponding SNR curves. The expected SNR of the optimized mask is 13.7 dB, whereas the random mask SNR is 9.8 dB, averaged over 100 random masks, an SNR gain of nearly 4 dB.

Fig. 17 shows an example reconstruction for an SNR of 20 dB. We can see that using the optimized masks, the true image is better reconstructed compared to a set of random masks. From Fig. 18 we observe an MSE improvement for all SNRs. Since the random masks show a very narrow MSE distribution, the optimized mask outperforms nearly all random masks in terms of the MSE (13). Although the MSE gain may not seem like much, the example reconstruction in Fig. 17 visualizes that a small MSE difference can have a non-negligible impact on the visual reconstruction quality. We observe that the multiple mask setup exhibits better imaging performance than a rotating mask with the same amount of measurement data. Finally, we remark that the imaging performance for this setup is better than the rotating mask configuration, since it does not restrict all masks to be related by rotation (although the algorithm is free to output such a set of masks).

### E. Mask Design for a Rotating Mask With Pixels on a Circle

In this subsection we analyse a particular example that admits an intuitive physical interpretation of the results. We use a single mask ($Q = 1$), and obtain new measurements by rotating the mask ($K = 30$). To make the results easily interpretable, we only consider pixels uniformly spaced on a circle with a radius of 1.8 mm ($M = K = 30$), where the center of the circle coincides with the center of rotation of the masked transducer.

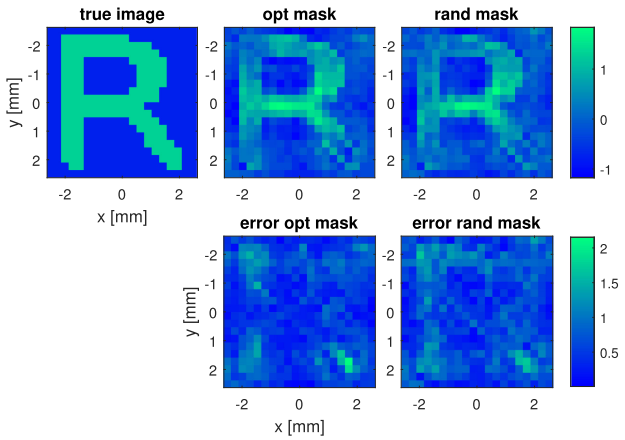


Fig. 17. Reconstruction of a letter R image combining the measured data of 40 masks obtained using the greedy optimization scheme, for an output SNR of 20 dB, for the simulation scenario of Section IV-D. The top figures show the true image and its reconstructions using both the optimized mask and a random mask. The bottom figures show the error maps for each reconstruction result.

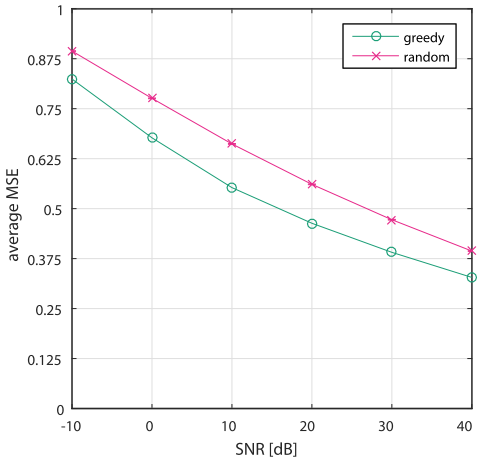


Fig. 18. MSE performance distribution for various SNR scenarios, optimizing for $Q = 40$ masks. The optimized mask was designed assuming an SNR of 10 dB, but performs relatively well across the entire SNR range. The crosses represent the mean MSE of 100 randomly designed masks, with the vertical bars showing the standard deviation. Simulation scenario of Section IV-D.

Consider the measurement matrix $\mathbf{A}_{\omega_k}$, containing only the measurements for frequency $\omega_k$. Assuming the angle of rotation is equivalent to the angle between pixels, each row or column is a circular shift of the previous row. As a result, each $\mathbf{A}_{\omega_k}$ is a circulant matrix, and the total measurement matrix is a concatenation of circulant matrices:

$$\mathbf{A} = \left[\mathbf{A}_{\omega_0}^{\mathsf{T}}\ \mathbf{A}_{\omega_1}^{\mathsf{T}}\ \ldots\ \mathbf{A}_{\omega_{K-1}}^{\mathsf{T}}\right]^{\mathsf{T}}. \tag{35}$$

