Deep Task-Based Analog-to-Digital Conversion

Nir Shlezinger, Ariel Amar, Ben Luijten, *Graduate Student Member, IEEE*, Ruud J. G. van Sloun, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Analog-to-digital converters (ADCs) allow physical signals to be processed using digital hardware. Their conversion consists of two stages: Sampling, which maps a continuous-time signal into discrete-time, and quantization, i.e., representing the continuous-amplitude quantities using a finite number of bits. ADCs typically implement generic uniform conversion mappings that are ignorant of the task for which the signal is acquired, and can be costly when operating in high rates and fine resolutions. In this work we design task-oriented ADCs which learn from data how to map an analog signal into a digital representation such that the system task can be efficiently carried out. We propose a model for sampling and quantization that facilitates the learning of non-uniform mappings from data. Based on this learnable ADC mapping, we present a mechanism for optimizing a hybrid acquisition system comprised of analog combining, tunable ADCs with fixed rates, and digital processing, by jointly learning its components end-to-end. Then, we show how one can exploit the representation of hybrid acquisition systems as deep networks to optimize the sampling and quantization rates given the task by utilizing Bayesian meta-learning techniques. We evaluate the proposed deep task-based ADC in two case studies: the first considers synthetic multi-variate symbol detection, where multiple analog signals are simultaneously acquired in order to recover a set of discrete symbols. The second application is beamforming of analog channel data acquired in ultrasound imaging. Our numerical results demonstrate that the proposed approach achieves performance which is comparable to operating with high sampling rates and fine resolution quantization, while operating with reduced overall bit rate. For instance, we demonstrate that deep task-based ADCs enable accurate reconstruction of ultrasound images while using 12.5% of the overall number of bits used by conventional ADCs.**

***Index Terms*—Analog-to-digital conversion, deep learning, sampling, quantization.**

Manuscript received 31 January 2022; revised 30 October 2022; accepted 27 November 2022. Date of publication 16 December 2022; date of current version 5 January 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David I Shuman. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 646804-ERC-COG-BNYQ, and in part by the Israel Science Foundation under Grant 0100101. An earlier version of this paper was presented in part at the 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [DOI: 10.1109/ICASSP40776.2020.9053855]. *(Corresponding author: Nir Shlezinger.)*

Nir Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: nirshlezinger1@gmail.com).

Ariel Amar and Yonina C. Eldar are with the Faculty of Math and CS, Weizmann Institute of Science, Rehovot 7632706, Israel (e-mail: arielamar123@gmail.com; yonina.eldar@weizmann.ac.il).

Ben Luijten and Ruud J. G. van Sloun are with the EE Department, Eindhoven University of Technology 5612 AE Eindhoven, The Netherlands, and also with the Phillips Research, Eindohoven, The Netherlands (e-mail: R.J.G.v.Sloun@tue.nl).

Digital Object Identifier 10.1109/TSP.2022.3229947

I. Introduction

A MULTITUDE of electronic systems process physical signals using digital hardware. Digital signal processors represent analog quantities as a set of bits using analog-to-digital conversion. Converting a continuous-time (CT) signal taking continuous-amplitude values into a finite-bit representation consists of two steps: The analog signal is first sampled into a discrete-time process, which is then quantized into discrete-amplitude values, such that it can be digitally processed [2].

The acquisition of analog signals is commonly carried out using scalar analog-to-digital convertors (ADCs) [3]. These devices sample the CT signal in uniformly spaced time-instances and obtain a digital representation using a uniform mapping of the real line. While this acquisition strategy is simple to implement, it is limited in its ability to accurately represent signals in digital [4], especially when operating under constrained sampling rate and low quantization resolution, due to, e.g., cost, power, or memory constraints. Furthermore, this procedure is carried out regardless of the task for which the analog signal is acquired into a digital representation.

In practice, analog signals are often acquired in order to extract some underlying information, namely, for a task other than recovering the analog process. One example is multiple-input multiple-output (MIMO) communications receivers, which recover a transmitted discrete message from their observed channel output. MIMO receivers typically operate under strict power and cost constraints, which are particularly relevant when operating in high frequency bands [5]. Another relevant example is ultrasound imaging, where large amounts of analog channel data are acquired to form an image, which notably affects the hardware cost and complexity [6]. While designing energy efficient uniform ADCs is an on-going area of research [7], [8], [9], [10], a natural approach to relieve the effects of high resolution acquisition is to restrict the sampling rate and quantization resolution of the ADCs.

When acquiring for a specific task, it was shown in [11], [12], [13], [14], [15], [16], [17], [18], [19] that the distortion induced by sample and bit limitations can be notably reduced by accounting for the task in acquisition. In particular, the works [11], [12], [13] analytically designed task-based quantization systems for estimation tasks by introducing analog processing and tuning the quantization rule, assuming ideal (Nyquist rate) samplers; A data-driven approach for designing task-based quantizers under generic setups was considered in [14], which utilized machine learning (ML) tools. The works [15], [16] also used ML to learn sampling mechanisms assuming error-free quantization, while [17] and [18] analytically designed samplers for

maximizing capacity in MIMO systems and for audio classification, respectively. However, none of these works study the full acquisition process. For acquisition involving both sampling and quantization, [19] studied the analytical design of hybrid analog/digital acquisition systems with uniform ADCs for recovering linear functions of their observations. The analysis of joint sampling and quantization systems was considered in [20], [21], which focused on complex source coding instead of scalar quantization, and derived bounds on the reconstruction accuracy in the absence of a task. The works [22], [23] designed acquisition systems with non-uniform sampling and one-bit ADC for MIMO receivers. The design of acquisition systems with possibly non-uniform ADCs for a (possibly analytically intractable) task, has not yet been studied.

In this work we propose a task-based acquisition system utilizing scalar ADCs for signals obeying a finite basis expansion model. As analytically deriving task-based methods is difficult and commonly requires imposing a limited structure, such as assuming uniform ADC mappings and linear operations [11], [13], [19], we adopt a data-driven approach. We design the system to learn its sampling and quantization mappings along with its analog and digital processing from training data, such that it can reliably carry out its task. A major challenge in designing ML-based ADCs and incorporating such devices into deep neural networks (DNNs), stems from the continuous-to-discrete nature of sampling and quantization: These operations are either non-differentiable or nullify the gradient [14], [15], limiting the application of conventional training based on backpropagation. To overcome this, we adopt a soft-to-hard approach based on gradient estimation through relaxation of the discrete process [24], also utilized in [14] for optimizing quantization mappings. We propose a differentiable approximation of sampling, which can be trained to learn non-uniform sampling methods, and is combined with the trainable quantizer of [14] into a dynamic data-driven ADC. We incorporate this adaptive ADC into a DNN architecture resulting in a deep task-based acquisition system, which can be trained using conventional training methods, e.g., stochastic gradient descent (SGD) with backpropagation.

By representing the hybrid analog/digital acquisition system as a trainable DNN with non-conventional layers encapsulating the sampling and quantization operations, we are also able to optimize the parameters dictating the overall bit rate. We treat the sampling rate, the quantization resolution, and the number of scalar ADCs, as hyperparameters of the DNN, and propose a meta-learning scheme to optimize these key quantities from data. We utilize a variation of Bayesian optimization based meta-learning [25] in order to minimize these quantities while preserving the ability to accurately carry out the task. Our proposed deep task-based hybrid acquisition system is evaluated in two applications: A synthetic scenario of detection from linear observations, and an ultrasound beamforming setup. For the first application, we demonstrate the ability of deep task-based ADCs to achieve comparable performance to the maximum a-posteriori probability (MAP) estimator *without quantization constraints*, and to outperform the common approach of processing only in the digital domain with uniform ADCs. Our proposed meta-learning scheme is shown to notably reduce the overall bit rate without degrading performance, and in some cases even improving the detection accuracy. For ultrasound beamforming, we show how the conversion of analog channel data into a set of pixels specializes a task-based acquisition setup. Then, we demonstrate that deep task-based ADCs can use as low as 12.5% the number of bits compared to conventional acquisition while hardly affecting the quality of the recovered image.

The rest of this paper is organized as follows: Section II presents the system model. Section III details the proposed deep task-based ADC system, while Section IV presents the Bayesian meta-learning scheme for optimizing its configuration. The synthetic case study and the application to ultrasound beamforming are discussed in Sections V and VI, respectively. Finally, Section VII concludes the paper.

Throughout the paper, we use boldface lower-case letters for vectors, e.g., $\boldsymbol{x}$; the ith element of $\boldsymbol{x}$ is written as $(\boldsymbol{x})_i$. Boldface upper-case letters denote matrices, e.g., $\boldsymbol{M}$; and $(\boldsymbol{M})_{i,j}$ is its (i,j)th element. Finally, $\mathcal{R}$ and $\mathcal{Z}$, $\text{sign}(\cdot)$ and $\delta(\cdot)$ are the sets of real numbers, integers, sign, and Dirac delta function, respectively.

II. Task-Based Signal Acquisition Setup

In this section we formulate the task-based acquisition setup. We first present the system model in Subsection II-A, after which we formulate the problem of designing such systems in a data-driven manner in Subsection II-B.

A. System Model

The task-based acquisition setup is modeled using the hybrid system illustrated in Fig. 1. The system consists of analog filtering, analog-to-digital conversion, and digital processing. Our goal, as detailed in Subsection II-B, is to propose a mechanism for learning these components from data. We focus on scenarios where a set of n analog signals $\{x_i(t)\}_{i=1}^n$ are converted into a digital representation in order to recover an unknown vector $\boldsymbol{s} \in \mathcal{S}^k$, referred to as *the system task*. The task vector $\boldsymbol{s}$ is statistically related to the multivariate analog signal $\boldsymbol{x}(t) \triangleq [x_1(t), \dots x_n(t)]^T$ via a conditional distribution $f_{\boldsymbol{x}|\boldsymbol{s}}$. Such scenarios represent, for example, measurements taken from sensor arrays to detect some physical phenomenon, acoustics echos acquired in order to form an image in ultrasound beamforming, or channel outputs acquired by a MIMO receiver for decoding a transmitted message.

