Compressed Sensing-Based Channel Estimation and Prediction for Underwater Acoustic Communications

by

©Yi Zhang, B.Eng., M.A.Sc

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

July 13, 2017

St. John's Newfoundland

Abstract

This thesis develops approaches for estimating and predicting sparse shallow-water acoustic communication channels. The broadband shallow-water channel has three characterizations: a large dimension of channel impulse response caused by excessively long delay spread, fast temporal variability induced by scattering from the moving sea surface, and a sparse channel structure due to the resolvable paths. Traditional least square estimation techniques fail to utilize the sparse channel structure, and suffer from the limitations on the capability of estimating large-dimensional channels with rapid fluctuations.

Compressed sensing, also known as compressive sensing (CS), has been intensively studied recently. It has been applied in various areas such as imaging, radar, speech recognition, and data acquisition. Recently, applying CS to sparse channel estimation has been largely accepted. This thesis details the application of CS to sparse estimation of both time-invariant and time-varying shallow-water acoustic channels. Specifically, various reconstruction algorithms are used to find the sparse channel coefficients. However, *a priori* knowledge of channel sparsity is often not available in practice. The first part of the thesis proposes an improved greedy pursuit algorithm which iteratively identifies the sparse channel coefficients without requiring *a priori* knowledge of channel sparsity. Then, the proposed algorithm is employed to estimate both time-invariant and time-varying sparse channels. In addition, a comparative study of the state-of-the-

art of various CS-based signal reconstruction algorithms is performed to gain better understanding of the mathematical insights.

Furthermore, based on the CS theory, different pilot placement choices will directly affect the performance of the channel estimation algorithm. The second part of the thesis investigates the pilot pattern design in sparse channel estimation. Unlike the equally spaced pilots for conventional channel estimation, randomly placed pilot tones are most used in existing CS-based channel estimation methods. In order to improve the efficiency of the optimal pilot pattern searching, a novel pilot pattern selection scheme is proposed based on the concatenated cyclic difference set. The performance of the proposed design is also compared with the existing search-based pilot placement methods. It should be noted that the proposed reconstruction algorithm and the pilot placement scheme are not restricted to underwater acoustic communication systems, but they can be applied to sparse channel estimation in other communication systems.

Finally, an outdated channel estimation will lead to severe performance degradation when the channel varies rapidly. Hence, to predict future channel state information, an efficient sparse channel prediction scheme is proposed which does not require any statistical *a priori* knowledge of channels and noise. A receiver structure which combines a sparse channel estimator and a decision feedback based adaptive channel predictor is developed to further improve the prediction accuracy. Simulation results are shown to demonstrate the performance of the proposed algorithms and schemes. The study of this thesis contributes to a better understanding of the channel physical constraints on algorithm design and potential performance improvement.

Acknowledgement

This thesis would not have been possible without the support and inspiration of a number of individuals. First and foremost, I wish to express my deepest gratitude and respect to my supervisors Professors Ramachandran Venkatesan, Octavia A. Dobre and Cheng Li. Without their guidance, support and continuous encouragement this thesis would hardly have been completed. They have shown me the way to become an independent researcher through their tremendous experience in research and supervision. I greatly appreciate all their contributions of time, advices and funding which make my Ph.D. experience valuable and stimulating.

My sincere thanks also go to the members of thesis advisory and comprehensive exam committee: Professors Howard Heys and Cecilia Moloney who have generously devoted their time to offer comments towards improving the quality of my work. I would also like to take this opportunity to thank PhD thesis examiners: Professors Hai Jiang from University of Alberta, Howard Heys and Telex Magloire Ngatched Nkouatchah, for their very helpful comments and suggestions in improving the presentation of the thesis.

I gratefully acknowledge the financial support from the Research and Development Corporation (RDC) Newfoundland and Labrador and Natural Science and Engineering Research Council (NSERC). I am also thankful to the Faculty of Engineering and Applied Science, the Department of Electrical and Computer Engineering, the Graduate Student Union, the School of Graduate Studies and the Teaching Assistant's Union

of Memorial University of Newfoundland for their financial support during my Ph.D. program.

I would like to give special thanks to Professor Leonard M. Lye who have shared his knowledge about various nonparametric measures of rank correlation, and his advice has been a valuable input for this thesis. I am also very grateful to Dr. Yemi Omomukuyo for his scientific advice and many insightful suggestions towards the thesis writing. His untiring support and continuous encouragement have been a crucial component in my Ph.D. journey.

I am forever thankful to all the past and present colleagues for lending me their expertise and intuition to my scientific and technical problems. Their friendship, valuable support and continuous encouragement meant a lot to me. Also, I must thank many faculty staffs: Adrian Dobre, Moya Crocker, Colleen Mahoney, and etc. Their kind assistance in a number of ways has been very beneficial for my Ph.D. experience.

Finally, my heartfelt gratitude goes to my loving family who always love, help and support unconditionally. Their unfailing emotional support and endless patience helped me went through hard times and kept me moving forward towards the goal. This journey would not have been possible if not for them, and I dedicate this milestone to them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMP	Approximate message passing
AR-1	First-order autoregressive
AS-SaMP	Adaptive step size SaMP
AUV	Autonomous underwater vehicle
AWGN	Additive white Gaussian noise
BER	Bit error rate
BP	Basis pursuit
C-CDS with TS	Concatenated CDS with an iterative tail search
CDS	Cyclic difference set
CIR	Channel impulse response
CoSaMP	Compressed sampling matching pursuit
CP	Cyclic prefix
CS	Compressive/compressed sensing
CSI	Channel state information

CTD ..... Conductivity-temperature-depth sensors

DDSF ..... Delay-Doppler spreading function

DFE ..... Decision feedback equalization

DFT ..... Discrete Fourier transform

DP ..... Dynamic programming

EKF ..... Extended Kalman filter

EM-BG ..... Expectation maximization-Bernoulli-Gaussian

EWRLS ..... Exponential weighted recursive least square

FIR ..... Finite impulse response

ICI ..... Intercarrier interference

ISI ..... Intersymbol interference

KAM ..... Kauai Acomms MURI Experiment

LDPC ..... Low-density parity-check

LE ..... Linear equalizer

LMS ..... Least mean square

LP ..... Linear programming

LS ..... Least square

LTE ..... Long term evolution

MACE ..... Mobile Acoustic Communications Experiment

MIMO ..... Multiple-input multiple-output

MIP ..... Mutual incoherence property

MMSE ..... Minimum mean-square error

MP ..... Matching pursuit

MRI ..... Magnetic resonance imaging

MSE ..... Mean squared error

NGDC ..... National geophysical data center

NP ..... Non-deterministic polynomial

OFDM ..... Orthogonal frequency division multiplexing

OMP ..... Orthogonal matching pursuit

ORLSMP ..... Order-recursive least square matching pursuit

PS ..... Pacific Storm Experiment

PSK ..... Phase-shift keying

QAM ..... Quadrature amplitude modulation

RF ..... Radio frequency

RIP ..... Restricted isometry property

RLS ..... Recursive least square

RPE-DFE ..... Residual prediction error decision feedback equalizer

SACM ..... Statistical acoustic channel model

SaMP/SAMP ..... Sparsity adaptive matching pursuit

SISO ..... Single-input single-output

SNR ..... Signal-to-noise ratio

SP ..... Subspace pursuit

SPACE ..... Surface Processes and Acoustic Communications Experiment

SpaRSA ..... Separable approximation

SSS ..... Stochastic sequential search

TV-APM ..... Time variable acoustic propagation model

TwIST ..... Two step iterative shrinkage/thresholding

UWA ..... Underwater acoustic

UWB ..... Ultra-wideband

WLAN ..... Wireless local network

Notation

Lower case letters	Scalars
Bold face letters	Vectors or matrices
$\boldsymbol{x}^H$	The Hermitian form of $\boldsymbol{x}$
$\boldsymbol{x}^T$	The transpose form of $\boldsymbol{x}$
$\boldsymbol{x}^*$	The complex conjugate of $\boldsymbol{x}$
$\boldsymbol{x}^\dagger$	The Moore-Penrose pseudo-inverse matrix of $\boldsymbol{x}$
$\|\boldsymbol{x}\|_p$	The p-norm form of $\boldsymbol{x}$ which equals to $(\Sigma\|\boldsymbol{x}\|^p)^{1/p}$
$\hat{x}$	An estimation of x
$\tilde{x}$	A prediction of x
$\langle \boldsymbol{x}\boldsymbol{y} \rangle$	Inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$
$inf(\boldsymbol{x})$	The infimum of $\boldsymbol{x}$ which is the largest real number that is smaller than or equal to every number in $\boldsymbol{x}$
$O(f(x))$	The big O notation which describes the limiting behavior of function $f(x)$ when the argument x tends towards a particular value or infinity

