Radar Target Profiling and Recognition Based on TSI-Optimized Compressive Sensing Kernel

Yujie Gu, *Member, IEEE*, Nathan A. Goodman, *Senior Member, IEEE*, and Amit Ashok, *Member, IEEE*

***Abstract*—The design of wideband radar systems is often limited by existing analog-to-digital (A/D) converter technology. State-of-the-art A/D rates and high effective number of bits result in rapidly increasing cost and power consumption for the radar system. Therefore, it is useful to consider compressive sensing methods that enable reduced sampling rate, and in many applications, prior knowledge of signals of interest can be learned from training data and used to design better compressive measurement kernels. In this paper, we use a task-specific information-based approach to optimizing sensing kernels for highresolution radar range profiling of man-made targets. We employ a Gaussian mixture (GM) model for the targets and use a Taylor series expansion of the logarithm of the GM probability distribution to enable a closed-form gradient of information with respect to the sensing kernel. The GM model admits nuisance parameters such as target pose angle and range translation. The gradient is then used in a gradient-based approach to search for the optimal sensing kernel. In numerical simulations, we compare the performance of the proposed sensing kernel design to random projections and to lower-bandwidth waveforms that can be sampled at the Nyquist rate. Simulation results demonstrate that the proposed technique for sensing kernel design can significantly improve performance.**

***Index Terms*—Compressive sensing (CS), Gaussian mixture (GM), optimal sensing matrix, radar profiling, task-specific information (TSI).**

I. Introduction

With the continued development of radar technology, more wideband signals are being used to improve system performance and range resolution. A typical radar system transmits a modulated wideband pulse, such as a linear frequency modulation (LFM) waveform, and then correlates the received signal with the transmit waveform via an implementation of the matched filter, which achieves pulse compression. However, for a wideband waveform, Nyquist-rate pulse compression generally requires a high-rate analog-to-digital (A/D) converter. Alternative approaches such as LFM waveforms with stretch processing or the use of stepped frequency waveforms exploit waveforms that trace out a wide bandwidth over time. Therefore, reduced-rate sampling can be performed, but at the expense of a large overhead in collection time. Hence, achieving adequate A/D conversion for an instantaneously wideband radar signal imposes severe challenges on the acquisition hardware in terms of high sampling rate and high dynamic range. In many cases, the required A/D technology is too costly and consumes too much power. In some extreme cases, the required combination of sample rate and bit depth may not exist; therefore, it is beneficial to consider acquisition of radar signals at lower, sub-Nyquist sampling rates.

Motivated by sampling limitations across many domains, compressive sensing (CS) [1] theory has received wide attention in the past decade because it can be used to acquire and accurately recover signals at a far lower sampling rate than the standard Nyquist-Shannon sampling rate. In general, signal reconstruction from undersampled measurements is a well-known ill-conditioned problem, and extra constraints must be adopted in order to regularize the solution. Broadly speaking, CS theory states that a signal can be accurately reconstructed if the signal is sparse in some basis, which is generally characterized by the fact that the signal has only a few non-zero coefficients when represented in that basis. For such a signal, a basic requirement of the measurements in the framework of CS theory is that the sensing kernel must be incoherent with the representation basis of the signal, which leads to accurate signal reconstruction with high probability [1], [2]. It is well known by now that random measurement kernels, such as Gaussian and Bernoulli kernels, meet this incoherence requirement [1], [2], and measurements taken with random kernels are now routinely called random projections. However, random projections do not exploit any prior knowledge of the signal beyond sparsity. In practical radar applications, prior knowledge of the signal is usually available. For example, training samples can be used to obtain statistical characteristics of the signal, which then introduce the possibility of optimizing the sensing kernels according to a metric that is appropriate for the radar's exploitation task.

Task-specific information (TSI) is a metric for sensor analysis and design that was first proposed in [3], [4] as an analysis tool for feature extraction in imaging systems. The TSI formulation maximizes the Shannon mutual information between the measurements and a task-specific source variable. In an estimation task, the source variable can be a vector representing the signal to be recovered or a parameter vector that parameterizes the received signals. In hypothesis testing problems, the source variable is typically a discrete random variable representing the hypothesis class labels. Unlike the common CS approach of using random sensing kernels, in this paper we develop and apply the TSI metric to the problem of sensing kernel design for radar range-only profiling and target recognition. The TSI metric is

Manuscript received September 06, 2013; revised January 15, 2014 and April 21, 2014; accepted April 22, 2014. Date of current version May 20, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Tabrikian. This work was supported in part by the Defense Advanced Research Projects Agency via grant #N66001-10-1-4079.

Y. Gu and N. A. Goodman are with the School of Electrical and Computer Engineering and the Advanced Radar Research Center, University of Oklahoma, Norman, OK, 73019 USA (e-mail: guyujie@ou.edu; goodman@ou.edu).

A. Ashok is with the College of Optical Sciences, University of Arizona, Tucson, AZ, 85721 USA (e-mail: ashoka@optics.arizona.edu).

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

Digital Object Identifier 10.1109/TSP.2014.2323022

valid and useful for sensor analysis and design because of its relationship to mean-square error (MSE) via conditional entropy [5] and its relationship to probability of error (PE) via Fano's inequality [6]. In many cases, MSE/PE cannot be used directly as an optimization metric without knowing and having the ability to compute the optimal estimator/detector; therefore, information-based approaches can sometimes lead to more useful expressions for sensor performance.

Our goal is to maximize range profiling fidelity via kernel design and prior signal models. Many real signals are better described by a mixture probability distribution than by a single probability distribution. Among mixture distributions, the Gaussian mixture (GM) distribution is a typical one because of its flexibility and tractability. A GM distribution describes signals using a collection of Gaussians, and in most cases the mixture can be made arbitrarily accurate by increasing the number of components. In previous research, GM models have been effective in describing practical signals for various applications, including image inverse problems and statistical CS [7]–[11]. In our case, the GM enables modeling of a compressible target structure through low-rank covariance matrices while allowing that low-rank structure to vary across components in the mixture. By relating different components in the mixture to different intervals of unknown target parameters such as orientation angle, the GM distribution enables modeling of compressible target profiles that are parameterized by physical parameters. Therefore, the GM representation also allows prior distributions on target parameters (such as range or orientation) to be used in the kernel optimization and profiling processes. Such prior information might be derived, for example, from previous track information. Finally, other representations of structured sparse signals are available [12], such as tree-based sparsity [13] and block sparse models [14]. However, the GM mixture approach is more suitable for the purposes here due to the fact that it enables measurement kernel optimization through an information-based metric as well as closed-form signal reconstruction.

In order to optimize the sensing kernel, we derive the approximate gradient of TSI with respect to the sensing matrix using a first-order Taylor series expansion of the logarithm of the probability distribution. The approximate gradient is then used in a gradient-based search to find the optimal sensing kernel. Our simulation results are generated from high-fidelity modeling of three different targets: F-16 and F-18 fighter aircraft, and an A-10 attack aircraft. These simulation results demonstrate the performance improvement of the GM optimized-based sensing kernel design compared with that of random sensing kernels both in radar range profiling and target recognition [15], [16].

The rest of the paper is organized as follows. In Section II, we briefly describe the radar signal model and present the minimum mean-square error (MMSE) range profile estimator based on the GM assumption. In Section III, we propose an optimal sensing kernel design method based on GM priors, and in Section IV, we compare the performance of the proposed sensing kernel with that of random sensing kernels and Nyquist performance. We make our conclusions in Section V.

II. Signal and Target Models

In a radar system, the complex baseband signal $\mathbf{r}(t)$ reflected from a target can be modeled as a convolution [17] between the transmitted baseband radar pulse $\boldsymbol{\psi}(t)$ and the baseband target impulse response $\mathbf{x}(t)$ as

$$\begin{aligned}\mathbf{r}(t) &= \boldsymbol{\psi}(t) * \mathbf{x}(t) + \mathbf{n}(t) \\ &= \int \boldsymbol{\psi}(t-\tau)\mathbf{x}(\tau)d\tau + \mathbf{n}(t),\end{aligned} \tag{1}$$

where $\mathbf{n}(t)$ is complex additive Gaussian noise. Traditionally, a band-limited measurement of the radar target $\mathbf{x}(t)$ can be obtained by pulse compression, that is, by correlating the received signal $\mathbf{r}(t)$ with the transmit pulse $\boldsymbol{\psi}(t)$ in a matched filter. This matched filtering can be performed in either the analog or digital domains, but both implementations require a high-rate A/D converter, as seen in Figs. 1(a) and (b) [18].

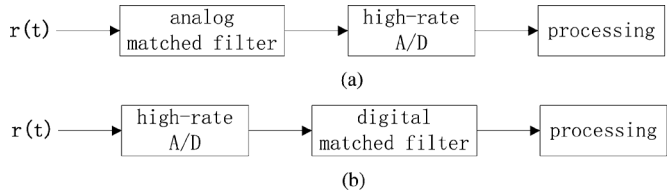


Fig. 1. Traditional radar receiver perform matched filtering either in the (a) analog or (b) digital domain.

We note that the target signals $\mathbf{x}(t)$ for man-made targets are typically compressible in some representation basis. For example, the target impulse response might be dominated by only a few dominant reflectors, resulting in a target response that is compressible in the time domain. CS theory indicates that sparse signals can be acquired and recovered with many fewer samples than suggested by the Nyquist sampling rate. Although our signals of interest are compressible, not sparse, CS theory still indicates the potential for measuring and encoding only M linear projections of the received signal $\mathbf{r}(t)$ using a set of measurement kernels $\{\boldsymbol{\phi}_m, m = 1, \cdots, M\}$, where M is much less than the number of samples required at the Nyquist rate. The mth data sample is obtained through a projection of the noisy received signal $\mathbf{r}(t)$ onto the mth measurement kernel $\boldsymbol{\phi}_m(t)$ as

$$\begin{aligned}y_m &= \langle \boldsymbol{\phi}_m(t), \mathbf{r}(t) \rangle \\ &= \int_0^T \boldsymbol{\phi}_m(t)\mathbf{r}(t)dt \\ &\approx \Delta \sum_{l=1}^{L} \phi_m[l] r[l] \\ &= \Delta \sum_{l=1}^{L} \phi_m[l] \left[\int_0^{T_x} \boldsymbol{\psi}(l\Delta - \tau)\mathbf{x}(\tau)d\tau + \mathbf{n}[l] \right] \\ &\approx \Delta \sum_{l=1}^{L} \phi_m[l] \left[\Delta \sum_{n=1}^{N_x} \psi[l-n]x[n] + \mathbf{n}[l] \right],\end{aligned} \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product, T is the maximum duration of the reflected signal (which depends on both the waveform and target impulse response durations), $L = T/\Delta$, T_x is the maximum duration of the target impulse response, $N_x = T_x/\Delta$, and the integrals have been approximated by dividing into discrete intervals of width Δ. If the interval Δ is much smaller than the reciprocal of the bandwidths of both the measurement kernel and the waveform, then the signals do not vary

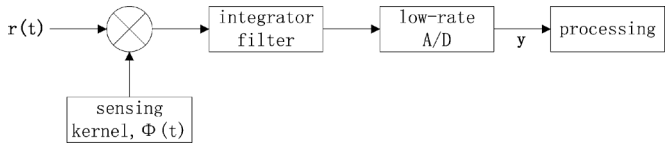


Fig. 2. Compressive radar receiver.

significantly within a discrete interval and the approximation will be accurate.