Signal Model: We focus on the case where each of the CT signals known to be spanned by a set of B basis functions, i.e., for each $x_i(t)$ there exists a set of functions $\{u_{i,j}(t)\}_{j=0}^{B-1}$ and coefficients $\{\tilde{x}_{i,j}\}_{j=0}^{B-1}$ such that $x_i(t) = \sum_{j=0}^{B-1} \tilde{x}_{i,j} u_{i,j}(t)$. By defining the $n \times n$ diagonal matrix $\boldsymbol{U}_j(t)$ such that $(\boldsymbol{U}_j(t))_{i,i}(t) \triangleq u_{i,j}(t)$ and the vector $\tilde{\boldsymbol{x}}_j \triangleq [\tilde{x}_{0,j}, \dots, \tilde{x}_{n-1,j}]^T$, we can write the multivariate signal as

$$\boldsymbol{x}(t) = \sum_{j=0}^{B-1} \boldsymbol{U}_j(t) \tilde{\boldsymbol{x}}_j. \tag{1}$$

The generic formulation in (1) accommodates a broad family of signals, as exemplified next:

Example 1: Let $\boldsymbol{x}(t)$ be a set of finite rate of innovation (FRI) signals, i.e., there exists a function $u(t)$ and a set $\{\tau_{i,j}\}$ such that

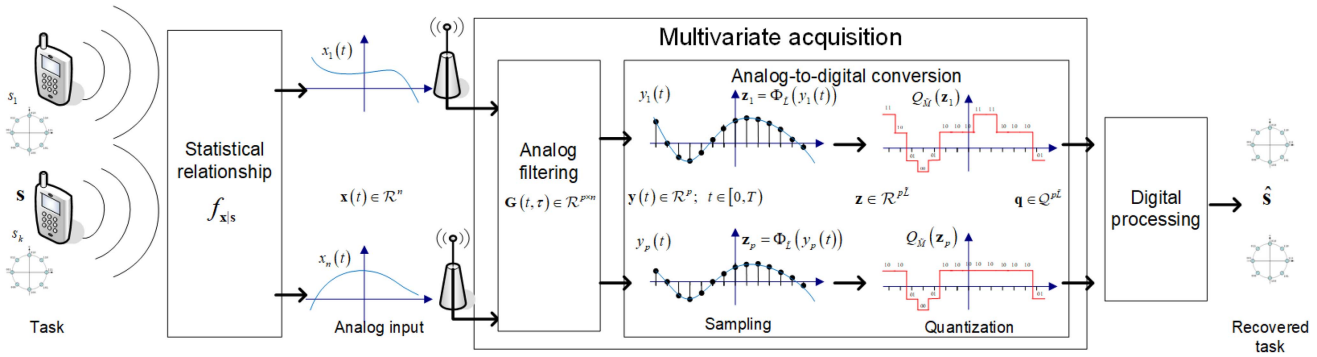


Fig. 1. Hybrid task-based acquisition system illustration. The task here is recovering a set of constellation symbols in MIMO communications.

$x_i(t)=\sum_{j=0}^{B-1}\tilde{x}_{i,j}u(t-\tau_{i,j})$ [16]. Such signals are a special case of (1), obtained by setting $u_{i,j}(t)=u(t-\tau_{i,j})$.

The FRI model of Example 1 represents signals encountered in various applications. For instance, in uplink MIMO communications settings, which is the setting illustrated in Fig. 1, when the statistical relationship $f_{\boldsymbol{x}|\boldsymbol{s}}$ represents a multipath channel, the received signal typically follows an FRI model with $u(t)$ being the pulse-shaping functions [26].

Example 2: Let $\boldsymbol{x}(t)$ be a set of signals defined over a finite time interval $t\in[0,P)$ for some $P>0$, and thus each $x_i(t)$ can be written using its Fourier series expansion. Such signals with finite Fourier series are a special case of (1) obtained when $u_{i,j}(t)$ is the jth Fourier basis function for $t\in[0,P)$.

Example 3: When $\boldsymbol{x}(t)$ is periodic with period $P>0$ and has a finite Fourier series expansion, then it can be written via (1) with $u_{i,j}(t)$ being the jth Fourier basis function $\forall t\in\mathcal{R}$.

The motivation for this model is that it allows to rigorously express the multivariate CT signal $\boldsymbol{x}(t)$ using the $nB\times 1$ vector $\boldsymbol{x}\triangleq[\tilde{\boldsymbol{x}}_0^T,\ldots,\tilde{\boldsymbol{x}}_{B-1}^T]^T$, which encapsulates the information needed to recover the task vector $\boldsymbol{s}$.

Analog Filtering: The observed $\boldsymbol{x}(t)$ is first mapped into a set of p CT signals $\{y_i(t)\}_{i=1}^p$, representing the processing carried out in analog. We focus on linear analog processing, in which $\boldsymbol{y}(t)\triangleq[y_1(t),\ldots y_p(t)]^T$ is obtained from $\boldsymbol{x}(t)$ via multivariate filtering with a matrix impulse response $\boldsymbol{G}(t,\tau)\in\mathcal{R}^{p\times n}$, i.e.,

$$\boldsymbol{y}(t)=\int\boldsymbol{G}(t,t-\tau)\boldsymbol{x}(\tau)d\tau. \tag{2}$$

We do not restrict the filter to be linear time-invariant (LTI) or causal, allowing it to represent a broad range of acquisition systems. We focus on linear operations as these are a common model for feasible analog processing; linear analog filters were shown to facilitate the sampling and recovery of multivariate signals in [27], while LTI memoryless combiners are commonly used for RF chains reduction in MIMO systems [28], [29], [30].

Based on the signal model in (1), it holds that for each $t\in\mathcal{R}$, the vector $\boldsymbol{y}(t)$ is given by a linear function of $\boldsymbol{x}$, as

$$\begin{aligned}\boldsymbol{y}(t)&=\int\boldsymbol{G}(t,t-\tau)\sum_{j=0}^{B-1}\boldsymbol{U}_j(\tau)\tilde{\boldsymbol{x}}_j d\tau\\&=\sum_{j=0}^{B-1}\left(\int\boldsymbol{G}(t,t-\tau)\boldsymbol{U}_j(\tau)d\tau\right)\tilde{\boldsymbol{x}}_j=\tilde{\boldsymbol{G}}(t)\boldsymbol{x},\end{aligned} \tag{3}$$

where $\tilde{\boldsymbol{G}}(t)$ is an $p\times nB$ block-matrix comprised of a row of B sub-matrices, i.e., $\tilde{\boldsymbol{G}}(t)=[\tilde{\boldsymbol{G}}_0(t),\ldots,\tilde{\boldsymbol{G}}_{B-1}(t)]$ where the jth submatrix is

$$\tilde{\boldsymbol{G}}_j(t)\triangleq\int\boldsymbol{G}(t,t-\tau)\boldsymbol{U}_j(\tau)d\tau. \tag{4}$$

Example 4: When $\boldsymbol{x}(t)$ is periodic as in Example 3 and the analog filter is LTI, then $\tilde{\boldsymbol{G}}_j(t)=\mathcal{F}\{\boldsymbol{G}\}(\frac{2\pi j}{P})\boldsymbol{U}_j(t)$, where $\mathcal{F}\{\boldsymbol{G}\}(\omega)$ is the multivariate frequency response of the filter, and $\boldsymbol{U}_j(t)$ is the jth Fourier basis as in Example 3.

Analog-to-Digital Conversion: Next, $\boldsymbol{y}(t)$ observed over the interval $t\in[0,T)$ is converted into a digital representation using a set of scalar ADCs. Each signal $y_i(t)$ undergoes the same ADC mapping, which consists of arbitrary non-uniform sampling and quantization: Sampling is represented by the operator $\Phi_{\tilde{L}}(\cdot)$, such that $\boldsymbol{z}_i=\Phi_{\tilde{L}}(y_i(t))$ is a $\tilde{L}\times 1$ vector whose entries are $(\boldsymbol{z}_i)_j=y_i(t_j)$, where $\{t_j\}_{j=1}^{\tilde{L}}\subset[0,T)$, i.e.,

$$\left(\Phi_{\tilde{L}}(\alpha(t))\right)_j=\int\alpha(\tau)\delta(\tau-t_j)d\tau,\quad j\in\{1,\ldots,\tilde{L}\}. \tag{5}$$

The parameters $\{t_j\}_{j=1}^{\tilde{L}}$ determine the sampling times, which are not restricted to represent uniform sampling. Quantization is carried out using a continuous-to-discrete mapping $Q_{\tilde{M}}:\mathcal{R}\mapsto\mathcal{Q}$ applied to each entry of $\boldsymbol{z}_i$, where $\tilde{M}=|\mathcal{Q}|$ is the resolution, i.e., it uses $\lceil\log_2\tilde{M}\rceil$ bits. The mapping is given by (almost everywhere on $\mathcal{R}$)

$$Q_{\tilde{M}}(\alpha)=a_0+\sum_{i=1}^{\tilde{M}-1}a_i\text{sign}\left(\alpha-b_i\right). \tag{6}$$

In (6), $\{b_i\}$ and $\{a_i\}$ determine the decision regions and their assigned values, i.e., the set $\mathcal{Q}$. For the special case of uniform quantization, the difference b_i-b_{i-1} and the values of a_i are constant, and do not depend on the decision region index i. By defining $\boldsymbol{z}\triangleq[\boldsymbol{z}_1^T,\ldots,\boldsymbol{z}_p^T]^T$, the output of the ADCs is the vector $\boldsymbol{q}\in\mathcal{Q}^{p\tilde{L}}$ whose entries are $(\boldsymbol{q})_l=Q_{\tilde{M}}((\boldsymbol{z})_l)$. The overall number of bits used for acquisition is $p\cdot\tilde{L}\cdot\lceil\log_2\tilde{M}\rceil$.

Digital Processing: The discrete vector $\boldsymbol{q}$ is processed in digital to estimate the task vector $\boldsymbol{s}$ as $\hat{\boldsymbol{s}}\in\mathcal{S}^k$.

B. Problem Formulation

Designing task-based acquisition systems using model-based methods, namely, analytically setting the filter $\boldsymbol{G}(t,\tau)$ and the

operators $\Phi_{\tilde{L}}(\cdot)$ and $Q_{\tilde{M}}(\cdot)$ based on $f_{\boldsymbol{x}|\boldsymbol{s}}$, is very difficult. Consequently, previous model-based studies assumed some specific conditional distribution $f_{\boldsymbol{x}|\boldsymbol{s}}$ with either fixed sampling rule and uniform quantizers [11], [13], [19], or alternatively, considering error-free quantization [17]. Furthermore, accurate knowledge of $f_{\boldsymbol{x}|\boldsymbol{s}}$ may not be available in practice. Consequently, our goal is to design task-based acquisition systems in a data-driven fashion using ML methods.

To formulate the setup such that it can be designed using ML tools, which typically operate on vectors and not on CT quantities, we restrict the sampling times $\{t_j\}$, which dictate $\Phi_{\tilde{L}}(\cdot)$, to be a subset of the some dense uniform grid. We divide the observation interval $[0, T)$ into L sub-intervals of duration $T_L = \frac{T}{L}$, and select our sampling points from the set $\{lT_L\}_{l=0}^{L-1}$. The grid is set such that the number of samples taken by the acquisition system $\tilde{L}$ is smaller, and preferably much smaller, than the size of the dense grid L. Here, by writing the dense samples as the vector $\boldsymbol{y} \triangleq [\boldsymbol{y}^T(0), \ldots, \boldsymbol{y}^T((L-1)T_L)]^T$, it holds that the input to the quantizer $\boldsymbol{z}$ consists of entries of $\boldsymbol{y}$. Constraining the acquisition system to sample from a discretized grid facilitates its design using ML methods, as done in [15]. Furthermore, by defining the $Lp \times nB$ matrix $\bar{\boldsymbol{G}}$ which is comprised of a column of L sub-matrices with $\tilde{\boldsymbol{G}}(lT_L)$ being the lth sub-matrix, it follows from (3) that the candidate samples are expressed as

$$\boldsymbol{y} = \bar{\boldsymbol{G}}\boldsymbol{x}. \tag{7}$$

It is noted that the matrix may be restricted to take a given structure, depending on the constrains imposed on the filter $\boldsymbol{G}(t, \tau)$ as well as the specification of the basis functions in (1). Accordingly, once such a structure is imposed, one can trace $\bar{\boldsymbol{G}}$ to the setting of the CT filter $\boldsymbol{G}(t, \tau)$. To how such structures are obtained, consider the following example:

Example 5: Consider FRI signals where the delays lie on the sampling grid, such that $\boldsymbol{x}(t)$ in Example 1 can be written with $\tau_{i,j} = jT_L$ and $B = L$. When the analog filter is LTI, i.e., $\boldsymbol{G}(t, \tau) \equiv \boldsymbol{G}(\tau)$, it holds by (4) that

$$\begin{aligned}\tilde{\boldsymbol{G}}_j(iT_L) &= \int \boldsymbol{G}(iT_L - \tau)u(\tau - jT_L)d\tau \\ &= \boldsymbol{G}_{\text{conv}}((i-j)T_L),\end{aligned} \tag{8}$$

where we define $\boldsymbol{G}_{\text{conv}}(t)$ as the convolution $\int \boldsymbol{G}(t-\tau)u(\tau)d\tau$. Under the given constraints, $\bar{\boldsymbol{G}}$ is a block-Toeplitz matrix.

