Ben-Gurion University of the Negev

Department of Electrical & Computer Engineering

The Relationship between Cyclostationarity, Sampling and Synchronization in Modern Communications Problems

היחסים בין ציקלוסטציונריות, דגימה וסנכרון בבעיות תקשורת מודרניות

Research Proposal for a PhD Thesis

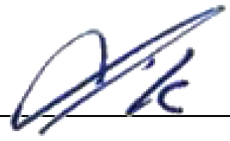
Zikun Tan

Under the supervision of **Prof. Ron Dabora**

Feb 2023

Prof. Ron Dabora
Advisor

Prof. Ilan Shalish
Head of Departmental Committee



Abstract

Man-made communications signals are typically generated via periodic operations, by which periodicity is induced into the mean and the autocorrelation function (AF) of such signals. These signals are statistically modeled as wide-sense cyclostationary (WSCS) processes. Processing of communications signals is typically performed in discrete-time (DT), which requires first sampling the continuous-time (CT) received signal. However, sampling a CT WSCS process does not necessarily result in a DT WSCS process: When the sampling interval and the cyclostationarity period of the CT WSCS process are related via a rational factor, the resulting DT process is WSCS. This situation is referred to as synchronous sampling. When these two parameters are related via an irrational factor, the resulting DT process is wide-sense almost cyclostationary (WSACS). This situation is referred to as asynchronous sampling. In practice, the later situation is more prevailing due to the variability inherent to physical clocks. Therefore, DT WSACS processes widely exist and investigating communications scenarios involving such processes is important. Accordingly, our first research topic is the rate-distortion function (RDF) characterization of asynchronously sampled WSCS Gaussian sources with memory. This topic is related to both information storage systems and cooperative transmission schemes. We note that because DT WSACS processes are not information stable, conventional information-theoretic tools are not applicable to the RDF characterization of such processes and instead the information-spectrum framework needs to be employed. In the second research topic, we plan to study the automatic modulation classification (AMC) problem. As AMC algorithms are typically computationally intensive, our objective in this topic is to apply reinforcement learning (RL) to obtain an efficient self-training modulation classifier which is trained with the actual environment, hence it is expected to achieve a better performance. We also plan to investigate the application of the multiple-antenna reception assisted via machine learning (ML) in the AMC algorithmic design. Lastly we plan to perform AMC with the temporal information (e.g., eye patterns). In the third research topic, we plan to investigate the sampling frequency synchronization problem in non-orthogonal multiple access (NOMA) networks. The base-station (BS) in a NOMA network simultaneously receives multiple signals from different users over the same frequency band and different signals are generated from different transmitters with their own sampling intervals. In this topic, the challenge is that for optimal digital processing, the BS needs to estimate the sampling frequency offset (SFO) between the sampling clock at the transmitter and the sampling clock at the receiver for each received signal.

Contents

List of Abbreviations

ACAF autocorrelated cyclic autocorrelation function.

AdaGrad adaptive gradient.

ADCs analog-to-digital converters.

AF autocorrelation function.

AMC automatic modulation classification.

AWGN additive white Gaussian noise.

BP backpropagation.

BPSK binary phase-shift keying.

BRNNs bidirectional recurrent neural networks.

BS base-station.

BW Baum-Welch.

CAF cyclic autocorrelation function.

CBLASSO concomitant Beurling LASSO.

CB-SFS cyclostationarity-based sampling frequency synchronization.

CC cyclic cumulant.

CCF cyclic correntropy function.

CDMA code-division multiple access.

CDP cyclic domain profile.

CF compress-and-forward.

CFO carrier frequency offset.

CNNs convolutional neural networks.

CP cyclic prefix.

CRC cyclic redundancy check.

CSI channel state information.

CT continuous-time.

DCD decimated component decomposition.

DCT discrete cosine transform.

DL deep learning.

DMC discrete memoryless channel.

DMS discrete memoryless source.

DNNs deep neural networks.

DRF distortion-rate function.

DRNNs deep recurrent neural networks.

DT discrete-time.

DWT discrete wavelet transform.

FB feature-based.

FC fully-connected.

FCNNs fully-connected neural networks.

FDMA frequency-division multiple access.

FLOCAF fractional lower-order cyclic autocorrelation function.

FNNs feedforward neural networks.

FSK frequency-shift keying.

GRUs gated recurrent units.

GSNR geometric signal-to-noise ratio.

HMM hidden Markov model.

i.i.d. independently and identically distributed.

ICI inter-carrier interference.

ILC interference-limited communications.

ISI inter-symbol interference.

JSCC joint source-channel coding.

LB likelihood-based.

LMMSE linear minimum mean square error.

LPF low-pass filter.

LS least squares.

LSTM long short-term memory.

LTF long training field.

MDP Markov decision process.

ML machine learning.

MSK minimum-shift keying.

MUSIC multiple signal classification estimator.

NNs neural networks.

NOMA non-orthogonal multiple access.

OFDM orthogonal frequency-division multiplexing.

OFDMA orthogonal frequency-division multiple access.

OMA orthogonal multiple access.

PC polyphase component.

POMDP partially observable Markov decision process.

QAM quadrature amplitude modulation.

QPSK quadrature phase-shift keying.

RDF rate-distortion function.

ResNets residual networks.

RL reinforcement learning.

RMSProp root mean square propagation.

RNNs recurrent neural networks.

RV random variable.

SCD subband component decomposition.

SDMA space-division multiple access.

SFO sampling frequency offset.

SGD stochastic gradient descent.

SIC successive interference cancellation.

SL supervised learning.

SNR signal-to-noise ratio.

SRNNs simple recurrent neural networks.

SSBAM single sideband amplitude modulation.

SSCS strict-sense cyclostationary.

STF short training field.

STO sampling time offset.

TDMA time-division multiple access.

UAV unmanned aerial vehicle.

UL unsupervised learning.

WSACS wide-sense almost cyclostationary.

WSCS wide-sense cyclostationary.

List of Figures

Chapter 1

Introduction

1.1 Cyclostationarity-based Signal Parameter Estimation

It is very common in statistical signal processing and in communications systems to assume that many involved processes are *stationary*. However, man-made signals are typically generated using periodic operations, by which periodicity is induced in the statistical characteristic of the signal, which is refered to as *cyclostationarity*. Clearly, cyclostationary signals are not stationary as their statistics depend on the time within the period as well as on the lag. Specifically, in communications systems, cyclostationarity is induced by the periodicity inherent to the modulation operation at the transmitter, which superimposes multiple symbols to only a continuous-time (CT) signal. In addition to signal processing and communications, cyclostationary statistics are also observed in some other fields, such as econometrics, therefore this statistic is very common.

A *strict-sense cyclostationary (SSCS)* process is characterized by a periodically varying probability distribution [3, Sec. 17.2]. However, in many practical signal analysis problems, the probability distribution does not vary periodically, but the mean and the autocorrelation function (AF) vary periodically with time. Such processes are referred to as *wide-sense cyclostationary (WSCS)* processes [4, Sec. 1]. Due to its periodicity, under mild regularity conditions, the AF of the WSCS process can be expanded into a Fourier series whose coefficients and frequencies are named as *cyclic autocorrelation functions (CAFs)* and *cyclic frequencies*, respectively [4, Sec. 1]. It is noted that the *fundamental cyclic frequency* is the reciprocal of the cyclostationarity period of the WSCS process and all cyclic frequencies are multiples of the fundamental cyclic frequency [4, Sec. 1]. The concept of wide-sense cyclostationarity can be generalized to cases in which the AF is an *almost periodic function*. In such cases, the AF can be expanded into a generalized Fourier series. Such processes are referred to as *wide-sense almost cyclostationary (WSACS)* processes [4, Sec. 1].

Cyclostationary processes exist in CT as well as in discrete-time (DT). When sampling a CT WSCS process, if the sampling interval is related with the cyclostationarity period via a *rational* factor, then, the resulting DT process is WSCS [4, Sec. 3.9]. This situation is refered to as *synchronous sampling* [4, Sec. 3.9]. If, however, the two parameters are related via an *irrational* factor, then the resulting DT process is WSACS [4, Sec. 3.9]. This situation is refered to as *asynchronous sampling* [4, Sec. 3.9]. A cyclostationary process can be represented as a multivariate stationary process via the *decimated component decomposition (DCD)* in the time domain [3, Sec. 17.2]. An alternative transformation between cyclostationary processes and stationary processes is the *subband component decomposition (SCD)* [3, Sec. 17.2]. This transformation can be realized in the frequency domain by passing modulated versions of a cyclostationary process through an ideal low-pass filter (LPF) with a certain narrow passband, resulting in a set of narrowband stationary subprocesses [3, Sec. 17.2]. We note that, as ideal LPFs cannot be practically realized, the SCD method is less practical than the DCD method [3, Sec. 17.2].

In communications systems, *synchronization* is a key step in recovering the originally transmitted information. For example, the two main advantages of orthogonal frequency-division multiplexing (OFDM) systems, namely, the high spectral efficiency and the robustness to frequency-selective fading [5], can be achieved only if the receiver is synchronized with the transmitter, since sampling frequency offset (SFO) and carrier frequency offset (CFO), which are two major synchronization errors, seriously degrade the system performance [6].

Cyclostationary properties can be very useful in signal parameter estimation. Using elements from the works [7] and [8], the work [9] implemented an efficient and blind SFO estimation based on cyclostationary properties in OFDM systems, which is refered to as the cyclostationarity-based sampling frequency synchronization (CB-SFS) algorithm. The CB-SFS algorithm uses the fact that both the received OFDM signal and its sampled version are WSCS processes (the sampled signal can be WSACS, but for sufficiently short sequences it can be approximated as a WSCS process). Using the finite-length sequence of received signal samples, a cost function corresponding to the sum of autocorrelated cyclic autocorrelation functions (ACAFs) over a range of time lags is empirically computed at every possible cyclic frequency. The location of the local maxima of the ACAF corresponds to the cyclic frequency value of the sampled received signal, thus the fundamental cyclic frequency can be obtained. Then, using the ratio relationship between the fundamental cyclic frequency and the sampling interval, the sampling interval at the transmitter is acquired. With the known sampling interval at the receiver, the value of SFO is finally computed. Computer simulations demonstrated that the CB-SFS algorithm yielded excellent estimation performance over a wide range of SFO and was computationally feasible compared with the major existing SFO estimation approaches.

1.2 Deep Learning-based Signal Parameter Estimation

Conventionally, signal processing algorithms are based on modelling the relationship between the parameters of interest and the measured samples. Then, the architecture and parameters of the estimator are designed based on the model. For example, as discussed in Sec. 1.1, the property of cyclostationarity is utilized in the design of algorithms for signal parameter estimation in OFDM systems. In recent years, due to the increasing complexity of signal processing and the unprecedented increase of computing capabilities, processing large volumes of data became feasible, which motivated the application of *machine learning (ML)* for solving signal processing problems.