Stacking the measurements $\{y_m, m = 1, \ldots, M\}$ gives

$$\mathbf{y} = \mathbf{\Phi}(\mathbf{\Psi}\mathbf{x} + \mathbf{n}), \tag{3}$$

where $\mathbf{y} \in \mathbb{C}^M$ is the measurement vector, $\mathbf{\Phi} \in \mathbb{C}^{M\times L}$ is the sensing matrix with mth row comprising the sampled measurement kernel $\boldsymbol{\phi}_m[l]$, $\mathbf{\Psi} \in \mathbb{C}^{L\times N_x}$ is the discrete waveform matrix with nth column comprising the sampled waveform delayed by $(n-1)\Delta$, $\mathbf{x} \in \mathbb{C}^{N_x}$ is the unknown target profile (henceforth, the signal) to be estimated, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_{nn}})$ is zero-mean complex Gaussian noise independent of signal $\mathbf{x}$. Fig. 2 is the system block diagram of a compressive radar receiver. Our goal is to design the sensing matrix $\mathbf{\Phi}$ (and equivalently the analog compression kernels $\boldsymbol{\phi}_m(t)$) to recover the signal $\mathbf{x}$ as accurately as possible from the measurement $\mathbf{y}$. Defining the waveform bandwidth as B, the underlying dimensionality of the received signal $\mathbf{r}(t)$ is $N = BT$, and the compressive sampler should have $M \ll N$. We also note that the sensing matrix operates on the noise in addition to the signal, which must be the case for a radio frequency (RF) sub-Nyquist sampling system where the signal is carried on transmission lines and the compression is implemented in analog electronics. One significant consequence of (3) is that RF sub-Nyquist systems suffer a signal-to-noise ratio (SNR) penalty compared to Nyquist sampling due to noise folding [19], [20].

The sensing kernel $\mathbf{\Phi}$ must enable the estimation of the length-N_x signal $\mathbf{x}$ from the length-M measurement $\mathbf{y}$, which is generally ill-conditioned because $M \ll N_x$. However, in the case where $\mathbf{x}$ is known to be sparse, the sensing kernel $\mathbf{\Phi}$ can be selected as a random matrix [1], [2], [21]. In such a case, both the restricted isometry property (RIP) [1] and the required incoherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$ can be achieved with high probability. When the RIP/incoherence holds for a given level of signal sparsity, the signal $\mathbf{x}$ can be recovered by solving a sparsity-constrained optimization problem that has a unique solution [22].

These fundamental theorems of CS do not make assumptions regarding prior knowledge of the signal $\mathbf{x}$ except for its sparsity. However, in some practical applications including radar range profiling, the distribution of the scattered signal for targets of interest can be learned from training samples; hence, the distribution can be regarded as (approximately) known *a priori*. Specifically, in this paper we model the target range profiles with a Gaussian mixture distribution where each component in the mixture has a low-rank covariance matrix. Each component of the mixture represents a compressible target profile learned over some small window of target orientation and range translation. Taken together, the components in the mixture can describe the compressible target over a wide range of target parameters. This compressible representation then admits compressive sampling followed by MMSE reconstruction.

Let the probability density function (pdf) of signal $\mathbf{x}$ be modeled by a K-component GM given by:

$$\mathbf{x} \sim \sum_{k\in\mathcal{K}} p_k \mathcal{CN}\left(\mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)}\right), \tag{4}$$

where $\sum_{k\in\mathcal{K}} p_k = 1$, and the cardinality (denoted by $|\cdot|$) of the set $\mathcal{K} = \{1, 2, \ldots, K\}$ is $|\mathcal{K}| = K$. The GM distribution means that the kth component is activated with probability $p_k \geq 0$, and when activated, that component generates a complex-valued Gaussian signal with distribution law $\mathcal{CN}(\mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)})$. The set $\{p_k, \mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)}; k \in \mathcal{K}\}$ defines the GM parameters of $\mathbf{x}$. Under (4), it can be shown that $\mathbf{y}$ in (3) is also a GM with $|\mathcal{K}|$ components, and its pdf has the form

$$f(\mathbf{y}) = \sum_{k\in\mathcal{K}} p_k f^{(k)}(\mathbf{y}), \tag{5}$$

where the kth component $f^{(k)}(\mathbf{y}) = \mathcal{CN}(\mathbf{u}_{\mathbf{y}}^{(k)}, \mathbf{C}_{\mathbf{yy}}^{(k)})$ is a Gaussian density with mean and covariance

$$\begin{aligned} \mathbf{u}_{\mathbf{y}}^{(k)} &= \mathbf{\Phi}\mathbf{\Psi}\mathbf{u}_{\mathbf{x}}^{(k)} \\ \mathbf{C}_{\mathbf{yy}}^{(k)} &= \mathbf{\Phi}\left(\mathbf{\Psi}\mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{\Psi}^H + \mathbf{C_{nn}}\right)\mathbf{\Phi}^H, \end{aligned} \tag{6}$$

where $(\cdot)^H$ denotes the Hermitian transpose.

When the signal $\mathbf{x}$ is a GM, the MMSE estimator defined by

$$\min_{\mathbf{x}} \; E\left\{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right\} \tag{7}$$

is given by [23]

$$\hat{\mathbf{x}} = E\{\mathbf{x}|\mathbf{y}\} = \sum_{k\in\mathcal{K}} p_{k|\mathbf{y}} \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k)} \tag{8}$$

where $E\{\cdot\}$ denotes statistical expectation,

$$p_{k|\mathbf{y}} = \frac{p_k f^{(k)}(\mathbf{y})}{f(\mathbf{y})} \tag{9}$$

is the posterior probability of the kth component given the measurement $\mathbf{y}$, and

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k)} = \mathbf{u}_{\mathbf{x}}^{(k)} + \mathbf{C}_{\mathbf{xx}}^{(k)}(\mathbf{\Phi}\mathbf{\Psi})^H \left(\mathbf{C}_{\mathbf{yy}}^{(k)}\right)^{-1}\left(\mathbf{y} - \mathbf{u}_{\mathbf{y}}^{(k)}\right) \tag{10}$$

is the kth MMSE estimator component given the measurement $\mathbf{y}$ [24].

Because we assume a prior distribution on the signal $\mathbf{x}$, we use the above MMSE estimator in our performance evaluations. Unlike the common solutions in the CS theory, the MMSE-based estimator (8) is analytical for a GM-distributed target signal; furthermore, the estimator and its performance clearly depend on the sensing kernel $\mathbf{\Phi}$. In contrast to randomly generated sensing kernels, the sensing kernel $\mathbf{\Phi}$ can now be optimized based on *a priori* knowledge of the signal.

Finally, we comment on structure imposed on the sensing matrix by the architecture in Fig. 2. At any given sample instant of the A/D converter, the resulting sample is a consequence of the

most recent T_{int} seconds of data where T_{int} is the effective integration time of the preceding filter. If the sampling interval is matched to this integration time, then successive samples result from approximately non-overlapping intervals of time, which translates to non-overlapping rows of the sensing matrix. In other words, the sensing matrix corresponding to the implementation in Fig. 2 should have an envelope with block structure such as in

$$\begin{bmatrix} \underbrace{1\cdots1}_{\lfloor\frac{L}{M}\rfloor} & \underbrace{0\cdots0}_{\lfloor\frac{L}{M}\rfloor} & \underbrace{0\cdots0}_{L-2\lfloor\frac{L}{M}\rfloor} \\ \underbrace{0\cdots0}_{\lfloor\frac{L}{M}\rfloor} & \underbrace{1\cdots1}_{\lfloor\frac{L}{M}\rfloor} & \underbrace{0\cdots0}_{L-2\lfloor\frac{L}{M}\rfloor} \\ \vdots & \ddots & \vdots \\ \underbrace{0\cdots0}_{L-2\lfloor\frac{L}{M}\rfloor} & \underbrace{0\cdots0}_{\lfloor\frac{L}{M}\rfloor} & \underbrace{1\cdots1}_{\lfloor\frac{L}{M}\rfloor} \end{bmatrix},$$

where $\lfloor\cdot\rfloor$ denotes the floor function. Later in the paper, we use this structure as a mask to enforce physical hardware constraints in our gradient search optimization of the measurement kernel. Because of the tighter constraints on structure, performance of kernels subject to this masking operation may be lower than for parallel kernels where the structure is not enforced. Implementation of parallel kernels requires the signal to be split, which has varying impact on SNR depending on the number of branches in the receiver and the quality of the initial low-noise amplifier. An analysis of these architecture tradeoffs in the context of radar target detection has been performed in [25].

III. Optimal Sensing Kernel Based on *a priori* Knowledge

In order to design an optimal sensing kernel, we define TSI for the range profiling problem as the Shannon mutual information $I(\mathbf{y};\mathbf{x})$ between the source signal $\mathbf{x}$ and the measurement $\mathbf{y}$ according to [6]

$$\text{TSI} \equiv I(\mathbf{y};\mathbf{x}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}), \tag{11}$$

where $H(\mathbf{y}) = -E\{\log[f(\mathbf{y})]\}$ denotes the entropy of the measurement $\mathbf{y}$, and $H(\mathbf{y}|\mathbf{x}) = -E\{\log[f(\mathbf{y}|\mathbf{x})]\}$ denotes the entropy of the measurement $\mathbf{y}$ conditioned on the signal $\mathbf{x}$. It is intractable to compute the TSI in (11) due to the dimensionality of the problem and non-Gaussianity of parameters.

