# Learning-Based Joint Waveform Optimization and Receiver Design for Dual-Functional MIMO Radar and Communications

Ziwei Kang and Qian He*, *Senior Member, IEEE*
University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

***Abstract*—For the joint multiple-input multiple-output (MIMO) radar and MIMO communications system with widely separated antennas, designing the waveforms properly plays a pivotal role, especially when dual-functional is requested. Considering target detection and decoding as the main radar and communications tasks, for the purpose of improving both performance, in this paper the dual-functional waveforms are designed jointly with the radar and communications receivers. Supervised learning is employed for the radar detector and communications decoder, while reinforcement learning is employed for the waveform design. We propose a joint training framework, where the detection and decoding performance metrics are used as the loss function of the supervised learning and the reward of the reinforcement learning. The reinforcement learning-based transmitter and supervised learning-based receivers are trained in an iterative way to meet the desired performance. Numerical examples are provided to show the efficiency and effectiveness of the proposed method.**

***Index Terms*—Dual-functional radar and communications (DFRC), multiple-input multiple-output (MIMO), reinforcement learning, supervised learning.**

## I. Introduction

With the number of electronic devices growing, spectrum resources become scarce, leading to overlapping operating bands between radar and communications. Dual-functional radar and communications (DFRC) systems can be an effective solution [1]. The DFRC system can reduce system overhead and realize tight integration in which dual-functional waveforms emitted from the shared transmitters perform both radar and communications functionalities simultaneously [2].

Waveforms play a pivotal role in the DFRC system, the design [3]–[7] of which hence is of great importance. Recently, orthogonal chirp-division multiplexing (OCDM) signals, which have the robustness to multipath propagation, high processing gain, and high data rate, have been used in the DFRC system for target detection [5] and decoding [6], [7]. Spatially orthogonal OCDM (SO-OCDM) waveforms are suitable for multiple-input multiple-output (MIMO) systems, in which the OCDM waveforms contributed from different transmitters are spatially orthogonal and subcarriers of the OCDM signal in the same bandwidth are mutually orthogonal. In this paper, we study the SO-OCDM waveform design for the MIMO DFRC system, with target detection and symbol decoding as the main tasks for the radar and communications subsystems, respectively.

Towards further improving the system performance of the DFRC system, the waveforms are optimized jointly with receiver design [8]–[10]. The authors in [9] and [10] joint designed waveforms and radar receiver filters based on the alternative direction method of multipliers and successive convex approximation algorithm. In [11] and [12], a supervised learning (SL) based autoencoder framework was proposed for detection, estimation, and decoding in a co-located MIMO DFRC system. Co-design of the waveforms and receivers is studied in this paper to enhance the target detection and symbol decoding performance.

For the joint waveform and receiver design problem, we employ a framework combining reinforcement learning (RL) and SL. The advantage of similar framework has been shown in [13] for joint encoding and decoding for a single-input signal-output (SISO) communications system and in [14] for joint waveform design and detection for a SISO radar system in combatting complex channels or clutter statistics. Inspired by [13] and [14], the joint SO-OCDM waveforms, detector, and decoder design based on the RL and SL is studied in this work for the MIMO DFRC system. RL and SL are adopted to optimize the waveforms and execute the detection and decoding tasks, and meanwhile the rewards to the RL-based transmitter are provided by SL. The optimization problem is handled by an iterative training process of the SL and RL.

## II. Reinforcement Learning Based Dual-Functional Transmitter

### A. Dual-Functional SO-OCDM Waveforms

Consider a MIMO DFRC system with $M$ transmitters. Assume one frame of a transmitted signal consists of $P_m$ SO-OCDM symbols, the sampling period is $T_s$, the index of the samples is $k$ ($k = 1, ..., K$), and the total number of samples is $K$. The SO-OCDM waveform emitted from the $m$th ($m = 1, ..., M$) transmitter is

$$x_m[k] = \sqrt{E_m} \sum_{p=0}^{P_m-1} \sum_{i=0}^{I_m-1} \frac{1}{I_m} a_{p,i,m} s_{p,i,m}[k] \operatorname{rect}\left(\frac{kT_s - pT_{\mathrm{o},m}}{T_{\mathrm{o},m}}\right) \quad (1)$$

*Correspondence: qianhe@uestc.edu.cn

This work is supported by the Municipal Government of Quzhou under Grant No.2022D002, and by the National Nature Science Foundation of China under Grant No. U21B2014.