$\arg\min$	The argument of the minimum
$\otimes$	Kronecker product
$\mathbb{C}^N$	The space of $N \times 1$ complex vectors
$\mathbb{R}^N$	The space of $N \times 1$ real vectors
$trace(\boldsymbol{x})$	The sum of the elements on the main diagonal of $\boldsymbol{x}$

List of Symbols

$x(t)$	Transmitted signal in time
$y(t)$	Received signal in time
$X(f)$	Transmitted signal in frequency
$Y(f)$	Received signal in frequency
T	Transmitted signal duration
f_c	Carrier frequency
a	Doppler scaling factor
v	Relative speed between a transmitter and a receiver
c	Acoustic wave propagation speed in water
A	Path loss
l	Transmission distance
l_r	Referential distance
k	Spreading factor
C_a	Absorption coefficient
$H_p(f)$	Transfer function for a time-invariant multipath UWA channel
Γ_p	Cumulative reflection coefficient for the pth path
N_p	Number of multipaths
γ_s	Surface reflection coefficients
γ_b	Bottom reflection coefficients

n_s	Times of surface reflections
n_b	Times of bottom reflections
θ_p	Grazing angle for the pth path
ρ_b	Density of the ground at the bottom of the ocean
ρ	Density of water
c_b	Speed of sound on the ocean floor
l_p	Length of the pth path
$\bar{l}_p$	Length of the nominal path
Δ_{l_p}	Deviation of the pth path from the nominal one
h_p	Gain of the pth path
τ_p	Delay of the pth path
$h_{p,i}$	The ith intrapath gain
$\tau_{p,i}$	The ith intrapath delay
$\Delta_{\tau_{p,i}}$	Deviation of the ith intrapath delay
$\sigma^2_{\Delta_p}$	Variance of $\Delta_{\tau_{p,i}}$
$\gamma_p(f)$	Small-scale coefficient
$H(f,t)$	Transfer function for a time-varying UWA channel
$\bar{H}_0(f)$	Transfer function for the reference path
B_{Δ_p}	3 dB width of the power spectrum density of $\Delta_{\tau_{p,i}}(t)$
$C_{\gamma_p}(\Delta_f, \Delta_t)$	Small-scale correlation function
η_p	Time invariant channel gain for the pth path
τ_p	Time invariant delay for the pth path
$\mathbf{D}$	DFT matrix
$\mathbf{Y_P}$	Received pilot subcarriers
$\mathbf{X_P}$	Transmitted pilot subcarriers
$\mathbf{S}$	Matrix for selecting pilot subcarriers

$\mathbf{W_p}$	Frequency response vector of AWGN corresponding to the pilot subcarriers
σ	Standard deviation of AWGN
Δf	Subcarrier spacing
$\mathbf{C^i}$	Candidate support set in the ith iteration
$\mathbf{F^i}$	Final support set in the ith iteration
$\mathbf{r^i}$	Residual vector in the ith iteration
s_I	Initial step size in the SaMP and AS-SaMP algorithm
s	Step size in the SaMP and AS-SaMP algorithm
K	The sparsity level of the target signal
Γ	The threshold in the AS-SaMP algorithm
ϵ	The tolerance in the halting conditions of the considered MP-type algorithms
J	The total number of stage in AS-SaMP
δ_{3K_i}	The RIP parameter in the ith stage
C_{K_i}	The RIP constant depends on δ_{3K_i}
K_i	The size of final support size in the ith stage
μ	Mutual coherence of the measurement matrix
Ω	The set of pilot location indices
E_p	Power of the pilot symbols
$\mathbf{Q}$	A cyclic difference set
$\mathbf{G}$	The list of difference for a cyclic difference set $\mathbf{Q}$
λ_g	Repetition times of g in $\mathbf{G}$
$r_s(\mathbf{A}, \mathbf{B})$	The Spearman's rank correlation
$h(k, n)$	Response of channel at time n to a unit impulse input at time $n - k$
$u(l, k)$	DDSF for the lth sampled Doppler frequency and the kth sampled delay
M	Training symbol length
f_s	Sampling frequency

ν	Doppler shift
$\Delta\nu$	The Doppler sample interval
$\mathbf{h_{lin}}$	Finite impulse response of a linear equalizer
$\mathbf{h_{ff}}$	Feedforward filter coefficients
$\mathbf{h_{fb}}$	Feedback filter coefficients
L_f	Length of $\mathbf{h_{lin}}$ and $\mathbf{h_{ff}}$
L_{fb}	Length of $\mathbf{h_{fb}}$
$\mathbf{R_w}$	Covariance matrix of the noise
$\mathbf{R_e}$	Effective noise correlation matrix
L_e	Number of time instances to average the correlation matrix of the RPE
ϵ_s	Soft decision error
λ	Forgetting factor in the EWRLS algorithm
$\mathbf{D}(b,j)$	The prediction coefficients for the bth block and jth DDSF element
N_b	The total number of data blocks
$\mathbf{K}(b,j)$	The RLS gain vector
$\mathbf{P}(b,j)$	The inverse correlation matrix of the input signal
d	Prediction horizon
T_c	Coherence time of the large-scale variation
A_w	Amplitude of the displacement caused by surface save
f_w	Frequency of the displacement caused by surface wave
σ_s	Standard deviation of the surface variation
σ_b	Standard deviation of the bottom variation

Chapter 1

Introduction

The underwater acoustic (UWA) channel is commonly considered as one of the most challenging communication channels. In the past two decades, we have witnessed significant advances in the development of high data UWA communications for an increased number of both military and commercial applications. Extensive research on communications in underwater environments is still underway. The success of such research depends on a thorough understanding of the physical constraints of UWA channels so as to develop effective signal processing techniques based on these constraints. The main focus in this thesis is on channel estimation for acoustic communications in shallow-water environments, of which the distinct challenges and features will be discussed in the following section.

1.1 Challenges in UWA communications

UWA communications represents techniques of sending and receiving messages below water using sound waves [2,3]. The past two decades have witnessed significant advances towards the development of UWA communications. In the underwater environment, the possibility of maintaining signal transmission enables data gathering from submerged

instruments without human intervention. The scope of UWA communications, which was initially limited to military applications, has now expanded to diverse commercial fields, with applications like surveillance for environmental systems, information collection for oceans, lakes and rivers, underwater resource exploration, and remote control in off-shore oil industry. The most common way of employing UWA communications is by using hydrophone, which converts the electromagnetic signal into a sound signal. In fact, the acoustic waves are not the only means for underwater wireless communications. However, due to the conductive sea water, electromagnetic waves that are used in radio frequency (RF) bandwidth experience high attenuation. Optical waves, which have lower attenuation than RF signals, are not suitable for long distance underwater wireless communication because they suffer from severe scattering and absorption in water. Therefore, in comparison with RF and optical waves, acoustic waves remain the primary carrier in underwater wireless communication applications, despite its low propagation speed and extremely limited bandwidth[1] [3].

In the past few years, significant techniques in UWA communications have been developed to improve the performance in terms of operational range, data throughput and error rate. These advances have made possible a 100 to 1000 fold increase in achievable data rates. As efficient communication systems have been developing, both military and commercial industries are now calling for real-time communications with submarines and autonomous underwater vehicles (AUVs)[2], not only for point-to-point links, but also for networks consisting of multiple mobile and stationary nodes. These types of networks will provide exchange of various data types ranging from control to

[1]The typical frequency range of an acoustic link is between 10 and 15 kHz. Though the total communication bandwidth is very low, the UWA system is inherently wideband in the sense that the occupied bandwidth is on the order of its center frequency [3].