Our objective is to maximize TSI over all allowable sensing matrices, i.e.,

$$\max_{\mathbf{\Phi}} I(\mathbf{y};\mathbf{x}).$$

Because the optimization variable $\mathbf{\Phi}$ is a matrix, dimensionality is high and a full numerical search is numerically intractable. We prefer a gradient-based search method, which requires the calculation of

$$\nabla_{\mathbf{\Phi}} I(\mathbf{y};\mathbf{x}) = \nabla_{\mathbf{\Phi}} H(\mathbf{y}) - \nabla_{\mathbf{\Phi}} H(\mathbf{y}|\mathbf{x}). \tag{12}$$

Unfortunately, the entropy calculation also does not have a closed form for most pdf's of interest in practical radar applications, including the GM model we have assumed in this paper. Instead, our approach is to substitute in the GM pdf and to perform a Taylor series expansion of the logarithm of the pdf required in the entropy definition. After this expansion, we can compute the gradient of the approximated entropy expression with respect to the sensing matrix, which will be used in a gradient-based sensing kernel optimization process.

First, substituting the GM pdf (5) of the measurement vector $\mathbf{y}$ into the definition of entropy for the measured data, we have

$$\begin{aligned} H(\mathbf{y}) &= -\int \sum_{q=1}^{K} p_q f^{(q)}(\mathbf{y}) \log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})\right] d\mathbf{y} \\ &= -\sum_{q=1}^{K} p_q \int f^{(q)}(\mathbf{y}) \log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})\right] d\mathbf{y} \\ &\approx -\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y_0})\right], \end{aligned} \tag{13}$$

where $\mathbf{y}_0$ is the mean value of the measurement and the full derivation can be found in the Appendix.

Next, we make a zero-mean assumption specific to the radar range profiling application. In this application, changes in the target's range on the order of a wavelength cause the phase of the reflected signal to vary dramatically. Although it may be possible to estimate a target's range to within the radar's range resolution, the radar's bandwidth is often at least an order of magnitude smaller than the center operating frequency. Thus, even if the target's position is estimated to sub-range-resolution accuracy, the range uncertainty is still on the order of a wavelength or more, making the overall global phase of the reflected signal completely uncertain. For completely random phase, the mean of the received signal is zero; consequently, we set $\mathbf{u}_{\mathbf{y}}^{(k)} = \mathbf{0}$ for all individual Gaussian components of $\mathbf{y}$. For individual Gaussian components that are all zero mean, it makes sense to set the Taylor Series expansion point to $\mathbf{y}_0 = \mathbf{0}$, resulting in

$$\begin{aligned} f^{(k)}(\mathbf{y_0}) &= \frac{1}{\pi^M \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|} \left.\mathrm{e}^{-\left(\mathbf{y}-\mathbf{u}_{\mathbf{y}}^{(k)}\right)^H \left[\mathbf{C}_{\mathbf{yy}}^{(k)}\right]^{-1}\left(\mathbf{y}-\mathbf{u}_{\mathbf{y}}^{(k)}\right)}\right|_{\mathbf{y}=\mathbf{u}_{\mathbf{y}}^{(k)}=\mathbf{0}} \\ &= \frac{1}{\pi^M \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|}. \end{aligned} \tag{14}$$

We now have

$$H(\mathbf{y}) \approx -\log\left[\sum_{k=1}^{K} p_k \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|^{-1}\right] + M\log\pi \tag{15}$$

where the second term is a constant independent of $\mathbf{\Phi}$.

Now that we have an approximation to the entropy of $\mathbf{y}$, we take the gradient of the resulting expression in (15) with respect to the sensing matrix $\mathbf{\Phi}$. Using the chain rule for derivatives of logarithms, we have

$$\begin{aligned} &\nabla_{\mathbf{\Phi}} H(\mathbf{y}) \\ &\quad\approx \nabla_{\mathbf{\Phi}}\left\{-\log\left[\sum_{k=1}^{K} p_k \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|^{-1}\right]\right\} \\ &\quad= -\frac{\sum_{k=1}^{K} p_k \nabla_{\mathbf{\Phi}}\left\{\left|\mathbf{\Phi}\left(\mathbf{\Psi}\mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{\Psi}^H + \mathbf{C}_{\mathbf{nn}}\right)\mathbf{\Phi}^H\right|^{-1}\right\}}{\sum_{k=1}^{K} p_k \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|^{-1}}. \end{aligned} \tag{16}$$

Finally, using properties of matrix derivatives [26], (16) can be written as (17), shown at the bottom of the page, where the denominator is a positive, real scalar that impacts the overall magnitude of the gradient, but not the direction. Thus, as approximated with the Taylor expansion, the gradient of the entropy of the measurement with respect to the sensing matrix for a GM prior is a weighted set of gradients. The individual gradient components are scaled by a combination of a) the probability of that component in the mixture, and b) the interaction between the current value of the sensing matrix with that component's structure (via the determinant term).

Considering the second term $H(\mathbf{y}|\mathbf{x})$ in (11), when the source signal $\mathbf{x}$ is given, the only random contribution is due to noise. That is to say,

$$H(\mathbf{y}|\mathbf{x}) = H\left((\mathbf{\Phi\Psi x} + \mathbf{\Phi n})|\mathbf{x}\right) = H(\mathbf{\Phi n}|\mathbf{x}) = H(\mathbf{\Phi n}), \quad (18)$$

because $\mathbf{\Phi n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi C_{nn} \Phi}^H)$ is independent of $\mathbf{x}$. The entropy of this term is

$$\begin{aligned} H(\mathbf{\Phi n}) &= \log\left|\pi e \mathbf{\Phi C_{nn} \Phi}^H\right| \\ &= \log\left|\mathbf{\Phi C_{nn} \Phi}^H\right| + M\log(\pi e) \end{aligned} \quad (19)$$

where the second term is a constant independent of $\mathbf{\Phi}$. Using the chain rule for derivatives of logarithms, the gradient of the entropy of the noise term with respect to the sensing kernel is

$$\begin{aligned} \nabla_{\mathbf{\Phi}} H(\mathbf{\Phi n}) &= \nabla_{\mathbf{\Phi}} \left\{\log\left|\mathbf{\Phi C_{nn} \Phi}^H\right|\right\} \\ &= \left[\mathbf{\Phi C_{nn} \Phi}^H\right]^{-1} \mathbf{\Phi C_{nn}}. \end{aligned} \quad (20)$$

Substituting (17) and (20) into (12), the approximated gradient of TSI with respect to the sensing kernel is (21), shown at the bottom of the page, which can be used in a gradient-based search process according to

$$\tilde{\mathbf{\Phi}} = \mathbf{\Phi} + \gamma \nabla_{\mathbf{\Phi}} I(\mathbf{y}; \mathbf{x}), \quad (22)$$

where $\gamma > 0$ is a small stepsize. In the gradient-based search method, an appropriate stepsize is important for the convergence of the iterations. The stepsize can be fixed or adaptive. During the sensing kernel optimization process, the updated sensing kernel $\tilde{\mathbf{\Phi}}$ is substituted for $\mathbf{\Phi}$ in the next iteration of calculating (17) and (20) and then updated using (22) to achieve an iterative search procedure.

When the complex additive Gaussian noise $\mathbf{n}$ is added prior to compression as in the RF model of (3), then the sensing matrix operates on the noise as well as the signal. Therefore, there is no benefit to having non-orthonormal measurement kernels (even in the case of colored noise), and henceforth we assume $\mathbf{\Phi\Phi}^H = \mathbf{I}$). If we further assume that the additive noise is white (namely, $\mathbf{C_{nn}} = \mathbf{I}$), then $H(\mathbf{\Phi n})$ in (19) is a constant independent of $\mathbf{\Phi}$. In the white-noise case, any additional structure of $\mathbf{\Phi}$ other than the orthonormal assumption has no impact on the entropy of the output noise [27]. Hence, the gradient of the entropy of the noise term in (20) can be simplified as

$$\nabla_{\mathbf{\Phi}} H(\mathbf{\Phi n}) = \mathbf{\Phi}. \quad (23)$$

Meanwhile, the gradient-based update equation (22) can be rearranged as

$$\tilde{\mathbf{\Phi}} = (1-\gamma)\mathbf{\Phi} + \gamma \nabla_{\mathbf{\Phi}} H(\mathbf{y}), \quad (24)$$

which is a convex combination between the current sensing kernel and the gradient of the entropy of the measurement with respect to the current sensing kernel. From (23) and (24), we can conclude that TSI cannot be increased simply by scaling $\mathbf{\Phi}$ to be larger. Because the noise is added prior to compression, scaling $\mathbf{\Phi}$ results in proportional scaling of both the signal and noise.

In summary, the sensing matrix optimization method performs the following iterative process:

Step 1: Initialize the sensing matrix $\mathbf{\Phi}$;

$$\begin{aligned} \nabla_{\mathbf{\Phi}} H(\mathbf{y}) &= \frac{\sum_{k=1}^{K} p_k \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|^{-1} \left[\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right]^{-1} \mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})}{\sum_{k=1}^{K} p_k \left|\mathbf{C}_{\mathbf{yy}}^{(k)}\right|^{-1}} \\ &= \frac{\sum_{k=1}^{K} p_k \left|\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right|^{-1} \left[\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right]^{-1} \mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})}{\sum_{k=1}^{K} p_k \left|\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right|^{-1}}, \end{aligned} \quad (17)$$

$$\nabla_{\mathbf{\Phi}} I(\mathbf{y}; \mathbf{x}) = \frac{\sum_{k=1}^{K} p_k \left|\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right|^{-1} \left[\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right]^{-1} \mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})}{\sum_{k=1}^{K} p_k \left|\mathbf{\Phi}(\mathbf{\Psi C}_{\mathbf{xx}}^{(k)} \mathbf{\Psi}^H + \mathbf{C_{nn}})\mathbf{\Phi}^H\right|^{-1}} - \left[\mathbf{\Phi C_{nn} \Phi}^H\right]^{-1} \mathbf{\Phi C_{nn}}, \quad (21)$$

Step 2: Calculate $\nabla_{\mathbf{\Phi}} H(\mathbf{y})$ and $\nabla_{\mathbf{\Phi}} H(\mathbf{\Phi n})$ via (17) and (20) to obtain an approximate gradient of TSI with respect to the sensing matrix $\nabla_{\mathbf{\Phi}} I(\mathbf{y};\mathbf{x})$ (21), evaluated at the current iteration of the sensing matrix;

Step 3: Calculate the updated sensing matrix $\tilde{\mathbf{\Phi}}$ via (22);

Step 4: Re-orthonormalize the rows of $\tilde{\mathbf{\Phi}}$, set $\mathbf{\Phi}$ to the result, and go to **Step 2**.