where $E_m$ is the energy of the waveform, $I_m$ the number of chirps, $a_{p,i,m}$ the modulated data for the $i$th ($i = 1, ..., I_m$) chirp in the $p$th ($p = 1, ..., P_m$) symbol, rect$(\cdot)$ denotes the rectangle function,

$$s_{p,i,m}[k] = \begin{cases} 0, & kT_s \in [pT_{o,m}, pT_{o,m} + T_{GI,m}) \\ s_{p,i,m}^{w/o\ GI}[k], & kT_s \in [pT_{o,m} + T_{GI,m}, (p+1)T_{o,m}) \end{cases} \quad (2)$$

the chirp signal,

$$s_{p,i,m}^{w/o\ GI}[k] = e^{\frac{j\pi}{4}} e^{-j\pi \frac{I_m}{T_m^2}\left(kT_s - i\frac{T_m}{I_m}\right)^2} e^{j2\pi m f_{\Delta,m} kT_s} \quad (3)$$

$T_{o,m} = T_m + T_{GI,m}$, $T_m$ is time duration of one SO-OCDM symbol, $T_{GI,m}$ is the guard interval (GI), and $f_{\Delta,m}$ denotes the frequency spacing between waveforms of two adjacent transmitters. From (3), the chirp rate and bandwidth can be calculated as $I_m/{T_m}^2$ and $B_m = I_m/T_m$, respectively. Let $f_{\Delta,m} > B_m, m = 1, ..., M$ to ensure the waveforms from different transmitters are spatially orthogonal, such that $\sum_{k=1}^{K_m} x_m^*[k] x_{m'}[k] = 0$ for $m \neq m'$, where $(\cdot)^*$ denotes conjugation. The orthogonality of chirps in the $p$th SO-OCDM symbols means $\sum_{k=K_{GI,m}}^{K_{o,m}} (s_{p,i,m}^{w/o\ GI}[k])^* s_{p,i',m}^{w/o\ GI}[k] = 0$ for $i \neq i'$, where $K_{GI,m}$ and $K_{o,m}$ are the sample number for the GI and for a SO-OCDM symbol. To facilitate future use, we denote the modulated data matrix by $\boldsymbol{A} = [\boldsymbol{a}_{1,1}, \boldsymbol{a}_{1,2}, \cdots \boldsymbol{a}_{P_M,I_M}]^T$, in which $\boldsymbol{a}_{p,i} = [a_{p,i,1}, \cdots, a_{p,i,M}]^T$ and $(\cdot)^T$ denotes transpose.

It can be seen from (1) that the SO-OCDM waveform $x_m[k]$ is determined by the associated parameters, which are collected together in a parameter vector,

$$\boldsymbol{\lambda}_m = [E_m, T_m, T_{G,m}, I_m, P_m]^T. \quad (4)$$

The parameter vectors of the waveforms from all the transmitter are

$$\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1^T, \cdots, \boldsymbol{\lambda}_M^T]^T \quad (5)$$

Thus, waveform design of the MIMO DFRC system employing SO-OCDM is equivalent to the optimization of $\boldsymbol{\lambda}$.

### B. RL-Based Approach for Waveform Optimization

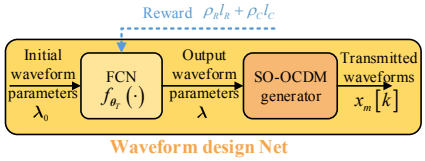


Fig. 1. Architectures of the waveform design net.

For the RL-based waveform optimization problem, SO-OCDM waveform designing is viewed as the action. The rewards provided from the SL-based receivers to the waveform design net is defined as $\rho_R l_R + \rho_C l_C$, where $l_R$ and $l_C$ denote the performance metric for the radar and communications, $\rho_R$ and $\rho_C$ denoted the weighted factors for radar and communications. As illustrated in Fig. 1, a fully connected network (FCN) $f_{\boldsymbol{\theta}_T}(\cdot)$ with trainable variables $\boldsymbol{\theta}_T$ takes the initial waveform parameter $\boldsymbol{\lambda}_0$ as the input. The network is adopted to adjust waveform parameters, namely, outputing designed waveform parameters $\boldsymbol{\lambda} = f_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_0)$. Finally, a SO-OCDM generator is applied to form a set of SO-OCDM waveforms $x_m[k]$.

## III. Supervised Learning Based Receivers

In this section, received signal models used for synthesizing training data and testing date are given. The SL-based approach is applied to design the detector and decoder.