The task-based acquisition system is thus required to learn to recover $\boldsymbol{s}$ from $\boldsymbol{x}(t)$ based on a training set $\{\boldsymbol{s}^{(j)}, \boldsymbol{x}^{(j)}\}_{j=1}^N$ consisting of N realizations of the densely-sampled inputs and their corresponding task vectors. In particular, the system parameters, i.e., the analog filter $\bar{\boldsymbol{G}}$, sampling operator $\Phi_{\tilde{L}}(\cdot)$, quantization rule $Q_{\tilde{M}}(\cdot)$, and the processing of the digital vector $\boldsymbol{q}$ into $\hat{\boldsymbol{s}}$, are learned from training. The hybrid acquisition system is restricted to utilize at most $\mathcal{B}$ bits i.e., $p \cdot \tilde{L} \cdot \lceil \log_2 \tilde{M} \rceil \leq \mathcal{B}$. Once these parameters are tuned, they can be configured into the task-based acquisition system detailed in the previous subsection, which operates on CT analog signals.

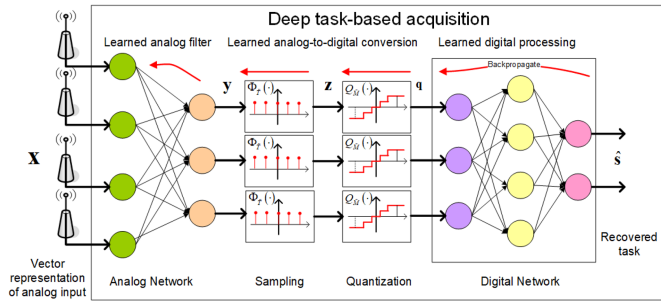


Fig. 2. Deep task-based analog-to-digital conversion illustration.

III. Learning Task-Based Acquisition

In this section we present our deep task-based acquisition system, which learns how to map analog signals into an estimate of the task vector. We focus here on optimizing the hybrid acquisition system when number of ADCs p, the amount of samples acquired $\tilde{L}$, and the quantizer resolution $\tilde{M}$, are fixed, and embodied in the hyperparemeter vector $\boldsymbol{\theta} = [p, \tilde{L}, \log_2 \tilde{M}]$. The optimization of these parameter subject to an overall bit constraint $\mathcal{B}$ is discussed in Section IV. We begin by detailing the analog and digital networks in Subsection III-A. Then, in Subsection III-B we present how ADC mappings are trained, and discuss the resulting structure in Subsection III-C. Throughout this section we consider generic tasks and DNN architectures. Specific models are detailed and evaluated in the case studies presented in Sections V–VI.

A. Learned Analog and Digital Processing

Our proposed deep task-based acquisition system implements the analog filtering and digital processing using dedicated DNNs, denoted as the *analog network* and the *digital network*, respectively. An illustration of such a system is depicted in Fig. 2. The input to the system is the vector representation of the observed signal $\boldsymbol{x}$. To realize linear filters, as detailed in our system model in Section II, the analog network should consist only of linear layers, while introducing non-linear activations yields non-linear analog processing.

While the network architecture illustrated in Fig. 2 is generic, the recovery of the vector $\boldsymbol{s}$ of interest can be broadly divided into two types of tasks: classification and estimation (regression). When $\mathcal{S}$ is a finite set, which is the case in the MIMO detection application presented in Section V the recovery of $\boldsymbol{s}$ can be viewed as classifying from $|\mathcal{S}|^k$ possible categories. In such cases, the output layer of the digital network is a softmax layer with $|\mathcal{S}|^k$ outputs, each representing the conditional distribution of the corresponding label given the input. The overall network is trained end-to-end to minimize the cross-entropy loss. By letting $\boldsymbol{\psi}$ be the set of system trainable parameters and $\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}(\boldsymbol{x}; \boldsymbol{\alpha})$ be the output corresponding to $\boldsymbol{\alpha} \in \mathcal{S}^k$ with network weights $\boldsymbol{\psi}$ and ADC hyperparameters $\boldsymbol{\theta}$, the loss function is given by

$$\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{\psi}) = \frac{1}{N}\sum_{j=1}^{N} -\log \gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}\left(\boldsymbol{x}^{(j)}; \boldsymbol{s}^{(j)}\right). \tag{9}$$

When $\mathcal{S}$ is a continuous set, which is the case in the ultrasound beamforming application discussed in Section VI, then the system task is estimation over a continuous domain. Here, the output layer of the digital network is comprised of k nodes, each estimating a single entry of $\boldsymbol{s}$. In such setups, the network output $\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}(\boldsymbol{x})$ takes values in $\mathcal{S}^k$, and is used as the estimate of $\boldsymbol{s}$. A common loss function for such tasks is the empirical mean-squared error (MSE), given by

$$\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{\psi}) = \frac{1}{N}\sum_{j=1}^{N}\left\|\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}\left(\boldsymbol{x}^{(j)}\right) - \boldsymbol{s}^{(j)}\right\|^2. \tag{10}$$

Once the system is trained, the learned parameters are used to configure the task-based acquisition system detailed in Subsection II-A, which operates on CT signals. The matrix representation of the linear analog network is used to set the filter $\boldsymbol{G}(t,\tau)$. The adaptation of the ADC mapping, trained along with the overall system assuming a fixed number of samples $\tilde{L}$ and quantization resolution $\tilde{M}$, is detailed in the sequel, while a method for optimizing the acquisition hyperparameters $\boldsymbol{\theta}$ is discussed in Section IV.

B. Learned Analog-to-Digital Conversion

The analog and digital processing parts of the hybrid task-based acquisition system are learned as conventional DNN models. However, the ADC mapping, determined by the sampling and quantization rules whose adjustable parameters are the sampling times $\{t_j\}$ in $\Phi_{\tilde{L}}(\cdot)$ and the decision regions $\{a_i, b_i\}$ in $Q_{\tilde{M}}(\cdot)$, cannot be represented using standard layers or activation functions. In particular, both the sampling mapping (5) and the quantization function (6) are non-differentiable or have a zero-valued gradient with respect to their input and/or the adjustable parameters. Consequently, one cannot use straight-forward application of backpropagation with SGD-based optimization to train the system end-to-end.

Following [14], which considered quantization without sampling, we adopt a *soft-to-hard* approach. This approach approximates non-differentiable mappings *during training* by smooth functions that faithfully capture their operation. Recall that Kronecker delta functions, from which the sampling mapping in (5) is comprised, can be obtained as the limit of a sequence of Gaussian functions with decaying variance. Substituting this into (5) while replacing integration with a summation over the discretized grid yields the following relaxation:

$$\left(\phi_{\tilde{L}}(\alpha(t))\right)_j = \sum_{i=0}^{L-1}\alpha(iT_L)\exp\left(\frac{(iT_L - t_j)^2}{\sigma_i^2}\right), \tag{11}$$

where the parameters $\{\sigma_i^2\}$ control the resemblance of $\phi_{\tilde{L}}(\cdot)$ to the non-differentiable sampling function $\Phi_{\tilde{L}}(\cdot)$. Approximating $\Phi_{\tilde{L}}(\cdot)$ with (11) during training allows the system to learn the sampling time instances $\{t_j\}$ along with the analog and digital networks. As samples are taken from the grid $\{lT_L\}$, the learned sampling times are projected onto this grid after training, i.e., t_j is replaced with its nearest grid point, while each point can only be assigned to a single entry of $\{t_j\}$. Similarly, as suggested in [14], the quantizer (6) is also approximated with a differentiable function. Since sign functions can be approached almost everywhere on $\mathcal{R}$ by a sequence of hyperbolic tangents, $Q_{\tilde{M}}(\cdot)$ is approximated during training as

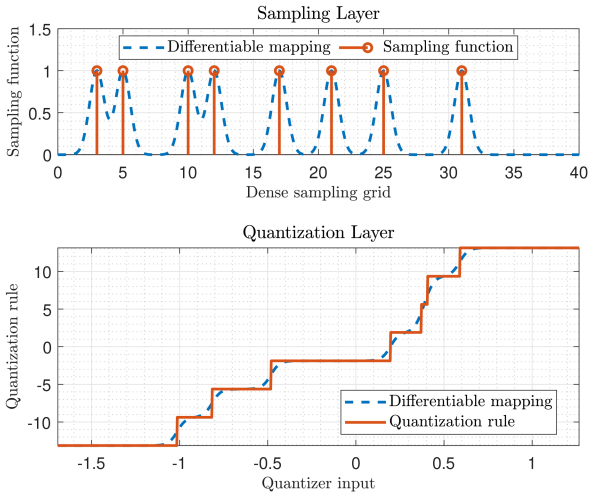


Fig. 3. Learned sampling and quantization rules illustration.

$$q_{\tilde{M}}(\alpha) = a_0 + \sum_{i=1}^{\tilde{M}-1} a_i \tanh\left(c_i \cdot \alpha - b_i\right), \tag{12}$$

where $\{c_i\}$ is a set of real-valued parameters. As c_i increases, its corresponding $\tanh$ function approaches a sign mapping as in (6). Using (11)–(12), the system can tune its parameters by backpropagating the gradient through the ADC mapping during training, while learning non-uniform quantization mappings by tuning $\{a_i, b_i\}$. Once training is concluded, the learned set $\{\frac{b_i}{c_i}\}$ is used to determine the borders of the decision regions of the true (non-differentiable) quantizer. Fig. 3 illustrates how $\phi_{\tilde{L}}(\cdot)$ and $q_{\tilde{M}}(\cdot)$ are converted into sampling and quantization rules.

In (11) and (12), the parameters $\{\sigma_i^2\}$ and $\{c_i\}$, respectively, balance the smoothness of the mapping and the accuracy in representing the non-differentiable function from which it originates. Consequently, they can be either fixed, or modified during training using annealing optimization [31], making the differentiable mapping gradually approach the actual non-differentiable function during training.

C. Discussion

The proposed deep task-based acquisition system jointly adapts its analog filter, sampling function, quantization rule, and digital processing based on training data. This is achieved by identifying smooth trainable approximations of the sampling and quantization mappings. While our system model is formulated for real-valued signals in Subsection II-A, our design can also be applied to complex-valued signals. This is achieved by separating complex quantities into their real and imaginary parts, as is the conventional operation of complex ADCs [32], and representing complex vectors as real-valued vectors of extended dimensions. The number of ADCs p, average sampling rate $\frac{\tilde{L}}{T}$, and the quantization resolution $\tilde{M}$ are all assumed to be fixed here. Yet, the fact that they can be treated as hyperparameters of

a DNN motivates their setting via hyperparameter optimization, as we explore in Section IV.