The convergence criterion of the iterative process is chosen as:

$$0 \leq \delta^{(j)} \leq \varepsilon \text{ and } \delta^{(j)} < \delta^{(j-1)}, \tag{25}$$

where

$$\delta^{(j)} = \left\| \nabla_{\mathbf{\Phi}}^{(j-1)} I(\mathbf{y};\mathbf{x}) \right\| - \left\| \nabla_{\mathbf{\Phi}}^{(j)} I(\mathbf{y};\mathbf{x}) \right\| \tag{26}$$

is the norm difference of the gradient of TSI, and the convergence parameter $\varepsilon > 0$ is a predefined small positive number. The convergence criterion requires that the gradient of TSI should converge to zero and the iterations $\delta^{(j)}$ should decrease step by step.

The present sensing matrix optimization process is not adaptive, which means the optimization of sensing kernel $\mathbf{\Phi}$ does not depend on the signal $\mathbf{x}$ and can be performed off-line. Finally, if the masking structure discussed earlier should be applied to $\mathbf{\Phi}$ in order to represent physical hardware constraints, then the updated sensing matrix $\tilde{\mathbf{\Phi}}$ (24) should be *masked* before its re-orthonormalization in **Step 4**. The masking operation is performed by element-wise multiplication of $\tilde{\mathbf{\Phi}}$ with a masking matrix between **Step 3** and **Step 4**. The row re-orthonormalization of $\tilde{\mathbf{\Phi}}$ in **Step 4** is an orthogonal Procrustes problem [28], which is easy to solve using the singular value decomposition (SVD) of $\tilde{\mathbf{\Phi}}$ to determine the row orthonormal matrix closest to $\tilde{\mathbf{\Phi}}$. In the case of the masking structure, the rows of $\tilde{\mathbf{\Phi}}$ are already orthogonal and only need normalization.

From (21), the computational complexity of calculating the TSI gradient with respect to the sensing kernel is $\mathcal{O}(KN_xL^2)$ because $N_x < L$. Hence, the overall computational complexity of the proposed compressive sensing kernel optimization is $\mathcal{O}(JKN_xL^2)$, where J denotes the number of iteration.

IV. Simulation

We now present simulation results that demonstrate the performance improvement achievable through the proposed sensing kernel design method. A GM model for targets of interest is learned from a target template library by grouping together templates with similar target parameters (i.e., pose angle) into Gaussian components. The GM distribution is then used for both the sensing matrix optimization and for the minimum mean-square estimator.

We used finite-difference time-domain (FDTD) electromagnetic modeling software to perform time-domain modeling of several targets. We used publicly available CAD models for an F-16 fighter, F-18 fighter, and A-10 attack aircraft [29]. These aircraft have different physical sizes and average radar cross sections (RCS). We used FDTD to calculate wideband time-domain responses from these targets at 0.1° intervals in azimuth from head-on (0°) to sidelooking (90°). The aspect elevation angle was held constant; therefore, we obtained 901 profiles per target. Because the FDTD software automatically sets the simulation timestep, we normalized the timescales of the various FDTD results such that each profile in the target template library was generated with the same effective bandwidth. We also normalized the magnitudes of the responses to correspond to published RCS levels: 1 m^2 for the F-18, 2 m^2 for the F-16, and 5 m^2 for the A-10. Fig. 3 shows the target templates for each target at 901 different azimuth angles. The total number of templates over the three targets was 2703.

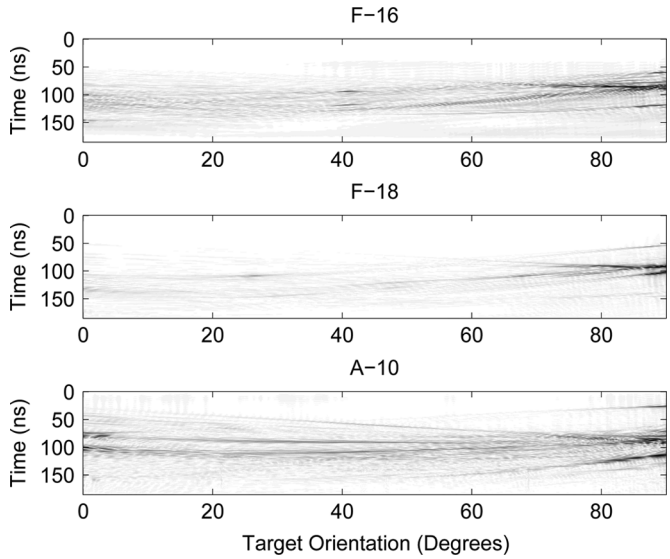


Fig. 3. Time-domain impulse responses of three aircraft.

In order to learn an approximate GM distribution from the 2703 target templates, we assumed that target profiles from the same target at similar aspect angles could be grouped together into individual components of the Gaussian mixture. The discrete-time target impulse response for a particular target at a particular azimuth angle can be viewed as a vector in N_x-dimensional space. As the azimuth angle varies continuously, this vector traces out a one-dimensional non-linear manifold in N_x-dimensional space. A piecewise linear approximation to the non-linear manifold can be obtained by grouping together profiles corresponding to similar aspect angles. For example, as the target azimuth angle varies from 0° to 1°, the target impulse response vector traces out a short, contiguous piece of the overall target manifold. If we use the impulse responses from this short piece of manifold to compute a sample covariance matrix, and then use that covariance matrix for the first Gaussian component in the mixture, then we can view the first component as a linear approximation to the first segment of the non-linear manifold [9]. Likewise, if we repeat the process to obtain a second Gaussian component created from profiles covering angles from 1° to 2°, we have a linear approximation to the second piece of the target manifold. Continuing this process, we can obtain a piecewise linear approximation to the overall target manifold. For our set of profiles ranging from 0° to 90°, we have divided the target manifold into segments corresponding to 1-degree increments of the target pose angle. Therefore, we obtained 90 components $(K = 90)$ per target in the Gaussian mixture, for a total of 270 components in the overall mixture. Although this training setup means that only

TABLE I
PARAMETERS OF THE FOUR TRANSMIT WAVEFORMS WITH 250 MHZ BANDWIDTH

Waveform Types	Bandwidth	Pulse Width	Nyquist Samples	Compressive Samples
Narrowband unmodulated pulse (NB AO)	16.7 MHz	0.06 μs	4	
Wideband unmodulated pulse (WB AO)	83.3 MHz	0.012 μs	17	3
LFM pulse (LFM)	83.3 MHz	0.132 μs	27	5
Length-11 Barker code (BC)	83.3 MHz	0.132 μs	27	5

10 training samples are used to estimate each Gaussian component, the effective rank of each component is much smaller due to the compressible nature of the target, and the number of training samples is sufficient.

In some cases the target templates may vary sharply for small changes in aspect angle, and the method of locally grouping templates into mixture components may not be the most accurate method for fitting the GM model to the training set. However, the locally grouped approach allows for prior knowledge on the physical parameters to be mapped directly to the mixture weights. For example, suppose that the target has been detected and tracked previous to transmitting a wideband profiling waveform. Based on the track history, it may be possible to estimate the target's pose angle. Using this estimate, and knowing that there is some uncertainty to the estimate, it is possible to describe the target's orientation with a probability distribution, which can then be mapped into the probability weights of the mixture. If the target is believed to be oriented at a 45° angle, then the Gaussian mixture components around 45° should have the highest probability while mixture components corresponding to 0° or 90° should have weak probability. Therefore, target parameter estimates can be directly used to fine tune the GM target model for better sensing kernel design and profiling results.

For the Monte Carlo simulation study, our test signals were not drawn from the GM prior directly; instead, we picked one of the three target types with equal probability and then generated a target profile with randomly selected azimuth angle. Because the angle realization generally did not fall on one of the grid points where we had an existing target response, we used linear interpolation to obtain the test profile from the two nearest templates in the target library. While this approach resulted in model mismatch between the actual statistics of the target profile and the GM prior, this mismatch is an important factor to include in analyzing the proposed method. Finally, in order to model small range translations on the order of the radar wavelength, we assigned each test profile a random global phase generated from a uniform distribution.

After normalizing all target templates to correspond to the same effective bandwidth, the final templates were defined on a sampling interval of $T_s = 4$ ns. In other words, the target template library corresponded to wideband responses generated with an effective bandwidth of 250 MHz, which limited the maximum waveform bandwidth that could be simulated. We also wanted to ensure that any results we obtained represented the effects of kernel design and prior distribution modeling, not effects due to the fidelity of the input signal models. For example, if we used waveform bandwidths of 250 MHz in our simulations, any super-resolving capability of the compression kernels might be artificially limited by the fact that the source signals were limited to only 250 MHz. Therefore, we decided to limit bandwidth of the transmit waveform to one-third of the bandwidth of the input source, for a maximum waveform bandwidth of approximately 83.3 MHz.

In our simulations, we compared four different waveforms as described in Table I. The first waveform is a narrowband reference waveform, selected such that no compression is needed for a sampling rate of 16.7 MHz. This narrowband waveform is a simple, unmodulated pulse (NB AO), used as our Nyquist reference. The fundamental issue is whether performance can be improved by transmitting more bandwidth than the receiver can sample, such that compressive sampling must be used. To evaluate this issue, we simulated three additional waveforms having bandwidth five times higher than the receiver sampling rate. The compression ratio for these waveforms, obviously, is five. The first wideband waveform is a simple, unmodulated pulse (WB AO). The second is an LFM pulse (LFM), and the third is a phase-coded waveform according to a length-11 Barker code (BC).