### A. The SL-based Radar Receiver

Assume the MIMO DFRC system has $N_R$ radar receivers. The signal received at the $n_R$th ($n_R = 1, ..., N_R$) receiver is

$$y_{n_R}[k] = \sum_{m=1}^{M} \zeta_{m,n_R}^t x_m(kT_s - \tau_{m,n_R}^t) + \sum_{m=1}^{M} \zeta_{m,n_R}^d x_m(kT_s - \tau_{m,n_R}^d) + z_{n_R}[k]$$

where $\tau_{m,n_R}^t$ and $\tau_{m,n_R}^d$ are the delays associated with the $mn_R$th target reflection path and the direct path, respectively. The radar propagation loss corresponding to target reflected path and direct path are defined as $\zeta_{m,n_R}^t = \delta_{m,n_R}\sqrt{\alpha_R^t/(d_{mt}^2 d_{tn_R}^2)}$ and $\zeta_{m,n_R}^d = \sqrt{\alpha_R^d/d_{mn_R}^2}$, where $\delta_{m,n_R}$ is the target reflection coefficient, $d_{mt}$ is the distance between the $m$th transmitter and the target, $d_{tn_R}$ is the distance between the target and the $n_R$th radar receiver, and $d_{mn_R}$ is the distance between the $m$th transmitter and the $n_R$th radar receiver. The $\alpha_R^t$ and $\alpha_R^d$ are the ratio of received energy to transmitted energy at $d_{mt} = d_{tn_R} = 1$ and $d_{mn_R} = 1$, respectively. The term $z_{n_R}[k] \sim \sum_{g_R=1}^{G_R} \kappa_{g_R} \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{g_R}\right)$ is the Gaussian mixture clutter-plus-noise, where $\mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{g_R}\right)$ represents the zero-mean complex Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{g_R}$, $\kappa_{g_C}$ is the coefficient of the $g_R$th Gaussian mixture component, and $\sum_{g_R=1}^{G_R} \kappa_{g_R} = 1$ [15].

Stacking the signals from all receivers, the signal vector is

$$\begin{aligned} \boldsymbol{y}_R &= [\boldsymbol{y}_1^T, \cdots \boldsymbol{y}_{N_R}^T]^T \\ &= \boldsymbol{x}^t \boldsymbol{\zeta}^t \boldsymbol{I}_{MN_R} + \boldsymbol{x}^d \boldsymbol{\zeta}^d \boldsymbol{I}_{MN_R} + \boldsymbol{z}_R \end{aligned} \quad (6)$$

where $\boldsymbol{x}^t = \text{Diag}\{\boldsymbol{x}_1^t, \cdots, \boldsymbol{x}_{N_R}^t\}$, $\boldsymbol{x}_{n_R}^t = [\boldsymbol{x}_{n_R}^t[1], \cdots, \boldsymbol{x}_{n_R}^t[K]]^T$, $\boldsymbol{x}_{n_R}^t[k] = [x_1^t[k], \cdots, x_M^t[k]]^T$, $\boldsymbol{\zeta}^t = \text{Diag}\{\boldsymbol{\zeta}_1^t, \cdots, \boldsymbol{\zeta}_{N_R}^t\}$, $\boldsymbol{\zeta}_{n_R}^t = \text{diag}\{\zeta_{1,n_R}^t, \cdots, \zeta_{M,n_R}^t\}$, where Diag$\{\cdot\}$ and diag$\{\cdot\}$ denote block diagonal operator and diagonal operator. The terms $\boldsymbol{x}^d$ and $\boldsymbol{\zeta}^d$ in (6) are defined similarly. The $\boldsymbol{I}_M$ is a column vector with all the elements being 1, $\boldsymbol{y}_{n_R} = [\boldsymbol{y}_{n_R}[1], \cdots, \boldsymbol{y}_{n_R}[K]]^T$, $\boldsymbol{z} = \left[\boldsymbol{z}_1^T, \cdots, \boldsymbol{z}_{N_R}^T\right]^T$, and $\boldsymbol{z}_{n_R} = [\boldsymbol{z}_{n_R}[1], \cdots, \boldsymbol{z}_{n_R}[K]]^T$.

To detect whether the target is present or not in the cell-under-test (CUT), considering the hypothesis testing problem

$$\begin{cases} H_1 : \boldsymbol{y}_R = \boldsymbol{x}^t \boldsymbol{\zeta}^t \boldsymbol{I}_{MN_R} + \boldsymbol{x}^d \boldsymbol{\zeta}^d \boldsymbol{I}_{MN_R} + \boldsymbol{z}_R & (7a) \\ H_0 : \boldsymbol{y}_R = \boldsymbol{x}^d \boldsymbol{\zeta}^d \boldsymbol{I}_{MN_R} + \boldsymbol{z}_R & (7b) \end{cases}$$

where $H_1$ and $H_0$ indicate the target present hypothesis and absence hypothesis. The labels associated with target present hypothesis for (7a) and absence hypothesis for (7b) are $\varepsilon_R = 1$ and $\varepsilon_R = 0$.