Unlike conventional methods which require underlying physical models, ML attempts to learn mappings between the data and the algorithm's parameters. With the help of training data, ML systems can automatically make decisions or predictions without being explicitly programmed to do so [10, Sec. 3]. ML algorithms can be classified into three main paradigms: *Supervised learning (SL)*, *unsupervised learning (UL)* and *reinforcement learning (RL)*. Inspired by the structure of human brains, *deep learning (DL)* is an important subfield of ML, which uses *neural networks (NNs)* to learn the underlying mapping between the input and the output of a signal processing system. One of the fundamental architectures of NNs is the *fully-connected neural network (FCNN)*, which consists of a series of fully-connected (FC) layers [11, Ch. 4]. A FC layer means that, all neurons in one layer are connected with neurons in the next layer and each neuron has its own learnable parameters and a nonlinear activation function [11, Ch. 4]. Besides FCNNs, additional types of NN architectures include the *convolutional neural networks (CNNs)* and the *recurrent neural networks (RNNs)*. CNNs accept tensor signals as the input and consist of multiple convolutional layers, pooling layers and FC layers while in RNNs, all network units are concatenated in a temporal order and each unit accepts both the output and the hidden states from the previous unit as the input information [12]. CNNs are commonly used in tensor signal processing, such as image processing while RNNs are usually applied in time-series signal processing, such as audio processing. There are some special-purpose architectures of CNNs widely used in engineering applications, such as LeNet-5 [2], AlexNet [13], VGG-16 [14] and residual networks (ResNets) [15]. As two commonly used enhanced versions of RNNs, *long short-term memory (LSTM)* [16] and *gated recurrent units (GRUs)* [17,18] are frequently used in practical applications to capture long-term dependencies more efficiently and avoid suffering from the vanishing gradient problem [19, Sec. 10.10].

During the past several decades, the development of communications systems has greatly relied on theoretical models, from information theory to channel modelling [20]. Nevertheless, due to the increasing complexity of communications systems design and application scenarios, limitations of conventional methods gradually became pronounced [20]. Therefore, the introduction of ML in many communications problems, such as channel estimation, resource allocation and standardization, has gained much attention [20]. In this proposal, two of the proposed research topics may involve applying ML to address challenging problems.

In the following, a brief review of some related works is presented:

The works [21] and [22] focused on estimating the frequency of pure sinusoidal waves. In the work [21], a novel frequency estimator based on deep neural networks (DNNs) was designed for a single sinusoidal wave signal observed with additive white Gaussian noise (AWGN) using one-bit analog-to-digital converters (ADCs). The input of the DNN is the sign vector of the real and the imaginary parts of the sampled received signal while

the estimated digital frequency is the final output. The authors stated that this estimator can be implemented using FCNNs, CNNs or RNNs and the CNN-based estimator can also be implemented via ResNets to improve the performance. Computer simulations suggested that the proposed estimator with any of the four aforementioned structures achieved a better performance with a shorter execution time using low signal-to-noise ratio (SNR) training data and short pilot sequences, compared with three conventional signal processing methods: The classic periodogram [23], Welch's method [24] and multiple signal classification estimator (MUSIC) [25]. In the work [22], a fast sinusoidal frequency estimator, for noisy multisinusoidal signals based on DNNs was proposed. The frequency estimator consists of two separate NNs: A frequency-representation module and a frequency-counting module. The architecture of the frequency-representation module was inspired by a frequency-representation NN called PSnet [26]. The frequency-representation module accepts the real and the imaginary parts of the sampled noisy multisinusoidal signal and outputs a pattern of peaks corresponding to a superposition of Gaussian kernels, such that the locations of the peaks correspond to the estimated frequencies. The output of the frequency-representation module is passed to the frequency-counting module, which outputs the number of sinusoidal components. This number is then passed to a peak location detector based on the frequency representation which determines the estimated frequency values as the locations of the peaks. The frequency-representation module and the frequency-counting module are trained, respectively, as two separate regression tasks, which are subsequently concatenated into an integrated system. Computer simulations suggested that the proposed method outperformed several learning-based and conventional approaches, including PSnet and concomitant Beurling LASSO (CBLASSO) [27], especially at medium-to-high noise levels. However, it is unclear that the proposed methods [21] and [22] can be extended for modulated signals, as in such signals the frequencies are not presented as explicit impulses in the frequency domain [28, Sec. 6.1].

The widespread use of OFDM in 5G communications has motivated a large body of works into signal parameter estimation for OFDM systems using DL. The work [29] proposed a DNN-based novel orthogonal frequency-division multiple access (OFDMA) receiver for the uplink in multiuser networks which can handle CFO. The proposed receiver consists of two main components: The first component applies CFO compensation within a certain CFO range and subsequently applies channel equalization. The second component implements signal classification. It was claimed that this receiver can work properly in both blind and non-blind (i.e., following coarse CFO estimation) scenarios. In blind scenarios, the received signal after cyclic prefix (CP)-removal is fed into all CFO compensators while in non-blind scenarios the signal after CP-removal is fed into only one or two specific CFO compensators. The DNN-based receiver is trained offline in a supervised manner with the Adam algorithm. Computer simulations suggested that the proposed DNN-based receiver was able to operate properly over a wide range of SNRs for both single-user and multiuser systems to provide CFO-free performance without implementing dedicated CFO estimation.

OFDM can also be implemented in unmanned aerial vehicle (UAV) communications systems. In the work [30], the authors proposed an OFDM-based UAV communications system, using NNs to compensate for the impact of the channel and the CFO on the detection performance. The NN for CFO compensation is applied at the frontend of the receiver and the channel equalization NN is applied just before the demodulator. Each network uses a pair of parallel CNNs for separately processing the real and the imaginary parts of the complex-valued received signal, then skip connections are used to combine the processed results together at the output port. The two NNs are trained in a supervised manner. Unlike conventional methods, the proposed approach does not need the use of a preamble or pilot sequences, thus it is more bandwidth efficient. Computer simulations suggested that the proposed communications network performed better than the existing CFO compensation method [31] with the least squares (LS) or the linear minimum mean square error (LMMSE) channel estimation methods [32].

The IEEE 802.11ah [33] preamble sequences can be used to design signal parameter estimators with DL. In the work [34], the performance of CFO estimation using IEEE 802.11ah preamble sequences, which is implemented via FC-DNNs, GRU-based RNNs and LSTM-based RNNs, respectively, were investigated and compared against the conventional CFO estimation algorithm inspired from [32, 35–42] for the IEEE 802.11ah standard scenario. The preamble sequence specified in the IEEE 802.11ah standard is utilized in both the conventional and DNN-based CFO estimation methods. The conventional method is divided into two steps: The first step coarsely estimates the CFO using the long training field (LTF) and the second step implements the fine CFO estimation based on the short training field (STF). Both the LTF and the STF are parts of the preamble defined in the IEEE 802.11ah standard. In the DL method, the input is the phase of the received STF samples, the output of the final linear layer is the estimated CFO, and the network is trained in a supervised manner. From the computer simulations, it was observed that under both AWGN and the indoor multipath channel, the RNN-based network was the best-performing architecture, which attained the accuracy of conventional method at low-to-medium SNRs while the FC-DNN based architecture performed much worse. For the two types of

RNN-based architectures, the GRU-based one performed similarly to the LSTM-based one but with a shorter execution time.

1.3 Source Coding and Data Compression

Established by the early works of Nyquist, Hartley and Shannon, information theory mainly concerns with two fundamental problems: What is the minimum data rate for compressing an information source up to a certain distortion level, and what is the maximum information transmission rate over a communications channel with an arbitrarily small probability of error [43, Ch. 1]. Within an integrated communications system, at the transmitter's side, the source encoder maps a sequence of source symbols from the information source into an index, which is commonly refered to as a *message*, which belongs to a finite set of indices, refered to as the *message set*. The message can be fed into the channel encoder for further processing for transmission. At the receiver's side, the received signal from the communications channel is first decoded by the channel decoder to obtain the decoded message, which is subsequently passed to the source decoder to generate a sequence of reconstructed source symbols within a specified distortion level from the original source symbol sequence.

The output of a information source can belong either to the analog domain, such as an audio signal, or to the digital domain, such as the output of computing systems [44, Ch. 6]. The amplitude of an analog domain source is continuous while that of a digital domain source is discrete. For a digital domain source, if its alphabet of symbols is finite, then *lossless compression* is possible, where by the *lossless source coding theorem (i.e., Shannon's first theorem)*, first stated in [45], the number of required bits per source symbol (i.e., source code rate) should not be less than the source entropy [46, Thm. 3.4]. However, for a digital domain source with an infinite alphabet of symbols or an analog domain source, only *lossy compression* is feasible, where by the *lossy source coding theorem (i.e., Shannon's third theorem)*, first stated in [47], the source can be encoded at a source code rate higher than the corresponding rate-distortion function (RDF) value in order for the distortion not to exceed a certain level [46, Thm. 3.5]. Typical lossless compression algorithms include the Huffman coding algorithm [48] and the Lempel-Ziv algorithm [49, 50] while the discrete cosine transform (DCT) [51] algorithm and the discrete wavelet transform (DWT) algorithm [52] are two frequently-used lossy compression algorithms.

The rate-distortion theory characterizes lossy compression in source coding. In lossy compression, the distortion between the original information symbols from the information source and the reconstructed information symbols can be quantified via an appropriate measurable function, such as the squared-error distortion function and the Hamming distortion function. Given a source distribution and a measure of distortion, the RDF provides the minimum required rate of compression, measured in bits per source symbol, to achieve a particular distortion level [43, Ch. 10]. In this proposal, the first research topic considers the rate-distortion theory for the compression of asynchronously sampled WSCS Gaussian sources with memory.

Next, we give a brief review of two relevant works in the following:

In the work [53], the authors derived the distortion-rate function (DRF) of both CT and DT WSCS Gaussian sources with memory. The derivation is based on orthogonalizing polyphase components (PCs) from the statistics of a WSCS source. The resulting DRF is presented in terms of the eigenvalues of a spectral density matrix and the rate is obtained by a reverse waterfilling over these eigenvalues. It was noted by the authors that the DRF of a CT WSCS Gaussian source may not have a closed-form solution since it is given in terms of a limit of a function over these eigenvalues. A lower bound of the resulting DRF is derived by averaging the minimal distortion obtained in compressing each of the PCs.

Based on the DRF of DT WSCS Gaussian sources with memory given in the work [53], the work [54] was the first work which studied the RDF characterization of DT WSACS memoryless Gaussian sources obtained by asynchronously sampling CT WSCS Gaussian sources. Unlike DT WSCS Gaussian sources (obtained by synchronously sampling CT WSCS Gaussian sources), DT WSACS sources are not information stable, thus the information-spectrum framework [55] was employed in the RDF characterization instead of conventional information-theoretic tools. The authors first derived the relationship between some relevant information-spectrum quantities for uniformly convergent sequences of random variables (RVs), then used this relationship to characterize the RDF of a DT WSACS Gaussian source as the limit superior of a sequence of RDFs, each corresponding to a DT WSCS Gaussian source. The model considered in the work [54] assumed that the sampling interval is larger than the maximal correlation length of the original CT source and the RDF characterization is valid for scenarios in which the target distortion constraint is less than the minimum variance of the original CT WSCS source process. Accordingly, the DT source process is assumed memoryless and the reconstruction must

be within a relatively small distortion. Computer simulations suggested that the RDF varied significantly with minor variations of the sampling interval and the sampling time offset (STO). Specifically, when the sampling interval changed between synchronous sampling and asynchronous sampling, which resulted in the DT source switching between a WSCS process and a WSACS process, the RDF significantly changed. The results from this work can be applied for the accurate and efficient signal compression in digital communications systems.

1.4 Automatic Modulation Classification

Automatic modulation classification (AMC) refers to the identification of the modulation scheme used for generating a communications signal via analysis of the received signal at the receiver. Serving an intermediate step between signal detection and demodulation, AMC is widely used in modern communications systems and in spectrum monitoring systems [56], [57, Sec. I].