Because of the different waveform bandwidths and because some waveforms are modulated while some aren't, the waveform pulse widths vary. Therefore, the maximum time duration of the received signal, which is the sum of the radar pulse width and the maximum duration of the target impulse response, varies for different waveforms. Consequently, the effective dimensionality of the received signal varies with waveforms, and the number of samples required at Nyquist is not constant. In order to compare waveforms, we chose to keep the sampling rate (and compression ratio) constant, which results in varying number of compressive samples as shown in Table I. Unfortunately, with different combinations of bandwidth and pulse width, there was no straightforward way to compare exactly the same compression ratio and number of compressive samples. Finally, in the process of sensing matrix optimization, the convergence parameter was chosen as $\varepsilon = 1 \times 10^{-6}$. For the stepsize parameter, we found both $\gamma = 0.1$ and $\gamma = 0.01$ to be good values, providing the same performance with reasonable convergence time. When we set a small stepsize, e.g., $\gamma = 0.001$, the optimization algorithm still had not converged after several thousand iterations. As Fig. 4 shows, the length of the target source signal $\mathbf{x}$ at the simulation interval was $N_x = 48$. For each data point (SNR), $N_{MC} = 10\,000$ Monte Carlo trials were performed. Fig. 4 shows a sample set of range profile reconstruction results for the narrowband reference and the Barker-coded waveform with both random and optimized compression. The magnitude of reconstructed profiles are shown in comparison to the true profile for the case where the target aspect angle was uniformly distributed from 0° to 90° and the target's range translation was uniformly distributed over five range resolution cells. It is apparent that the low-bandwidth reference waveform does not provide enough resolution to separate closely spaced target peaks in the presence of range uncertainty, but compres-

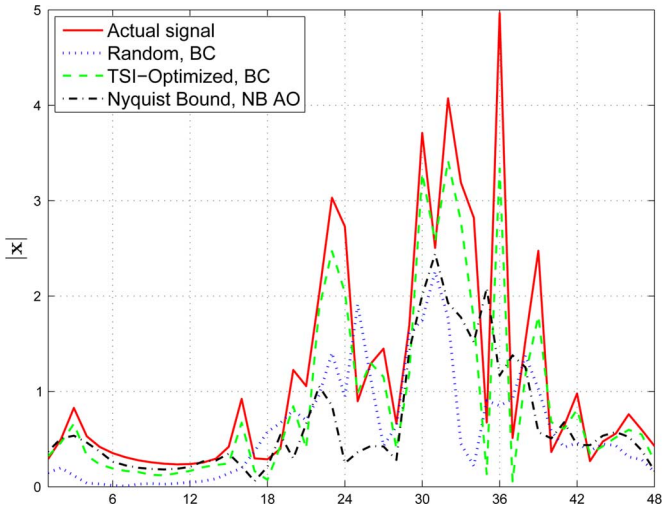


Fig. 4. Sample set of profile reconstruction results (shown in magnitude) for the case with range uncertainty.

sive sampling of the wider-bandwidth Barker Code improves detail in the reconstruction.

Fig. 5 shows the relative MSE of target range-only profiling for the case where the target aspect angle was uniformly distributed from 0° to 90°. All 270 components in the Gaussian mixture (90 per target) have the same weight $p_k = \frac{1}{270}, k \in \{1, 2, \ldots, 270\}$. The relative MSE is defined as

$$\frac{1}{N_{MC}} \sum_{d=1}^{N_{MC}} \frac{(\mathbf{x}(d) - \hat{\mathbf{x}}(d))^H (\mathbf{x}(d) - \hat{\mathbf{x}}(d))}{\mathbf{x}^H(d)\mathbf{x}(d)}, \qquad (27)$$

where $\hat{\mathbf{x}}(d)$ is the estimator of the dth Monte Carlo trial $\mathbf{x}(d)$. The MSE is normalized in this way in order to remove RCS differences between different targets as well as effects of target SNR scaling on absolute error levels. The normalization is independent of waveform and compression scheme; therefore, it does not impact conclusions about their relative performance. The MSE performance is shown as a function of SNR for both the TSI-optimized and random sensing matrices, where SNR is defined as a scale factor on the target profiles - starting from the normalized impulse responses, the templates were scaled by $\sqrt{SNR}$ to model varying receive power for varying range to the target. The results for the narrowband reference pulse with Nyquist sampling are depicted with the dashed line. In order to justify an operational scenario where additional bandwidth is transmitted, but sampled at the same rate as the narrowband waveform, the compressive results should produce lower errors than the narrowband reference. When the sensing matrix is allowed to have arbitrary structure, the TSI-optimized results do indeed outperform the narrowband reference. But when the sequential mask structure of the sensing matrix is enforced, the compressive LFM results lose much of their benefit and the compressive Barker-coded waveform actually performs worse than the narrowband reference. In both cases (with and without the masking structure), random compression performs worse than the narrowband reference. Fig. 6 shows the same type of results, but with target orientation limited uniformly between 40° and 70°. In Fig. 7, the target orientation is further limited to between 40° and 50°. As expected, performance improves with stronger prior knowledge of the scenario, but the qualitative conclusions regarding different waveforms and compression kernels remains unchanged.

The performance variation with type of waveform has to do with the amount of overlap between the received signal and various rows of the compression kernel. For relatively long, modulated waveforms the optimized compression kernel is able to sample variations in the waveform-target interaction. For example, the LFM waveform begins at the low end of its frequency range and gradually chirps faster. If one part of the target impulse response contains particularly useful information, the compressive receiver can observe the interaction of this part of the target with low frequencies at the beginning of the pulse and with high frequencies at the end of the pulse. In other words, multiple samples produced by the compressive receiver can result from different parts of the waveform structure interacting with the same part of the target. The unmodulated wideband pulse, however, is very short and does not produce much variation in the waveform-target interaction for the compression kernels to exploit. Furthermore, the absolute number of samples is very low. These limitations are especially noticeable in the results with sequential mask structure where each row of the compression kernel has a short time duration in which to capture the short pulse. In general, longer modulated pulses lend more flexibility to the measurement kernel design, especially when the structure of the measurement kernel is restricted.

In general, because the wideband transmit waveforms have higher bandwidth than the narrowband waveform, they should achieve improved resolution. But in Figs. 5 through 7, the wideband waveforms with compression perform only marginally better than the fully sampled narrowband waveform. The reason for the limited improvement is that the target's absolute range is treated as perfectly known, which minimizes the need for range resolution. In practical scenarios, existing track information may be able to approximately locate the target in range, but the range will always have uncertainty. Therefore, we now also consider range as a nuisance parameter. In the following simulations, range uncertainty is set to five range resolution cells, and target profiles are added to the target library in 0.1-resolution-cell increments. Therefore, the target library now has 51 templates for every aspect angle, each offset in range by 0.1 resolution cells. As before, we group templates into Gaussian mixture components according to similar parameters. Templates with up to one resolution cell in relative delay are grouped together, creating five mixture components in the range dimension. Combining the orientation and range uncertainties together, there are now 450 Gaussian components per target for the Gaussian mixture. For the results shown below, the width of the component groups in range and orientation have been chosen through experiment—we are currently researching the potential for using clustering algorithms to minimize the number of components in the mixture while retaining good profiling performance.

Similar to the results shown previously, Figs. 8 and 9 show the MSE performances for range-only profiling, but now with uncertainty in both orientation and range. In both cases, the total range uncertainty is five range resolution cells while the orientation uncertainty is $[0^\circ, 90^\circ]$ and $[40^\circ, 50^\circ]$, respectively. The

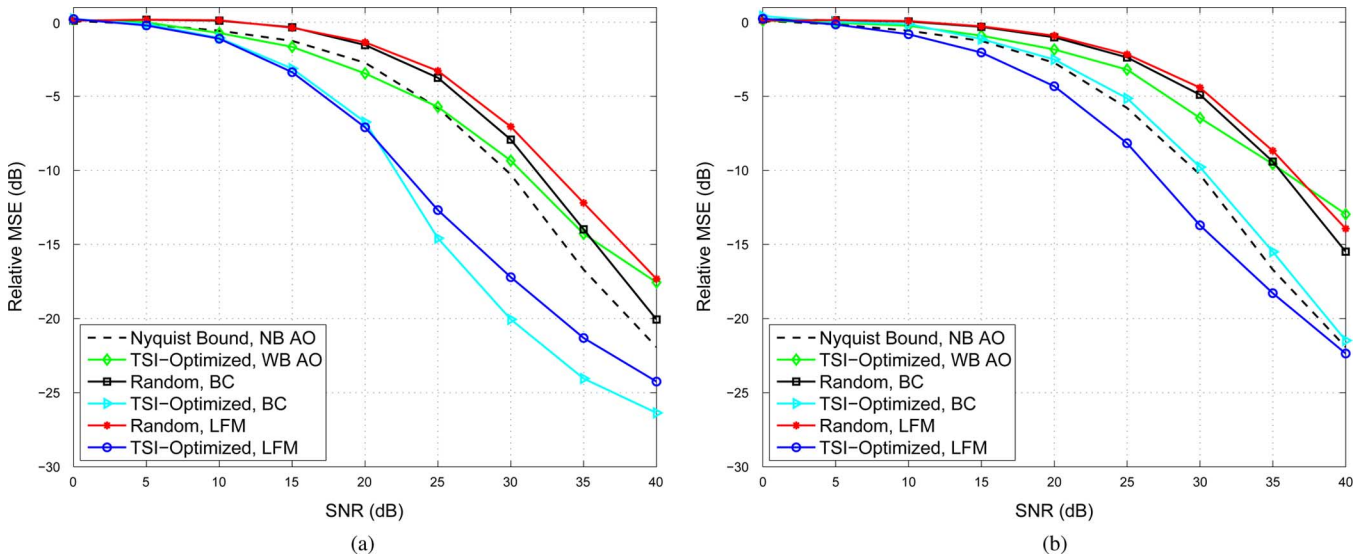


Fig. 5. Target range-only profiling performance, [0°, 90°]. (a) Without mask, (b) With mask.