The system model detailed in Section II bears some similarity to that considered in [22], [23], where non-uniform sampling is implemented by selecting the sampling times from a dense uniform (oversampled) grid, and acquisition is carried out for a task other than reconstructing the analog signal. Nonetheless, our work, which considers a data-driven design for possibly intractable statistical models, is fundamentally different from the works [22], [23] that focused on fully known linear Gaussian MIMO settings with one-bit quantization. Furthermore, our system model allows for a pre-acquisition analog processing, which can implement any linear processing. In particular, the learned matrix $\bar{\boldsymbol{G}}$ can be any real-valued $L \cdot n \times L \cdot p$ matrix. Nonetheless, in some cases, a specific family of filters, such as causal, LTI, memoryless, or phase-shifters may be preferred. Such restrictions can be incorporated into our model by imposing a specific structure on $\bar{\boldsymbol{G}}$. For example, for LTI filters, $\bar{\boldsymbol{G}}$ is block-Toeplitz as shown in Example 5, while for memoryless filters, $\bar{\boldsymbol{G}}$ can be written as $\boldsymbol{I}_L \otimes \tilde{\boldsymbol{G}}$ for some $\tilde{\boldsymbol{G}} \in \mathcal{R}^{n \times p}$. Alternative forms of constrained analog processing, such as those induced by the inherent controllable combining of dynamic metasurface antennas [33], [34], [35] or by using phase shifter networks [28], [29], result in different constraints on $\bar{\boldsymbol{G}}$. Here, we focus on generic linear analog mappings, and leave these special cases to future investigation.

As detailed in Subsection III-A, our system is designed to be trained offline, and the learned parameters are configured in the acquisition system once training is concluded. Nonetheless, one can also envision an adjustable acquisition hardware device, which is capable of learning its multivariate analog-to-digital conversion mapping online. Such a system can utilize configurable ADCs combined with neuromorphic circuits [36] based on, e.g., memristors [37], for realizing the trainable analog network. However, for a deep task-based acquisition system to adjust its parameters online using conventional methods such as SGD, the digital processor must have access to the vector $\boldsymbol{x}$ during training, namely, it must process a high-resolution version of the vector representation of its observed signals during the periods in which it has knowledge of the task $\boldsymbol{s}$. This requirement can be satisfied by, e.g., utilizing additional dedicated high-resolution ADCs which are employed only during the specific periods where the analog input can be used for training along with its label.

IV. Acquisition Hyperparameters Optimization

So far, we have utilized ML methods to jointly learn the analog filtering, ADC mappings, and digital processing, in an end-to-end manner. This learning stage is carried out while fixing some of the key parameters of analog-to-digital conversion: the number of ADCs p, the number of samples taken $\tilde{L}$, and the quantization resolution $\tilde{M}$. The fact that these hyperparameters directly affect some important aspects of acquisition, such as power consumption and memory usage [3], motivates their learning as part of the training procedure. In this section we detail how the proposed framework of learned task-based acquisition can be extended to tune these key parameters using Bayesian meta-learning tools. To that aim, we first introduce some basics in Bayesian meta-learning in Subsection IV-A, after which we present a method for optimizing the acquisition parameters in Subsection IV-B.

A. Preliminaries in Meta-Learning

Meta-learning is a subfield of ML which deals with optimizing hyperparameters. Unlike conventional parameters, e.g., the weights of a neural network, that are learned in the training process, hyperparameters are parameters of ML algorithms that control the model class, e.g., the network architecture [38], or the learning process, e.g., the learning rate [39] and the optimization rule [40]. Those can be either chosen from a discrete set or from a continuous range; the space of hyperparameters is referred to henceforth as the *search space*.

Some hyperparmeters, such as the initial weights used during training and the learning rate, can be optimized using gradient based methods, as done in model-agnostic meta-learning [41]. However, computing the gradient of the loss with respect to architecture-related hyperparameters, such as the acquisition parameters of our deep task-based acquisition system, is often infeasible. For such settings where gradient-based methods cannot be applied, several methods have been proposed in the literature for hyperparmeter optimization. To formulate the different strategies, we use $\boldsymbol{\theta} \in \Theta$ to denote the set of hyperparameters of the learning algorithm $\mathcal{A}$, where Θ is the search space. We also let $f_{\mathcal{A}}(\boldsymbol{\theta})$ be the evaluation function of these hyperparmeters. In the context of meta-learning the architecture of a neural network, computing this function involves training a network to obtain the empirical loss of the learning algorithm with hyperparmaters $\boldsymbol{\theta}$. Thus, computing $f_{\mathcal{A}}(\boldsymbol{\theta})$ tends to be costly and time-consuming to evaluate. Since for our task-based acquisition system, the search space Θ can be very large, while evaluating $f_{\mathcal{A}}(\boldsymbol{\theta})$ involves re-training a DNN and is thus costly to compute, we adopt the Bayesian optimization approach for meta-learning [25].

Bayesian meta-learning involves a controllable amount of evaluations of $f_{\mathcal{A}}(\cdot)$, and is based on assuming a prior distribution on it. Bayesian meta-learning models $f_{\mathcal{A}}(\boldsymbol{\theta})$ as a Gaussian process over Θ with postulated mean $\mu_0(\boldsymbol{\theta}) = \mathbb{E}[f_{\mathcal{A}}(\boldsymbol{\theta})]$ and autocovariance function $\Sigma_0(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{cov}(f_{\mathcal{A}}(\boldsymbol{\theta}), f_{\mathcal{A}}(\boldsymbol{\theta}'))$. The method sequentially samples the evaluation function $f_{\mathcal{A}}(\cdot)$, iteratively refining the selected $\boldsymbol{\theta}$ assuming an underlying Gaussian model. Given the samples of $f_{\mathcal{A}}(\cdot)$ measured at the i vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i$, the next sampling vector $\boldsymbol{\theta}_{i+1}$ is selected as the one maximizing the expected improvement [25]

$$\mathrm{EI}_i(\boldsymbol{\theta}) \triangleq \mathbb{E}\left[\left(f_{\mathcal{A}}(\boldsymbol{\theta}) - f_{\mathcal{A}}(\boldsymbol{\theta}_i^*)\right)^+ | \{f_{\mathcal{A}}(\boldsymbol{\theta}_j)\}_{j \le i}\right], \tag{13}$$

where $\boldsymbol{\theta}_i^* \triangleq \arg\max_{j \le i} f_{\mathcal{A}}(\boldsymbol{\theta}_j)$ is the current best observation, and $a^+ \triangleq \max(a, 0)$.

Under the assumption that $f_{\mathcal{A}}(\cdot)$ is a Gaussian process, the distribution of $f_{\mathcal{A}}(\boldsymbol{\theta})$ conditioned on $\{f_{\mathcal{A}}(\boldsymbol{\theta}_j)\}_{j \le i}$ is also Gaussian with mean value $\mu_i(\boldsymbol{\theta})$ and standard deviation $\sigma_i(\boldsymbol{\theta})$, obtained from $\mu_0(\cdot)$ and $\Sigma_0(\cdot, \cdot)$ via [25, (3)]:

$$\mu_i(\boldsymbol{\theta}) = \Sigma_0\left(\boldsymbol{\theta}, \boldsymbol{\theta}_{1:i}\right) \Sigma_0^{-1}\left(\boldsymbol{\theta}_{1:i}, \boldsymbol{\theta}_{1:i}\right)$$

Algorithm 1: Bayesian Hyperparameters Optimization.

Init: Randomly sample hyperparameter configuration $\boldsymbol{\theta}_1$, set $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1$, calculate $f_{\mathcal{A}}(\boldsymbol{\theta}_1)$

for $i = 1, 2, \ldots, i_{\max} - 1$ **do**
- Compute $\boldsymbol{\theta}_{i+1} = \operatorname{argmax}_{\boldsymbol{\theta}} \mathrm{EI}_i(\boldsymbol{\theta})$ via (14)
- Calculate $f_{\mathcal{A}}(\boldsymbol{\theta}_{i+1})$
- **if** $f_{\mathcal{A}}(\boldsymbol{\theta}^*) < f_{\mathcal{A}}(\boldsymbol{\theta}_{i+1})$ **then**
  - Update $\boldsymbol{\theta}^* = \boldsymbol{\theta}_{i+1}$
- **end**

end

Output: Hyperparameter configuration $\boldsymbol{\theta}^*$.

$$\times \left(f_{\mathcal{A}}(\boldsymbol{\theta}_{1:i}) - \mu_0\left(\boldsymbol{\theta}_{1:i}\right)\right) + \mu_0\left(\boldsymbol{\theta}\right),$$

$$\sigma_i^2\left(\boldsymbol{\theta}\right) = \Sigma_0\left(\boldsymbol{\theta}, \boldsymbol{\theta}\right) - \Sigma_0\left(\boldsymbol{\theta}, \boldsymbol{\theta}_{1:i}\right) \Sigma_0^{-1}\left(\boldsymbol{\theta}_{1:i}, \boldsymbol{\theta}_{1:i}\right) \Sigma_0(\boldsymbol{\theta}_{1:i}, \boldsymbol{\theta}).$$

Here, $f_{\mathcal{A}}(\boldsymbol{\theta}_{1:i})$ denotes the $i \times 1$ vector whose jth entry is $f_{\mathcal{A}}(\boldsymbol{\theta}_j)$; $\Sigma_0(\boldsymbol{\theta}_{1:i}, \boldsymbol{\theta}_{1:i})$ is an $i \times i$ matrix whose (m, j)th entry is $\Sigma_0(\boldsymbol{\theta}_m, \boldsymbol{\theta}_j)$; $\Sigma_0(\boldsymbol{\theta}, \boldsymbol{\theta}_{1:i})$ is a $1 \times i$ vector whose jth entry is $\Sigma_0(\boldsymbol{\theta}, \boldsymbol{\theta}_j)$; and $\mu_0(\boldsymbol{\theta}_{1:i})$ is an $i \times 1$ vector whose jth entry is $\mu_0(\boldsymbol{\theta}_j)$. By letting $F_G(\cdot)$ and $p_G(\cdot)$ are the cumulative distribution function and the probability density function of the standard normal distribution, respectively, the expected improvement is computed via [25, (8)]

$$\mathrm{EI}_i(\boldsymbol{\theta}) = (\mu_i(\boldsymbol{\theta}) - f_{\mathcal{A}}(\boldsymbol{\theta}_i^*)) F_G(Z_i) + \sigma_i(\boldsymbol{\theta}) p_G(Z_i), \tag{14}$$

where

$$Z_i = \begin{cases} \frac{(\mu_i(\boldsymbol{\theta}) - f_{\mathcal{A}}(\boldsymbol{\theta}_i^*).)}{\sigma_i(\boldsymbol{\theta})} & \text{if } \sigma_i(\boldsymbol{\theta}) > 0, \\ -\infty & \text{if } \sigma_i(\boldsymbol{\theta}) = 0. \end{cases} \tag{15}$$

The resulting sequential hyperparameter optimization is summarized as Algorithm 1. The term $f_{\mathcal{A}}(\boldsymbol{\theta}_i^*)$ is often replaced with $f_{\mathcal{A}}(\boldsymbol{\theta}_i^*) + \zeta$ for some $\zeta > 0$ to improve exploration and reduce the probability of yielding a local optima [42].