In the following, we review three closely related works:

The work [58] proposed an AMC architecture for classifying signals within a set of digital modulations contaminated by additive impulsive noise, based on cyclostationary features. The impulsive noise is modelled as an independently and identically distributed (i.i.d.) non-Gaussian α-stable process. The authors considered three cyclostationary statistical features: The CAF, the fractional lower-order cyclic autocorrelation function (FLOCAF) and the cyclic correntropy function (CCF). The authors applied these three methods to extract cyclostationary features from signals modulated by binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), 8-quadrature amplitude modulation (QAM), 16-QAM and 32-QAM. Computer simulations suggested that the CAF feature did not facilitate the desired AMC performance, since the cyclic information profiles of the signal modulated with the aforementioned modulation types are not sufficiently distinct to facilitate reliable classification decisions. The FLOCAF and the CCF seemed to result in sufficient distinction for distinguishing between the different modulations, thus these methods are useful for the design of AMC schemes. The authors also derived the values of the fractional lower-order parameters and the kernel size, which are free parameters of the FLOCAF and of the CCF respectively and can optimize the performance of corresponding AMC schemes. Computer simulations suggested that for the same geometric signal-to-noise ratio (GSNR) level, which is the quality indicator of the impulsive noise channel, and the same modulation type, the classification hit-rate of the CCF-based AMC scheme was always higher than that of the FLOCAF-based one. Therefore, it was concluded that the CCF is a better choice for AMC in such channel conditions. It is noted that the work [58] did not include an algorithmic complexity comparison between the considered AMC schemes, therefore, the best AMC scheme for the considered channel conditions in practical applications is still undecided.

The work [59] surveyed and reproduced the results from the works [60] and [61]. The work [60] proposed a learning-based AMC method while the work [61] proposed a hidden Markov model (HMM)-based AMC method. The learning-based AMC method of [60] is based on the MAXNET architecture. This architecture contains several mutually unconnected FCNNs, where the output value of each FCNN corresponds to the likelihood level of a certain modulation type. Each FCNN computes its own output value which varies between -1 and 1 and the largest among these outputs is selected as the final output value of the MAXNET architecture. During the training stage, different realizations of the cyclic domain profile (CDP) (i.e., peak values of the spectral coherence function) of signals generated with different modulation types and different SNRs are used to train the MAXNET via the backpropagation (BP) training algorithm. Computer simulations suggested that the trained MAXNET attained a high probability of correct classification even at low SNRs, where four modulation types, BPSK, QPSK, frequency-shift keying (FSK) and minimum-shift keying (MSK) were applied. In the HMM-based AMC method, the modulation classifier is also tasked with identifying the existence of the signal. In this scheme, a CDP-based signal detector based on [62] is used to determine whether a signal exists. If this is the case, the scheme passes the CDP to the modulation classifier. In the training of the HMM-based modulation classifier, the binary feature vectors generated by the peaks of the CDP with different modulation types are applied by using the Baum-Welch (BW) algorithm. In the testing setup, the CDP of the signal is passed from the signal detector to the trained HMM-based modulation classifier to generate its likelihood value and compare it with the likelihood of the original sequence. Then, the modulation type whose likelihood value is nearest to that of the received sequence is the selected modulation type. BPSK, QPSK, FSK, MSK and single sideband amplitude modulation (SSBAM) were used in the experiments. Computer simulations suggested that the HMM-based AMC method attained a probability of correct classification of nearly 100% for all modulation types when the number of symbols contained in the input modulated signal was around 300. However, the work [59] claimed that these two proposed methods are not capable of classifying higher-order modulation types, such as higher-order QAM, since these modulation types exhibit the same cyclic features as QPSK.

The work [63] proposed an AMC method for distinguishing between linear digital modulations based on higher-order cyclic cumulants (CCs) in flat-fading channels. The proposed AMC method can be implemented with either single-antenna or multi-antenna classifiers. For single-antenna CC-based modulation classifiers, the feature vector contains the absolute values of several CCs of the sampled baseband signal received over a flat-fading channel, at certain values of cyclic frequencies and delay vectors. Each modulation type has its own analytic feature vector. An estimate of the feature vector is computed for each modulation type using the received signal samples. The modulation type for which the Euclidean distance between the analytic feature vector and the estimated one is the smallest, is selected as the classification result. For multi-antenna CC-based modulation classifiers, spatial diversity can be utilized to combat the fading effect and improve the system performance: First, a selection combiner is applied to select the received signal with the highest SNR for the feature vector estimation. Then, the modulation format whose analytic feature vector is closest to the estimated one w.r.t. the Euclidean norm is determined to be the classification result. The average probability of correct classification is used as the classifier performance metric and nine different modulation formats were used in the numerical test. Computer simulations over the Rayleigh fading channel suggested that when the number of observed symbols was fixed, the performance improved as the SNR increased, while when the SNR was fixed, the performance improved as the number of observed symbols increased. It was also observed that using more antennas in the classifier resulted in better performance in both simulation scenarios, which validated the advantage of spatial diversity. An additional computer simulation compared the proposed single-antenna CC-based classifier with the likelihood-based (LB) classifier proposed in the work [64] and the feature-based (FB) classifiers proposed in the works [65] and [66], respectively, for single-antenna reception. This comparison suggested that when the SNR in the Rayleigh fading channel and the number of observed symbols were fixed, the proposed single-antenna CC-based classifier outperformed the other three classifiers. However, due to the lack of algorithmic complexity analysis in this work, it is unclear whether the proposed single-antenna CC-based classifier is the most appropriate selection in practical applications.

Chapter 2

Scientific Background

2.1 Notations and Symbols

In this proposal, calligraphic letters represent sets, e.g., $\mathcal{X}$, where the sets of real numbers, integers, positive integers, irrational numbers, nonnegative real numbers and positive real numbers are denoted by $\mathcal{R}$, $\mathcal{Z}$, $\mathcal{N}$, $\mathcal{P}$, $\mathcal{R}^+$ and $\mathcal{R}^{++}$, respectively. Scalar random processes are denoted with $X(t)$, $t \in \mathcal{R}$ for CT and with $X[n]$, $n \in \mathcal{Z}$ for DT. $\mathbb{E}\{\cdot\}$ denotes the statistical expectation, $(\cdot)^*$ denotes the complex conjugate, $|\cdot|$ represents the absolute value and $\odot$ represents the element-wise multiplication. Scalar RVs and scalar deterministic variables are denoted with X and x, respectively. Boldface lowercase letters denote deterministic column vectors, e.g., $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$, and boldface uppercase letters denote random column vectors, e.g., $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$, where in both cases $n \in \mathcal{N}$ denotes the number of elements in the vector. $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 and $Bern(p)$ represents a Bernoulli distribution with probability p for successful occurrence. Sans-serif uppercase letters denote matrices, e.g., X.

2.2 Cyclostationary Processes

In this section, cyclostationary properties are presented in detail. Begin by recalling the definition of WSCS processes:

Definition 2.2.1 (WSCS processes [3, Def. 17.1], [4, Sec. 3.2.1])**.** A complex CT (resp. DT) random process $X(t)$, $t \in \mathcal{R}$ (resp. $X[n]$, $n \in \mathcal{Z}$) is called *WSCS* if both its mean $m_X(t)$ (resp. $m_X[n]$) and its AF $R_X(t, \tau)$ (resp. $R_X[n, \Delta]$) are periodic in time t (resp. n) with some period $T_0 \in \mathcal{R}$ (resp. $N_0 \in \mathcal{Z}$) for any lag value $\tau \in \mathcal{R}$ (resp. $\Delta \in \mathcal{Z}$).

The mean $m_X(t)$ (resp. $m_X[n]$) and the AF $R_X(t, \tau)$ (resp. $R_X[n, \Delta]$) of the CT (resp. DT) WSCS process $X(t)$ (resp. $X[n]$) are expressed in the following [3, Def. 17.1], [4, Eqns. (3.4), (3.5)]:

$$m_X(t) \triangleq \mathbb{E}\{X(t)\} = m_X(t + T_0) \tag{2.1a}$$

$$R_X(t, \tau) \triangleq \mathbb{E}\{X(t + \tau)X^*(t)\} = R_X(t + T_0, \tau) \tag{2.1b}$$

$$m_X[n] \triangleq \mathbb{E}\{X[n]\} = m_X[n + N_0] \tag{2.1c}$$

$$R_X[n, \Delta] \triangleq \mathbb{E}\{X[n + \Delta]X^*[n]\} = R_X[n + N_0, \Delta] \tag{2.1d}$$

As the AF of a CT WSCS process is a periodic function in the time t and with a period of T_0, it can be expanded into a Fourier series, see [4, Eqn. (3.6)]:

$$R_X(t, \tau) = \sum_{k=-\infty}^{\infty} R_X^{\alpha_k}(\tau) e^{j2\pi\alpha_k t} \tag{2.2}$$

where $\alpha_k \triangleq \frac{k}{T_0},\ k \in \mathcal{Z}$. In Eqn. (2.2), the Fourier coefficients $R_X^{\alpha_k}(\tau)$ are refered to as the *CAF* and α_k is refered to as the *cyclic frequency*. Based on the Fourier series theory, the CAF $R_X^{\alpha_k}(\tau)$ can be obtained via [4, Eqn. (3.7)]:

$$R_X^{\alpha_k}(\tau) = \frac{1}{T_0} \int_{t=-\frac{T_0}{2}}^{\frac{T_0}{2}} R_X(t,\tau) e^{-j2\pi\alpha_k t} dt \tag{2.3}$$

For a DT WSCS process $X[n]$ with the period N_0, the AF can be expanded into a Fourier series via [3, Eqn. (17.1)]:

$$R_X[n,\Delta] = \sum_{k=0}^{N_0-1} R_X^{\tilde{\alpha_k}}[\Delta] e^{j2\pi\tilde{\alpha_k} n} \tag{2.4}$$

where $\tilde{\alpha_k} \triangleq \frac{k}{N_0},\ k \in \{0,1,\ldots,N_0-1\}$. In Eqn. (2.4), the Fourier coefficients $R_X^{\tilde{\alpha_k}}[\Delta]$ are the *CAF* and $\tilde{\alpha_k}$ is referred to as the *cyclic frequency*. Accordingly, the CAF $R_X^{\tilde{\alpha_k}}[\Delta]$ can be computed via [3, Eqn. (17.1)]:

$$R_X^{\tilde{\alpha_k}}[\Delta] = \frac{1}{N_0} \sum_{n=0}^{N_0-1} R_X[n,\Delta] e^{-j2\pi\tilde{\alpha_k} n} \tag{2.5}$$

Before proceeding to the definition of WSACS processes, we first give the definition of almost periodic functions:

Definition 2.2.2 (Almost periodic functions [54, Def. 2.1], [67, Def. 3])**.** A CT (resp. DT) function $X(t),\ t \in \mathcal{R}$ (resp. $X[n],\ n \in \mathcal{Z}$) is called *almost periodic*, if for any $\epsilon > 0$, there exists an associated number $l_\epsilon \in \mathcal{R}^{++}$ (resp. $\tilde{l_\epsilon} \in \mathcal{N}$) which satisfies that for any $\alpha \in \mathcal{R}$ (resp. $\beta \in \mathcal{Z}$), there exists $\delta \in [\alpha, \alpha + l_\epsilon]$ (resp. $\Delta \in [\beta, \beta + \tilde{l_\epsilon}]$), such that $|X(t+\delta) - X(t)| \leqslant \epsilon$ (resp. $|X[n+\Delta] - X[n]| \leqslant \epsilon$).

Next, we define WSACS processes as follows:

Definition 2.2.3 (WSACS processes [3, Def. 17.2], [4, Sec. 3.2.1])**.** A complex CT (resp. DT) random process $X(t),\ t \in \mathcal{R}$ (resp. $X[n],\ n \in \mathcal{Z}$) is called *WSACS* if both its mean $m_X(t)$ (resp. $m_X[n]$) and its AF $R_X(t,\tau)$ (resp. $R_X[n,\Delta]$) are almost periodic in time t (resp. n) for any lag value $\tau \in \mathcal{R}$ (resp. $\Delta \in \mathcal{Z}$).