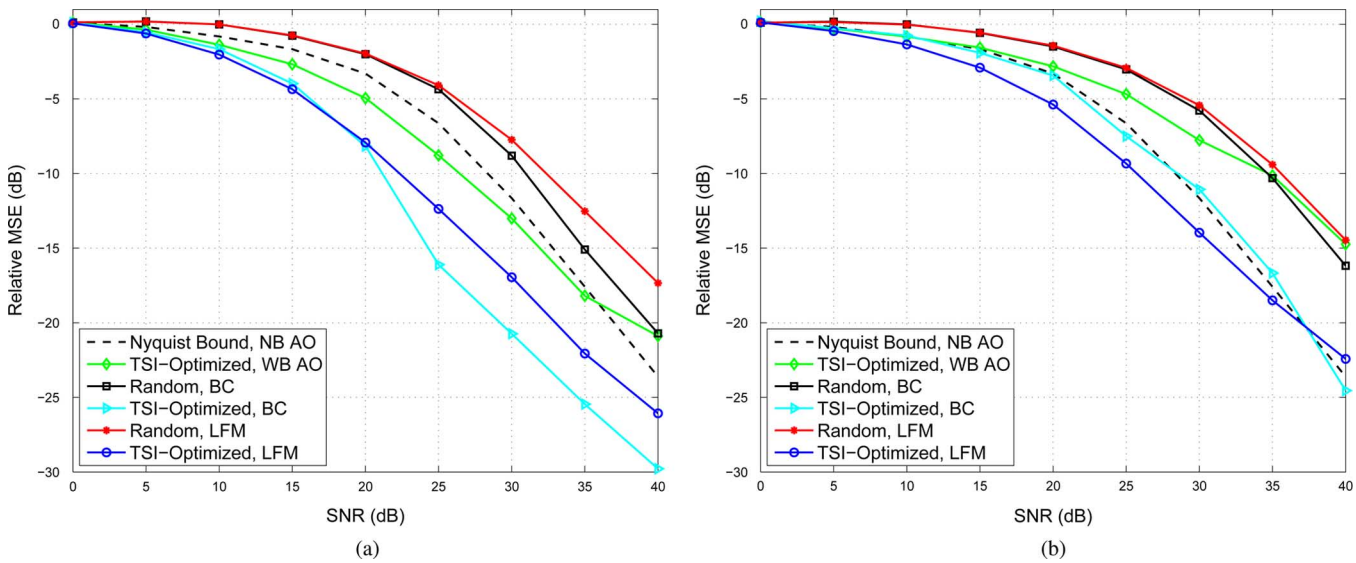


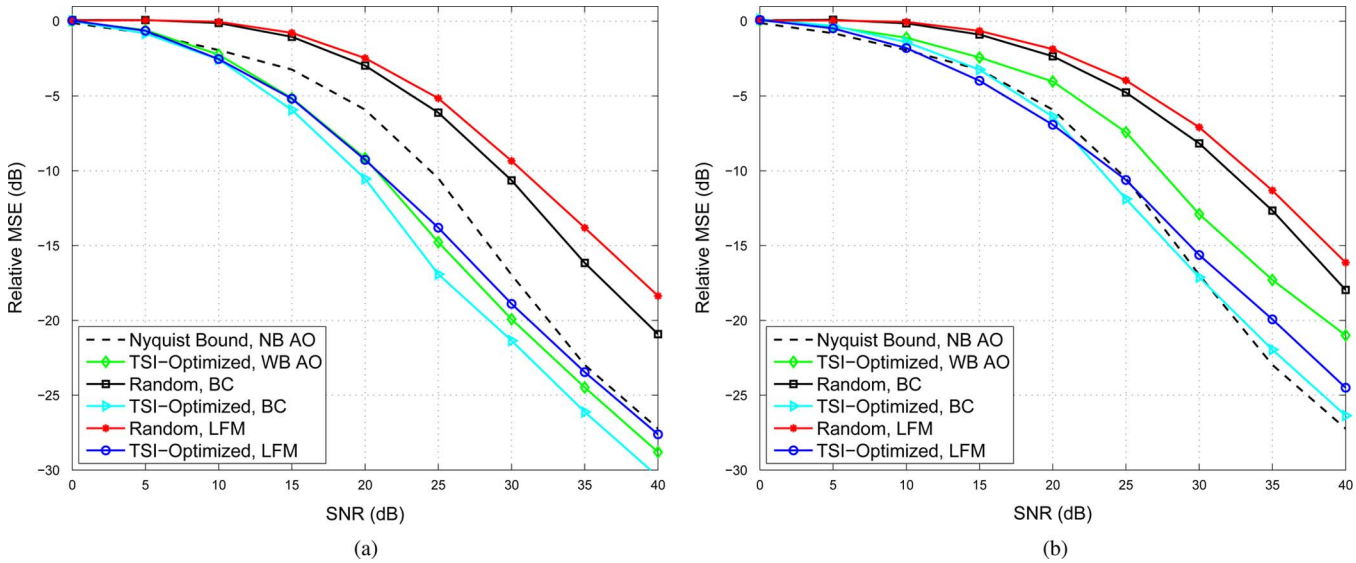
Fig. 6. Target range-only profiling performance, [40°, 70°]. (a) Without mask, (b) With mask.

performance difference between the TSI-optimized sensing kernels, the narrowband reference, and the random sensing kernels is now considerable. We can conclude that prior knowledge in the design of the measurement kernels has enabled exploitation of the resolution information inherent in the higher-bandwidth waveforms.

Although it is nice to produce quality high-resolution range profiles, in general these profiles would be used for some exploitation task such as target recognition. Therefore, we now consider recognition performance for the same compressive measurement design strategy. The same waveforms and measurement kernels are considered, but instead of computing a range profile, we use the mixture model to compute posterior probabilities for each of the three target types. The GM model admits a closed-form calculation for the posterior mixture weights, which can then be summed over all components corresponding to a particular target type. We choose the target that has the maximum posterior probability. In the Monte Carlo simulations, each target is chosen with equal probability, and the target's orientation is chosen from a uniform distribution on [40°, 70°] for Fig. 10 and [40°, 50°] for Fig. 11, respectively. Probability of recognition error is shown versus SNR, and again the modulated wideband waveforms with optimized compressive sampling outperform the other combinations.

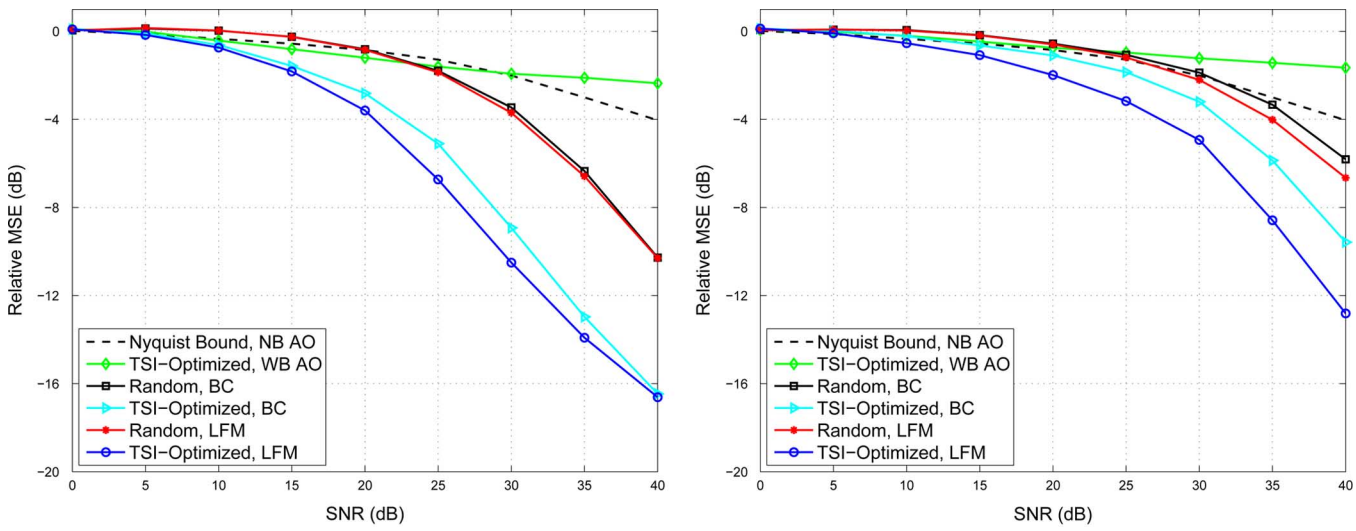
It is interesting that the tighter orientation information in Fig. 11 compared to Fig. 10 seems to make a much larger impact than it did for the profiling results. The reason is that the recognition problem depends on separation of target types rather than estimation of target profiles for the same target at different orientations. Apparently, the tighter orientation information in Fig. 11 is now enough to enable distinct separation of the target types in the optimized, compressed measurements.

For the sake of completeness, we consider a simulation with higher effective bandwidth of 1 GHz. In this simulation, the target templates were defined on a sampling interval of $T_s = 1$ ns, allowing transmit bandwidth to be increased without exceeding the underlying fidelity of the input target models. As before, the bandwidth of the transmit waveform is limited to one-third of the bandwidth of the input source, for a maximum waveform bandwidth of approximately 333.3 MHz. Waveform



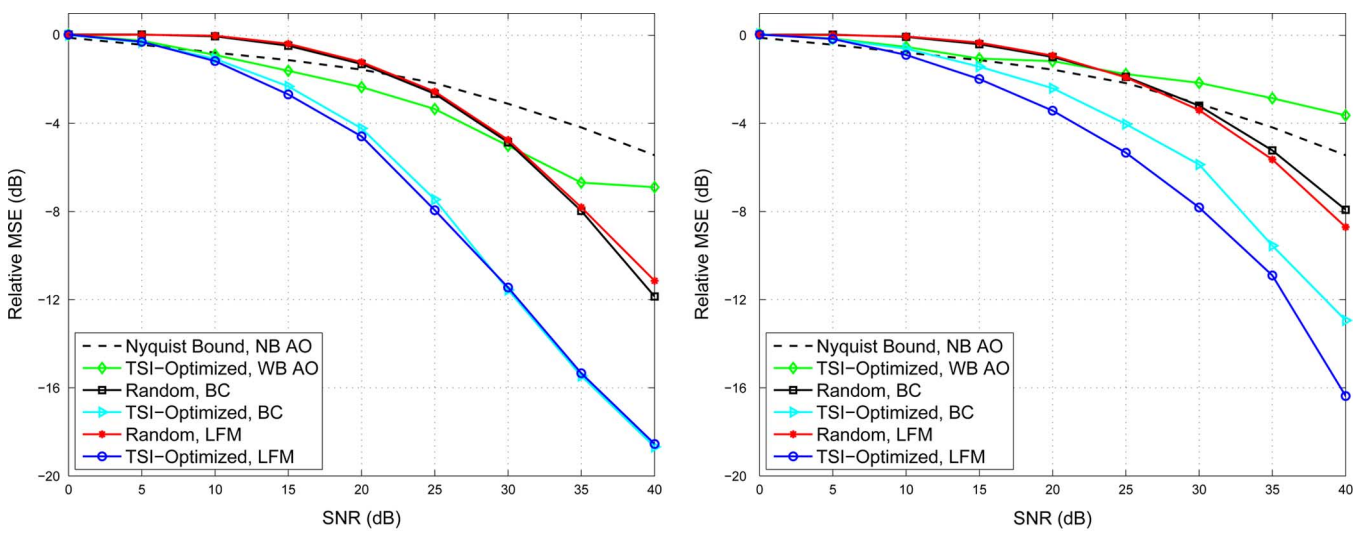
(a) (b)

Fig. 7. Target range-only profiling performance, [40°, 50°], (a) Without mask, (b) With mask.



(a) (b)

Fig. 8. Target range-only profiling performance with range uncertainty, [0°, 90°]. (a) Without mask, (b) With mask.



(a) (b)

Fig. 9. Target range-only profiling performance with range uncertainty, [40°, 50°]. (a) Without mask, (b) With mask.