The FCN $f_{\boldsymbol{\theta}_R}(\cdot)$ of detector net with a set of trainable variables $\boldsymbol{\theta}_R$ takes $\boldsymbol{y}_R$ as input, the output $p_R$ is compared with a threshold $\Lambda$ to make the final decision $\hat{\varepsilon}_R \in \{0, 1\}$, which is shown in Fig. 2. To realize the Neyman-Pearson (NP) detector using the SL, all the network output with label $\varepsilon_R = 0$ are ranked. Based on a required level of false alarm probability

$P_{FA}$, the corresponding value is set as the threshold $\Lambda$. For the SL-based detector, if the output $p_R > \Lambda$, $\hat{\varepsilon}_R = 1$, else $\hat{\varepsilon}_R = 0$.

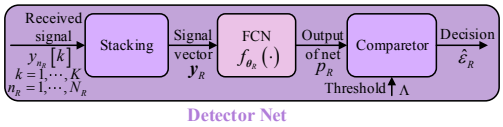


Fig. 2. Framework of SL-based detector.

### B. The SL-based Communications Receiver

Consider the system consisting of $N_C$ communications receivers. The signal received at the $n_C$th ($n_C = 1, ..., N_C$) receiver is

$$y_{n_C}[k] = \sum_{m=1}^{M} \zeta_{m,n_C}^{d} x_m(kT_s - \tau_{m,n_C}^{d}) + \sum_{m=1}^{M} \zeta_{m,n_C}^{t} x_m(kT_s - \tau_{m,n_C}^{t}) + z_{n_C}[k]$$

where $\tau_{m,n_C}^{d}$ and $\tau_{m,n_C}^{t}$ are the delays associated with the $mn_C$th direct path and the target reflected path, respectively. The communications propagation loss corresponding to direct path and target reflected path are $\zeta_{m,n_C}^{d} = \sqrt{\alpha_C^d / d_{mn_C}^2}$ and $\zeta_{m,n_C}^{t} = \delta_{m,n_C} \sqrt{\alpha_C^t / (d_{mt}^2 d_{tn_C}^2)}$, where $\delta_{m,n_C}$ is the target reflection coefficient, $d_{mn_C}$ is the distance between the $m$th transmitter and the $n_C$th communications receiver, and $d_{tn_C}$ is the distance between the target and the $n_C$th communications receiver. The terms $\alpha_C^t$ and $\alpha_C^d$ are the ratio of received energy to transmitted energy at $d_{mt} = d_{tn_C} = 1$ and at $d_{mn_C} = 1$. The Gaussian mixture clutter-plus-noise is $z_{n_C}[k] \sim \sum_{g_C=1}^{G_C} \kappa_{g_C} \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{g_C})$, where $\mathbf{\Sigma}_{g_C}$ is the covariance matrix, $\kappa_{g_C}$ is the coefficient of the $g_C$th Gaussian mixture component, and $\sum_{g_C=1}^{G_C} \kappa_{g_C} = 1$.

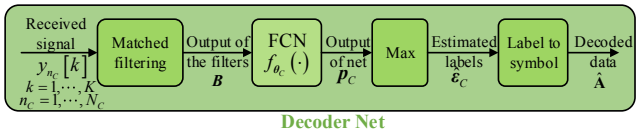


Fig. 3. Framework of SL-based decoder.

After matched filtering to the corresponding chirp defined in (3), the output of the filter is

$$b_{p,i,m} = \frac{T_{o,m}}{K_{o,m}} \sum_{k=K_{GI,m}}^{K_{o,m}} y_{n_C}^{*}[k + pK_{o,m} + K_{GI,m}] s_{p,i,m}^{\text{w/o GI}}[k] \quad (8)$$

The $b_{p,i,m}$ is the data including the effect of propagation loss, target, and noise. Assume the data in different transmitters are the same to acquire the multiplexing gain. The received noisy data matrix is $\boldsymbol{B} = [\boldsymbol{b}_{1,1}, \boldsymbol{b}_{1,2}, \cdots \boldsymbol{b}_{P,I}]^T$, where $\boldsymbol{b}_{p,i} = [b_{p,i,1}, \cdots, b_{p,i,M}]^T$ with $P_m = P$ and $I_m = I$ for $m = 1, ..., M$. The labels $\boldsymbol{\varepsilon}_C \in \{0, \cdots, J-1\}^{(PI)}$ are generating according to the modulated data matrix $\boldsymbol{A}$ with the modulation order $J$.