B. Acquisition Parameters Optimization

Our proposed deep task-based ADC system is modeled as a DNN such that its analog filter, acquisition mappings, and digital processing can be learned end-to-end from data. The analog-to-digital conversion configuration is dictated by the number of ADCs p, the number of samples taken in each interval $\tilde{L}$, and the quantization resolution $\tilde{M}$. The triplet $(p, \tilde{L}, \tilde{M})$, which dictates the number of bits used in acquisition $p \cdot \tilde{L} \cdot \lceil \log_2 \tilde{M} \rceil$, affects the architecture of the DNN, and is thus treated as the hyperparameters of the model. Consequently, in order to optimize the model under a given bit budget $\mathcal{B}$, as requested in the problem formulation in Subsection II-B, we utilize meta-learning via Bayesian optimization. By letting $\Theta_{\mathcal{B}}$ be the set of triplets of positive integers $(p, \tilde{L}, \tilde{M})$ such that $p \cdot \tilde{L} \cdot \lceil \log_2 \tilde{M} \rceil \leq \mathcal{B}$, meta-learning is expressed as the following optimization problem:

$$\underset{(p, \tilde{L}, \tilde{M}) \in \Theta_{\mathcal{B}}}{\arg\min} \; f_{\mathcal{A}}\left(p, \tilde{L}, \tilde{M}\right) \tag{16}$$

The objective in (16) is determined by the system task and the available data set. For example, $f_{\mathcal{A}}(p, \tilde{L}, \tilde{M})$ can represent the training loss as in (9), e.g.,

$$f_{\mathcal{A}}\left(\boldsymbol{\theta}\right) = -\min_{\boldsymbol{\psi}} \mathcal{L}_{\boldsymbol{\theta}}\left(\boldsymbol{\psi}\right), \tag{17}$$

or alternatively, the validation error of a deep task-based acquisition system as in Fig. 1 with hyperparameters $(p, \tilde{L}, \tilde{M})$ after trained using a given data set. Further, one can boost configurations with reduced number of overall bits by including in the formulation of the objective a regularization term which accounts for the number of bits acquired, as we do in the numerical study in Subsection V-C.

The objective in (16) satisfies the following conditions: 1) Computing $f_{\mathcal{A}}(\cdot)$ requires training the network anew, and thus involves a computationally expensive computation; 2) the cardinality of the search space $\Theta_{\mathcal{B}}$ grows with the size of the sampling grid L, and can thus be large. These conditions imply that naive search techniques may be computationally infeasible, hence, we utilize Bayesian meta-learning in Algorithm 1 for tuning the ADC configuration. The resulting overall learning procedure of the deep task-based ADC, including the learning of the both the acquisition hyperparameters $\boldsymbol{\theta}$ and the DNN weights $\boldsymbol{\psi}$, is summarized as Algorithm 2.

Algorithm 2: Deep Task-Based ADC Learning.

Init: Define loss measure $\mathcal{L}_{\boldsymbol{\theta}}(\cdot)$ and meta-learning objective $f_{\mathcal{A}}(\boldsymbol{\theta})$.

1. Obtain $\boldsymbol{\theta}$ via Algorithm 1.
2. Set $\boldsymbol{\psi}$ to minimize $\mathcal{L}_{\boldsymbol{\theta}}(\cdot)$ via DNN training.

Output: Deep task-based ADC system $\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}(\cdot)$.

This application of Algorithm 2 can be further facilitated by recasting the multiplicative formulation of the search space $\Theta_{\mathcal{B}}$ into an additive one by writing it as

$$\log_2 p + \log_2 \tilde{L} + \log_2 \left\lceil \log_2 \tilde{M} \right\rceil \leq \log_2 \mathcal{B}. \tag{18}$$

Expressing $\Theta_{\mathcal{B}}$ via (18) enables the application of existing Bayesian meta-learning toolboxes, such as BoTorch [43] and Ax [44]; the latter is used in our numerical evaluations. Algorithm 2 can be further simplified by noting that the Bayesian optimization procedure involves training the DNN with the optimized acquisition hyperparameters, and thus one can extract the network weights $\boldsymbol{\psi}$ from Algorithm 1 and avoid re-training in Step 2, as we do in our numerical study in Subsection V-C. In fact, such an approach is expected to yield improved performance as the networks weights are selected from a set of $i_{\max}$ independent training procedures. Nonetheless, as one may prefer to utilize different data sets or different optimization configuration (e.g., step size, number of epochs) in learning the DNN weights compared to those used when evaluating the meta-learning objective $f_{\mathcal{A}}(\cdot)$, we include a dedicated separate weights training step in the overall learning procedure in Algorithm 2.

V. Case Study: Synthetic Model

We next apply the deep task-based acquisition system for detection in a synthetic linear model. The aim of this study is to

demonstrate the ability of the proposed mechanism to jointly train the component of the hybrid analog/digital system, as well as to evaluate the hyperparameter optimization mechanism proposed in Section IV. Therefore, in this section we consider a relatively simple synthetic model for which we are able to, e.g., compute the model-based MAP rule as a benchmark. The application of the proposed deep task-based acquisition framework in a non-synthetic setup is detailed in next case study in Section VI.

Here, we first describe the task and the signal model and the experimental setup[1] in Subsection V-A. Then, we evaluate the deep task-based acquisition system for such setups with fixed acquisition hyperparameters in Subsection V-B and with optimized hyperparameters in Subsection V-C.

A. Experimental Setup

We consider the detection of a vector of binary-valued symbols $\boldsymbol{s}$ whose entries take value in a discrete set $\mathcal{S} = \{-1, 1\}$. The CT $n \times 1$ signal $\boldsymbol{x}(t)$ observed at time instance $t \in [0, T)$ is related to the task vector $\boldsymbol{s}$ via the following linear model:

$$\boldsymbol{x}(t) = \boldsymbol{H}(t)\boldsymbol{s} + \boldsymbol{w}(t). \tag{19}$$

Here, $\boldsymbol{H}(t) \in \mathcal{R}^{n \times k}$ is time-varying measurement matrix and $\boldsymbol{w}(t)$ is the noise vector, comprised of independent zero-mean Gaussian entries with variance $\sigma_w^2(t) > 0$. To obtain the signal model as in (1), we approximate the CT signal using dense sampling, such that $\tilde{\boldsymbol{x}}_j = \boldsymbol{x}(j \cdot T/L)$ and $\boldsymbol{U}_j(t)$ is the identity matrix for $t \in [j \cdot T, (j+1) \cdot T)$, and the all-zero matrix otherwise.

The task is thus given by the recovery of $\boldsymbol{s}$ from the observed $\boldsymbol{x}(t)$, and can thus be treated as acquisition for a classification task. Note that in the absence of noise, $\boldsymbol{s}$ can be often accurately recovered from a single sample of $\boldsymbol{x}(t)$, and thus the gain in processing multiple samples is in reducing the effect of noise quantization distortion. We set $n = 6$, $k = 4$, and the signal duration is $T = 1$ μSec. The noise in (19) satisfies $\sigma_w^2(t) \equiv 1$, while the measurement matrix $\boldsymbol{H}(t)$ represents spatial exponential decay with temporal variations, and its entries are

$$(\boldsymbol{H}(t))_{i,j} = \sqrt{\rho}\,(1 + 0.5\cos(2\pi f_0 t))\, e^{-|i-j|}, \tag{20}$$

where $\rho > 0$ is referred to as the signal-to-noise ratio (SNR) and $f_0 = 10^3$ Hz. Note that the fact that the measurement matrix (20) varies within the symbol duration motivates the usage of non-uniform sampling.

In our experimental study we implement the following architecture for the deep task-based analog-to-digital conversion system: The analog network is an $n \cdot L \times p \cdot L$ fully-connected layer, and the digital DNN is comprised of a $p \cdot \tilde{L} \times 32$ layer, a ReLU activation, a 32×16 layer, and a softmax output layer. Consequently, the network output is a probability vector over $\mathcal{S}^k$ (whose cardinality is $|\mathcal{S}|^k = 16$), and detection is carried out by taking the symbol vector corresponding to the maximal value of this probability vector. The network is trained to minimize (9) over $N = 10^4$ samples using the ADAM optimizer [45] with learning rate of 0.01.

[1]The source code used in our numerical studies is available online on https://github.com/arielamar123/ADC-Learning-hyperopt.

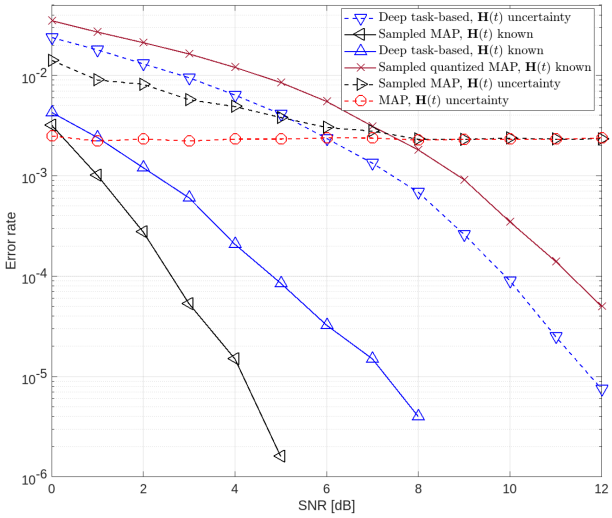


Fig. 4. Error rate versus SNR.

B. Fixed Hyperparameters Experiments

We begin by evaluating the deep task-based acquisition system with fixed acquisition hyperparameters. The simulated acquisition system uses $p = 4$ ADCs, while selecting $\tilde{L} = 4$ samples out of a grid of $L = 20$ time instances, and quantizing each sample using up to $\log_2 \tilde{M} = 3$ bits. We compare the error rate of our deep acquisition system to the following model-based detectors: The MAP rule for recovering $\boldsymbol{s}$ from a uniformly sampled version of $\boldsymbol{x}$ with sampling rate $\tilde{L}/T$ referred to as *sampled MAP*, namely, the minimal achievable error rate when using the same number of samples as our deep task-based system without quantization constraints; and the MAP rule for recovering $\boldsymbol{s}$ from $\boldsymbol{x}$ from a uniformly sampled and quantized version of $\boldsymbol{x}$ without analog processing, referred to as *sampled quantized MAP*. The resulting error rates, averaged over 10^6 Monte Carlo simulations, versus SNR are depicted in Fig. 4.

We note that while the data-driven acquisition system is ignorant of the statistical model relating $\boldsymbol{s}$ and $\boldsymbol{x}(t)$ and learns its mapping from training samples corresponding to this model, the MAP receivers require accurate knowledge of the underlying model. In particular, they rely on the fact that the underlying signal model is given by (19), and require knowledge of $\boldsymbol{H}(t)$. As accurate knowledge of the signal model may not be available in some scenarios, we also depict in Fig. 4 the error rate obtained by the sampled MAP receiver as well as the MAP rule for recovering $\boldsymbol{s}$ from $\boldsymbol{x}$ without sampling and quantization constraints, when these receivers have access to a noisy version of $\boldsymbol{H}(t)$, in which each entry is corrupted by additive i.i.d. Gaussian noise whose variance is 30% of its magnitude. This scenario is referred to as $\boldsymbol{H}(t)$ uncertainty. To evaluate our acquisition system under $\boldsymbol{H}(t)$ uncertainty, we compute its achievable error rate when trained using samples taken from the same inaccurate noisy signal model. In principle, one can also design dedicated robust model-based detectors that are aware of the presence of model mismatches, thus deviating from the standard MAP operation. Here, we focus on comparing the operation of algorithms based

on the MAP for Gaussian settings as in (19) to the deep task-based acquisition system with fixed architecture, exploring the ability to use the same algorithm while coping with uncertainty. Such studies were also considered in the context of DNN-aided symbol detection in, e.g., [46], [47].