The AF of a CT WSACS process $X(t)$, $R_X(t,\tau)$, can be expanded into a generalized Fourier series [4, Eqn. (3.10)]:

$$R_X(t,\tau) = \sum_{\alpha \in \mathcal{A}} R_X^{\alpha}(\tau) e^{j2\pi\alpha t} \tag{2.6}$$

In Eqn. (2.6), the Fourier coefficients $R_X^{\alpha}(\tau)$ are refered to as the *CAF* and $\mathcal{A}$ is a countable set of *cyclic frequencies* which are possibly incommensurate, [4, Eqn. (3.11)]:

$$R_X^{\alpha}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} R_X(t,\tau) e^{-j2\pi\alpha t} dt \tag{2.7}$$

For a DT WSACS process $X[n]$, its AF $R_X[n,\Delta]$ can be expanded into a generalized Fourier series as [3, Eqn. (17.2)]:

$$R_X[n,\Delta] = \sum_{\tilde{\alpha} \in \tilde{\mathcal{A}}} R_X^{\tilde{\alpha}}[\Delta] e^{j2\pi\tilde{\alpha} n} \tag{2.8}$$

In Eqn. (2.8), $\tilde{\mathcal{A}}$ is a countable set of *cyclic frequencies* which are possibly incommensurate and the Fourier coefficients $R_X^{\tilde{\alpha}}[\Delta]$ are refered to as the CAF, [3, Eqn. (17.2)]:

$$R_X^{\tilde{\alpha}}[\Delta] = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} R_X[n,\Delta] e^{-2j\pi\tilde{\alpha} n} \tag{2.9}$$

As communications signals are in CT, the application of digital processing requires first sampling the CT signal. Applying sampling to a CT WSCS process does not necessarily result in a DT WSCS process: Consider sampling a CT WSCS process, if its cyclostationarity period T_0 and the sampling interval T_{samp} satisfy a *rational* ratio, i.e.,:

$$\frac{T_0}{T_{samp}} = \frac{U}{V} \tag{2.10}$$

where $U, V \in \mathcal{N}$, then the resulting DT process is DT WSCS [4, Sec. 3.9]. This is refered to as *synchronous sampling*. However, if T_0 and T_{samp} are related by an *irrational* factor, i.e.,:

$$\frac{T_0}{T_{samp}} = \frac{U}{V} + \varepsilon;\ \varepsilon \in \mathcal{P} \text{ and } \varepsilon \in (0, 1) \tag{2.11}$$

then the resulting DT process is DT WSACS [4, Sec. 3.9]. This is refered to as *asynchronous sampling*.

There are two transformation approaches that can represent a scalar cyclostationary process as a multivariate stationary process, one in the time domain while the other in the frequency domain, as presented in the following two propositions:

Proposition 2.2.1 (DCD [3, Rep. 1])**.** A DT WSCS process $X[n]$ with a cyclostationarity period N_0 can be decomposed into an N_0-variate stationary process $\mathbf{X}[n]$ consisting of subprocesses $X_i[n] = X[nN_0 + i],\ i \in \{0, 1, \ldots, N_0 - 1\}$.

Proposition 2.2.2 (SCD [3, Rep. 2])**.** A DT WSCS process $X[n]$ with a cyclostationarity period N_0 can be expressed as a superposition of N_0 stationary narrowband subprocesses: $X[n] = \sum_{m=0}^{N_0-1} \overline{X}_m[n] e^{-j2\pi \frac{m}{N_0} n}$, where $\overline{X}_m[n]$ is obtained by passing $X[n]e^{j2\pi \frac{m}{N_0} n}$, the corresponding modulated version of $X[n]$, through an LPF whose passband is $(-\frac{\pi}{N_0}, \frac{\pi}{N_0}]$.

2.3 Deep Learning and Neural Networks

In this section we expand on DL and NNs, beginning with the FCNN architecture.

FCNNs are the most basic type of NNs, in which neurons in each layer are fully connected with neurons in the next layer. A simple example of a FCNN is presented in Fig. 2.1:

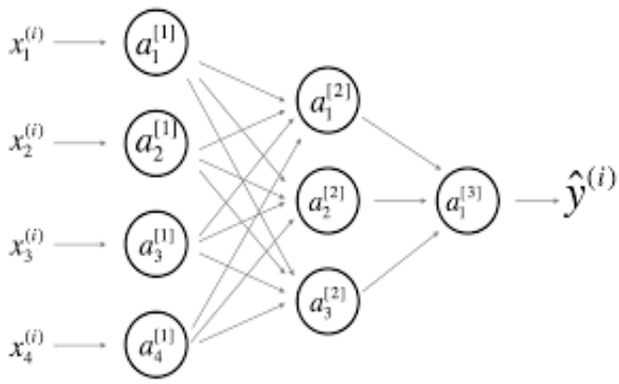


Figure 2.1: Architecture of a FCNN [1, Fig. 1]

The FCNN in Fig. 2.1 has an input layer, two hidden layers and an output layer. The relationship between quantities in each layer can be represented by vectorization in the following equations:

$$\mathsf{A}^{[1]} = g^{[1]}(\mathsf{W}^{[1]}\mathsf{X} + \mathbf{b}^{[1]}) \tag{2.12a}$$

$$\mathsf{A}^{[2]} = g^{[2]}(\mathsf{W}^{[2]}\mathsf{A}^{[1]} + \mathbf{b}^{[2]}) \tag{2.12b}$$

$$\hat{\mathsf{Y}} = \mathsf{A}^{[3]} = g^{[3]}(\mathsf{W}^{[3]}\mathsf{A}^{[2]} + \mathbf{b}^{[3]}) \tag{2.12c}$$

In Eqns. (2.12), $\mathsf{X} = \begin{bmatrix}\mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \cdots & \mathbf{x}^{(i)} & \cdots & \mathbf{x}^{(m)}\end{bmatrix}$ is the input matrix, where the superscripts (i) and (m) denote the index of the input vector example and the total number of input vector examples, respectively; $\mathsf{W}^{[l]}$ is the weight matrix of the l^{th} layer and there are three layers (i.e., only the hidden layers and the output layer are counted) in this FCNN; $\mathbf{b}^{[l]}$ denotes the bias vector of the l^{th} layer; $g^{[l]}(\cdot)$ denotes the nonlinear activation function (e.g., sigmoid function, tanh function, ReLu function, etc.) of the l^{th} layer and it is typically selected according to the application scenario; $\mathsf{A}^{[l]} = \begin{bmatrix}\mathbf{a}^{[l](1)} & \mathbf{a}^{[l](2)} & \cdots & \mathbf{a}^{[l](i)} & \cdots & \mathbf{a}^{[l](m)}\end{bmatrix}$ denotes the output matrix of the l^{th} layer. In this FCNN, the 3^{rd} layer is the output layer, therefore $\mathsf{A}^{[3]} = \hat{\mathsf{Y}} = \begin{bmatrix}\hat{\mathbf{y}}^{(1)} & \hat{\mathbf{y}}^{(2)} & \cdots & \hat{\mathbf{y}}^{(i)} & \cdots & \hat{\mathbf{y}}^{(m)}\end{bmatrix}$. In the training stage, $L(\mathbf{y}^{(i)}, \hat{\mathbf{y}}^{(i)})$ denotes the *loss function* for the i^{th} training example between the expected output label vector $\mathbf{y}^{(i)}$ and the actual output vector $\hat{\mathbf{y}}^{(i)}$ and $J(\mathsf{W}, \mathbf{b}) = \frac{1}{m}\sum_{i=1}^{m} L(\mathbf{y}^{(i)}, \hat{\mathbf{y}}^{(i)})$ denotes the *cost function* of the entire training set for the FCNN setting of W and $\mathbf{b}$, where W and $\mathbf{b}$ denote the concatenation of all $\mathsf{W}^{[l]}$ and the concatenation of all $\mathbf{b}^{[l]}$, respectively.

A common algorithm for training feedforward neural networks (FNNs) using gradient descent in SL is the *BP* algorithm [68]. In the BP algorithm, the gradients of the cost function w.r.t. the weights and biases for each network layer are computed using the chain rule, then the weights and biases are updated using gradient descent to minimize the cost function [19, Sec. 6.5]. *Adam* algorithm [69], which combines the benefits of both *adaptive gradient (AdaGrad)* [70] and *root mean square propagation (RMSProp)* [71], is widely used in engineering applications as an optimized version of *stochastic gradient descent (SGD)* to facilitate training the NN.

Next, we consider two special types of NNs: CNNs and RNNs. *CNNs*, which consist of multiple convolutional layers, pooling (i.e., subsampling) layers and FC layers, are commonly used in processing tensor signals. As an example, Fig. 2.2 illustrates the architecture of LeNet-5 for digit recognition [2], where the two convolutional layers containing filters and activation functions are the main layers of the CNN, the two pooling layers are applied to reduce the dimensions of the feature maps, the three FC layers are used to flatten the feature maps and output the final classification result. In Fig. 2.2, pooling layers are refered to as subsampling layers. It is noted that its final FC layer used Gaussian connection when this CNN was proposed, but nowadays softmax function is commonly used instead in this final FC layer in multiclass classification problems.

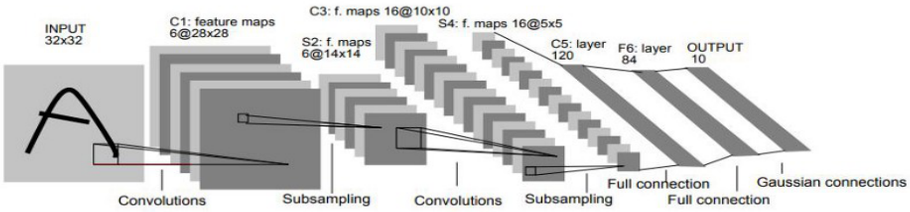


Figure 2.2: Architecture of LeNet-5 for digit recognition [2, Fig. 2]

RNNs is another common NN architecture, which can be used in many signal processing problems involving sequence models, such as speech recognition. The architecture of a unidirectional RNN is illustrated in Fig. 2.3, which depicts multiple RNN units concatenated in a temporal sequence with hidden states passed between each pair of consecutive units. There are many possible types for the internal structure of an RNN unit, such as the Elman network unit [72], the Jordan network unit [73], the GRU and the LSTM unit. Elman networks and Jordan networks are also refered to as *simple recurrent neural networks (SRNNs)*. An RNN inherently has a looping mechanism that allows information flowing from one unit to the next to keep the long-term dependency and the hidden states between adjacent units are just such information [74]. The content of the hidden states depends on the choice of the unit types (i.e., SRNN units, GRUs or LSTM units).

Figure 2.3: Architecture of a unidirectional RNN

For an *Elman network*, at the timestep t, the relationship between the input vector $\mathbf{x}^{\langle t \rangle}$ and the output vector $\hat{\mathbf{y}}^{\langle t \rangle}$ of the RNN unit is stated in the following equations [72]:

$$\mathbf{a}^{\langle t \rangle} = g_1(\mathsf{W}_a[\mathbf{a}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_a) \tag{2.13a}$$

$$\hat{\mathbf{y}}^{\langle t \rangle} = g_2(\mathsf{W}_y \mathbf{a}^{\langle t \rangle} + \mathbf{b}_y) \tag{2.13b}$$

In Eqns. (2.13), the input activation vector $\mathbf{a}^{\langle t-1 \rangle}$ and the output activation vector $\mathbf{a}^{\langle t \rangle}$ are the input hidden state and the output hidden state, respectively. W_a and W_y are internal weight matrices while $\mathbf{b}_a$ and $\mathbf{b}_y$ are internal bias vectors. The activation function $g_1(\cdot)$ is usually a tanh function while the choice of the output activation function type of $g_2(\cdot)$ depends on the application scenario. For example, in a binary classification task, $g_2(\cdot)$ can be a sigmoid function while in a multiclass classification task, $g_2(\cdot)$ can be a softmax function. *Jordan networks* are similar to Elman networks, see [73].