As shown in Fig. 3, the matrix $\boldsymbol{B} \in \mathbb{C}^{(PI,M)}$ is sent into the decoder FCN $f_{\boldsymbol{\theta}_C}(\cdot)$, where $\boldsymbol{\theta}_C$ is the set of trainable variables. The output $\boldsymbol{p}_C \in \mathbb{R}^{(PI,J)}$ is the probability matrix over the labels. Choosing the labels with the largest probability in the second dimension of $\boldsymbol{p}_C$ as estimated labels $\hat{\boldsymbol{\varepsilon}}_C \in \{0, \cdots, J-1\}^{(PI)}$, which is implemented by the max block. Finally, the decoded data matrix $\hat{\boldsymbol{A}}$ is obtained by utilizing the label to symbol block.

## IV. RL and SL Based Joint Waveform and Receiver Optimization

The radar performance metric $l_R$ is a function of $\boldsymbol{\lambda}$, $\boldsymbol{\theta}_T$, $\boldsymbol{\theta}_R$, and the communications performance metrics $l_C$ is a function of $\boldsymbol{\lambda}$, $\boldsymbol{\theta}_T$, $\boldsymbol{\theta}_C$. Thus, we denote $l_R$ and $l_C$ as $l_R(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R)$ and $l_C(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$, so the joint waveform optimization and receiver design problem can be written as

$$\begin{aligned} &\min_{\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R, \boldsymbol{\theta}_C} \rho_R l_R(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C) \\ &\text{s.t.} \quad \text{constraint on } \boldsymbol{\lambda}, \boldsymbol{\lambda} = f_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_0). \end{aligned} \quad (9)$$

The optimization involves training of the waveform design net, radar detector net, and communications decoder net. During the detector and decoder training, the parameter sets $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_C$ are updated with the fixed $\boldsymbol{\theta}_T$. The SO-OCDM waveforms are generated according to $\boldsymbol{\lambda}$ based on the deterministic policy $f_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_0)$. During the training of the waveform design net, fixing $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_C$, $\boldsymbol{\theta}_T$ is updated to minimize $\rho_R l_R(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$ as shown in Fig. 4. Additionally, to explore the waveform parameter space, a stochastic policy $\pi_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_p | \boldsymbol{\lambda}_0)$ is introduced to produce $\boldsymbol{\lambda}_p$ [16] which is depicted by the diagram with dotted black lines. For the continuous waveform parameter space, the Gaussian policy is adopted and the waveform parameters are sampled from the Gaussian distribution $\boldsymbol{\lambda}_p \sim \mathcal{N}(f_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_0), \sigma_{\boldsymbol{\lambda}}^2)$ with the mean $f_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_0)$ and the variance $\sigma_{\boldsymbol{\lambda}}^2$; for the discrete waveform parameter space, the output of the network is the probability of each discrete waveform parameter. The two training steps iterate alternatively until the entire network converges.

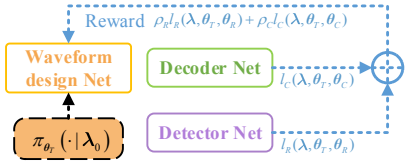


Fig. 4. Joint training process for RL and SL.

### A. Waveform Design Net Training

The training of waveform design net aims at minimizing the reward, that is, the loss function,

$$L_{\boldsymbol{\theta}_T} = \mathbb{E}_{\boldsymbol{\lambda}_p \sim \pi_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_p | \boldsymbol{\lambda}_0)} [\rho_R l_R(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C)].$$

where $\mathbb{E}(\cdot)$ denotes the ergodic average. Use $Q_T$ samples to compute the gradient of the loss function [17],

$$\nabla_{\boldsymbol{\theta}_T} L_{\boldsymbol{\theta}_T} = \frac{1}{Q_T} \sum_{q_T=1}^{Q_T} [\rho_R l_R^{(q_T)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C^{(q_T)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C)] \times \nabla_{\boldsymbol{\theta}_T} \log \pi_{\boldsymbol{\theta}_T}(\boldsymbol{\lambda}_p^{(q_T)} | \boldsymbol{\lambda}_0)$$

where $\rho_R l_R^{(q_T)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C^{(q_T)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$ is the loss and $\boldsymbol{\lambda}_p^{(q_T)}$ is the parameter vector sampled from the stochastic

policy for the $q_T$th training data. At the $u$th iteration, the $\boldsymbol{\theta}_T$ is optimized with learning rate $\eta_T$ to minimize the loss function,