[2]Unmanned underwater vehicles or AUVs, which are equipped with a wide variety of sensors and sonar systems, are widely used in UWA communications to perform missions like underwater survey, e.g., detecting and mapping submerged wrecks, rocks, and obstructions that pose a hazard to navigation for commercial and recreational vessels.

video signals, between AUVs and underwater robots which are equipped with various sensors, sonars, or cameras. Towards achieving these goals, researchers are currently focusing on the design of efficient modulation and coding schemes, and development of signal processing techniques for mobile underwater environments.

The UWA channel is considered as one of the most challenging communication media because of its high time-frequency selectivity [4]. When compared with terrestrial radio channels, UWA channels exhibit larger time dispersions (on the order of hundreds of milliseconds). As such, they could cause severe intersymbol interference (ISI), which requires sophisticated and computationally expensive equalization techniques [5–7]. Orthogonal frequency division multiplexing (OFDM), which is employed in several terrestrial radio communications standards, such as wireless local network (WLAN) and long term evolution (LTE) [8], is also considered for UWA communications due to its high data transmission rate, efficient spectral utilization and ability to cope with multipath fading [9]. As UWA channels have been regarded significantly different from wireless radio channels, due to the unique characteristics, such as abundance of transmission paths, large temporal variations, and wideband property in nature. To highlight the challenges of OFDM communications over UWA channels relative to those over wireless radio channels, three example systems are compared. Table 1.1 provides a comparison of parameters in these three systems, which are OFDM UWA communications [9], WLAN using IEEE 802.11a/g [10] and ultra-wideband (UWB) communication systems [11].

The major challenges noticed from the table are as follows:

1. Multipath delays in UWA communication systems are much larger than those in WLAN and UWB systems. Typical multipath spreads in UWA channels increase a thousandfold when compared to the spreads in commonly used radio channels, which leads to strong frequency selectivity. This has motivated the use of vari-

Table 1.1: Comparison of parameters for UWA, WLAN, and UWB systems

	UWA [9]	WLAN [10]	UWB [11]
Propagation speed	1500 m/s	3×10^8 m/s	3×10^8 m/s
Bandwidth	12 kHz	20 MHz	528 MHz
Carrier frequency	27 kHz	5.2 GHz	$3 - 10$ GHz frequency hopping
Bandwidth / Carrier frequency	0.44	0.0038	$0.0176 - 0.0528$
Doppler scaling factor a [1](waveform time compression) for moving terminals with a relative speed v	1.3×10^{-3} for $v = 2$ m/s	7×10^{-8} for $v = 20$ m/s	7×10^{-8} for $v = 20$ m/s
Typical multipath spread	$\sim 10 - 100$ ms	~ 500 ns	~ 100 ns
OFDM symbol duration	$\sim 50 - 150$ ms	4 μs	0.3 μs

[1] Doppler scaling factor $a = v/c$, where c is the wave propagation speed.

ous techniques at the receiver, such as decision feedback equalization (DFE) and multicarrier transmission schemes to cope with this multipath fading.

2. A major difference between UWA and RF wireless communications lies in the fast variability of the nature of the underwater acoustic medium. Time-variability of the UWA channels is caused by inherent environmental variations and changes occurred due to transmitter and receiver motion. As such, the significant Doppler effect becomes the most challenging characteristic of the UWA channels, especially in shallow water. The relative motion between a transmitter and a receiver results in changes in the signal duration and frequency, as shown in Figure 1.1. Particularly, the transmitted signal $x(t)$, whose duration is T, is modulated onto a carrier of frequency, f_c. $y(t)$ represents the received signal with the frequency response $Y(f)$. It can be observed that the changes include frequency (time) shift and frequency (time) spreading. For frequency shift, an offset, af_c, is introduced, while for frequency spreading, the bandwidth of the signal is scaled by the

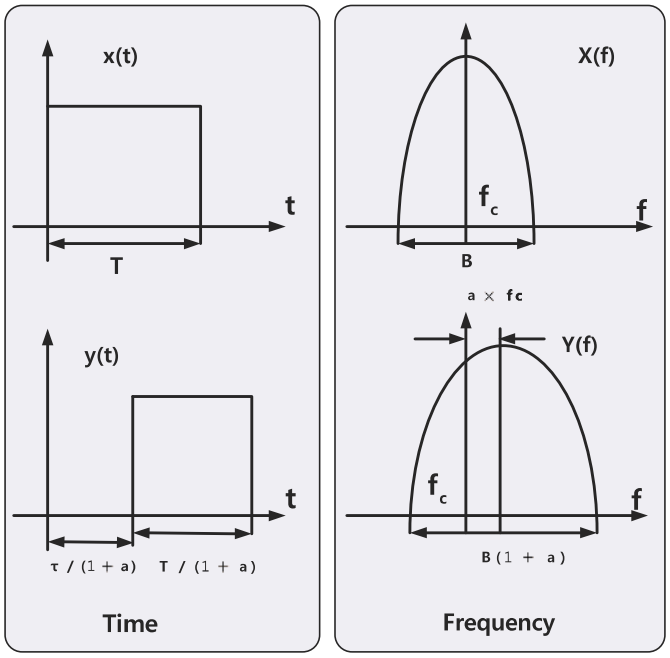


Figure 1.1: Motion-induced changes in the signal duration and frequency [12].

factor of a (also referred to as the Doppler scaling factor [12]). In other words, the distortion is proportional to the ratio between the relative speed between the transmitter and receiver and the propagation speed. For comparison, the Doppler scaling factor in a highly mobile radio system with a relative speed of more than 100 km/h is on the order of 10^{-7}. In contrast, for an AUV in an acoustic system moving at several meters per second (some submarines can move at greater velocities), the factor is on the order of 10^{-3} [13]. Furthermore, in spite of the intended motion-induced Doppler effects, different propagation paths can have different Doppler scaling due to the medium instability, e.g., the current induced platform instability, and the random displacement of surface reflecting points caused by waves. As a result, the Doppler scaling which indicates the amount of frequency (time) compression or expansion of the signal at the receiver side, needs to be considered for UWA communications.

3. The aforementioned challenges in UWA channels require sophisticated underwater

acoustic system designs which can be computationally complex from a practical implementation point-of-view. For instance, the computational complexities involved in dealing with ISI in UWA systems, which extends over several ten to hundreds of symbols, renders many of the methods used in RF wireless systems impractical. As such, more recent studies have focused on receiver algorithm and structure designs with reduced complexity.

1.2 Motivation

As previously mentioned, the UWA channel is characterized by a long delay spread and a significant Doppler effect [3, 9, 12, 13]. Various receiving techniques, e.g., DFE, have been found useful to combat severe ISI caused by the excessively long delay in single-carrier modulation systems [14–22]. In multicarrier modulation like OFDM, the cyclic prefix (CP) is used to compensate for the ISI effect to eliminate the need for a complex equalizer [3, 6, 7, 9, 23–30]. In coherent digital wireless systems, obtaining accurate estimates of the channel state information (CSI) is critical at the receiver [24, 31–33]. The data-aided channel estimation in OFDM communication systems can be performed by either inserting pilot tones into certain subcarriers of each OFDM symbol, or by using all subcarriers as pilots within a specific period [34]. Recently, studies have suggested that many UWA multipath channels tend to exhibit a sparse structure in the sense that the majority of the channel impulse response (CIR) taps end up being either zero or below the noise floor [3, 35, 36]. Conventional methods for CSI estimation, such as least square (LS) and minimum mean-square error (MMSE) [37], cannot exploit the sparsity of the wireless channels and they often lead to excessive utilization of spectral and energy resources. On the other hand, channel estimation exploiting the sparsity of channels reduces the required number of pilots, and thus,

effectively improves the spectral and energy efficiency [3, 23, 27, 29, 35, 38, 39].