We note that SRNNs typically do not properly capture long-term dependencies and suffer from the vanishing gradient problem [75]. Therefore, GRU-based RNNs and LSTM-based RNNs are usually implemented in engineering applications. If the RNN is based on *GRUs*, then at the timestep t, the relationship between the input vector $\mathbf{x}^{\langle t \rangle}$ and the output vector $\hat{\mathbf{y}}^{\langle t \rangle}$ is presented in the following equations [17, Sec. 2], [18, Sec. 3.2]:

$$\mathbf{\Gamma}_u = \sigma(\mathsf{W}_u[\mathbf{c}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_u) \tag{2.14a}$$

$$\mathbf{\Gamma}_r = \sigma(\mathsf{W}_r[\mathbf{c}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_r) \tag{2.14b}$$

$$\tilde{\mathbf{c}}^{\langle t \rangle} = \phi(\mathsf{W}_c[\mathbf{\Gamma}_r \odot \mathbf{c}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_c) \tag{2.14c}$$

$$\mathbf{c}^{\langle t \rangle} = \mathbf{\Gamma}_u \odot \tilde{\mathbf{c}}^{\langle t \rangle} + (1 - \mathbf{\Gamma}_u) \odot \mathbf{c}^{\langle t-1 \rangle} \tag{2.14d}$$

$$\mathbf{a}^{\langle t \rangle} = \mathbf{c}^{\langle t \rangle} \tag{2.14e}$$

$$\hat{\mathbf{y}}^{\langle t \rangle} = g(\mathsf{W}_y \mathbf{a}^{\langle t \rangle} + \mathbf{b}_y) \tag{2.14f}$$

Similarly, if the RNN is based on *LSTM* units, the relationship between the input vector $\mathbf{x}^{\langle t \rangle}$ and the output vector $\hat{\mathbf{y}}^{\langle t \rangle}$ is represented in the following equations [16, Sec. 4]:

$$\mathbf{\Gamma}_u = \sigma(\mathsf{W}_u[\mathbf{a}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_u) \tag{2.15a}$$

$$\mathbf{\Gamma}_f = \sigma(\mathsf{W}_f[\mathbf{a}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_f) \tag{2.15b}$$

$$\mathbf{\Gamma}_o = \sigma(\mathsf{W}_o[\mathbf{a}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_o) \tag{2.15c}$$

$$\tilde{\mathbf{c}}^{\langle t \rangle} = \phi_1(\mathsf{W}_c[\mathbf{a}^{\langle t-1 \rangle}, \mathbf{x}^{\langle t \rangle}] + \mathbf{b}_c) \tag{2.15d}$$

$$\mathbf{c}^{\langle t \rangle} = \mathbf{\Gamma}_u \odot \tilde{\mathbf{c}}^{\langle t \rangle} + \mathbf{\Gamma}_f \odot \mathbf{c}^{\langle t-1 \rangle} \tag{2.15e}$$

$$\mathbf{a}^{\langle t \rangle} = \mathbf{\Gamma}_o \odot \phi_2(\mathbf{c}^{\langle t \rangle}) \tag{2.15f}$$

$$\hat{\mathbf{y}}^{\langle t \rangle} = g(\mathsf{W}_y \mathbf{a}^{\langle t \rangle} + \mathbf{b}_y) \tag{2.15g}$$

In Eqns. (2.14) and (2.15), $\mathbf{\Gamma}_u$, $\mathbf{\Gamma}_r$, $\mathbf{\Gamma}_f$ and $\mathbf{\Gamma}_o$ are refered to as the update gate vector, the reset gate vector, the forget gate vector and the output gate vector, respectively. $\tilde{\mathbf{c}}^{\langle t \rangle}$ denotes the candidate memory cell, $\mathbf{c}^{\langle t-1 \rangle}$ denotes the input memory cell and $\mathbf{c}^{\langle t \rangle}$ denotes the output memory cell. $\mathbf{a}^{\langle t-1 \rangle}$ denotes the input activation vector while $\mathbf{a}^{\langle t \rangle}$ denotes the output activation vector. $\sigma(\cdot)$ typically denotes the sigmoid function, $\phi(\cdot)$, $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are usually tanh functions, where the choice of the output activation function $g(\cdot)$ depends on the

application scenario. W_u, W_r, W_f, W_o, W_c and W_y are internal weight matrices and $\mathbf{b}_u$, $\mathbf{b}_r$, $\mathbf{b}_f$, $\mathbf{b}_o$, $\mathbf{b}_c$ and $\mathbf{b}_y$ are internal bias vectors. In both types of RNNs, $\mathbf{c}^{\langle t \rangle}$ is the cell state and $\mathbf{a}^{\langle t \rangle}$ is the hidden state. The cell state and the hidden state are always equal in GRU-based RNNs but they are not necessarily equal in LSTM-based RNNs. It is noted that for generalizing an RNN model for input sequences with different lengths and reducing the model complexity, RNN units across all timesteps share the same values of internal weight matrices and the internal bias vectors [19, Ch. 10].

There is no straightforward rule about the choice of GRUs and LSTM in many engineering application problems. However, generally, GRU-based RNNs are much simpler to implement and need shorter time and less computational power to train while LSTM-based RNNs are much more powerful in capturing long-term dependencies [76]. Unidirectional RNNs can also be extended to be bidirectional recurrent neural networks (BRNNs) and deep recurrent neural networks (DRNNs) to solve more sophisticated sequence model problems.

2.4 Rate-Distortion Theory

In this section we review relevant definitions and notions from the rate-distortion theory, beginning with the definition of a source coding scheme:

Definition 2.4.1 (Source coding scheme [43, Sec. 10.2], [46, Sec. 3.5])**.** A *source coding scheme* $(2^{nR}, n)$ with the *code rate* R and the *blocklength* n consists of:

- An *encoding function* (also refered to as *encoder*) $f_n(\cdot)$ that maps a sequence of n source symbols $\{x_i\}_{i=1}^n \equiv x^n$, over corresponding alphabets $\{\mathcal{X}_i\}_{i=1}^n \equiv \mathcal{X}^n$, into an index selected from a set containing 2^{nR} indices, i.e., $f_n(\cdot) : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$.

- A *decoding function* (also refered to as *decoder*) $g_n(\cdot)$ that assigns a sequence of n reconstruction symbols $\{\hat{x}_i\}_{i=1}^n \equiv \hat{x}^n$, over corresponding alphabets $\{\hat{\mathcal{X}}_i\}_{i=1}^n \equiv \hat{\mathcal{X}}^n$, to each received index, i.e., $g_n(\cdot) : \{1, 2, \ldots, 2^{nR}\} \to \hat{\mathcal{X}}^n$.

The set of *codewords* $\{g_n(i)\}_{i=1}^{2^{nR}}$ is referred to as the *codebook* while $\{f_n^{-1}(i)\}_{i=1}^{2^{nR}}$ are the corresponding *assignment regions* [43, Sec. 10.2].

In lossless source coding, the probability of error of a $(2^{nR}, n)$ source code is defined as $P_e^{(n)} \triangleq \Pr\{\hat{X}^n \neq X^n\}$ [46, Sec. 3.5]. If there exists a source code $(2^{nR}, n)$ which achieves $\lim\limits_{n\to\infty} P_e^{(n)} = 0$, then the code rate R is said to be *achievable* and the optimal code rate R^* is the infimum of all achievable code rates [46, Sec. 3.5]. The *lossless source coding theorem* for discrete memoryless sources (DMSs) is given as follows:

Theorem 2.4.1 (Lossless source coding theorem for DMSs [77, Sec. 4.4])**.** A DMS X with entropy $H(X)$ can be encoded with a code rate not less than $H(X)$ with a negligible probability of error as $n \to \infty$; conversely if the code rate is less than $H(X)$, having reconstruction errors has a finite positive probability.

Lossless source coding can be used to compress digital sources whose symbols are drawn from finite alphabets. Nevertheless, lossless compression of digital sources whose symbols are drawn from infinite alphabets and of analog sources requires an infinite code rate, which is practically impossible [46, Sec. 3.6]. Therefore, for such sources, compression is applied with a bounded reconstruction error.

This reconstruction error is characterized by a *distortion measure function*, which for a single symbol can be understood as a mapping from the pair of source symbol alphabet and reconstruction symbol alphabet into the set of nonnegative real numbers, i.e., $d(\cdot, \cdot) : \mathcal{X} \times \hat{\mathcal{X}} \to \mathcal{R}^+$ [43, Eqn. (10.2)]. The *Hamming distortion function* and the *squared-error distortion function* are two common distortion measure functions used to quantify the cost of representing the source symbol x by the reconstruction symbol $\hat{x}$:

- **Hamming distortion function**:

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases} \tag{2.16}$$

- **Squared-error distortion function**:

$$d(x,\hat{x}) = (x-\hat{x})^2 \tag{2.17}$$

The definition of distortion measure functions can be extended to sequences, by considering pairwise average distortions between the elements in a sequence of source symbols x^n and the elements in the corresponding sequence of reconstruction symbols $\hat{x}^n$, see [43, Eqn. (10.6)]:

$$d(x^n,\hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i) \tag{2.18}$$

The expected distortion associated with a lossy source coding scheme $(2^{nR},n)$ is defined as [43, Eqns. (10.9), (10.10)]:

$$\mathbb{E}\{d(X^n,\hat{X}^n)\} = \mathbb{E}\Big\{d\Big(X^n,g_n\big(f_n(X^n)\big)\Big)\Big\} = \sum_{x^n\in\mathcal{X}^n} p_{X^n}(x^n)d\Big(x^n,g_n\big(f_n(x^n)\big)\Big) \tag{2.19}$$

A *rate-distortion pair* (R,D) is defined to be *achievable* if there exists a sequence of source code $(2^{nR},n)$ which satisfies $\lim\limits_{n\to\infty}\mathbb{E}\Big\{d\Big(X^n,g_n\big(f_n(X^n)\big)\Big)\Big\}\leqslant D$. The *rate-distortion region* is the closure of the set of achievable rate-distortion pairs [43, Sec. 10.2]. The *RDF*, denoted as $R(D)$, is the infimum of the code rates R for which the rate-distortion pair (R,D) is achievable [46, Sec. 3.6]. Before presenting the lossy source coding theorem, we first define the information RDF:

Definition 2.4.2 (Information RDF [43, Sec. 10.2])**.** For a DMS X defined over $\mathcal{X}$, a reconstruction RV $\hat{X}$ defined over $\hat{\mathcal{X}}$, and a distortion measure function $d(x,\hat{x})$, the *information RDF* $R^{(I)}(D)$ is defined as:

$$R^{(I)}(D) = \min_{p_{\hat{X}|X}(\hat{x}|x):\mathbb{E}\{d(X,\hat{X})\}\leqslant D} I(X;\hat{X}) \tag{2.20}$$

where $I(\cdot;\cdot)$ denotes the mutual information between two RVs and the minimization is over all conditional distributions $p_{\hat{X}|X}(\hat{x}|x)$ which satisfies the preset distortion constraint $\mathbb{E}\{d(X,\hat{X})\}\leqslant D$.

For DMSs, the RDF is equal to the information RDF [78, Thm. 25.1]. Therefore the *lossy source coding theorem* can be expressed as:

Theorem 2.4.2 (Lossy source coding theorem [43, Thm. 10.2.1], [46, Thm. 3.5])**.** For a DMS X defined over $\mathcal{X}$, a reconstruction RV $\hat{X}$ defined over $\hat{\mathcal{X}}$, and a distortion measure function $d(x,\hat{x})$, the RDF is equal to the respective information RDF:

$$R(D) = R^{(I)}(D) = \min_{p_{\hat{X}|X}(\hat{x}|x):\mathbb{E}\{d(X,\hat{X})\}\leqslant D} I(X;\hat{X}) \tag{2.21}$$

where $R(D)$ is the infimum of all code rates for which $\mathbb{E}\{d(X,\hat{X})\}\leqslant D$.