$$\boldsymbol{\theta}_T^{(u+1)} = \boldsymbol{\theta}_T^{(u)} - \eta_T \nabla_{\boldsymbol{\theta}_T} L_{\boldsymbol{\theta}_T}(\boldsymbol{\theta}_T^{(u)}) \tag{10}$$

where $\nabla_{\boldsymbol{\theta}_T} L_{\boldsymbol{\theta}_T}(\boldsymbol{\theta}_T^{(u)})$ is the gradient of the loss function $L_{\boldsymbol{\theta}_T}(\boldsymbol{\theta}_T^{(u)})$ with respect to the parameter set $\boldsymbol{\theta}_T$ evaluated at $\boldsymbol{\theta}_T = \boldsymbol{\theta}_T^{(u)}$.

### B. Decoder Net Training

For the decoder, the symbol error rate (SER) is [11]

$$P_e = 1 - \sum_{p=1}^{P} \sum_{i=1}^{I} p[(\varepsilon_C)_{p,i}] p[(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}] \tag{11}$$

where $p[(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}]$ is the probability corresponding to $(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}$ and $p[(\varepsilon_C)_{p,i}]$ is the probability for $(\varepsilon_C)_{p,i}$. Assuming each label appears equally and adding $\log(\cdot)$ to $p[(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}]$ to facilitate the derivation of the network, the loss for the $q_C$th data can be written as

$$l_C^{(q_C)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) = -\sum_{p=1}^{P} \sum_{i=1}^{I} \log p[(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}] \tag{12}$$

where $p[(\hat{\varepsilon}_C)_{p,i} = (\varepsilon_C)_{p,i}]$ is obtained by choosing the corresponding probability in $\boldsymbol{p}_C$. Using $Q_R$ synthetic training data and lables to compute the loss function, we can obtain

$$L_{\boldsymbol{\theta}_C} = \frac{1}{Q_C} \sum_{q_C=1}^{Q_C} l_C{}^{(q_C)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C). \tag{13}$$

The parameter set $\boldsymbol{\theta}_C$ is updated with learning rate $\eta_C$,

$$\boldsymbol{\theta}_C^{(u+1)} = \boldsymbol{\theta}_C^{(u)} - \eta_C \nabla_{\boldsymbol{\theta}_C} L_{\boldsymbol{\theta}_C}(\boldsymbol{\theta}_C^{(u)}) \tag{14}$$

where $\nabla_{\boldsymbol{\theta}_C} L_{\boldsymbol{\theta}_C}(\boldsymbol{\theta}_C^{(u)})$ is defined similarly as $\nabla_{\boldsymbol{\theta}_T} L_{\boldsymbol{\theta}_T}(\boldsymbol{\theta}_T^{(u)})$.

### C. Detector Net Traning

For the radar detector net, the loss for the $q_R$th data is calculated with the output of the net $p_R^{(q_R)}$ and the lable $\varepsilon_R^{(q_R)}$

$$l_R{}^{(q_R)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) = -[\varepsilon_R^{(q_R)} \log p_R^{(q_R)} + (1 - \varepsilon_R^{(q_R)}) \log(1 - p_R^{(q_R)})].$$

Assume $Q_R$ synthetic training data and corresponding lables are generated based on the radar received signal model. The loss function can be computed as

$$L_{\boldsymbol{\theta}_R} = \frac{1}{Q_R} \sum_{q_R=1}^{Q_R} l_R{}^{(q_R)}(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R). \tag{15}$$

By minimizing the loss function with learning rate $\eta_R$, the parameter set $\boldsymbol{\theta}_R$ is updated,

$$\boldsymbol{\theta}_R^{(u+1)} = \boldsymbol{\theta}_R^{(u)} - \eta_R \nabla_{\boldsymbol{\theta}_R} L_{\boldsymbol{\theta}_R}(\boldsymbol{\theta}_R^{(u)}). \tag{16}$$

where $\nabla_{\boldsymbol{\theta}_R} L_{\boldsymbol{\theta}_C}(\boldsymbol{\theta}_R^{(u)})$ is defined similarly as $\nabla_{\boldsymbol{\theta}_T} L_{\boldsymbol{\theta}_T}(\boldsymbol{\theta}_T^{(u)})$.