Consequently, the singular values $\mathbf{s} \in \mathbb{R}_+^M$ of the SVD of $\mathbf{A}$ can be expressed as the square root of the element-wise sum of the squared magnitude of the eigenvalues over all circulant submatrices:

$$\mathbf{s} = \sqrt{\sum_{k=0}^{K-1} |\mathbf{s}_{\omega_k}|^2}, \tag{36}$$

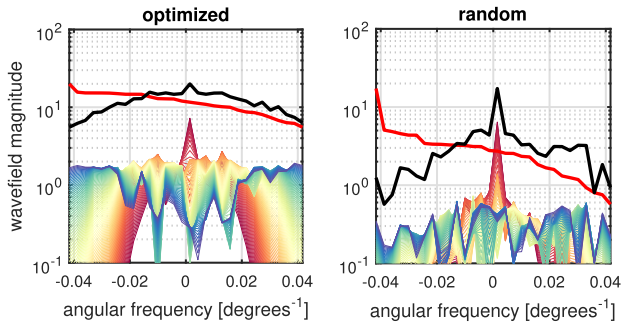


Fig. 19. Spatial frequency and singular value spectra for the optimized mask and a particular random mask with pixels on a circle (Section IV-E). The color-coded curves show the spatial frequency spectrum of the complex pressure field of each temporal pulse frequency $\omega_k$ (starting with low temporal frequencies in red, ending with high frequencies in blue). The black curve is the compounded spatial frequency spectrum of all temporal frequencies, as computed per (37), which is equivalent to the singular values of $\mathbf{A}$. The red curve shows the singular values of $\mathbf{A}$, sorted by magnitude.

where $\mathbf{s}_{\omega_k}$ is the vector of eigenvalues of submatrix $\mathbf{A}_{\omega_k}$. Since each submatrix is circulant, its eigenvalues are equivalent to the DFT spectrum of its first column. Therefore, we can replace each $\mathbf{s}_k$ by the DFT of the first column of $\mathbf{A}_{\omega_k}$:

$$\mathbf{s} = \sqrt{\sum_{k=0}^{K-1} |\hat{\mathbf{a}}_{\omega_k}|^2}, \tag{37}$$

where $\hat{\mathbf{a}}_{\omega_k}$ is such a DFT. Note that in (36) and (37), the square root and magnitude operators are assumed to be taken element wise.

This yields an intuitive way of comparing two masks. Due to the reciprocity principle, the columns/rows of $\mathbf{A}_{\omega_k}$ can be interpreted in two ways. First, they can be viewed as the transducer output signal for each pixel, if only that single pixel is transmitting a wavefield $p(\omega)$. Alternatively, they can be seen as the complex pressure field measured at the pixel locations if the transducer is *transmitting* the pulse $p(\omega)$ through the mask towards the pixel locations. Since the singular values of $\mathbf{A}$ can be expressed as a function of the Fourier transformed columns of the sub-matrices $\mathbf{A}_{\omega_k}$, the mask with the best (i.e., slowest decrease of the) singular value spectrum is the mask that generates a pressure field that has high spatial bandwidth, i.e., has a balanced spectrum with its energy spread across all the spatial frequencies as evenly as possible and has a high total energy.

For this particular case, the mask can also be analyzed from another point of view: the measured signal (per temporal frequency) equals a circular convolution of the image map, and a good mask results in a filter that preserves as much frequency content of the image as possible.

We illustrate the preceding with a simulated example. In addition to the parameters mentioned, we use the same transmit pulse as in Fig. 4 ($N = 101$ or $\Delta f = 40$ kHz). All channels ($S = 1245$) are located on a circle of 5 mm radius. We define $R = 6$ mask thickness levels up to a thickness of 0.65 mm, and $M = 30$ pixels on a ring of 5 mm diameter.