More recently, advances in the field of compressed/compressive sensing (CS) [40–42] have led to numerous applications such as imaging, radar, data acquisition, and communications [28]. In communications, it is largely accepted that CS can be applied to sparse channel estimation [3, 28, 35]. One of the most noteworthy results in CS theory[3] is that the sparse channel coefficients can be reconstructed using optimization algorithms, and thus a proper reconstruction/recovery algorithm is crucial to obtaining an accurate or even exact reconstruction of the target signal. Especially, among those reconstruction algorithms, the matching pursuit (MP) algorithm and its variants are favoured in single-carrier systems [43] and multicarrier systems [29, 44, 45] because of their low computational complexity. Besides the reconstruction algorithms, the performance of CS-based channel estimation also strongly depends on the design of the measurement matrix. This issue has led to a large amount of work in CS-based sparse channel estimation for the OFDM systems focused on the design of pilot patterns with tractable recovery complexity. It should be noted that such a pilot placement problem can be viewed as a search problem for the optimum CS measurement matrix which minimizes the errors in channel estimation. Although an exhaustive search guarantees the optimum pilot pattern, it is impractical in real applications due to the prohibitively high complexity. Motivated by the facts stated above, the work in this thesis focuses on designing a reliable and efficient reconstruction algorithm as well as an efficient pilot placement scheme for CS-based sparse channel estimation.

The successful application of CS-based signal reconstruction algorithms to the sparse time-invariant channel estimation demonstrated an improved performance due to the reduced number of taps to be estimated [26, 30]. In order to combine the sparse structure and rapid fluctuations, a time-varying UWA channel can be represented by its

[3] An overview on CS theory can be found in Chapter 2.

delay-Doppler spreading function (DDSF) [46], which tends to be sparse because the multipath and Doppler effects are limited to a few dominant paths [43, 47]. An advantage of the DDSF representation is that the time-varying channel estimation can have reduced computational complexity, memory storage requirement, and more importantly, can track the first-order channel dynamics without the need for explicit dynamic modeling [43]. On the other hand, time-varying multipath fading and an extremely limited bandwidth are the main constraints on the UWA communication data rate. Consequently, the tremendous growth in demand for achievable throughput in UWA communications has created the need for adaptive transmission techniques including adaptive modulation, adaptive coding, adaptive power control, etc. [48, 49]. Thus, not only the receiver requires the up-to-date CSI for reliable symbol detection, but also the transmitter requires the fedback CSI from receiver to adaptively adjust modulation, coding, and power. Considering the feedback and related processing delays, the future CSI needs to be predicted accurately. Especially, when the channel varies rapidly, an outdated channel estimation will lead to severe performance degradation, and thus result in the failure of using those adaptive transmission techniques. To this end, applying CS into sparse channel estimation approach at the receiver side, which can track time-varying channels, and developing a reliable and efficient sparse channel prediction scheme, which can predict future CSI based on current channel estimates, are both desired.

1.3 Thesis Contributions

The main contributions of this thesis are summarized as follows:

- **1**- An overview of UWA channel modeling is provided in chronological order, which consists of three representative models. In addition, a survey of the state-

of-the-art of different UWA channel simulation software has been conducted. In order to obtain UWA channel realizations, a recently-proposed channel simulator, which incorporates acoustic signal propagation laws and random environmental effects, is employed with a collection of real experimental data.

- **2**- An improved reconstruction algorithm for CS-based time-invariant sparse channel estimation in UWA-OFDM systems is proposed, and its theoretical performance analysis is performed. Without requiring *a priori* knowledge of the sparsity, the proposed algorithm adjusts the step size adaptively to approach the true sparsity, thereby improving the estimation accuracy. Simulation results are shown to demonstrate a better trade-off of the proposed algorithm between the mean squared error (MSE)/ bit error rate (BER) performance and complexity when compared with the conventional LS estimation method and other published MP-based algorithms. The theoretical analysis is provided to show the reconstruction performance of the proposed algorithm.

- **3**- Comparative performance analyses of the existing CS-based signal reconstruction methods and the state-of-the-art of different pilot placement schemes, in terms of estimation accuracy and computational complexity, are presented.

- **4**- A novel efficient scheme for near-optimal pilot placement is developed to meet the requirement of the measurement matrix for a satisfactory reconstruction. Simulation results show that the proposed pilot allocation scheme with the proposed reconstruction algorithm provides the best MSE and BER performance among all the other considered pilot schemes and reconstruction algorithms. It should be noted that the proposed reconstruction algorithm and the pilot placement scheme are not restricted to underwater acoustic communication systems, but they can also be applied to sparse channel estimation in other communication systems

(cellular, WiFi, etc.).

- **5**- An adaptive sparse channel prediction is developed which does not require any statistical *a priori* knowledge of channel and noise. The proposed channel prediction can operate in both normal and decision feedback modes. The normal mode indicates that prediction is made solely based on the current estimated channel, while the decision feedback mode means that prediction is obtained using both the estimated channel and the detected symbols. Numerical results show that the decision feedback prediction scheme, along with the proposed CS-based channel estimation algorithm, offers a good error performance without a significant increase in computational complexity.

1.4 Thesis Organization

In this thesis, based on the preliminary discussion in the previous sections, we identify the challenges and motivations related to CS-based sparse channel estimation and channel prediction for UWA communications. The algorithms and results presented in Chapter 3-6 have been disseminated in peer-reviewed publications. Following the introduction, which is presented in this chapter, the organization of this thesis is as follows:

- Background knowledge on CS theory and the literature review are provided in Chapter 2.

- An overview of UWA channel modeling and a survey of several representative simulation software packages are presented in Chapter 3. This was published in proceedings of International Wireless Communications and Mobile Computing Conference (IWCMC) [50].

- Chapter 4 proposes a new reconstruction algorithm for sparse time-invariant channel estimation. A comparative analysis with the existing reconstruction algorithms and simulation results show the improved performance of the proposed algorithm. This research was published in proceedings of IEEE on Wireless Communications and Networking Conference (WCNC) [26].

- Chapter 5 addresses the pilot placement problem described in the previous section. Specifically, a novel pilot placement scheme is proposed for a near-optimum sparse signal reconstruction. The material included in Chapters 5 and 6 was published in the IEEE Transactions on Wireless Communications [30].

- Chapter 6 deals with the CS-based time-varying sparse channel estimation and channel prediction. Particularly, an adaptive sparse channel prediction scheme is developed and evaluated along with the previously proposed channel estimation algorithm. This work is under review for publication.

- Chapter 7 consists of concluding remarks and potential future research directions based on the results of the thesis.

Chapter 2

Background

This chapter provides an overview on the fundamental CS theory, which emerged mainly in the works [3, 40–42, 51, 52], and presents its key mathematical underpinnings. Furthermore, the existing research papers in the related fields of channel estimation, pilot placement for CS-based sparse channel estimation, and channel prediction for UWA communications are reviewed.

2.1 CS Fundamentals

CS is a novel paradigm for efficiently acquiring and reconstructing a signal which is sparse in a pre-defined transform domain. The underlying principle of CS lies in finding the conditions under which the sparsity of a signal can be exploited to recover the signal from far fewer samples. Two basic conditions for a possible recovery are known as the existence of a sparse signal representation and incoherence between the sensing and the transforming process which achieves the sparse representation. In this section, four fundamental components in CS will be investigated.

2.1.1 The Sensing Problem

Consider a real-valued, one-dimensional, discrete signal $\mathbf{x} \in \mathbb{R}^N$ which can be represented by an $N \times 1$ vector. CS directly acquires a low-dimensional measurement $\mathbf{y} \in \mathbb{R}^M$ ($M \ll N$) by computing M inner products between $\mathbf{x}$ and a collection of sensing vectors $\{\boldsymbol{\phi}_i\}_{i=0}^{M-1}$, as shown below:

$$y_i = \langle \mathbf{x} \cdot \boldsymbol{\phi}_i \rangle. \tag{2.1}$$

Let each row of the sensing matrix $\boldsymbol{\Phi}$ be the sensing vector $\boldsymbol{\phi}_i^T$. Then, the matrix form of (2.1) is written as below

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}. \tag{2.2}$$

Equation (2.2) describes a general sensing process. For example, if the sensing bases are Dirac delta functions, then $\mathbf{y}$ is a vector of sampled values of $\mathbf{x}$. For some applications, such as neutron scattering imaging process, magnetic resonance imaging (MRI), etc., the sensing process may be very expensive or time-constrained. This new sensing paradigm can be efficient and useful. Admittedly, one needs to solve (2.2) which is an underdetermined system, and thus, raises important questions. Is it possible to accurately reconstruct $\mathbf{x}$ from fewer measurements $\mathbf{y}$? How would a sensing matrix be designed in order to capture almost all information about $\mathbf{x}$?