Observing Fig. 4, we note that for accurate training, our deep task-based acquisition system achieves comparable performance to the sampled MAP which operates without quantization constraints. Furthermore, our data-driven system notably outperforms the quantized MAP rule, which utilizes uniform quantizers of lesser resolution, as it does not reduce the dimensionality in analog and must thus assign less bits for each ADC. In the presence of model uncertainty, the performance of our proposed system is degraded by approximately 5 dB in SNR compared to accurate training, yet it is still capable of achieving error rate below 10^{-4} for SNRs above 10 dB. The model-based MAP rule operating without quantization constraints, whether processing a uniformly sampled input or even on the densely discretized $\boldsymbol{x}$, reaches an error floor of above 10^{-3}. These results demonstrate the ability of the proposed deep task-based acquisition framework in jointly optimizing the analog and digital mappings along with the ADC rule in a manner which allows to accurately carry out the desired task.

C. Meta-Learned Acquisition Hyperparameters

Next, we numerically evaluate the meta-learning procedure detailed in Section IV for optimizing the acquisition hyperparameters. We do so by applying Algorithm 1 to optimize the parameters of the acquisition system, such that the overall number of bits utilized is minimized without degrading the overall performance. The network weights are selected as the ones trained along with the selected hyperparameters in the Bayesian optimization procedure, as discussed in Subsection IV-B.

Here, we fix the maximal overall bit budget to $\mathcal{B} = 20$ bits. Since hyperparameters optimization does not involve computing the gradient of its objective $f_{\mathcal{A}}(\boldsymbol{\theta})$, we use the error rate objective, being the desired performance measure, rather than the cross entropy loss (9) which is used when training the weights. In particular, we design the Bayesian optimization procedure to both tune $p, \tilde{L}, \tilde{M}$ to get low error rate as possible, and use lowest possible number of bits under the constraint of our bit budget. The objective is thus set to

$$\boldsymbol{\theta}^* \triangleq \left(p^*, \tilde{L}^*, \tilde{M}^*\right) = \underset{p\cdot\tilde{L}\cdot\lceil\log_2(\tilde{M})\rceil\leq\mathcal{B}}{\arg\min} \alpha\cdot\left(p\cdot\tilde{L}\cdot\lceil\log_2(\tilde{M})\rceil\right) + \sum_{\rho\in\mathcal{P}} \mathrm{ER}_{\rho}\left(\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}}\right), \tag{21}$$

where $\mathcal{P}$ is the set of SNR values for each channel we are testing; $\mathrm{ER}_{\rho}(\gamma_{\boldsymbol{\psi},\boldsymbol{\theta}})$ is the error rate achieved using an acquisition system with hyperparameters $\boldsymbol{\theta}$ that was trained with channel with SNR ρ; and α balances the contribution of two measures one is interested in minimizing: the number of bits the acquisition system is using and the model performance respectively. Specifically, we utilize Ax [44] for Bayesian optimization. Since Ax supports Bayesian optimization with additive constraints, we replace $(p, \tilde{L}, \tilde{M})$ with their logarithms as our variables, converting the multiplicative constraint in (21) into an additive one.

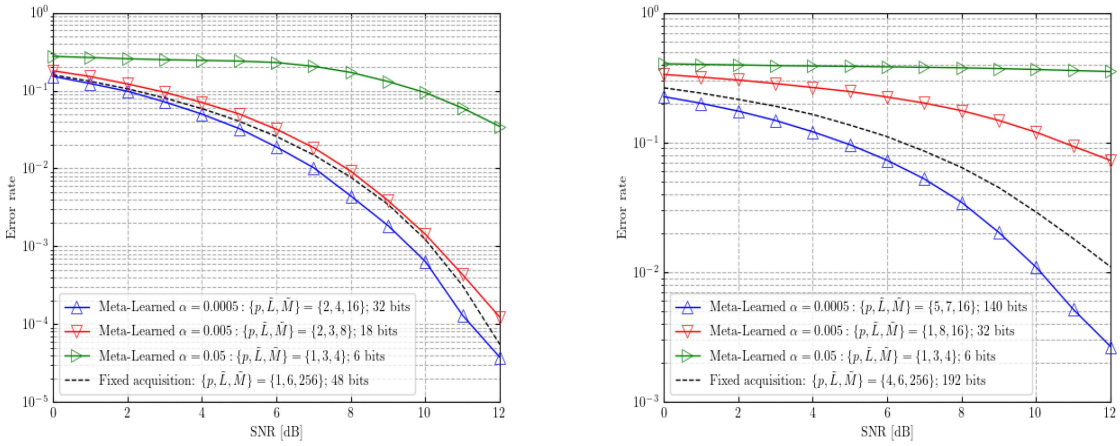
In the setting with fixed hyperparameters detailed in Subsection V-B, the acquisition system uses $p = 1$ ADC, $\tilde{L} = 6$ samples out of a grid of $L = 20$ time instances, and samples are quantized using $\log_2 \tilde{M} = 8$ bits. Thus, an overall of $1 \cdot 6 \cdot 8 = 48$ bits are used for acquisition. As observed in Fig. 5(a), the proposed Bayesian meta-learning scheme allows to achieve error rates results within a minor gap of that of the original configuration while using 62.5% less bits. In Fig. 5(b), we illustrate that for larger scale settings, with $k = 8$ and $n = 16$, hyperparameter optimization via Algorithm 2 manages to reduce the number of bits by 83% with a relatively small loss of approximately 4 dB in SNR. In both Fig. 5(a) and Fig. 5(b) we observe that the proposed meta-learning even allows to achieve improved accuracy while reducing the number of bits by 33% and 27%, respectively. This is due to its inherent training of multiple systems and the selection of the most accurate one. These results indicate that the proper combination of Bayesian meta-learning with learning of the overall mapping via deep task-based acquisition allows to improve both performance and bit efficiency.

In Fig. 6 we depict a contour plot representing the objective $f_{\mathcal{A}}(\theta)$ as a function of p and L when fixing the quantization resolution to be $\tilde{M} = 4$. This plot shows the relations between number of ADCs and number of samples taken when the quantization resolution is relatively small. We can see from Fig. 6 that Algorithm 2 is likely to prefer hyperparameter configurations taking small number of ADCs with relatively large amount of samples. Another option shown in the plot is taking relatively large amount of ADCs with low amount of samples, shown to be a local minimum of the objective function. However, when looking in the standard deviation contour plot we can see that a configuration like this will not be stable, because the standard deviation is high.

VI. Case Study: Ultrasound Beamforming

In this section we apply the deep task based acquisition framework to a real-world case study of ultrasound image reconstruction, which can be modeled as a regression problem. Here, we show that joint training of task-based ADCs combined with deep learning based adaptive beamforming can achieve high quality imaging at low data rates, improving over competing approaches.

Ultrasound imaging is based on the transmission and reflection of high frequency sound waves in tissue. These reflections, denoted $\{x_i(t)\}_{i=1}^n$, are recorded by an array of n transducer elements, and are used to form a brightness mode (B-mode) image of the tissue by applying beamforming. Typical ultrasound devices use probes with $n = 128$ or more transducer elements, each of which contributing a separate data-stream. Additionally, multiple consecutive recordings might be required for a single frame in order to achieve an SNR that leads to the desired image properties. The transfer of these recordings involves hardware that supports a large data bandwidth, which



(a) Synthetic signal model, $k = 4$ and $n = 6$

(b) Synthetic signal model, $k = 8$ and $n = 16$

Fig. 5. Error rate versus SNR for deep task-based acquisition with meta-learned acquisition parameters versus fixed parameters. (a) Synthetic signal model, $k = 4$ and $n = 6$. (b) Synthetic signal model, $k = 8$ and $n = 16$.

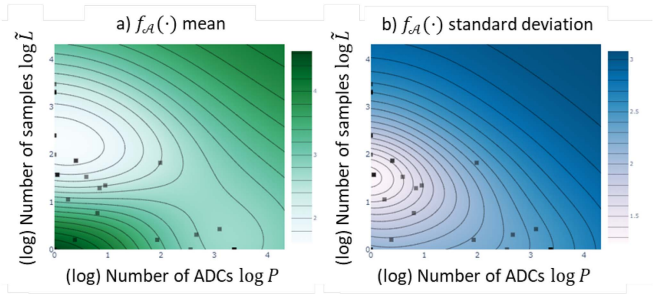


Fig. 6. Contour plots of the (a) mean and (b) standard deviation of $f_{\mathcal{A}}$ as function of number of ADCs, and number of samples when quantization resolution is fixed $\tilde{M} = 4$. Squares marked values of $f_{\mathcal{A}}$ queried in hyperparameter optimization.

is expensive and not always feasible to implement. In recent developments such as 3D ultrasound ($n > 512$) or portable scanners (bandwidth constrained), these limitations are especially problematic. The need for compression of data early on in the signal chain motivates the use of our proposed deep task based signal acquisition in ultrasound imaging, which is studied in this section, beginning the description of the experimental setup in Subsection VI-A, and followed by the statement of the results in Subsection VI-B.

A. Experimental Setup

1) Data: We consider plane-wave (PW) imaging, in which a planar wavefront is transmitted, energizing the whole imaging medium with a single pulse. For training we use 1000 in-vivo recordings, acquired using a Verasonics Vantage system with the L11-4v linear probe. Additionally, 100 images are obtained for testing purposes. High quality target images were generated similar to [48], by computationally intensive minimum variance beamforming.

2) Filtering and Quantization: Following the task-based acquisition model depicted in Fig. 1, we consecutively filter, sample and quantize $\boldsymbol{x}(t)$. During the forward pass, we follow (2), (5) and hard quantization as in (6). The input to the ADCs is given by

$$\boldsymbol{y}(t) = \boldsymbol{G}_1 \boldsymbol{x}(t) \tag{22}$$

where $\boldsymbol{G}_1 \in \mathbb{R}^{p \times n}$ is a trainable matrix. Note that (22) specializes the generic formulation of (2) by restricting the time-varying analog filter $\boldsymbol{G}(\cdot, \cdot)$ to represent time-invariant spatial combining, i.e., take on the form $\boldsymbol{G}(t, \tau) = \boldsymbol{G}_1 \delta(\tau)$. The filtering operation transforms the n channel signals into a linear combination of p channels, such that $p \leq n$, to further reduce data rates. After quantization, the signals are expanded again into a set of n signals through

$$\boldsymbol{q}[i] = \boldsymbol{G}_2 Q_{\tilde{M}} \left(\boldsymbol{y} \left(i T_L \right) \right), \tag{23}$$

where $\boldsymbol{G}_2 \in \mathbb{R}^{n \times p}$ is another trainable matrix that maps back to the original array geometry.

To simplify notation, we summarize the sampling, quantization and filtering operations in (22) and (23) as a single function $g_{\boldsymbol{\phi}}(\cdot)$ such that

$$\boldsymbol{q}[i] = g_{\boldsymbol{\phi}} \left(\boldsymbol{x}(i T_L) \right). \tag{24}$$

The vector $\boldsymbol{\phi}$ constitutes the trainable parameters, corresponding to the quantization levels and the filters $\boldsymbol{G}_1, \boldsymbol{G}_2$; the sampling instances in this experimental study represent fixed uniform sampling. This is typical in ultrasound systems due to dense scattering in tissue, and the need to cover a broad imaging region.