In Fig. 19, we show the spatial frequency spectra of the columns of all $\mathbf{A}_{\omega_k}, \omega_k \in \Omega$, in addition to $\mathbf{s}$ as computed in (37),

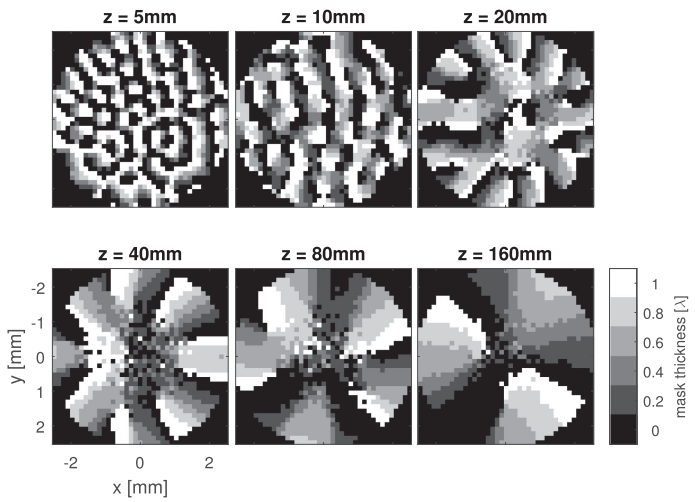


Fig. 20. Optimized masks for various depths. For simplicity, pixels are positioned on a ring, and the excitation signal is a single frequency.

and the singular values s sorted according to their magnitude. We do this for the greedily optimized mask as well as for a particular random mask. From Fig. 19 we observe that the optimization algorithm tries to find a mask such that the total spatial frequency bandwidth over the ring of pixels is increased. As a result, the singular values of the optimized matrix decrease more slowly, and are better conditioned. This is achieved due to each temporal frequency generating energy in a specific spatial frequency. For example, in the left panel of Fig. 19, the blue temporal frequencies generate high spatial frequencies, whereas temporal frequencies in the red spectrum focus energy in the lower spatial frequencies. The random mask is unable to accomplish the same result, and depressions in the blue curve are not compensated by any other temporal frequencies. We further observe that the total energy is much higher for the optimized mask, resulting in an increased SNR.

## V. Discussion

In Figs. 14 and 19 we showed that an optimized mask (for a rotating measurement configuration) acts as a spatial filter, enhancing echoes coming from within the ROI. In the angular frequency domain this means that the mask should act as an angular filter with a passband for directions within the ROI. Hence, as a rule of thumb, the structure within the mask should contain spatial features within the same angular frequency band. For example, if pixels are located in the far-field at an incident angle of 0 degrees, the mask can coherently sum signals from this ROI using a flat mask. On the other hand, if pixels are located at an incident angle of 90 degrees, a sinusoidal profile with the same period as the pulse wavelength will coherently sum echoes from that direction, and average out echoes originating from 0 degrees. This is more complicated in the near field, but the same principle holds. To illustrate this point, we show several masks optimized for ROIs parallel to the transducer (for simplicity, we only define pixels on a circle with the same $(x, y)$-coordinates as the edge of the mask) at different depths (Fig. 20). As the ROI is located further away, its angular range decreases, and consequently, spatial features in the mask profile contain a lower range of spatial frequencies.

Beside angular filtering, the mask should decorrelate echoes from different pixels. Optimized masks thus don't consist of pure harmonics, but of more complex patterns, in order to bring about this decorrelation. As discussed in Section IV-E, for a ring of pixels this means that the optimized mask should cause high variation of the wavefield w.r.t. this ring. Generalizing this to a fully sampled ROI by adding more rings is more complicated, but one could imagine that different wavelength intervals in the transmitted pulse can focus on different rings. One should, roughly speaking, try to design a mask that if used for transmit, causes a high spatial variation in the transmitted wavefield. Hence, an alternative design methodology would consist of first defining a wavefield at the ROI with high spatial bandwidth, and propagate it back in time towards the transducer to obtain a phase mask (similar to [36], [37]). Needless to say, this should be co-designed with the desired angular filtering discussed before, and algorithms presented in this paper try to find a balance between these two criteria.

Still, we have explored this alternative design strategy as suggested in the above study. First, we designed a pressure field for a single frequency ($f = 4$ MHz at $z = 6$ mm), with high spatial bandwidth, and which has most energy focused inside the pixel ROI within this $z$-plane. We then use the Iterative Angular Spectrum Approach (IASA) to approximate a phase mask that generates this pressure plane, following the method proposed in [36]. This method starts with a flat mask, finds the associated pressure field in the image plane, and compares it to the desired pressure field. It then changes the complex amplitudes in the obtained forward field to the desired amplitudes, but keeping the phases unchanged. The field is then propagated back in time to the transducer plane, where the mask profile is adapted to cause the required phase changes to generate the back-propagated pressure field. This is repeated until the algorithm converges to a phase mask. For further details on this method, we refer to the original paper [36]. As a desired pressure field we generated a white complex Gaussian field, and applied a Gaussian window in both the space and k-space domain to obtain a focused bandlimited target field for the image plane. The exact details for the target field are shown in Appendix A.