After training, the optimized parameter sets $\boldsymbol{\theta}_T^*$, $\boldsymbol{\theta}_R^*$, $\boldsymbol{\theta}_C^*$ are obtained. In the testing process, a batch of testing waveforms associated with $\boldsymbol{\lambda}^* = f_{\boldsymbol{\theta}_T^*}(\boldsymbol{\lambda}_0)$ can be obtained. Based on the radar and communications received signal models, testing data and labels are generated. Then, the testing data and labels for communications and radar are fed into the decoder net the and detector net to evaluate the performance.

## V. Numerical Results

In this section, the performance of detection and decoding is presented. Instead of designing $\boldsymbol{\lambda}$ in (4), the energy and the time duration of SO-OCDM symbol are considered to design to simplify the simulation. Thus, the waveform parameters can be reduced to $\boldsymbol{E} = [E_1, ..., E_m]^T$, $\boldsymbol{T} = [T_1, ..., T_m]^T$, the waveform design problem is rewritten as

$$\min_{\boldsymbol{E}, \boldsymbol{T}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \boldsymbol{\theta}_R} \rho_R l_R(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_R) + \rho_C l_C(\boldsymbol{\lambda}, \boldsymbol{\theta}_T, \boldsymbol{\theta}_C) \tag{17a}$$

$$s.t. \quad \sum_{m=1}^{M} E_m = E_{total}, E_m \geq E_{min} \tag{17b}$$

$$T_{\min} \leq T_m \leq T_{\max}, m = 1, ..., M \tag{17c}$$

where (17b) is the energy constraint with the total energy $E_{total}$ and the minimal energy $E_{min}$, (17c) denoted the time duration constraint. The $T_{\min} = 1$ ms, $T_{\max} = 2$ ms, $E_{total} = 10^6$, and $E_{min} = 2 \times 5^6$.

Suppose the MIMO DFRC system has $M = 4$ transmitters located at (70, 0) km, (-70, 0) km, (0, 70) km, (0, -70) km, $N_C = 4$ communications receivers located at (41, 57) km, (-28, 64) km, (54, -45) km, (59, -36) km, and $N_R = 4$ radar receivers located at (64, 25) km, (35, 61) km, (-54, -45) km, (59, -36) km. The target is located at (8, 20) km. There are $P = 2$ symbols in one frame and the 4-quadrature amplitude modulation is adopted, namely, $J = 4$. The frequency spacing $f_{\Delta,m} = 16$ kHz, for $m = 1, ..., M$. The target reflection coefficients are $\delta_{m,n_C} = 500$, $\delta_{m,n_R} = 1000$ and the ratio of fading coefficients are setting as $\alpha_C^t = 5.03 \times 10^6$, $\alpha_C^d = 1.58 \times 10^3$, $\alpha_R^t = 2.02 \times 10^7$, $\alpha_R^d = 1.6 \times 10^4$.

The waveform design net has two hidden layers and one output layer. Each hidden layer has 30 neurons and the output neuron number is 8. The exponential linear units (ELU) is the activation fucntions for the hidden layer. A normalization layer is employed to ensure the output parameters satisfy the constraints. The detector FCN is composed of a complex-to-real (C2R) layer which converts the complex value into real value, 2 hidden layers with the activation function rectified linear unit (ReLU). The numbers of hidden neurons are 100, 20. For the output layer with 1 output neurons, the activation function is sigmoid. The decoder net comprises of a C2R layer, 2 hidden layers, and a output layer with $J$ output neurons. The neuron number of each hidden layers is 30. The ELU function is utilized in the hidden layers, while the softmax is employed for the output layer. The variance of Gaussian policy is $\sigma_{\boldsymbol{\lambda}}^2 = 0.01$. The learning rate is set as 0.005. The optimizers used for all nets are Adam due to its fast convergence speed and the whole network converges after 60 epochs.

The signal-to-clutter-plus-noise ratio (SCNR) for radar is $-1$ dB and for communications is $-0.9$ dB during training. During each training iteration, we generate $Q_R = Q_C = 1024$ data for receiver training and $Q_T = 2048$ for transmitter training. The number of testing data is 40000, the testing SCNR is -2.8 db for the radar, and the SCNR changes from -4.8 dB to 2.9 dB for communications.

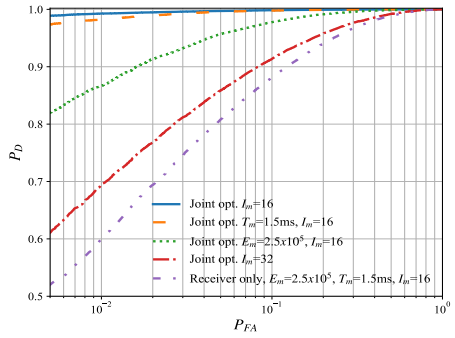


Fig. 5. ROC curves for joint waveform and receiver design and receiver only design with different waveform parameters.