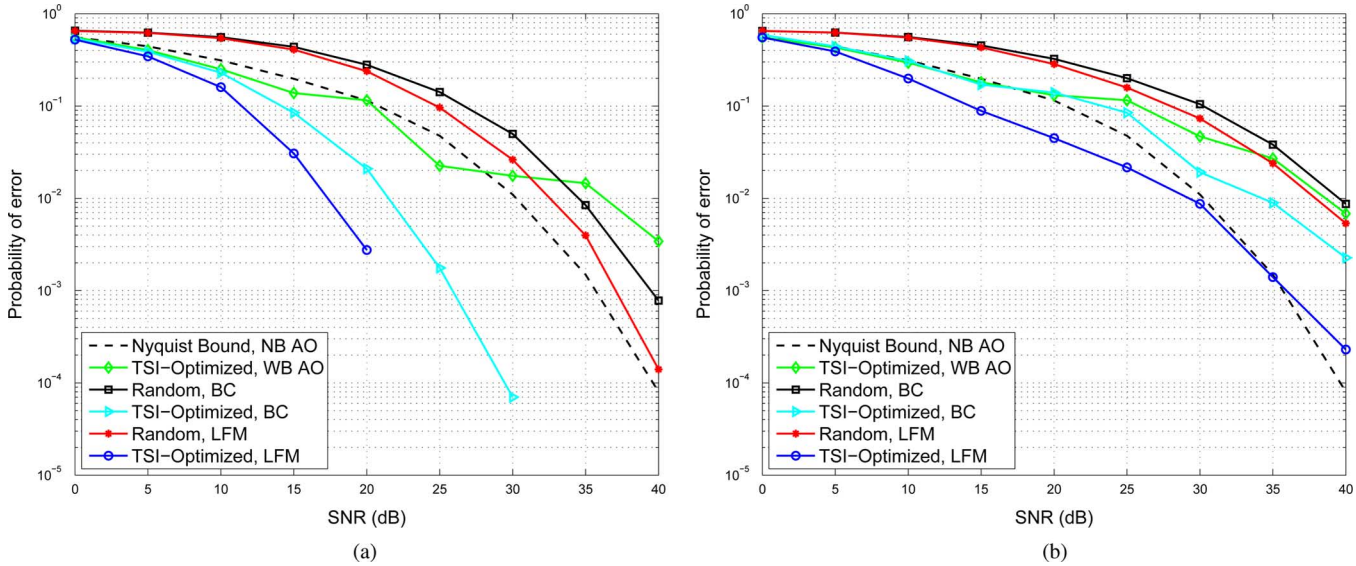


Fig. 10. Target recognition of three aircraft, [40°, 70°]. (a) Without mask, (b) With mask.

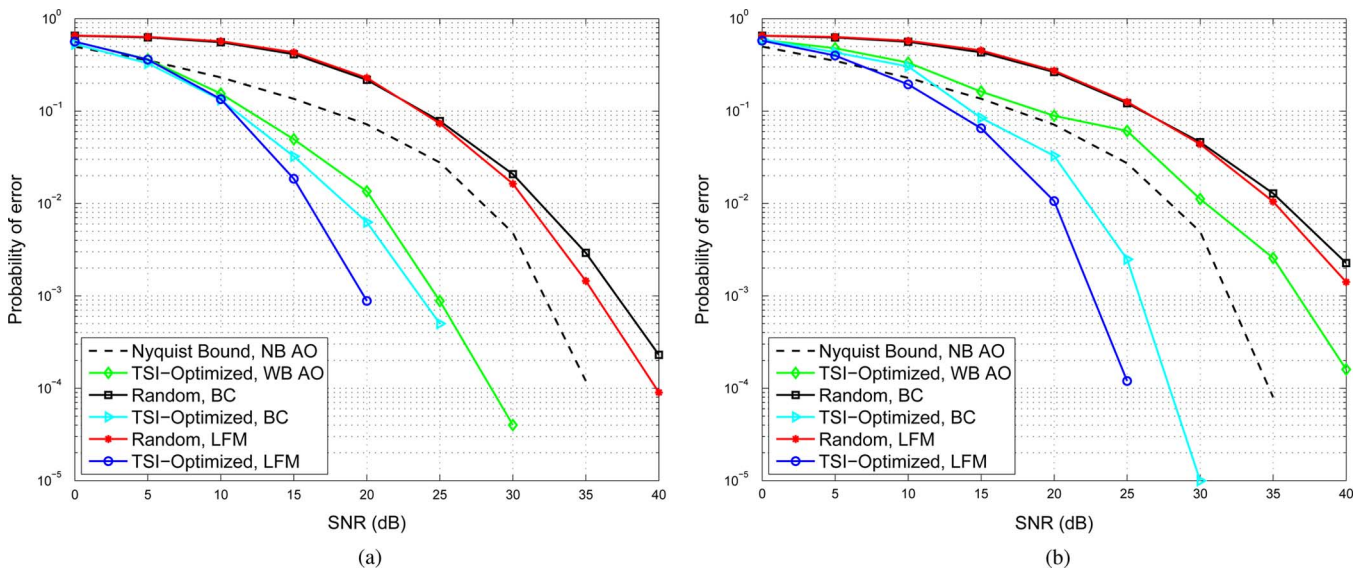


Fig. 11. Target recognition of three aircraft, [40°, 50°]. (a) Without mask, (b) With mask.

TABLE II
PARAMETERS OF THE FOUR TRANSMIT WAVEFORMS WITH 1 GHZ BANDWIDTH

Waveform Types	Bandwidth	Pulse Width	Nyquist Samples	Compressive Samples
NB AO	66.7 MHz	0.015 μs	14	
WB AO	333.3 MHz	0.003 μs	63	13
LFM	333.3 MHz	0.132 μs	106	21
Pseudo-random code (PRC)	333.3 MHz	0.132 μs	106	21

and sampling parameters are described in Table II. Because the length of the longest Barker code is 13, in this simulation the Barker code is replaced by a pseudo-random code (PRC).

Fig. 12 shows the MSE performance for range-only profiling, with five resolution cells of range uncertainty and [0°, 90°] orientation uncertainty. The advantage of wideband waveforms combined with TSI-optimized compression kernels is again obvious when compared to random sensing kernels and to Nyquist sampling of a reduced-bandwidth waveform. One important change is that the performance of the unmodulated wideband pulse is now comparable to the other wideband pulses, which is unlike the result in Fig. 8. Initially, we concluded that the modulated waveforms performed better because their ratio of pulsewidth to target length was higher. With the higher bandwidth results of Fig. 12, we conclude instead that the unmodulated wideband pulse was previously hindered by the fact that the compressed sampling only allowed for three samples. This number was too small in an absolute sense to provide sufficient flexibility on the compressive kernel design.

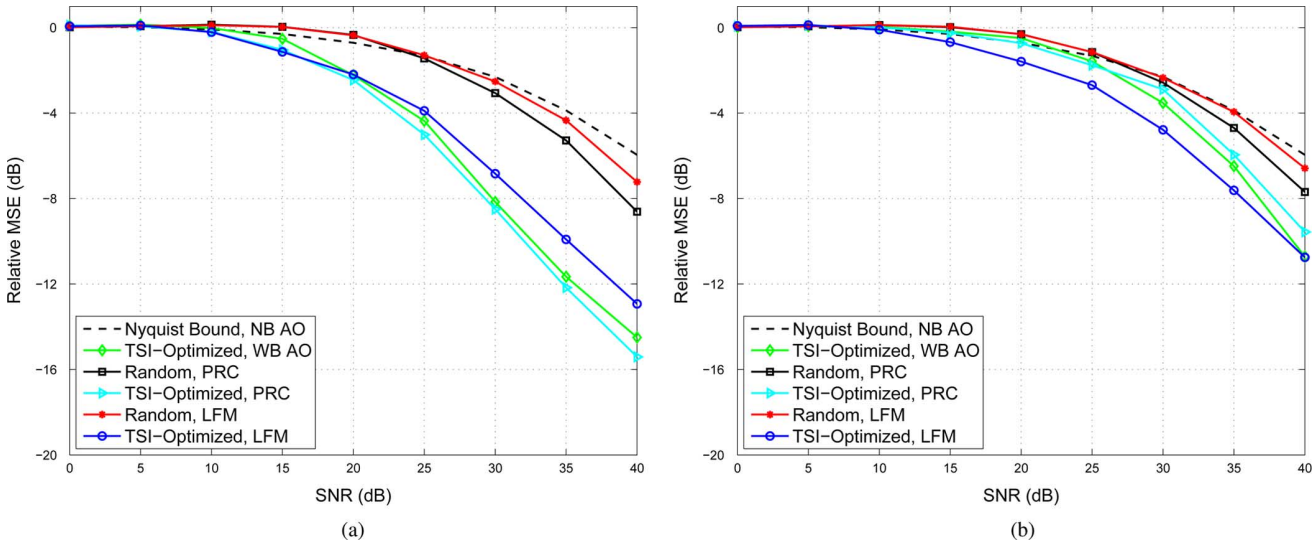


Fig. 12. Target range-only profiling performance with 1 GHz bandwidth with range uncertainty, [0°, 90°], (a) Without mask, (b) With mask.

V. Conclusion

RF applications of compressive sensing suffer from SNR loss due to noise folding. Therefore, the best applications of RF compressive sensing will be in certain, non-noise-limited circumstances, and prior knowledge should be used to optimize the compressive measurement kernels for the given task. One such application is radar profiling in high-SNR scenarios, and toward that end, we have presented a compressive sensing kernel optimization method based on the concept of task-specific information. The information-based approach requires a prior distribution for the signal to be reconstructed; hence, we have used a Gaussian mixture distribution to model target template training data over aspect angle and range translation. Although the Gaussian mixture model can have high complexity in the presence of multiple template parameters, we have also shown how the model can be used to obtain a gradient-based search procedure for measurement kernel design. Continuing research includes efforts to reduce model complexity via clustering algorithms. Examples of performance improvement achievable through kernel optimization have been demonstrated in a high-resolution range profiling application. The simulated results show that improved profiling and target recognition performance can be achieved by transmitting additional bandwidth, even when the receiver must perform compressive sampling of the reflected signals. Random projections actually performed worse than a Nyquist-sampled narrowband signal; therefore, we conclude that performance improvement in an profiling application is most likely to be achieved when prior knowledge is exploited in the measurement kernel design. Designed kernels are more effective at capturing relevant information and overcoming SNR loss. We also observed varying performance for different waveforms having the same bandwidth and time duration—further understanding of this behavior and potential joint waveform/kernel optimization are additional potential areas of future work.