The general performance of the resulting mask for Gaussian images is included in Fig. 11. We also visualize the reconstructed image for the resulting mask in Fig. 21, where it is compared to the mask obtained with the proposed greedy method. The IASA-based design is superior to random mask design, but is not able to surpass the method proposed in this work, although the difference in reconstruction quality for the non-Gaussian (sparse) phantom in Fig. 21 is relatively small (the reconstruction MSEs for this particular phantom are 0.65 and 0.70 for the greedy and IASA-based method, respectively). Of course, the crucial design step in this strategy amounts to choosing a proper target field for the phase mask to produce, and in this experiment we have tried to find a good target field by trial-and-error. However, it is not clear what the 'best' forward field would be. Although initial investigation agrees with our suggestions

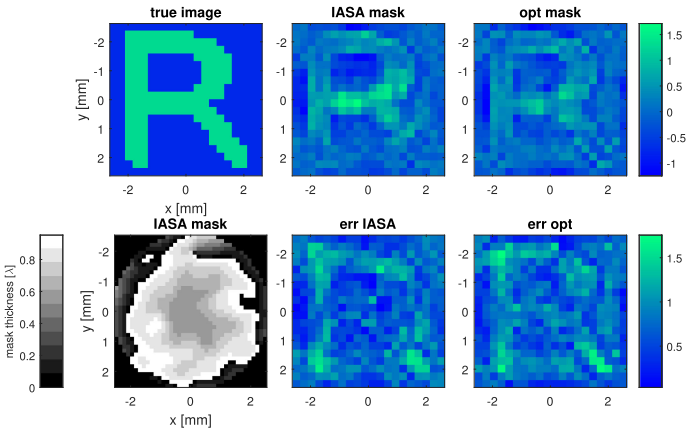


Fig. 21. Bottom left: Coding mask obtained using the IASA-method [36], by finding a phase mask that gives a focused pressure field in the image ROI with high spatial bandwidth. Top/bottom: Image reconstruction comparison between the IASA-based mask design method and the proposed method in this work.

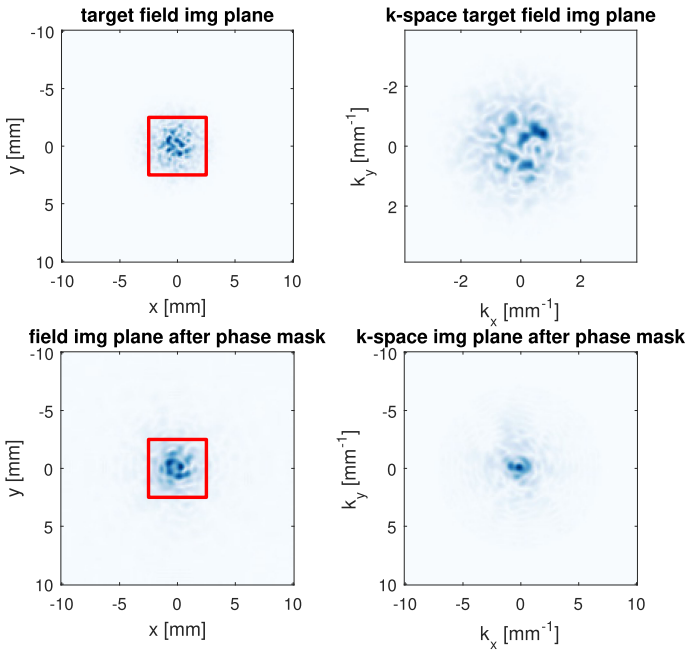


Fig. 22. Top: Target pressure field for $f = 4$ MHz in the image plane. Bottom: The field in the image plane produced with the mask obtained through the IASA method. All figures show the magnitude of the complex pressure fields. The red box indicates the pixel ROI used for optimization in Section IV-C.

in this paragraph, (e.g., a focused forward field seemed to give better results than an unfocused beam), broad-bandedness of the spatial target distribution does not seem to be sufficient to acquire a good mask with this method. Hence, this design strategy can probably be significantly improved by choosing a more intelligent pressure distribution than the one used in this section.