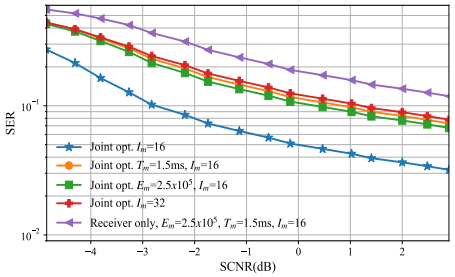


Fig. 6. SER for joint waveform and receiver design and receiver only design with different waveform parameters.

The receiver operating characteristic (ROC) curves show the detection probability $P_D$ when varying $P_{FA}$ in Fig. 5. The decoding performance when designing different waveform parameters are compared in Fig. 6. The joint training of waveform design net, detector, and decoder results in significant performance improvement of ROC and SER when comparing with the receivers only training case. Furthermore, the combining design of the energy and the time duration of SO-OCDM waveforms outperforms the performance of optimizing the energy or the time duration alone in both ROC and SER performance. When the number of chirps increases, the SER increases as one SO-OCDM symbol contains more modulated data and the detection performance decreases due to the randomness causes by the data.

## VI. Conclusion

The joint design of the SO-OCDM waveforms, radar detector, and communications decoder in the learning-based MIMO DFRC system was studied. We presented the SO-OCDM signal model and the framework for the RL-based waveform design net. The received signal models for synthesizing training and testing data, as well as the SL-based detector and decoder, were presented. A joint training process of the RL and SL was proposed to solve the problem of joint waveform design optimization and receiver design. It showed that the proposed joint design of waveforms and receivers outperforms the receiver-only design. We showed that the design of different waveform parameters, together with the receiver design, can lead to different impact on the MIMO DFRC system performance.

## References


[1] A. Hassanien, M. G. Amin, E. Aboutanios, and B. Himed, "Dual-function radar communication systems: A solution to the spectrum congestion problem," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 115–126, 2019.

[2] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1295–1315, 2021.

[3] M. F. Keskin, V. Koivunen, and H. Wymeersch, "Limited feedforward waveform design for OFDM dual-functional radar-communications," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2955–2970, 2021.

[4] M. Bekar, C. J. Baker, E. G. Hoare, and M. Gashinova, "Joint MIMO radar and communication system using a PSK-LFM waveform with TDM and CDM approaches," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6115–6124, 2021.

[5] Y. Wang, Z. Shi, X. Ma, and L. Liu, "A joint sonar-communication system based on multicarrier waveforms," *IEEE Signal Processing Letters*, vol. 29, pp. 777–781, 2021.

[6] L. G. d. Oliveira, B. Nuss, M. B. Alabd, A. Diewald, M. Pauli, and T. Zwick, "Joint radar-communication systems: Modulation schemes and system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 3, pp. 1521–1551, 2021.

[7] S. Bhattacharjee, K. V. Mishra, R. Annavajjala, and C. R. Murthy, "Evaluation of orthogonal chirp division multiplexing for automotive integrated sensing and communications," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8742–8746.

[8] L. Chen, Z. Wang, Y. Du, Y. Chen, and F. R. Yu, "Generalized transceiver beamforming for DFRC with MIMO radar and MU-MIMO communication," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1795–1808, 2022.

[9] R. Liu, M. Li, Q. Liu, and A. L. Swindlehurst, "Joint waveform and filter designs for STAP-SLP-based MIMO-DFRC systems," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1918–1931, 2022.

[10] C. Wen and T. N. Davidson, "Transceiver design for MIMO-DFRC systems," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[11] J. M. Mateos-Ramos, J. Song, Y. Wu, C. Häger, M. F. Keskin, V. Yajnanarayana, and H. Wymeersch, "End-to-end learning for integrated sensing and communication," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 1942–1947.

[12] C. Muth and L. Schmalen, "Autoencoder-based joint communication and sensing of multiple targets," *arXiv preprint arXiv:2301.09439*, 2023.

[13] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 298–303.

[14] W. Jiang, A. M. Haimovich, and O. Simeone, "End-to-end learning of waveform generation and detection for radar systems," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1672–1676.

[15] Z. Zhang, Z. Du, and W. Yu, "Mutual-information-based OFDM waveform design for integrated radar-communication system in gaussian mixture clutter," *IEEE Sensors Letters*, vol. 4, no. 1, pp. 1–4, 2019.

[16] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2503–2516, 2019.

[17] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in neural information processing systems*, vol. 12, 1999.