Next, we provide two examples for the application of Eqn. (2.21):

- The RDF for compressing a DMS $\sim Bern(p)$ with the Hamming distortion function is expressed as [43, Eqn. (10.13)]:

$$R(D) = \begin{cases} H_b(p) - H_b(D) & 0\leqslant D\leqslant \min\{p,1-p\} \\ 0 & D>\min\{p,1-p\} \end{cases} \tag{2.22}$$

  where $H_b(\cdot)$ denotes the *binary entropy function*.

- The RDF for compressing a DMS $\sim\mathcal{N}(0,\sigma^2)$ with the squared-error distortion function is expressed as [43, Eqn. (10.24)]:

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D} & 0\leqslant D\leqslant\sigma^2 \\ 0 & D>\sigma^2 \end{cases} \tag{2.23}$$

As an extension of Eqn. (2.23), when compressing m independent (but not necessarily identical) Gaussian sources $X_1, X_2, \ldots, X_m$, $X_i \sim \mathcal{N}(0, \sigma_i^2)$ and using the squared-error distortion function, the RDF is [43, Eqn. (10.51)]:

$$R(D) = \begin{cases} \frac{1}{2} \sum_{i=1}^{m} \log \frac{\sigma_i^2}{D_i} & 0 \leqslant D \leqslant \sum_{i=1}^{m} \sigma_i^2 \\ 0 & D > \sum_{i=1}^{m} \sigma_i^2 \end{cases} \tag{2.24}$$

where $D_i = \min\{\sigma_i^2, \lambda\}$ and λ is chosen for $\sum_{i=1}^{m} D_i = D$. Eqn. (2.24) is also called *reverse waterfilling bit-allocation* [43, Sec. 10.3.3].

Finally, we consider the transmission of source symbols to a destination over a channel considered in [45]. For this transmission, source coding can be jointly designed with channel coding, which is refereed to as *joint source-channel coding (JSCC)* [46, Sec. 3.9]. In a JSCC scheme, the transmitter transmits k source symbols over a channel using n channel symbols, where the *joint source-channel code rate* is said to be $R = \frac{k}{n}$ and the *bandwidth expansion factor* is $\rho = \frac{n}{k}$ [79, Def. 6.32], [78, Sec. 26.3]. The receiver reconstructs the source symbols with a expected distortion less than or equal to the prescribed constraint D [46, Sec. 3.9]. The JSCC transmission scheme can be simplified by considering two separate steps: First, use a source code to encode the source sequences into messages, and then use a channel code to convey the messages to the destination for reconstruction [43, Thm. 10.4.1]. Therefore, the source code and the channel code are designed separately. It turns out that the separation of source coding and channel coding leads to an asymptotically optimal performance when sending a DMS over a discrete memoryless channel (DMC) [46, Sec. 3.9]. The *source-channel separation theorem* for transmitting a DMS over a DMC is stated as follows:

Theorem 2.4.3 (Source-channel separation theorem [43, Thm. 10.4.1], [46, Thm. 3.7], [78, Sec. 26.3])**.** For a DMS S, with a code rate R and the RDF $R(D)$ transmitted over a DMC with the capacity C, if $R(D) \leqslant \frac{C}{R}$, then the *rate-distortion pair* (R, D) is said to be *achievable* via separate source and channel codes; conversely, if (R, D) is achievable with separate source and channel codes, then $R(D) \leqslant \frac{C}{R}$.

We note that Thm. 2.4.3 can also be specialized to lossless JSCC scenarios that the transmitter transmits a sequence of i.i.d. source symbols, each with entropy $H(S)$, through a DMC with capacity C, with a code rate R, so that the receiver can reconstruct the source symbols with a negligible probability of error. Thm. 2.4.3 implies that such a transmission requires $H(S) \leqslant \frac{C}{R}$ [46, Rmk. 3.14].

Chapter 3

Future Research Topics

3.1 Source Coding for Asynchronously Sampled Wide-Sense Cyclostationary Gaussian Sources with Memory

In Sec. 1.1, it was explained that cyclostationarity is inherent to many communications signals as it is induced by periodic operations applied in the generation process of communications signals. As observed in previous works, e.g., [54] and [80], sampling a CT WSCS process can result in one of two situations: When the sampling interval is related to the cyclostationarity period of the CT WSCS process via a rational number, the resulting DT process is WSCS. This situation is refered to as synchronous sampling. When the sampling interval and the cyclostationarity period of the CT WSCS process are related via an irrational number, the resulting DT process is WSACS. This situation is refered to as asynchronous sampling. The lossy compression of sampled WSCS Gaussian sources was considered in two previous works: The work [53] characterized the DRF of DT WSCS Gaussian sources *with memory*, which corresponds to the synchronous sampling of a CT WSCS Gaussian source, where the sampling interval is *shorter* than the maximal correlation length of the CT source. The work [54] derived the RDF of DT WSACS *memoryless* Gaussian sources, which corresponds to the asynchronous sampling of a CT WSCS Gaussian source, where the sampling interval is *larger* than the maximal correlation length of the CT source. In this proposal, we propose to study the RDF of asynchronously sampled WSCS Gaussian sources *with memory*, which corresponds to the asynchronous sampling of a CT WSCS Gaussian source, where the sampling interval is *smaller* than maximum correlation length of the CT source.

Typical applications of the considered source compression model include information storage systems and cooperative transmission schemes. As an example, consider the transmission process at the relay in *compress-and-forward (CF) relay networks* [81]. In a single CF relay network, in addition to a direct channel between the source and the destination, there is also a source-relay-destination channel, which consists of a source-relay channel and a relay-destination channel [82, Sec. 15.3]. In such a network, the signals received at the relay and at the destination are different noisy versions of the same source signal, thus they are correlated. Then, the signal received at the relay is compressed and subsequently forwarded to the destination [83, Sec. 4.2.2]. The destination decodes the information based on the joint output obtained from the source-destination channel and the relay-destination channel. In practice, due to the inherent jitter in physical clocks, it is reasonable to assume that the sampling interval at the relay is related to the cyclostationarity period of the signal received at the relay via an irrational number. Moreover, in practical scenarios, the sampling interval at the relay is shorter than the maximal correlation length of the received signal. Therefore, the sampled received signal at the relay can be modelled as a DT WSACS Gaussian process with memory. This application scenario is an important motivation of the proposed research.

As introduced in Sec. 1.3, the work [54] characterized the RDF of DT WSACS memoryless Gaussian sources using the information-spectrum framework, for scenarios in which the distortion is not larger than the minimal source variance. Another related recent work is [84], which studied the capacity of asynchronously sampled linear interference-limited communications (ILC) channels with additive CT WSCS Gaussian noise with memory, again using the information-spectrum framework. In the work [84], the model used in the work [80] was generalized to add memory to the asynchronously sampled WSCS Gaussian interference model. This was motivated by the fact that memory is inherent in the statistical characterization of many sampled communications signals.

The work [84] also elaborated on the rationale for using Gaussianity as an appropriate statistical characteristic in many modulation types. Using the works [54] and [84] as starting points, the proposed research aims to derive the RDF of DT WSACS Gaussian sources with memory. We note that this characterization is the dual problem for the channel capacity characterization described in the work [84], therefore it is estimated that we can adapt some of the tools and analysis methods derived in [84] to the RDF characterization problem, following the general outline of the work [54]. We note that the analysis in the proposed research topic is expected to require significant modifications and additional analysis w.r.t. [84], given that each of the works [54], [80] and [84] imposed different constraints and introduced different assumptions to facilitate analysis.

3.2 Novel Algorithmic Design of Automatic Modulation Classification

In this research, we propose to improve upon current AMC schemes by utilizing the inherent cyclostationary properties of communications signals combined with ML techniques. In the following, we briefly discuss this topic and elaborate on the planned research.

The design of a modulation classifier generally involves two steps: The first step is *preprocessing*, and the second step is the *modulation classification rule* applied based on the preprocessed information [57, Sec. I]. Typical operations in the preprocessing step include noise reduction, carrier frequency estimation and equalization. Preprocessing typically depends on the design of the modulation classification rule in the second step [57, Sec. I]. AMC algorithms can be generally classified into *LB* methods and *FB* methods [57, Sec. I]. LB methods are based on the likelihood function of the received signal and the final classification decision is reached by comparing the likelihood ratio with a threshold level, while FB methods are based on extracting useful characteristic features from the received signal which are the basis for the classification decision [57, Sec. I]. A modulation classifier should either decide which modulation scheme is used to generate the received signal, or conclude that the received modulation scheme is not within the predefined classification catalog [85, Sec. I]. As detailed in Sec. 1.1, the modulation process induces cyclostationary properties upon the statistics of communications signals, therefore one approach to implement AMC is based on extracting cyclostationary features from the received signal. ML can also be introduced to optimize modulation classifiers design in terms of both performance and complexity with FB methods [86, Ch. 6].

Within the proposed AMC research, there are three possible initial directions that can be pursued in order to improve upon current algorithms:

The first direction is implementing AMC via RL. As introduced in Sec. 1.2, RL is one of the three major paradigms of ML. In RL, the agent learns to map environmental observations to different actions, in order to maximize a cumulative reward [87, Sec. 1.1]. Unlike SL and UL, RL implements goal-directed learning from trial-and-error experiments with the environment [87, Sec. 1.1]. The RL optimization problem is typically modelled as a Markov decision process (MDP), or more commonly a partially observable Markov decision process (POMDP) and in both cases dynamic programming techniques can be adopted for problem-solving [87, Ch. 4], [88]. RL has been extensively applied in many sequential decision-making problems, such as industrial automation. In our future research, we plan to design an RL-based AMC algorithm, in which a DNN makes the classification decision based on the extracted features. One of the main challenges in the implementation of RL is how to automatically determine its reward, i.e., determine whether the classification decision is correct or not. We shall start with a binary modulation classification problem, where the agent consists of only one detector, which is used to distinguish between two modulations. The agent computes the decision metric for the two modulations via a DNN. If the metric is higher than a certain threshold, the agent selects the corresponding modulation as the classification result; otherwise the classification result is the other modulation. After the classification, the received signal is demodulated using the decided modulation type. Assuming a non-adversarial scenario, we propose to check the integrity of the demodulated information using a cyclic redundancy check (CRC). If the CRC test is passed, the agent receives a positive reward; otherwise, the agent receives a negative reward. Next, we propose to consider a generalization of the binary classification problem to a multiclass classification problem, in which more than two modulation types are involved. Now the agent consists of a multiclass detector which assigns a confidence level to each modulation type. The agent selects the modulation scheme with the highest level as the classification result, and the received signal is subsequently demodulated using the selected modulation type. Then, CRC is applied to check the integrity of the demodulated information and the reward is accordingly determined. One possible approach here is to incorporate the MAXNET architecture considered in [59] and apply it within the RL setup, rather than the SL setup. For the next step, we can consider more aspects in this research, such as determining the reward

without decoding the received signal to protect information privacy.

The second direction is based on the multi-antenna modulation classifier proposed in [63]. It seems that using a weighted combination of the output of each receiving antenna or a weighted combination of the classification evaluation metric from each receiving antenna will improve the AMC system performance. Following this intuition, we propose to train a DNN to determine the weight parameters by DL.