### A. Mask Design for Multiple Sensors

The optimization framework discussed in this work could also be extended to include multiple sensors, each with possibly their own mask. In that case, one would have the following measurement equation for $L$ sensors, combining equations (10) and (11):

$$\mathbf{y}(\mathcal{W}) = \begin{bmatrix} \mathbf{y}_0(\mathbf{w}_0) \\ \mathbf{y}_1(\mathbf{w}_1) \\ \vdots \\ \mathbf{y}_{L-1}(\mathbf{w}_{L-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0(\mathbf{w}_0) \\ \mathbf{A}_1(\mathbf{w}_1) \\ \vdots \\ \mathbf{A}_{L-1}(\mathbf{w}_{L-1}) \end{bmatrix} \mathbf{x} = \mathbf{A}(\mathcal{W})\mathbf{x}, \tag{38}$$

where each $\mathbf{A}_l, (l = 0, 1, \ldots, L)$ represents the measurement equations for sensor $l$. The equations are still linear with respect to $\mathcal{W}$, and the same proposed techniques can be applied.

## VI. Conclusion

Using an approximate linear signal model, we formulated mask optimization as an *additive* sensor selection problem by discretizing each mask channel in the thickness dimension, and optimizing for the imaging MSE, which is a function of the measurement matrix $\mathbf{A}$. Each candidate sensor corresponds to a full measurement matrix, and the final measurement model is the *summation* of the individual sensors' measurement matrices. We have shown that this is not a conventional sensor selection problem, and proposed to solve this problem by transforming the original variables into a larger amount of 'lifted' variables. By further relaxing the problem, and by exploiting the structure among the lifted optimization variables, a convex problem is formulated. To get a solution that is feasible in the original problem formulation, a random rounding technique was proposed. In addition to the convex approach, we also proposed a greedy algorithm to solve the same problem.

Although the proposed algorithms are not guaranteed to provide the globally optimal mask, a small numerical experiment showed that our mask design algorithms come close to the true optimum, and all experiments showed that they outperform randomly generated masks in terms of imaging MSE, even for images with non-white and non-Gaussian statistics. The distribution of masks using the combination of convex optimization and random rounding has a lower MSE range than the distribution of completely random masks. Furthermore, the proposed greedy coordinate-descent algorithm obtained good results close to the global optimum. Our experiments demonstrate that a good mask is able to filter in the angular domain, enhancing echoes coming from within the ROI, while strongly decreasing correlation between the pulse-echo signals of pixels within the ROI.

## Appendix A
## IASA-Based Mask Design

To obtain a mask using the IASA method, one has to first choose a target pressure field in the image plane that the phase mask is supposed to produce. We designed a focused wavefield with maximum spatial bandwidth, by first generating a completely complex Gaussian random pressure field. Next, we filtered this field by applying a circular, symmetric Gaussian window in the 2D DFT of this field. The radius of this window

was chosen approximately equal to the wavenumber in water, choosing $f = 4$ MHz, the central frequency of the transmit pulse. After transforming back to the spatial domain, we again applied a circular Gaussian window, so that most energy is focused in the pixel ROI. We show the resulting pressure field in Fig. 22. The pressure field produced by the resulting mask is limited in spatial bandwidth due to the limited size of the aperture.

## REFERENCES


[1] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[5] H. Liebgott, R. Prost, and D. Friboulet, "Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing," *Ultrasonics*, vol. 53, no. 2, pp. 525–533, 2013.

[6] C. Quinsac, A. Basarab, and D. Kouamé, "Frequency domain compressive sampling for ultrasound imaging," *Adv. Acoust. Vib.*, vol. 2012, 2012, Art. no. 231317.

[7] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4643–4657, Sep. 2012.

[8] T. Chernyakova and Y. C. Eldar, "Fourier-domain beamforming: The path to compressed ultrasound imaging," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 61, no. 8, pp. 1252–1267, Aug. 2014.

[9] M. F. Schiffner and G. Schmitz, "Fast pulse-echo ultrasound imaging employing compressive sensing," in *Proc. IEEE Int. Ultrason. Symp.*, 2011, pp. 688–691.

[10] G. David, J.-L. Robert, B. Zhang, and A. F. Laine, "Time domain compressive beam forming of ultrasound signals," *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2773–2784, 2015.

[11] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, "Compressed delay-and-sum beamforming for ultrafast ultrasound imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2509–2513.

[12] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[13] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Appl. Phys. Lett.*, vol. 93, no. 12, 2008, Art. no. 121105.

[14] L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," *Nature*, vol. 516, no. 7529, pp. 74–77, 2014.

[15] A. C. Fannjiang, "Exact localization and superresolution with noisy data and random illumination," *Inverse Problems*, vol. 27, no. 6, 2011, Art. no. 065012.