2.1.2 Signal Sparse Representation

Many natural signals are sparse or compressible in the sense that they have concise representations when expressed in certain bases. Mathematically speaking, the signal $\mathbf{x}$

of length N can be expanded in an orthonormal basis $\mathbf{\Psi} = [\boldsymbol{\psi}_0 \boldsymbol{\psi}_1 \cdots \boldsymbol{\psi}_{N-1}]$ as follows:

$$\mathbf{x} = \sum_{j=0}^{N-1} \alpha_j \boldsymbol{\psi}_j = \mathbf{\Psi} \boldsymbol{\alpha}, \tag{2.3}$$

where $\boldsymbol{\alpha}$ is the $N \times 1$ coefficients vector and $\boldsymbol{\psi}_j$ is a $N \times 1$ vector. $\mathbf{x}$ and $\boldsymbol{\alpha}$ are equivalent representations of the target signal in the time or space domain and the $\mathbf{\Psi}$ domain, respectively. The signal $\mathbf{x}$ is K-sparse if only K coefficients ($K \ll N$) in $\boldsymbol{\alpha}$ are non-zero and the remaining $N - K$ are zero. Moreover, the signal $\mathbf{x}$ is compressible if $\boldsymbol{\alpha}$ contains a few significant entries and many insignificant entries. As is well-known, most lossy compression techniques such as JEPG, MEPG, or MP3 rely on such sparse structures in the coefficients on a certain transform domain of the signal. Considering the observation noise $\mathbf{n}$, which is zero mean and complex Gaussian distributed with variance σ, and substituting $\mathbf{x}$ in (2.2) with (2.3), the signal acquisition is written as below:

$$\mathbf{y} = \mathbf{\Phi} \mathbf{\Psi} \boldsymbol{\alpha} + \mathbf{n} = \mathbf{A} \boldsymbol{\alpha} + \mathbf{n}, \tag{2.4}$$

where $\mathbf{A}$ is referred to as a *measurement matrix*. Fig. 2.1 shows the compressive sensing process for an example noiseless case. Here, the signal $\boldsymbol{\alpha}$ is sparse with $K = 4$. There are four columns in $\mathbf{A}$ corresponding to the non-zero entries in $\boldsymbol{\alpha}$, thus $\mathbf{y}$ is a linear combination of these four columns[4]. The novelty in CS is that with the knowledge of the sparse structure of $\boldsymbol{\alpha}$, finding an estimator $\hat{\boldsymbol{\alpha}}$, whose estimation error is proportional to σ becomes possible [41, 42].

2.1.3 Measurement Matrices

Another premise which enables CS to accurately recover the target signal is the incoherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$. We can define the coherence between the sensing basis $\mathbf{\Phi}$

[4]In Fig. 2.1, different colors are randomly chosen for different elements in the vectors and matrix.

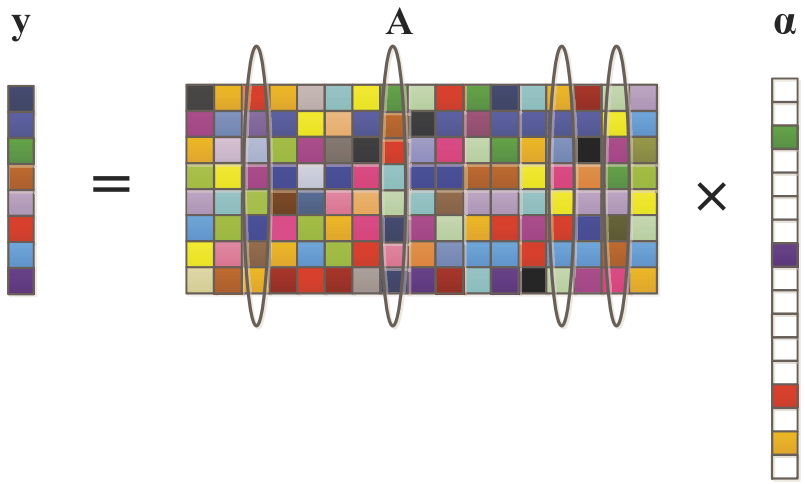


Figure 2.1: Compressive acquisition with a measurement matrix $\mathbf{A}$.

and the representation basis $\mathbf{\Psi}$ as follows:

$$\eta(\mathbf{\Phi}, \mathbf{\Psi}) = N \max_{i,j} |\langle \boldsymbol{\phi}_i \boldsymbol{\psi}_j \rangle|^2. \tag{2.5}$$

CS is applicable with small η (low coherence) which requires that $\boldsymbol{\phi}_i$ can not sparsely represent $\boldsymbol{\psi}_j$ and vice versa. An intuitive explanation of this can be revealed through a counterexample. If we construct $\mathbf{\Phi}$ in coherence with $\mathbf{\Psi}$, for example, extracting the first M columns of $\boldsymbol{\psi}$ as the rows in $\mathbf{\Phi}$, then from (2.4) $\widehat{\boldsymbol{\alpha}} = [\mathbf{y_0}, \mathbf{y_1}, \cdots, \mathbf{y_{M-1}}, 0, \cdots, 0]^T$. Obviously, this is probably a wrong estimation as we assume the first M elements of $\boldsymbol{\alpha}$ are significant. However, the positions of K significant entries are unknown and are different for different signals. Hence the probability to find the $\mathbf{A}$ that corresponds the positions of K significant entries is $1/\binom{N}{K}$. $\mathbf{\Phi}$ is expected to be extremely incoherent with $\mathbf{\Psi}$.

Additionally, the linear measurements should preserve the stable energy of the target signal. One key property used in demonstrating the reliability of a CS reconstruction process is referred to as *restricted isometry property*(*RIP*) [42], which is defined as follows:

Definition 1. RIP: Define ℓ_p norm of the vector $\mathbf{x}$ as $(\|\mathbf{x}\|)_p = \sum_{i=1}^{N} |x_i|^p$, and $\|\mathbf{x}\|_2 = \|\mathbf{x}\|$ for simplicity. A matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ satisfies the RIP of order $K \leq M$, and $\mathbf{A_I}$ is a submatrix of $\mathbf{A}$ consisting of the columns associated with the index set $\mathbf{I}$. The RIP constant $\delta_K \in (0, 1)$, if for $\mathbf{I} \in (0, N)$ such that $|\mathbf{I}| \leq K$ and for all $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathbf{I}|}$, it holds that

$$(1 - \delta_K) \parallel \boldsymbol{\alpha} \parallel^2 \leq \parallel \mathbf{A_I}\boldsymbol{\alpha} \parallel^2 \leq (1 + \delta_K) \parallel \boldsymbol{\alpha} \parallel^2 .$$

For a K-sparse signal $\boldsymbol{\alpha}$, δ_K is defined as:

$$\delta_K := \inf(\delta : (1 - \delta) \parallel \boldsymbol{\alpha} \parallel^2 \leq \parallel \mathbf{A_I}\boldsymbol{\alpha} \parallel^2 \leq (1 + \delta) \parallel \boldsymbol{\alpha} \parallel^2, \quad \forall |\mathbf{I}| \leq K),$$

where inf represents the infimum of the set δ which satisfies the inequality above, and δ represents the set of RIP parameters δ of order K.

Assuming that the estimated $\hat{\boldsymbol{\alpha}}$ has the least distance to the measurements, which means $\|\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\alpha}}\|^2$ is minimum, and based on the fact that the estimation error, i.e., $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$ is at most $2K$-sparse, the estimation error is bounded by

$$\mathbb{E}\{\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\|^2\} \leqslant \frac{2N\sigma^2}{1 - \delta_{2K}}$$

From above equation, it can be easily seen $\delta_{2K} < 1$ is required for a possible exact recovery. As $\delta_K \leqslant \delta_{K+1}$, it is commonly believed that $\mathbf{A}$ satisfies the RIP if δ_K is not too close to 1. In other words, $\mathbf{A}$ which satisfies the $2K$-RIP is able to approximately preserve the distance between any pair of K-sparse vector; this has fundamental implications concerning robustness to noise [3].