The third direction is the implementation of AMC in the time domain via eye patterns. In digital communications, an eye pattern is obtained by displaying a received digital signal on the vertical input with the horizontal sweep rate set at the data rate [44, Sec. 9.2]. Eye patterns are typically used to evaluate the combined effects of inter-symbol interference (ISI) and channel noise in baseband pulse-transmission systems. Minimal ISI and channel noise lead to the "opening" of an eye pattern while the "closure" of an eye pattern is the result of severe ISI and channel noise [44, Sec. 9.2], [89]. Different modulation types lead to different eye patterns. We propose to use this characteristic to design AMC algorithms, where the DNN is applied to perform the classification task, possibly based on a processed version of the eye pattern of the received signal.

3.3 Sampling Frequency Synchronization in Non-Orthogonal Multiple Access Networks

In multi-user communications, multiple-access techniques are applied to facilitate sharing the common and finite wireless radio resources [90, Ch. 8]. Conventionally, to reduce the mutual interference between users, communications systems have been designed based on the *orthogonal multiple access (OMA)* paradigm, which separates different users in at least one domain, e.g., frequency (frequency-division multiple access (FDMA)), time (time-division multiple access (TDMA)), code (code-division multiple access (CDMA)) or space (space-division multiple access (SDMA)). From 1G to 4G communications, FDMA, TDMA, CDMA and OFDMA have been used, respectively. However, such separation restricts the flexibility of resource allocation, and therefore is inherently suboptimal. This suboptimality has motivated the interest in *non-orthogonal multiple access (NOMA)*. Existing analysis has shown that the NOMA paradigm carries multiple advantages over OMA techniques, including larger system capacity, higher energy/spectral efficiency and lower latency [91, Sec. 6], [92]. Therefore, NOMA has been a promising multiple-access technique in 5G communications and beyond.

In a NOMA network, all users can transmit their signals simultaneously over the same frequency band to their respective destination receivers. The signals intended for a certain receiver can be distinguished in the power domain: Different users are allocated different power levels for signal transmission, based on their channel state information (CSI). The transmitted signals are superimposed at the receiver, which then applies a multiuser decoding scheme for decoding these simultaneously received signals. To motivate the proposed research, consider a single-cell uplink NOMA network with two users and one base-station (BS). Both users transmit OFDM signals simultaneously over the same frequency band to the BS with different transmitter sampling intervals, and the BS receives the superposition of the two signals. There are two fundamental approaches for decoding the signals. The first one is joint decoding, namely decoding both signals simultaneously. Joint decoding typically results in good performance, but its implementation usually leads to a high computational complexity. Another possible decoding approach is successive interference cancellation (SIC), in which the BS decodes the stronger signal first while regarding the other signal as the noise, subsequently subtracts the stronger signal from the received signal, and finally decodes the weaker signal [93, Sec. 6.1.1]. As such schemes operate in DT, the CT received signal has to be sampled first, hence the sampling frequency estimation for each user's signal is a necessary step before implementing either decoding approach. It is noted that the SFO between the sampling clock at the transmitter and the sampling clock at the receiver typically induces inter-carrier interference (ICI) in OFDM systems and ISI in linear modulations.

To the best of our knowledge, the problem of sampling frequency synchronization in NOMA networks has not been considered previously. Previous works on sampling frequency synchronization only considered estimating the transmitter's sampling interval for a single received communications signal. However, as in a NOMA network, at the receiver, there are multiple simultaneously received communications signals, then estimating for a single sampling interval will inevitably result in at least one signal (and likely both) being sampled at an interval which is different from its sampling interval at its transmitter, which results in the ICI/ISI. The proposed research is to estimate the appropriate sampling interval for each signal received at the receiver in a NOMA network. This will facilitate, for example, splitting the received signal into two decoders such that each decoder receives a DT signal sampled with the appropriate sampling interval for decoding its desired information, thereby minimizing the ICI/ISI.

There are three works about the sampling frequency synchronization in OMA communications networks, which could be starting points for our research.

The first work is [9], which proposed the CB-SFS algorithm and has been discussed in Sec. 1.1. The second work is [94], which proposed an unbiased LS algorithm with a low computational complexity for the joint estimation of the CFO and the SFO in an OFDM system. This algorithm uses continual pilots, which are non-uniformly distributed pilot symbols embedded implemented in the OFDM symbols in the DVB-C2, which is the second generation of the digital transmission system for cable communications. The objective of the continual pilots is to assist with the frequency synchronization task. The algorithm derived in [94] computes the phase-difference dependent signals between a pair of successive OFDM symbols for a specifically selected set of continual pilots. Then, two principal phase angles are obtained as two sums of these phase differences. The estimates of the CFO and the SFO are finally computed from the two principal phase angles via some simple operations. Due to the non-uniformity of the distribution of pilot subcarriers, the pilot subcarrier indices need to be properly selected to eliminate the bias in the SFO/CFO estimation.

The third work is [95], which proposed a sequential and pilot-assisted algorithm for the joint estimation of the SFO and the CFO in OFDM systems, without the knowledge of the CSI. One SFO estimator and two CFO estimators were proposed in the work. The proposed algorithm estimates the SFO at first, then it estimates the CFO with either of the CFO estimators assuming successful SFO compensation. For SFO estimation, the algorithm first makes some processing of the pilot subcarriers from any two consecutive OFDM symbols to form a cost function containing only one auxiliary parameter. Then the SFO estimate is obtained by substituting two designated values of the auxiliary parameter into the cost function and applying some simple operations. For CFO estimation, another cost function containing only one auxiliary parameter is formed by processing the pilot subcarriers from any two consecutive OFDM symbols assuming the impact of the SFO has been eliminated. Two CFO estimators are derived from this cost function: The first CFO estimator is obtained by substituting three specific values of the auxiliary parameter into the cost function and applying some simple operations, while the other CFO estimator is derived without substituting values into the cost function.

In our research, we plan to first analyze the statistics of the received OFDM signals and design a parallel sampling frequency estimator utilizing the fact that the cyclic properties of the received signals are directly related to their symbol periods. Initial analysis supports the feasibility of such idea. The work planned within this research includes the design of a robust algorithm for extracting such cyclic information, as well as the analysis and the simulation of the proposed algorithm.

Chapter 4

Research Plan

The planned research will be carried out sequentially.

- The first research topic we plan to carry out is the RDF characterization of DT WSACS Gaussian sources with memory as detailed in Sec. 3.1. The works [54] and [84] can be used as starting points and we plan to adopt some tools and analysis methods from the work [84], following the general outline of the work [54]. We shall also specifically consider the application of this result to CF relay networks. This work is expected to lead to the first journal paper publication.

- The second research topic is the design of novel AMC methods assisted by ML as introduced in Sec. 3.2. We shall first investigate the application of RL to AMC. Subsequently we shall investigate the incorporation of the multiple-antenna reception and eye patterns assisted by ML in AMC algorithmic design. We will first start with a binary modulation classification problem, then generalize to a multiclass classification problem. This work is expected to produce the second journal paper publication.

- The third research topic is the sampling frequency synchronization in NOMA networks as detailed in Sec. 3.3. In a NOMA network, the BS simultaneously receives multiple communications signals from different users over the same frequency band with different transmitter's sampling intervals, and the challenge in this topic is to estimate the SFO between the sampling clock at the transmitter and the sampling clock at the receiver for each received signal. We will use existing sampling frequency synchronization methods in OMA networks as starting points for our proposed research topic. This work is expected to produce the third journal paper publication.

- The final step is to write the PhD thesis for submission.

References

[1] A. Ng. "CS230 Deep Learning: Standard notations for deep learning". Available online: `https://cs230.stanford.edu/files/Notation.pdf`. (Accessed on 2022-08-03).

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[3] G. B. Giannakis. "Cyclostationary signal analysis". In V. K. Madisetti and D. B. Williams, Editors, *Digital Signal Processing Handbook*, Chapter 17. CRC Press, Boca Raton, FL, USA, 1999.

[4] W. A. Gardner, A. Napolitano, and L. Paura. "Cyclostationarity: Half a century of research". *Signal Processing*, 86(4):639–697, 2006.

[5] R. van Nee and R. Prasad. *OFDM for Wireless Multimedia Communications*. Artech House, Norwood, MA, USA, 2000.

[6] M. Speth, S. Fechtel, G. Fock, and H. Meyr. "Optimum receiver design for OFDM-based broadband transmission .II. A case study". *IEEE Transactions on Communications*, 49(4):571–578, 2001.

[7] G. K. Yeung and W. A. Gardner. "Search-efficient methods of detection of cyclostationary signals". *IEEE Transactions on Signal Processing*, 44(5):1214–1223, 1996.

[8] Y. Hu, B. Yu, Z. Deng, G. He, and H. Zhou. "Efficient cycle frequency acquisition of a cyclostationary signal with the FACA method". *Radioengineering*, 28(2):447–455, 2019.

[9] M. Kumar and R. Dabora. "A novel sampling frequency offset estimation algorithm for OFDM systems based on cyclostationary properties". *IEEE Access*, 7:100692–100705, 2019.

[10] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane. "Automated design of both the topology and sizing of analog electrical circuits using genetic programming". In J. S. Gero and F. Sudweeks, Editors, *Artificial Intelligence in Design '96*, Chapter 9. Springer Netherlands, Dordrecht, Netherlands, 1996.

[11] B. Ramsundar and R. B. Zadeh. *TensorFlow for Deep Learning*. O'Reilly Media, Sebastopol, CA, USA, 2018.

[12] IBM Cloud Education. "Recurrent neural networks". Available online: `https://www.ibm.com/cloud/learn/recurrent-neural-networks`. (Accessed on 2022-09-06).

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks". In *Advances in 2012 Neural Information Processing Systems (NeurIPS 2012)*, volume 25, pp 1097–1105, Lake Tahoe, NV, USA, Dec 2012.

[14] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In *2015 International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, May 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp 770–778, Las Vegas, NV, USA, Jun 2016.

[16] S. Hochreiter and J. Schmidhuber. "Long short-term memory". *Neural Computation*, 9(8):1735–1780, 1997.

[17] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio. "On the properties of neural machine translation: Encoder–decoder approaches". In *Proceedings of 2014 Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST 2014)*, pp 103–111, Doha, Qatar, Oct 2014.

[18] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In *2014 Neural Information Processing Systems (NeurIPS 2014) Workshop on Deep Learning*, Montreal, QC, Canada, Dec 2014.

[19] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016.

[20] M. C. Valenti. "Best readings in machine learning in communications". Available online: `https://www.comsoc.org/publications/best-readings/machine-learning-communications`. (Accessed on 2022-07-06).

[21] R. M. Dreifuerst, R. W. Heath, M. N. Kulkarni, and J. Charlie. "Deep learning-based carrier frequency offset estimation with one-bit ADCs". In *2020 IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2020)*, pp 1–5, Atlanta, GA, USA, May 2020.

[22] G. Izacard, S. Mohan, and C. Fernandez-Granda. "Data-driven estimation of sinusoid frequencies". In *Proceedings of the 2019 International Conference on Neural Information Processing Systems (NeurIPS 2019)*, volume 32, Vancouver, BC, Canada, Dec 2019.

[23] A. Schuster. "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena". *Terrestrial Magnetism*, 3(1):13–41, 1898.

[24] P. Welch. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[25] R. Schmidt. "Multiple emitter location and signal parameter estimation". *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[26] G. Izacard, B. Bernstein, and C. Fernandez-Granda. "A learning-based framework for line-spectra super-resolution". In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*, pp 3632–3636, Brighton, UK, May 2019.

[27] C. Boyer, Y. De Castro, and J. Salmon. "Adapting to unknown noise level in sparse deconvolution". *Information and Inference: A Journal of the IMA*, 6(3):310–348, 2017.

[28] A. Rotem. "Simultaneous synchronization in non-orthogonal multiple-access". MSc Thesis, Ben-Gurion University of the Negev, Be'er Sheva, Israel, 2022.