3) Beamforming: Next, the digital signals are focused towards each pixel position by applying delays, effectively transforming $\boldsymbol{q}$ from the time-domain to the pixel-domain. For each pixel, this yields the $n \times 1$ channel domain signal given by

$$\begin{aligned} \boldsymbol{u}(\boldsymbol{r}) &= [u_1(\boldsymbol{r}), u_2(\boldsymbol{r}), \ldots, u_n(\boldsymbol{r})] \\ &= [\boldsymbol{q}[\Delta_1(\boldsymbol{r})], \boldsymbol{q}[\Delta_2(\boldsymbol{r})], \ldots, \boldsymbol{q}[\Delta_n(\boldsymbol{r})]], \end{aligned} \tag{25}$$

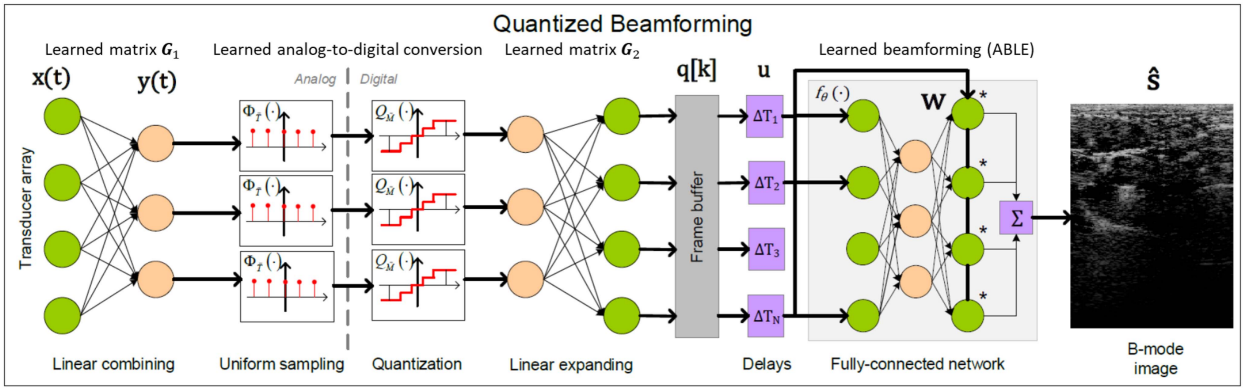


Fig. 7. Schematic overview of the task-based analog-to-digital conversion for ultrasound, jointly trained with adaptive beamforming by deep learning. The (analog) RF signals are consecutively compressed in p channels, uniformly sampled, quantized, and expanded back into n channels before being stored in a digital buffer. Trough delays, these RF lines can be focused to individual pixels, after which adaptive apodization (ABLE) is applied to yield a beamformed RF image.

where $\boldsymbol{r}$ is a coordinate vector towards that pixel position, and $\Delta_i(\boldsymbol{r})$ the corresponding time-delay.

In conventional delay-and-sum beamforming (DAS), these delayed signals are weighed according to a window $\mathbf{w}(\boldsymbol{r})$, favouring either contrast (Hanning window) or resolution (rectangular window), and subsequently summed. This operation is given by

$$s_{\mathrm{DAS}}(\boldsymbol{r}) = \boldsymbol{w}^{\mathrm{T}}(\boldsymbol{r})\mathbf{u}(\boldsymbol{r}). \tag{26}$$

where $\mathrm{s}_{\mathrm{DAS}}$ denotes the beamformed signal output at every pixel index. Note here that, while $\boldsymbol{r}$ can vary per pixel, it does not adapt to the received signals.

To improve upon such a fixed apodization scheme, an adaptive method can be employed in the digital domain. Here, we consider the trainable Adaptive Beamforming by Deep Learning (ABLE) [48], a model-based deep learning framework which can learn to predict optimal channel apodizations based on time-delayed RF data. ABLE can be written as an apodization function $f_{\boldsymbol{\theta}}(\cdot)$, which depends on a set of trainable parameters $\boldsymbol{\theta}$, and maps an input signal to a content-adaptive apodization pattern. The ABLE model combined with the mapping matrix $\boldsymbol{G}_2$ constitute the digital processing of the generic task-based acquisition system of Fig. 2, as illustrated for the considered ultrasound beamforming setup in Fig. 7.

To conclude, our beamformed output signal is given by

$$\hat{\boldsymbol{s}}(\boldsymbol{r}) = f_{\boldsymbol{\theta}}\left(\mathbf{u}\left(\boldsymbol{r}\right)\right)^{T}\mathbf{u}(\boldsymbol{r}), \tag{27}$$

where $\hat{\boldsymbol{s}}$ denotes the predicted beamformed signal. The trainable parameters here constitute the weights and biases of four fully-connected layers of which ABLE is comprised [48], along with the ADC parameters and the filters $\boldsymbol{G}_1, \boldsymbol{G}_2$. Finally, a B-mode image is obtained by envelope detection and logarithmic compression of the beamformed output.

4) Training: Because of the non-uniform distribution of ultrasound data, we initialize the ADCs with a logarithmic quantization rule (exponentially spaced), which is known to result in a higher quantization resolution in the low-intensity ranges [49]. Training is based on a signed-mean-squared-logarithmic-error loss function, defined as

$$\mathcal{L}\left(\hat{\boldsymbol{s}}, \boldsymbol{s}\right) = \frac{1}{2}\| \log_{10}\left(\hat{\boldsymbol{s}}^{+}\right) - \log_{10}\left(\boldsymbol{s}^{+}\right) \|_2^2 + \frac{1}{2}\| \log_{10}\left(-\hat{\boldsymbol{s}}^{-}\right) - \log_{10}\left(-\boldsymbol{s}^{-}\right) \|_2^2, \tag{28}$$

where $\hat{\boldsymbol{s}}$ and $\boldsymbol{s}$ denote the predicted and target frames, respectively, defined as $\hat{s}^{+} := \max(s, 0)$ and $\hat{s}^{-} := -\min(s, 0)$, and $(\cdot)^{\pm}$ denote the positive and negative signal components. Training the network parameters $f_{\boldsymbol{\theta}}$ and $g_{\boldsymbol{\phi}}$ can then be formulated as a regression problem, aiming at setting $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ to minimize the loss between a desired image $\boldsymbol{s}$ and its reconstructed one obtained via (27).

5) Evaluation: We train a set of models with different compression ratios by changing the rate of analog combining and the number quantization levels (bits) in the ADC. For each level of compression, numerical performance is assessed by measuring the contrast-to-noise ratio (CNR) over a simulated anechoic cyst phantom from the PICMUS dataset [50]. The CNR is defined as

$$\mathrm{CNR} = 20\log_{10}\left(\frac{|\mu_{\mathrm{low}} - \mu_{\mathrm{high}}|}{\sqrt{\left(\sigma_{\mathrm{low}}^2 + \sigma_{\mathrm{high}}^2\right)/2}}\right), \tag{29}$$

where $\mu_{\mathrm{low}}, \mu_{\mathrm{high}}, \sigma_{\mathrm{low}}^2$ and σ_{high}^2 represent the mean intensities and the variances of the anechoic and hyperechoic regions, respectively. Furthermore we evaluate the mean-absolute-error (MAE) between the baseline (uncompressed) and compressed reconstructions. It should be noted however, that this metric does not directly provide an measure of image quality, but gives a good indication of similarity to the training target.

B. Results

We evaluate deep task-based acquisition with the ABLE digital beamformer for different levels of data compression by varying the amount of analog combining and quantization in the model. To that end, we demonstrate 3 settings, at compression rations of 2.6, 4 and 8. These results are compared against a

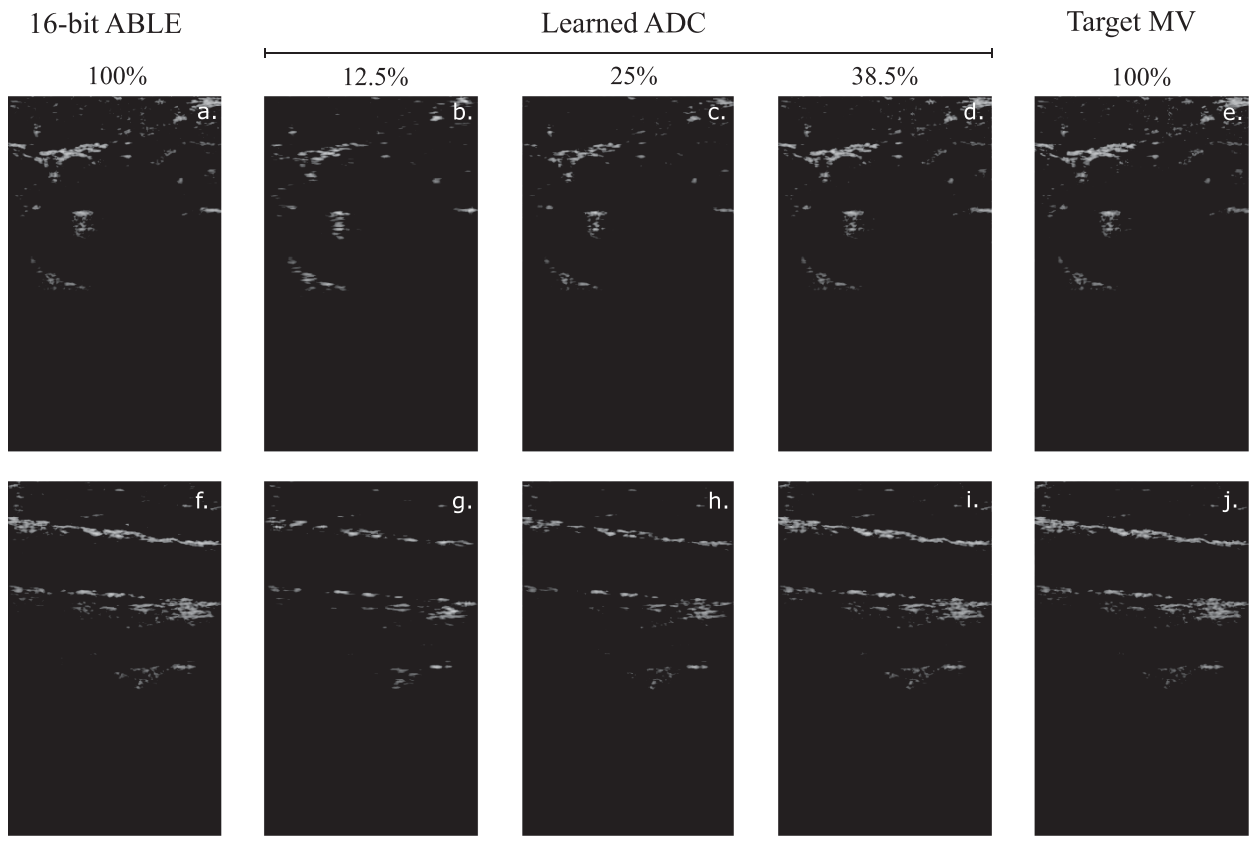


Fig. 8. Image reconstructions, with a dynamic range of 60 dB, for: (a)–(e) Carotid artery cross-section, (f)–(g) Carotid artery longitudinal cross-section. For each image, reconstructions are shown for: Uncompressed 16-bit ABLE (a) & (f), the proposed task based acquisition with compression ratios of 12.5% (b) & (g), 25% (c) & (h), and 38.5% (d) & (i), and finally the 16-bit minimum variance beamformed training targets. (e) & (j).

TABLE I
NUMERICAL METRICS

	Compression	CNR	MAE
Learned ADC	8x	7.01dB	4.17
	4x	10.39dB	2.99
	2.67x	12.25dB	1.3
Fixed ADC	8x	4.71dB	4.28
	4x	8.47dB	3.62
	2.67x	11.98dB	1.33
Baseline	1x	12.22dB	0

baseline model, which is ABLE without deep task-based acquisition [48]. Furthermore we compare the learned ADC strategy with a deterministic (non-learned) approach having the same bit-budget, to show the performance difference between the two strategies.