The RIP property can be used to theoretically evaluate the reliability of a certain measurement matrix. Unfortunately, the RIP evaluation for a particular matrix is a hard problem with non-deterministic polynomial (NP) complexity. Instead, a widely-

used framework for measurement matrix evaluation is the mutual incoherence property (MIP) introduced in [51], which requires the mutual coherence of the measurement matrix to be smaller than a threshold[5]. An underlying principle in MIP is previously discussed in 2.1.2. It should be noted that the MIP condition is stronger than the RIP, meaning that, while a matrix that satisfies the MIP will also satisfy the RIP, the reverse may not be true. Surprisingly, most random matrices, such as matrices with Gaussian or Bernoulli entries or randomly selected rows of a discrete Fourier transform (DFT) matrix, have been proven to obey RIP with high probability with a sufficient number of measurements. Reference [40] shows in detail that for a measurement matrix which are composed of random rows in an $N \times N$ DFT matrix, if $M > C_\delta K \log N$, any K-sparse signal can be reconstructed with a probability of at least $1 - O(N^{-\delta})$, where C_δ is approximately linear with δ.

2.1.4 Sparse Signal Reconstruction

Previously, we considered conditions under which the ill-posed problem in (2.4) becomes a well-posed one in which there exists an accurate, and sometimes exact, estimate of $\hat{\boldsymbol{\alpha}}$. The goal of reconstruction algorithms is to solve (2.4) normally by minimizing ℓ_0, ℓ_1, or ℓ_2 over the solution space. The conventional way to solve inverse problems is to find the vector with the least energy, i.e., the ℓ_2 minimization. It is well-known that the closed-form solution of the ℓ_2 minimization is $\hat{\boldsymbol{\alpha}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$. However, the solution of the ℓ_2 minimization can hardly be sparse, yielding an inaccurate estimate. In fact, ℓ_0 minimization aims to find the sparsest solution in the feasible solution set, and can recover the sparse signal exactly with high probability using only $(K+1)$ inde-

[5]For the acquisition process expressed in (2.4), the mutual coherence of the measurement matrix which measures the maximum correlation between any two columns has been widely used for analyzing the CS recovery performance. Particularly, [52] has shown that it is sufficient to exactly recover a K-sparse signal from the noiseless measurements if the mutual coherence is smaller than $\frac{1}{2K-1}$

pendent and identical distributed (i.i.d.) Gaussian measurements [53, 54]. We consider the noiseless situation for simplicity. The ℓ_0 optimization can be written as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min \parallel \boldsymbol{\alpha} \parallel_0, \; s.t. \, \mathbf{y} = \mathbf{A}\boldsymbol{\alpha}. \tag{2.6}$$

However, (2.6) is an NP-hard problem and computationally infeasible. Therefore, alternatives have been proposed to solve the problem: linear programming (LP) and dynamic programming (DP) [42]. Basis pursuit (BP), as a representative of LP algorithms, is used to solve the optimization problem formulated below:

$$\hat{\boldsymbol{\alpha}} = \arg\min \parallel \boldsymbol{\alpha} \parallel_1, \; s.t. \, \mathbf{y} = \mathbf{A}\boldsymbol{\alpha}. \tag{2.7}$$

Equation (2.7) is a convex optimization problem. It should be noted that under a particular condition the LP-type algorithms produce the same result of ℓ_0 minimization algorithms. The condition considers the minimum number of measurements M to perfectly reconstruct K-sparse non-zero coefficients with high probability given a certain measurement matrix. Solving (2.7) is computationally expensive ($\mathcal{O}(N^3)$) and is not suitable for real-time applications. Other convex ℓ_1-norm relaxation algorithms have been previously reported with modified cost functions and constrain conditions [55–60]. Recently, the principles of message passing and belief propagation have also been applied to sparse signal reconstruction [61, 62]. More details about these algorithms are presented in the next section.

Another type of approach, based on DP, aims to heuristically identify the K significant coefficients and solve the resulting constrained least-square problem at the cost of requiring slightly more measurements. The DP-based algorithms are favoured over those LP-based ones because the DP-based algorithms can be easily implemented and have low computationally complexity. The representative algorithms of this type are

greedy algorithms,[6] like orthogonal matching pursuit (OMP) [52]. The underlying idea of these algorithms is to find the strongest component in the measurement signal, remove it from the signal and search for the next strongest component which is in the residual signal. This procedure is repeated until the residual signal contains the insignificant information. Recently, variants of OMP like order recursive least square matching pursuit (ORLSMP) [43], compressed sampling matching pursuit (CoSaMP) [63] and sparsity adaptive matching pursuit (SaMP) [64][7] have been reported. This thesis focuses on MP-based algorithms due to their popularity in real applications, and presents a review of these sparse recovery algorithms in the next section.

2.2 Literature Review

In this section, previous work related to the thesis from the aspects of channel estimation, signal reconstruction algorithms, pilot placement schemes, and channel prediction is reviewed.

2.2.1 Channel Estimation for UWA Communication

The last two decades have witnessed many fast-growing advances in the research development of UWA communication analog and digital, non-coherent and coherent systems. However, the technology evolution remains challenging due to the unique acoustic propagation properties and environment in water and an increasing number of intelligent systems deployed in UWA applications. The complex medium characteristics such as the long multipath delay spread, the significant Doppler effects, and the time-varying channel property present challenges in designing high data rate UWA

[6] Greedy algorithms are a special case of DP algorithm under certain circumstances.

[7] The name SAMP was originally used by the authors who proposed the algorithm in [64]; we employ the abbreviation SaMP instead of SAMP in order not to be confused with the AMP (approximated message passing) algorithm.

communication systems [9, 12, 13, 65]. In coherent digital wireless systems, after passing through a channel, the effect of the channel on the transmitted signal must be estimated in order to recover the transmitted information [31]. Therefore, obtaining accurate estimate of CSI is integral and critical to the receiver design.

Extensive research efforts have been made toward developing reliable channel estimation techniques in UWA communication systems. Channel estimation for single-carrier UWA communication systems is carried out to obtain CSI in the time or frequency domain [20, 66, 67]. Specifically, a small training block is attached at the front of a data package for initial estimation and the CSI is re-estimated using a portion of the detected data and the received data for the next data package. In [66], the time-invariant channel impulse responses were estimated by eigenvalue decomposition algorithm. Later, a frequency-domain channel estimation was employed in [67] to cope with the frequency-domain equalization scheme. However, the above methods did not exploit the inherent sparse features of the UWA channels. Sparse estimation was also investigated for the single-carrier UWA systems [68–71]. The common strategy is using approximation schemes to solve the non-linear optimization problem which minimizes the MMSE as a function of the gain and delay location of the significant taps. In [68], an adaptive delay filter was used to estimate the delay values, and then the corresponding gains were sequentially calculated. Afterwards, a sparse DFE was proposed in [69], where the full-size equalizer coefficients (for the feedforward and feedback filters) were determined from an adaptive channel estimation firstly, then only the coefficients corresponding to the significant taps (the gains of the tap is greater than a pre-defined threshold) were kept. Similarly in [71], locations of the dominant taps were detected through blended least square algorithm[8], and then low-order filters were used to compute those tap gains.

It is noted that numerous work including conceptual system analysis, simulation-

[8]The name "blended least square" is adopted as the algorithm blends the LS estimation with the correlation and thresholding based estimation methods [71].

based studies [24, 72], and experiment-based investigations [9, 73, 74] suggested that OFDM is an appealing choice for high-rate transmission over UWA channels. Channel estimation techniques for OFDM based systems can be grouped into two main categories: blind and data-aided or otherwise referred to as pilot/training-based [75]. The former approach exploits the statistics of the received signals and requires higher computational complexity of receivers [76]. In the latter approach, channel estimation can be achieved by inserting pilot signals in the time or/and frequency domain. Two widely used methods for obtaining the estimated channel are LS [37,77] and MMSE [37,78,79]. Generally, the MMSE estimators provide better performance, in terms of MSE, than the LS estimators at the cost of higher complexity, provided that the MMSE receivers have knowledge of the second-order channel statistics [80].