[29] A. Li, Y. Me, S. Xue, N. Yi, and R. Tafazolli. "A carrier-frequency-offset resilient OFDMA receiver designed through machine deep learning". In *2018 IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2018)*, pp 1–6, Bologna, Italy, Sep 2018.

[30] S. Kumari, K. K. Srinivas, and P. Kumar. "Channel and carrier frequency offset equalization for OFDM based UAV communications using deep learning". *IEEE Communications Letters*, 25(3):850–853, 2021".

[31] S. Banerjee and K. Giridhar. "A novel method for non-stationary CFO estimation and tracking in inter-UAV OFDM links". In *2019 IEEE Vehicular Technology Conference (VTC 2019-Fall)*, pp 1–5, Honolulu, HI, USA, Sep 2019.

[32] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson. "On channel estimation in OFDM systems". In *1995 IEEE Vehicular Technology Conference (VTC 1995). Countdown to the Wireless Twenty-First Century*, volume 2, pp 815–819 vol.2, Chicago, IL, USA, Jul 1995.

[33] IEEE Computer Society. "IEEE standard for information technology–Telecommunications and information exchange between systems - Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 2: Sub 1 GHz license exempt operation". *IEEE Std 802.11ah-2016 (Amendment to IEEE Std 802.11-2016, as amended by IEEE Std 802.11ai-2016)*, pp 1–594, 2017.

[34] V. Ninkovic, A. Valka, D. Dumic, and D. Vukobratovic. "Deep learning-based packet detection and carrier frequency offset estimation in IEEE 802.11ah". *IEEE Access*, 9:99853–99865, 2021.

[35] Y. Li, L. J. Cimini, and N. R. Sollenberger. "Robust channel estimation for OFDM systems with rapid dispersive fading channels". *IEEE Transactions on Communications*, 46(7):902–915, 1998.

[36] T. M. Schmidl and D. C. Cox. "Robust frequency and timing synchronization for OFDM". *IEEE Transactions on Communications*, 45(12):1613–1621, 1997.

[37] E. Perahia and R. Stacey. *Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n*. Cambridge University Press, Cambridge, UK, 2008.

[38] K. S. Kim, S. W. Kim, Y. S. Cho, and J. Y. Ahn. "Synchronization and cell-search technique using preamble for OFDM cellular systems". *IEEE Transactions on Vehicular Technology*, 56(6):3469–3485, 2007.

[39] Y. Zhou, H. Wang, S. Zheng, and Z. Z. Lei. "Advances in IEEE 802.11ah standardization for machine-type communications in Sub-1GHz WLAN". In *2013 IEEE International Conference on Communications (ICC 2013) Workshops*, pp 1269–1273, Budapest, Hungary, Jun 2013.

[40] P. H. Moose. "A technique for orthogonal frequency division multiplexing frequency offset correction". *IEEE Transactions on Communications*, 42(10):2908–2914, 1994.

[41] A. van Zelst and T. C. W. Schenk. "Implementation of a MIMO OFDM-based wireless LAN system". *IEEE Transactions on Signal Processing*, 52(2):483–494, 2004.

[42] S. Nagaraj, S. Khan, C. Schlegel, and M. Burnashev. "On preamble detection in packet-based wireless networks". In *2006 IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA 2006)*, pp 476–480, Manaus, Brazil, Aug 2006.

[43] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, NJ, USA, 2006.

[44] J. G. Proakis and M. Salehi. *Digital Communications*. McGraw-Hill, New York, NY, USA, 2008.

[45] C. E. Shannon. "A mathematical theory of communication". *The Bell System Technical Journal (BSTJ)*, 27(3):379–423, 1948.

[46] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, Cambridge, UK, 2011.

[47] C. E. Shannon. "Coding theorems for a discrete source with a fidelity criterion". In N. J. A. Sloane and A. D. Wyner, Editors, *Claude E. Shannon: Collected Papers*, Chapter 22. Wiley-IEEE Press, Piscataway, NJ, USA, 1993.

[48] D. A. Huffman. "A method for the construction of minimum-redundancy codes". *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[49] J. Ziv and A. Lempel. "A universal algorithm for sequential data compression". *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.

[50] J. Ziv and A. Lempel. "Compression of individual sequences via variable-rate coding". *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.

[51] N. Ahmed, T. Natarajan, and K. R. Rao. "Discrete cosine transform". *IEEE Transactions on Computers*, C-23(1):90–93, 1974.

[52] I. Daubechies. "Orthonormal bases of compactly supported wavelets". *Communications on Pure and Applied Mathematics (CPAM)*, 41(7):909–996, 1988.

[53] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar. "The distortion rate function of cyclostationary Gaussian processes". *IEEE Transactions on Information Theory*, 64(5):3810–3824, 2018.

[54] E. Abakasanga, N. Shlezinger, and R. Dabora. "On the rate-distortion function of sampled cyclostationary Gaussian processes". *Entropy*, 22(3), 2020.

[55] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer, Berlin, Germany, 2003.

[56] J. Hamkins and M. K. Simon. "Deep Space Communications and Navigation Series: Chapter 9 Modulation classification". Available online: `https://descanso.jpl.nasa.gov/monograph/series9/Descanso9_09_rev.pdf`. (Accessed on 2022-11-02).

[57] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. "Survey of automatic modulation classification techniques: Classical approaches and new trends". *IET Communications*, 1:137–156, 2007.

[58] T. V. R. O. Câmara, A. D. L. Lima, B. M. M. Lima, A. I. R. Fontes, A. D. M. Martins, and L. F. Q. Silveira. "Automatic modulation classification architectures based on cyclostationary features in impulsive environments". *IEEE Access*, 7:138512–138527, 2019.

[59] B. Ramkumar. "Automatic modulation classification for cognitive radios using cyclic feature detection". *IEEE Circuits and Systems Magazine*, 9(2):27–45, 2009.

[60] A. Fehske, J. Gaeddert, and J. H. Reed. "A new approach to signal classification using spectral correlation and neural networks". In *2005 IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, pp 144–150, Baltimore, MD, USA, Nov 2005.

[61] C. Schreyogg, K. Kittel, U. Kressel, and J. Reichert. "Robust classification of modulation types using spectral features applied to HMM". In *Proceedings of 1997 IEEE Military Communications Conference (MILCOM 1997)*, volume 3, pp 1377–1381 vol.3, Monterey, CA, USA, Nov 1997.

[62] W. A. Gardner and C. M. Spooner. "Signal interception: Performance advantages of cyclic-feature detectors". *IEEE Transactions on Communications*, 40(1):149–159, 1992.

[63] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. "Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels". *Wireless Personal Communications*, 54(4):699–717, 2010.

[64] W. Wei and J. M. Mendel. "Maximum-likelihood classification for digital amplitude-phase modulations". *IEEE Transactions on Communications*, 48(2):189–193, 2000.

[65] A. Swami and B. M. Sadler. "Hierarchical digital modulation classification using cumulants". *IEEE Transactions on Communications*, 48(3):416–429, 2000.

[66] C. M. Spooner. "On the utility of sixth-order cyclic cumulants for RF signal classification". In *Conference Record of 2001 Asilomar Conference on Signals, Systems and Computers (ACSSC 2001)*, volume 1, pp 890–897 vol.1, Pacific Grove, CA, USA, Nov 2001.

[67] F. Chérif. "A various types of almost periodic functions on Banach spaces: Part I". *International Mathematical Forum*, 6(19):921–952, 2011.

[68] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". *Nature*, 323(6088):533–536, 1986.

[69] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In *2015 International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, May 2015.

[70] J. Duchi, E. Hazan, and Y. Singer. "Adaptive subgradient methods for online learning and stochastic optimization". *Journal of Machine Learning Research (JMLR)*, 12:2121–2159, 2011.

[71] G. Hinton. "CSC321 Introduction to Neural Networks and Machine Learning: Lecture 6a Overview of mini-batch gradient descent". Available online: `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`. (Accessed on 2022-08-03).

[72] J. L. Elman. "Finding structure in time". *Cognitive Science*, 14(2):179–211, 1990.

[73] M. I. Jordan. "Serial order: A parallel distributed processing approach". In J. W. Donahoe and V. P. Dorsel, Editors, *Neural-Network Models of Cognition*, Advances in Psychology, Chapter 25. North-Holland, Amsterdam, Netherlands, 1997.

[74] M. Phi. "Illustrated guide to recurrent neural networks". Available online: `https://towardsdatascience.com/illustrated-guide-to-recurrent-neural-networks-79e5eb8049c9`. (Accessed on 2022-08-08).

[75] Y. Bengio, P. Simard, and P. Frasconi. "Learning long-term dependencies with gradient descent is difficult". *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.

[76] K. Mani. "GRU's and LSTM's". Available online: `https://towardsdatascience.com/grus-and-lstm-s-741709a9b9b1`. (Accessed on 2022-08-14).

[77] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.

[78] Y. Polyanskiy and Y. Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, Cambridge, UK, 2022.

[79] F. Alajaji and P. N. Chen. *An Introduction to Single-User Information Theory*. Springer, Berlin, Germany, 2018.

[80] N. Shlezinger, E. Abakasanga, R. Dabora, and Y. C. Eldar. "The capacity of memoryless channels with sampled cyclostationary Gaussian noise". *IEEE Transactions on Communications*, 68(1):106–121, 2020.

[81] T. Cover and A. E. Gamal. "Capacity theorems for the relay channel". *IEEE Transactions on Information Theory*, 25(5):572–584, 1979.

[82] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski. *Cooperative Communications and Networking*. Cambridge University Press, Cambridge, UK, 2009.

[83] M. Dohler and Y. Li. *Cooperative Communications - Hardware, Channel & PHY*. John Wiley & Sons, Hoboken, NJ, USA, 2010.

[84] R. Dabora and E. Abakasanga. "On the capacity of communications channels with memory and sampled additive cyclostationary Gaussian noise". Submitted to *IEEE Transactions on Information Theory*, 2022.

[85] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. "Blind modulation classification: A concept whose time has come". In *2005 IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, pp 223–228, Princeton, NJ, USA, Apr 2005.

[86] Z. Zhu and A. K. Nandi. *Automatic Modulation Classification: Principles, Algorithms and Applications*. John Wiley & Sons, Hoboken, NJ, USA, 2015.

[87] R. S. Sutton and A. G. Barto. *Reinforcement learning - An Introduction*. MIT Press, Cambridge, MA, USA, 2018.

[88] W. S. Lovejoy. "A survey of algorithmic methods for partially observed Markov decision processes". *Annals of Operations Research*, 28(1):47–65, 1991.

[89] C. M. Miller. "High-speed digital transmitter characterization using eye diagram analysis". *Hewlett-Packard Journal*, 45(4):29–37, 1994.

[90] T. S. Rappaport. *Wireless Communications: Principles and Practices*. Pearson, London, UK, 2001.

[91] R. C. Kizilirmak. "Non-orthogonal multiple access (NOMA) for 5G networks". In H. K. Bizaki, Editor, *Towards 5G Wireless Networks: A Physical Layer Perspective*, Chapter 4. IntechOpen, London, UK, 2016.

[92] M. C. Valenti and X. Wang. "Best readings in non-orthogonal multiple access". Available online: `https://www.comsoc.org/publications/best-readings/non-orthogonal-multiple-access`. (Accessed on 2022-12-22).

[93] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, UK, 2005.

[94] Y.-A. Jung, J.-Y. Kim, and Y.-H. You. "Complexity efficient least squares estimation of frequency offsets for DVB-C2 OFDM systems". *IEEE Access*, 6:35165–35170, 2018.

[95] A. Rotem and R. Dabora. "A novel low-complexity estimation of sampling and carrier frequency offsets in OFDM communications". *IEEE Access*, 8:194978–194991, 2020.