[16] R. Fergus, A. Torralba, and W. T. Freeman, "Random lens imaging," MIT CSAIL, Tech. Rep. 2006-058, 2006.

[17] A. Liutkus *et al.*, "Imaging with nature: Compressive imaging using a multiply scattering medium," *Scientific Rep.*, vol. 4, 2014, Art. no. 5552.

[18] V. Boominathan *et al.*, "Lensless imaging: A computational renaissance," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 23–35, Sep. 2016.

[19] N. Antipa *et al.*, "DiffuserCam: Lensless single-exposure 3D imaging," *Optica*, vol. 5, no. 1, pp. 1–9, 2018.

[20] F. C. Meral, M. A. Jafferji, P. J. White, and G. T. Clement, "Two-dimensional image reconstruction with spectrally-randomized ultrasound signals," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 60, no. 12, pp. 2501–2510, Dec. 2013.

[21] N. Etaix, M. Fink, and R. K. Ing, "Acoustic imaging device with one transducer," *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. EL395–EL399, 2012.

[22] P. Kruizinga *et al.*, "Compressive 3D ultrasound imaging using a single sensor," *Sci. Adv.*, vol. 3, no. 12, 2017, Art. no. e1701423.

[23] J. Janjic *et al.*, "Structured ultrasound microscopy," *Appl. Phys. Lett.*, vol. 112, no. 25, 2018, Art. no. 251901.

[24] A. Austeng and S. Holm, "Sparse 2-D arrays for 3-D phased array imaging-design methods," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 49, no. 8, pp. 1073–1086, Aug. 2002.

[25] B. Diarra, M. Robini, P. Tortoli, C. Cachard, and H. Liebgott, "Design of optimal 2-D nongrid sparse arrays for medical ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3093–3102, Nov. 2013.

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

[27] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[28] F. Bian, D. Kempe, and R. Govindan, "Utility based sensor selection," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 11–18.

[29] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," *Am. Assoc. Artif. Intell.*, vol. 7, pp. 1650–1654, 2007.

[30] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 2572–2577.

[31] M. Coutino, S. Chepuri, and G. Leus, "Near-optimal greedy sensor selection for MVDR beamforming with modular budget constraint," in *Proc. 25th Eur. Signal Process. Conf.*, 2017, pp. 1981–1985.

[32] M. Coutino, S. P. Chepuri, and G. Leus, "Submodular sparse sensing for Gaussian detection with correlated observations," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4025–4039, Aug. 2018.

[33] O. M. Bushnaq, T. Y. Al-Naffouri, S. P. Chepuri, and G. Leus, "Joint sensor placement and power rating selection in energy harvesting wireless sensor networks," in *Proc. 25th Eur. Signal Process. Conf.*, 2017, pp. 2423–2427.

[34] P. van der Meulen, P. Kruizinga, J. G. Bosch, and G. Leus, "Spatial compression in ultrasound imaging," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, 2017, pp. 1016–1020.

[35] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5/6, pp. 877–905, 2008.

[36] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, "Holograms for acoustics," *Nature*, vol. 537, no. 7621, pp. 518–522, 2016.

[37] M. Brown, D. Nikitichev, B. Treeby, and B. Cox, "Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles," *Appl. Phys. Lett.*, vol. 110, no. 9, 2017, Art. no. 094102.

[38] B. E. Treeby and B. T. Cox, "k-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, 2010, Art. no. 021314.

[39] M. Verweij, B. Treeby, K. Van Dongen, L. Demi, and A. Brahme, "Simulation of ultrasound fields," *Comprehensive Biomed. Phys.*, vol. 2, pp. 465–499, 2014.

[40] J. T. Fokkema and P. M. van den Berg, *Seismic Applications of Acoustic Reciprocity*. New York, NY, USA: Elsevier, 1993.

[41] C. Helmberg, "Semidefinite programming for combinatorial optimization," Konrad-Zuse-Zentrum für Informationstechnik Berlin, Tech. Rep. ZIB-Report 00-34, Oct. 2000.

[42] A. dAspremont and S. Boyd, "Relaxations and randomized methods for nonconvex QCQPs," EE392o Class Notes, Stanford University, Stanford, CA, USA, 2003, no. 1, pp. 1–16.

[43] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[44] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[45] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optim. Methods Softw.*, vol. 9, no. 1–3, pp. 141–160, 1998.