Considering the inherent sparsity in the impulse response of the UWA channels due to the large number of delay taps and the small number of significant paths, sparse channel estimation can offer a better estimation performance than the conventional LS and MMSE methods [3, 23, 27, 29, 35, 38, 39]. As an emerging theory, CS has been successfully applied in reconstructing a sparse channel from fewer pilot symbols [3,26,28, 30,35,47,81]. Basically, CS allows accurate reconstruction of the signal which is sparse on a certain basis, from a small number of random linear projections/measurements [42]. To ensure an accurate or even exact reconstruction of the target signal, a proper reconstruction algorithm and a properly designed measurement matrix are essential, and are reviewed in the following sections.

2.2.2 Existing Signal Reconstruction Algorithms

Existing algorithms to recover a target sparse signal are generally grouped in two categories: LP and DP. On one hand, the BP method in LP achieves a good MSE performance; however, its high computational complexity makes it less attractive to real large-

scale applications. Several variations of BP (e.g., two step iterative shrinkage/thresholding (TwIST) [57], separable approximation (SpaRSA) [58], and YALL1 [59,60]) were reported to solve unconstrained ℓ_2-ℓ_1 optimization problems where the minimizing cost function is a weighted sum of a least-squares term with ℓ_1-norm regularization to yield sparse solutions with a comparable computational complexity to the DP-type algorithms. Recently, the approximate message passing (AMP) algorithm [61] and its variants, e.g., generalized AMP [82] and expectation maximization (EM)-Bernoulli-Gaussian (BG)-AMP [83], were reported to achieve the reconstruction performance[9] almost identical to the LP-type methods, while requiring less computational efforts [61]. However, the performances of these algorithms were evaluated only for certain statistics of the target signal, elements of measurement matrix, and noise. For example, the performance of EM-BG-AMP relies on the BG distribution of the target signal.

On the other hand, algorithms based on DP have been widely adopted due to the ease of implementation and low computational complexity. The OMP algorithm [52] is the most popular algorithm in DP [28, 29]. An OMP variant, namely ORLSMP, was introduced to estimate sparse UWA time-varying channels in [43]. Meanwhile, another OMP variant, referred to as CoSaMP was proposed in [63], with the MSE performance close to that of the BP algorithm. However, CoSaMP requires knowledge of the channel sparsity level, which is often not available in practical applications. Although OMP can be further improved to work in the cases where the sparsity level is not known [84], the MIP [51], which ensures the exact recovery in the bounded noise cases, cannot be easily satisfied in practice [84]. Subsequently, the SaMP algorithm was proposed to address this issue [64]. While CoSaMP requires the level of sparsity as *a priori* information to determine the number of iterations of the algorithms, SaMP uses a stage-based approach to estimate the sparsity level. In particular, the estimated

[9]The considered performance metric was the phase transition curve; for details, the reader is referred to [61].

sparsity level is accumulated with a fixed preset step size stage by stage. The results in [64] showed that SaMP can outperform the original OMP algorithm and its variants. However, the MSE performance and complexity of SaMP are affected by the choice of the step size. More recently, a stage-wise algorithm which uses different step sizes for different stages has been proposed in [85]. However, it is not a truly adaptive algorithm, as the change of step sizes depends on a specific relationship between the number of measurements and the sparsity level. In this thesis, we propose a novel CS-based reconstruction algorithm based on the SaMP algorithm, referred to as the adaptive step size SaMP (AS-SaMP), which can adaptively adjust the step size to achieve fast convergence and has demonstrated a better MSE and BER performance over the other considered algorithms [26, 30].

2.2.3 Pilot Placement Schemes for CS-based Channel Estimation in OFDM Systems

For CS-based channel estimation in OFDM systems, different pilot placement choices result in different CS measurements, and the result directly affects the performance of the channel estimation algorithms. With regard to the problem of pilot placement, equally spaced pilots are in general optimal for conventional channel estimation methods, which are, however, not true for CS-based methods [86]. As mentioned, the RIP is adopted to guarantee reliable and efficient reconstruction of sparse (or compressible) signals from fewer measurements obtained via a measurement matrix which satisfies the RIP. However, to find the matrix which satisfies the RIP is an NP-hard problem [42]. Instead of the RIP, most pilot designs aim to find the optimal pilot pattern which satisfies the MIP in order to make the evaluation process much easier. Although an exhaustive search of all possible combinations of the pilot indices would guarantee the optimal pilot pattern, the computational complexity increases exponentially as the

search space expands. For instance, if 16 out of 256 subcarriers are used as pilots, there are $\binom{256}{16} \approx 10^{25}$ different pilot patterns in the search space. Thus, the computational complexity is prohibitively high. In most of the existing studies related to CS, randomly and deterministically placed pilot tones have been reported [86–93]. In [87], a deterministic pilot pattern for a specific reconstruction method, named the Dantzig selector, was proposed, and this is particularly suitable for measurements corrupted by stochastic noise. Alternatively, reference [89] reported a procedure based on the discrete stochastic approximation to perform an offline search of the near-optimal pilot pattern. Moreover, randomly-placed pilot clusters were used in [88] to estimate the time-varying UWA channels. However, there is a lack of theoretical guarantee of the successful reconstruction, and implementing the random pilot pattern is more challenging in practical applications.

Furthermore, provided a partial DFT measurement matrix, it is known that if the pilot indices set is a cyclic difference set (CDS), the mutual coherence of the measurement matrix is minimized [90, 92]. However, it is not guaranteed that a CDS will exist for every pilot size. In this thesis, we investigate the problem of pilot placement based on CDS. When CDS does not exist, we propose a near-optimal pilot pattern selection scheme which relies on the concatenated CDS with an iterative tail search (C-CDS with TS) [30]. Because the proposed design is deterministic, it is generally more computationally efficient than any other search-based methods. An improvement in the MSE and BER performance has been demonstrated in this thesis, using the proposed pilot placement scheme, when compared to the randomly scattered pilots, and other pilot placement schemes in the literature.

2.2.4 Channel Prediction in UWA Communications

Inherent changes in the propagation medium and intended motion between the transmitter and receiver result in fast time-varying UWA channels which exhibit both dispersion in the delay and Doppler domains. Therefore the UWA channels are often characterized as doubly spread channels [94,95]. For the doubly spread channel estimation, one approach is to directly track channel variation from symbol to symbol using adaptive filters [94]. Alternatively, a linear time-varying channel can be represented by its first-order approximation, i.e., the DDSF [43,46], which accommodates time variations of the CIR within a short block of data. Experimental evidence and physical arguments showed that UWA channels tend to be sparse in both the delay and Doppler domains, and can be estimated using the CS-based algorithms [35,43,47,94,96–98]. Estimation of sparse DDSF coefficients has been investigated for both BP-based algorithms [94,96,97] and MP-based algorithms [43, 47, 98]. In this thesis, we focus specifically on assessing the performance of the aforementioned MP-based algorithms for the DDSF estimation.

A potential disadvantage of channel estimation is that the estimated CSI may become quickly outdated for the channels that correspond to rough sea conditions. This can cause significant performance degradation as the channel variation due to the severe Doppler effects results in different channels from time to time. Additionally, many applications of promising techniques (e.g., adaptive modulation) requires the up-to-date CSI, and an outdated CSI may lead to failure of the techniques [99]. To achieve the up-to-date CSI, the channel has to be reliably predicted. Although many researchers have addressed the problem in terrestrial RF wireless communication systems, the prediction of time-varying channels in UWA communication systems has not been addressed until the recent years. The approaches generally fall into two groups, i.e., model-dependent and model-independent. Model-dependent methods assume certain knowledge about the channel and its variation (e.g., statistic property and dynamic equation), and thus

can achieve an improved prediction accuracy if a properly defined model structure is assumed [100–105]. In [101], channel evolution was modelled as autoregressive progress, and the related parameters were tracked by an extended Kalman filter (EKF). Similarly, in [100], the channel dynamics were explicitly described using a state-space model which consists of clusters of moving point scatters, and an EKF was employed to track its dynamic model parameters. It is noted that both of the above approaches were based on the assumption of uncorrelated tap coefficients and independent transitions. Later, the work in [104] considered the correlated multipath arrivals and used recursive least square (RLS)-based algorithm with a postfilter, which takes advantage of the channel correlation, to improve the tracking capability of the channel.