In Table I we show the MAE and CNR for the different model settings. Additionally we provide the rate of analog combining and bitrate that achieve a specific compression ratio. As expected we see reduced CNR, and increased MAE compared to the target image, for higher compression ratios. Furthermore we see that in all cases, the task-based ADC framework outperforms the fixed scheme. That is, using a fixed logarithmic quantization rule, and ABLE as beamformer.

To demonstrate that the performance metrics in Table I are indeed translated into a clear ultrasound image acquired in a compressed manner, we show in Fig. 8 the reconstructed images of two in-vivo records for the baseline model, the deep task-based ADC, and the target algorithm. Here, we can see that the model can handle reconstruction at tight bit-budgets well, yielding similar to baseline images. From the different compression levels we see that at rates of 2.6x and even 4x compression, the images are not notably affected in terms of image quality. However, as can be expected, at more extreme rates (i.e. 8x) the images start to get more blurry and fine details are lost. These results demonstrate the ability of deep task-based acquisition to facilitate operation with reduced number of bits in practical applications involving analog-to-digital conversion.

VII. CONCLUSION

In this work we designed a deep task-based acquisition system which learns to map a set of analog signals into an estimate of an underlying task vector, obtained in the digital domain, in a data-driven manner. Our system adjusts its ADC mapping by approximating its continuous-to-discrete conversions using differentiable functions, allowing to learn non-uniform mappings and to train the overall system in an end-to-end fashion. The proposed system was evaluated in both a synthetic detection setup as well as for ultrasound beamforming scenario, demonstrating its gains over using uniform ADCs and digital processing.

REFERENCES

[1] N. Shlezinger, R. J. van Sloun, I. A. Huijben, G. Tsintsadze, and Y. C. Eldar, "Learning task-based analog-to-digital conversion for MIMO receivers," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 9125–9129.

[2] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge University Press, 2015.

[3] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[4] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Analog-to-digital compression: A new paradigm for converting signals to bits," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 16–39, May 2018.

[5] M. Xiao et al., "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

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

[7] R. T. Yazicigil, T. Haque, P. R. Kinget, and J. Wright, "Taking compressive sensing to the hardware level: Breaking fundamental radio-frequency hardware performance tradeoffs," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 81–100, Mar. 2019.

[8] N. Jain et al., "Esampling: Energy harvesting ADCs," 2020, *arXiv:2007.08275*.

[9] S. Lee, A. P. Chandrakasan, and H. Lee, "A 1 GS/s 10b 18.9 mW time-interleaved SAR ADC with background timing skew calibration," *IEEE J. Solid-State Circuits*, vol. 49, no. 12, pp. 2846–2856, Dec. 2014.

[10] Y. Zhou, B. Xu, and Y. Chiu, "A 12-b 1-GS/s 31.5-mW time-interleaved SAR ADC with analog HPF-assisted skew calibration and randomly sampling reference ADC," *IEEE J. Solid-State Circuits*, vol. 54, no. 8, pp. 2207–2218, Aug. 2019.

[11] N. Shlezinger, Y. C. Eldar, and M. R. Rodrigues, "Hardware-limited task-based quantization," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5223–5238, Oct. 2019.

[12] N. Shlezinger, Y. C. Eldar, and M. R. Rodrigues, "Asymptotic task-based quantization with application to massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3995–4012, Aug. 2019.

[13] S. Salamtian, N. Shlezinger, Y. C. Eldar, and M. Medard, "Task-based quantization for recovering quadratic functions using principal inertia components," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 390–394.

[14] N. Shlezinger and Y. C. Eldar, "Deep task-based quantization," *Entropy*, vol. 23, no. 1, 2021, Art. no. 104.

[15] I. A. M. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. J. G. van Sloun, "Learning sub-sampling and signal recovery with applications in ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3955–3966, Dec. 2020.

[16] S. Mulleti, H. Zhang, and Y. C. Eldar, "Learning to sample: Data-driven sampling and reconstruction of FRI signals," 2021, *arXiv:2106.14500*.

[17] G. Solodky and M. Feder, "Sampling a noisy multiple output channel to maximize the capacity," in *Proc. IEEE 26th Eur. Signal Process. Conf.*, 2018, pp. 445–449.

[18] X. Liu, E. Gönültaş, and C. Studer, "Analog-to-feature (A2F) conversion for audio-event classification," in *Proc. IEEE 26th Eur. Signal Process. Conf.*, 2018, pp. 2275–2279.

[19] P. Neuhaus, N. Shlezinger, M. Dörpinghaus, Y. C. Eldar, and G. Fettweis, "Task-based analog-to-digital converters," *IEEE Trans. Signal Process.*, vol. 69, pp. 5403–5418, 2021.

[20] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 401–429, Jan. 2016.

[21] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Fundamental distortion limits of analog-to-digital compression," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6013–6033, Sep. 2018.

[22] Z. Shao, L. T. Landau, and R. C. de Lamare, "Dynamic oversampling for 1-bit ADCs in large-scale multiple-antenna systems," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3423–3435, May 2021.

[23] Z. Shao, L. T. Landau, and R. C. De Lamare, "Channel estimation for large-scale multiple-antenna systems using 1-bit ADCs and oversampling," *IEEE Access*, vol. 8, pp. 85243–85256, 2020.

[24] E. Agustsson et al., "Soft-to-hard vector quantization for end-to-end learning compressible representations," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 1141–1151, 2017.

[25] P. I. Frazier, "A tutorial on bayesian optimization," 2018, *arXiv: 1807.02811*.

[26] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, "Performance bounds and design criteria for estimating finite rate of innovation signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 4993–5015, Aug. 2012.

[27] N. Shlezinger, S. Salamtian, Y. C. Eldar, and M. Medard, "Joint sampling and recovery of correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 385–389.

[28] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[29] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog–digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.

[30] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "RF chain reduction for MIMO systems: A hardware prototype," *IEEE Syst. J.*, vol. 14, no. 4, pp. 5296–5307, Dec. 2020.

[31] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1249–1257, Jul. 1992.

[32] T. Zirtiloglu, N. Shlezinger, Y. C. Eldar, and R. T. Yazicigil, "Power-efficient hybrid MIMO receiver with task-specific beamforming using low-resolution ADCs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2022, pp. 5338–5342.

[33] N. Shlezinger, O. Dicker, Y. C. Eldar, I. Yoo, M. F. Imani, and D. R. Smith, "Dynamic metasurface antennas for uplink massive MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6829–6843, Oct. 2019.

[34] H. Wang et al., "Dynamic metasurface antennas for MIMO-OFDM receivers with bit-limited ADCs," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2643–2659, Apr. 2021.

[35] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 106–113, Apr. 2021.

[36] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[37] L. Danial, N. Wainstein, S. Kraus, and S. Kvatinsky, "Breaking through the speed-power-accuracy tradeoff in ADCs using a memristive neuromorphic architecture," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 5, pp. 396–409, Oct. 2018.

[38] O. Vinyals et al., "Matching networks for one shot learning," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 3630–3638, 2016.

[39] D. Maclaurin, D. Duvenaud, and R. Adams, "Gradient-based hyperparameter optimization through reversible learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2113–2122.

[40] O. Wichrowska et al., "Learned optimizers that scale and generalize," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 3751–3760.

[41] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[42] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2010, *arXiv:1012.2599*.

[43] M. Balandat et al., "Botorch: Programmable Bayesian optimization in pytorch," 2019, *arXiv:1910.06403*.

[44] "Ax adaptive experimentation platform," [Online]. Available: https://ax.dev/

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[46] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1349–1362, Feb. 2021.

[47] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5663–5678, Nov. 2018.

[48] B. Luijten et al., "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020.

[49] J. Guilherme and J. Vital, "Logarithmic Analogue-to-Digital Converters," in *CMOS Telecom Data Converters*, A. Rodríguez-Vázquez, F. Medeiro, and E. Janssens, Eds., Boston, MA ,USA: Springer, 2003, pp. 241–275.

[50] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *Proc. IEEE Int. Ultrasonics Symp.*, 2016, pp. 1–4.

Nir Shlezinger received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from Ben-Gurion University, Beersheba, Israel, in 2011, 2013, and 2017, respectively. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Ben-Gurion University. From 2017 to 2019, he was a Postdoctoral Researcher with Technion, and from 2019 to 2020, he was a Postdoctoral Researcher with the Weizmann Institute of Science. His research interests include communications, information theory, signal processing, and machine learning. He was the recipient of the FGS prize for outstanding research achievements with the Weizmann Institute of Science.

Ariel Amar received the B.Sc. degree in electrical and computer engineering from The Hebrew University of Jerusalem, Jerusalem, Israel, in 2018. He is currently working toward the M.Sc. degree in mathematics and computer science with the Weizmann Institute of Science, Rehovot, Israel. His research interests include deep learning, medical imaging, signal processing, and deep learning for ultrasonic imaging.

Ben Luijten (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering in 2017 and 2019, respectively, from the Eindhoven University of Technology, Eindhoven, The Netherlands, where he is currently working toward the Ph.D. degree with Biomedical Diagnostics Laboratory. His research interests include deep learning for signal processing, ultrasound beamforming, model-based deep learning, and compressed sensing.

Ruud J. G. van Sloun received the B.Sc. and M.Sc. degrees (*cum laude*) in electrical engineering and the Ph.D. degree (*cum laude*) from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2012, 2014, and 2018, respectively. Since then, he has been an Assistant Professor with the Department of Electrical Engineering, Eindhoven University of Technology, and a Kickstart-AI Fellow with Philips Research, Eindhoven, since January 2020. From 2019 to 2020, he was also a Visiting Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. He is currently an NWO Rubicon laureate, and was the recipient of an ERC starting grant and Google Faculty Research Award. His research interests include deep learning for signal processing and imaging, active signal acquisition, model-based deep learning, compressed sensing, ultrasound imaging, and probabilistic signal and image reconstruction.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was previously a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair of Engineering. She is also a Visiting Professor with MIT, a Visiting Scientist with Broad Institute, and an Adjunct Professor with Duke University, Durham, NC, USA, and was a Visiting Professor at Stanford.

She is the author of the book *Sampling Theory: Beyond Bandlimited Systems* and co-author of five other books published by Cambridge University Press. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging and optics.

Dr. Eldar was the recipient of many awards for excellence in research and teaching, including IEEE Signal Processing Society Technical Achievement Award (2013), IEEE/AESS Fred Nathanson Memorial Radar Award (2014), IEEE Kiyo Tomiyasu Award (2016), Michael Bruno Memorial Award from the Rothschild Foundation, Weizmann Prize for Exact Sciences, Wolf Foundation Krill Prize for Excellence in Scientific Research, Henry Taub Prize for Excellence in Research (twice), Hershel Rich Innovation Award (three times), Award for Women with Distinguished Contributions, Andre and Bella Meyer Lectureship, Career Development Chair at the Technion, Muriel & David Jacknow Award for Excellence in Teaching, Technion's Award for Excellence in Teaching (two times), and best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize, UFFC Outstanding Paper Award, Signal Processing Society Best Paper Award, and IET Circuits, Devices and Systems Premium Award, was selected as one of the 50 most influential women in Israel and in Asia. She was a Member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a Member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. She was a Signal Processing Society Distinguished Lecturer, Member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing*, *SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*. She was the Co-Chair and Technical Co-Chair of several international conferences and workshops. She is a Member of the Israel Academy of Sciences and Humanities (elected 2017), and a EURASIP Fellow.