Appendix

In this Appendix, we derive the Taylor-expansion-based approximation of the entropy of the measurement $\mathbf{y}$. We begin by performing a first-order Taylor series expansion of $\log[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})]$ around $\mathbf{y}_0 = E[\mathbf{y}] = \mathbf{u}_\mathbf{y}$, the mean value of the measurement $\mathbf{y}$, which yields (28).

$$\begin{aligned}
&\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})\right] \\
&\quad\approx \log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right] \\
&\qquad + \left[\nabla_\mathbf{y}\left\{\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})\right]\right\}|_{\mathbf{y}=\mathbf{y}_0}\right]^H \\
&\qquad \cdot(\mathbf{y}-\mathbf{y}_0)+\cdots \\
&\quad\approx \log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right] \\
&\qquad + \left[\frac{\sum_{k=1}^{K} p_k \nabla_\mathbf{y}\left\{f^{(k)}(\mathbf{y})\right\}}{\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y})}|_{\mathbf{y}=\mathbf{y}_0}\right]^H \cdot(\mathbf{y}-\mathbf{y}_0) \\
&\quad= \log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right] + \mathbf{g}(\mathbf{y}_0)^H(\mathbf{y}-\mathbf{y}_0). \qquad (28)
\end{aligned}$$

$$\begin{aligned}
H(\mathbf{y}) &\approx -\sum_{q=1}^{K} p_q \int f^{(q)}(\mathbf{y}) \\
&\quad\times\left\{\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right] + \mathbf{g}(\mathbf{y}_0)^H(\mathbf{y}-\mathbf{y}_0)\right\} d\mathbf{y} \\
&= -\sum_{q=1}^{K} p_q \int \left\{f^{(q)}(\mathbf{y})\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right]\right. \\
&\qquad \left. + f^{(q)}(\mathbf{y})\mathbf{g}(\mathbf{y}_0)^H(\mathbf{y}-\mathbf{y}_0)\right\} d\mathbf{y} \\
&= -\sum_{q=1}^{K} p_q \left\{\int f^{(q)}(\mathbf{y})\log\left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)\right] d\mathbf{y}\right. \\
&\qquad \left. + \int f^{(q)}(\mathbf{y})\mathbf{g}(\mathbf{y}_0)^H(\mathbf{y}-\mathbf{y}_0)d\mathbf{y}\right\}. \qquad (29)
\end{aligned}$$

$$
\begin{aligned}
H(\mathbf{y}) &= -\sum_{q=1}^{K} p_q \left\{ \int f^{(q)}(\mathbf{y}) \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0) \right] d\mathbf{y} \right. \\
&\quad \left. + \mathbf{g}(\mathbf{y}_0)^H \int f^{(q)}(\mathbf{y})(\mathbf{y}-\mathbf{y}_0) d\mathbf{y} \right\} \\
&= -\sum_{q=1}^{K} p_q \left\{ \int f^{(q)}(\mathbf{y}) \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0) \right] d\mathbf{y} \right. \\
&\quad \left. + \mathbf{g}(\mathbf{y}_0)^H \left[\mathbf{u}_{\mathbf{y}}^{(q)} - \mathbf{u}_{\mathbf{y}} \right] \right\} \\
&= -\sum_{q=1}^{K} p_q \int f^{(q)}(\mathbf{y}) \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0) \right] d\mathbf{y} \\
&\quad - \mathbf{g}(\mathbf{y}_0)^H \sum_{q=1}^{K} p_q \left[\mathbf{u}_{\mathbf{y}}^{(q)} - \mathbf{u}_{\mathbf{y}} \right] \\
&= -\sum_{q=1}^{K} p_q \int f^{(q)}(\mathbf{y}) \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0) \right] d\mathbf{y}. \quad (30)
\end{aligned}
$$

Substituting (28) into the second equality of (13), we obtain (29). But, the function $\mathbf{g}(\mathbf{y}_0)$ is independent of $\mathbf{y}$ and can be pulled out of the integration, resulting in (30), because $\mathbf{u}_{\mathbf{y}} = \sum_{q=1}^{K} p_q \mathbf{u}_{\mathbf{y}}^{(q)}$.

In a similar way, we note that the term $\log[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y}_0)]$ is once again independent of $\mathbf{y}$. Therefore, we can express

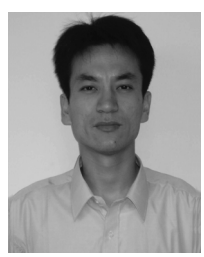
$$
\begin{aligned}
H(\mathbf{y}) &= -\sum_{q=1}^{K} p_q \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y_0}) \right] \int f^{(q)}(\mathbf{y}) d\mathbf{y} \\
&= -\sum_{q=1}^{K} p_q \log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y_0}) \right] \\
&= -\log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y_0}) \right] \sum_{q=1}^{K} p_q \\
&= -\log \left[\sum_{k=1}^{K} p_k f^{(k)}(\mathbf{y_0}) \right]. \quad (31)
\end{aligned}
$$

ACKNOWLEDGMENT

The authors would like to thank the associate editor Prof. Joseph Tabrikian and the anonymous reviewers for their helpful comments and suggestions. We also thank Remcom for providing their XFdtd software.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[2] 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.
[3] M. A. Neifeld, A. Ashok, and P. K. Baheti, "Task-specific information for imaging system analysis," *J. Opt. Soc. Amer. A*, vol. 24, no. 12, pp. B25–B41, Dec. 2007.
[4] P. K. Baheti and M. A. Neifeld, "Recognition using information optimal adaptive feature-specific imaging," *J. Opt. Soc. Amer. A*, vol. 26, no. 4, pp. 1055–1070, Apr. 2009.
[5] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[7] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
[8] G. Yu and G. Sapiro, "Statistical compressive sensing of Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3728–3731.
[9] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, May 2012.
[10] J. Wu *et al.*, "Multivariate compressive sensing for image reconstruction in the wavelet domain: Using scale mixture models," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3483–3494, Dec. 2011.
[11] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of Gaussian mixture models," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 585–600, Feb. 2013.
[12] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
[13] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
[14] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
[15] A. Zyweck and R. E. Bogner, "Radar target classification of commercial aircraft," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 598–606, Apr. 1996.
[16] S. P. Jacobs and J. A. O'sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 364–382, Apr. 2000.
[17] A. V. Oppenheim, A. S. Willsky, and S. Hamid, *Signals and Systems*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.
[18] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Waltham, MA, USA, Apr. 2007, pp. 128–133.
[19] B. Pollock and N. A. Goodman, "An examination of the effects of sub-Nyquist sampling on SNR," in *Proc. SPIE*, Baltimore, MD, USA, Apr. 2012, vol. 8365.
[20] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 478–481, Aug. 2011.
[21] M. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[23] J. T. Flåm, S. Chatterjee, K. Kansanen, and T. Ekman, "On MMSE: A linear model under Gaussian mixture statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3840–3844, Jul. 2012.
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[25] B. R. Pollock and N. A. Goodman, "Detection performance of multi-branch and multichannel compressive receivers," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Hoboken, NJ, USA, Jun. 2012, pp. 341–344.
[26] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Lyngby, Denmark: Technical Univ. of Denmark, Nov. 2012.
[27] Y. Gu and N. A. Goodman, "Compressed sensing kernel design for radar range profiling," in *Proc. 2013 IEEE Radar Conf.*, Ottawa, ON, Canada, Apr. 29–May 3, 2013, pp. 1–5.
[28] P. H. Schönemann, "A generalized solution of the orthogonal Procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966.
[29] High quality 3D Models Library [Online]. Available: [Online]. Available: http://www.3dcadbrowser.com



Yujie Gu (M'10) received the B.E. degree in mechanical engineering from Harbin Institute of Technique in 2001, M.S. degree in control theory and control engineering from Sichuan University in 2004, and the Ph.D. degree in electrical engineering from Zhejiang University in 2008, respectively.

Upon graduation, he joined the CETC 51, Shanghai, China as a R&D engineer. From 2009 to 2010, he was a postdoctoral fellow with the Department of Electrical and Computer Engineering, Concordia University, Montréal, Canada. From 2010 to 2011, he was a postdoctoral researcher with the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel. In 2012, he held a research associate

professor position at the Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China. He is currently with the School of Electrical and Computer Engineering and the Advanced Radar Research Center, University of Oklahoma, Norman, OK as a postdoctoral research associate. His research interests are in radar and array signal processing, robust adaptive beamforming.

Dr. Gu was a recipient of the 2009 ReSMiQ Postdoctoral Scholarship, Montréal, Canada. Dr. Gu currently serves on the editorial boards for *International Journal of Signal and Information Processing* and *International Journal of Communications, Network and System Sciences*.

Nathan A. Goodman (S'98–M'02–SM'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, in 1995, 1997, and 2002, respectively.

From 1996 to 1998, he was an RF systems engineer for Texas Instruments, Dallas, TX. From 1998 to 2002, he was a graduate research assistant in the Radar Systems and Remote Sensing Laboratory, University of Kansas. Dr. Goodman was a faculty member in the ECE Department of the University of Arizona, Tucson, from 2002 to 2011, and is now an associate professor in the School of Electrical and Computer Engineering and associate director of the Advanced Radar Research Center at the University of Oklahoma, Norman. His research interests are in radar and array signal processing.

Dr. Goodman was awarded the Madison A. and Lila Self Graduate Fellowship from the University of Kansas in 1998. He was also awarded the IEEE 2001 International Geoscience and Remote Sensing Symposium Interactive Session Prize Paper Award. Dr. Goodman has served as technical co-chair for the 2011 IEEE Radar Conference, finance chair for the 2012 SAM workshop, and associate editor-in-chief for Elsevier's *Digital Signal Processing* journal. He is currently an associate editor for IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.

Amit Ashok (M'07) received B.S. degree in electronics engineering from the University of Swaziland, M.S. degree in electrical engineering from the University of Cape Town, and Ph.D. degree in electrical and computer engineering from the University of Arizona in 1998, 2001 and 2008 respectively.

From 2002 to 2007 he was graduate research assistant in the Optical Computing and Processing Lab, University of Arizona. He worked as senior scientist in the research division of Omnivision CDM Optics from 2007 to 2009. He returned to University of Arizona as a research scientist in the ECE department from 2009 to 2011. Dr. Ashok is currently an assistant professor in the College of Optical Sciences and ECE department at the University of Arizona. His research interests are in statistical inference, information theory, machine learning, physical optics and image science.

Dr. Ashok was awarded the International Student Scholarship from the University of Cape Town, South Africa in 1999 and 2000. He received the best paper award in SPIE Conference on Three-dimensional Imaging, Visualization and Display Conference (2010) and best poster award in OSA Conference on Computational Optical Sensing and Imaging (2013). Dr. Ashok has served as program co-chair for SPIE Conference on Visual Information Processing in 2012 and program co-chair and general co-chair for the OSA Conference on Computational Optical Sensing and Imaging in 2013 and 2014 respectively.