The above methods suffer performance degradation when the channel evolution mismatches the assumed model. In contrast, model-independent methods (referred to as adaptive prediction hereafter) do not rely on *a priori* knowledge of the channel dynamics [94, 95, 106, 107], and thus are useful for various practical applications. Generally speaking, an adaptive predictor can adjust itself with a goal of tracking the channel variations. Two classical adaptation algorithms, namely least mean square (LMS) and RLS have been extensively used in channel prediction [99, 106–111]. It is well-known that LMS is less complex, while RLS has advantages of a lower MSE and a faster convergence [112]. However, the existing predictors do not exploit the sparse features of DDSF which requires a much lower prediction complexity. The computational savings are from two perspectives: 1) the number of significant elements in DDSF to be predicted is relatively small; 2) prediction is only performed once for a block of data, and not for every data symbol, thanks to the advantage of DDSF representation. Furthermore, a decision-directed channel prediction/estimation[10] can improve the MSE

[10]In [113], a decision-directed maximum *a posteriori* probability channel prediction was proposed. However, as clarified in its text, the term "estimation" is used to broadly refer to the procedure from which the CSI is obtained via either prediction or estimation.

performance with the aid of the detected symbols [107, 113]. Following these observations, this thesis proposes a decision-directed sparse adaptive predictor for time-varying UWA channels. To work with the CS-based sparse channel estimation, the predictor can operate in delay-Doppler domain, and by exploiting the sparse DDSF coefficients, the RLS adaption algorithm can be employed without worrying about the computational complexity.

2.3 Conclusion

The framework, concept and results of the thesis relate to several areas, including sparse channel estimation, reconstruction algorithms, pilot placement, and adaptive prediction, with focus on UWA communications. After reviewing the current advances in these relevant aspects, the work in this thesis aims to develop sparse channel estimation and prediction algorithms, for a resource-limited receiver. These should provide accurate CSI without significantly increasing the computational complexity for both time-invariant and time-varying UWA systems.

Chapter 3

Underwater Acoustic Channel Modeling

Acoustic propagation is under continued exploration for commercial and scientific applications. However, unlike terrestrial RF communication channels, where a number of models for probability and power spectral density of the fading process are widely accepted and standardized, there is no commonly agreed characterization of acoustic communication channels. Therefore, experimental measurements are often used to access the statistical properties of the channel in specific deployment sites. It is beyond the scope of this thesis to characterize random changes of the UWA channel responses in general. Nevertheless, considering that relatively little attention has been given to an overview of the development of UWA channel modeling, we provide a survey of UWA channel modeling which consists of three stages in chronological order, as well as different state-of-the-art UWA channel simulation software. Furthermore, simulated channel responses are obtained based on a recently-proposed model, referred to as the statistical acoustic channel model (SACM) in [114], using a collection of real environmental data, and the statistical characterizations are also investigated. Finally, combining the UWA

channel sparsity property with the channel modeling is discussed.

3.1 Characteristics of Acoustic Propagation Under Water

The UWA channel is considered as one of the most challenging communication channels because the difficulties introduced by characterizations of the terrestrial radio channel, e.g., attenuation, multipath, Doppler effect, and noise, are exacerbated in UWA communications. This is due to the low speed of the sound signal in water (1500 m/s) when compared to the speed of the RF signal (3×10^8 m/s). In this section, we review several basic UWA channel properties with regards to acoustic attenuation, noise, multipath, and temporal variability.

3.1.1 Attenuation and Noise

A characteristic feature of acoustic channels is that the path loss depends on the signal frequency. This path loss can be calculated using [115]

$$A(l, f) = (\frac{l}{l_r})^k C_a(f)^{(l-l_r)}, \tag{3.1}$$

where l is the transmission distance taken from the referential distance l_r, k is the spreading factor which is normally between 1 and 2, and $C_a(f)$ is the absorption coefficient which can be determined using the Thorp's empirical formula [115],

$$C_a(f) = 0.11\frac{f^2}{1+f^2} + 44\frac{f^2}{4100+f^2} + 2.75 \times 10^{-4} f^2 + 0.003, \tag{3.2}$$

in which f is the signal frequency in kHz. Then, the transfer function for a time-invariant multipath UWA channel [114] is

$$H_p(f) = \frac{\Gamma_p}{\sqrt{A(l_p, f)}}, p = 0, 1, ..., N_p - 1, \tag{3.3}$$

where N_p is the number of multipaths, $\Gamma_p = \gamma_s^{n_s}\gamma_b^{n_b}\theta_p$ is the cumulative reflection coefficient over n_s and n_b times of surface and bottom reflections, and θ_p is the pth path's grazing angle. Specifically, the surface reflection coefficients $\gamma_s = -1$ under ideal conditions [12, 114] and the bottom reflection coefficients can be modelled by [115]

$$\gamma_b = \begin{cases} \frac{\rho_b \sin\theta_p - \rho\sqrt{(\frac{c}{c_b})^2 - \cos^2\theta_p}}{\rho_b \sin\theta_p + \rho\sqrt{(\frac{c}{c_b})^2 - \cos^2\theta_p}} & \cos\theta_p \leq \frac{c}{c_b} \\ 1 & otherwise \end{cases} \tag{3.4}$$

where c is the speed of sound in water, ρ is the density of water, c_b is the speed of sound on the ocean floor, and ρ_b is the density of the ground at the bottom of the ocean. Acoustic noise observed in the ocean arises from numerous sources, e.g., marine life, shipping traffic, underwater explosions, etc. Generally, it is grouped into two categories: ambient and site-dependent noise. While ambient noise is omnipresent in the background of deep ocean, site-dependent noise exists in specific area. Most ambient noise spectra can be described using Gaussian distribution [115], on the contrary, site-dependent noise often contains significant non-Gaussian components [12]. As a result, noise, together with frequency-dependent transmission loss, limit the communication range, bandwidth and signal-to-noise ratio (SNR) at the receiver side [13].

3.1.2 Multipath

Multipath propagation in water is the result of combination of two effects: sound reflection and refraction. Reflection occurs at the surface, bottom, or any objects between transmitter and receiver, while refraction occurs as a consequence of variation in sound speed from one location to another[11], which is mostly seen in deep water

[11] As sound speed is a function depth (described in a so-called sound speed profile), ray bending, which indicates the rays of sound always bend towards the region of lower propagation speed, is the main cause of multipath propagation in deep water [115].

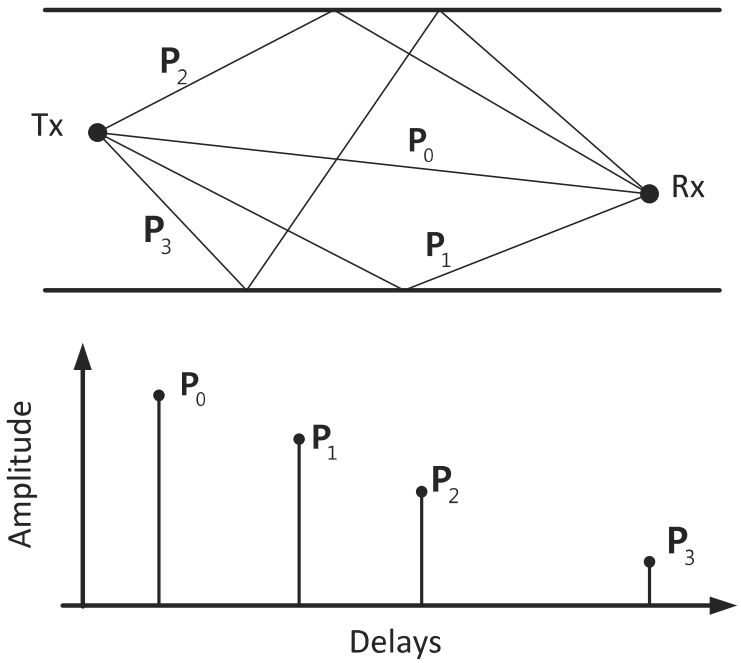


Figure 3.1: Multipath propagation in shallow water environment.

channels. According to Snell's law, a ray of sound bends toward the region where the signal propagates at a lower speed [115]. The modeling of multipath propagation is commonly based on ray-tracing theory. When a sound signal is transmitted, a beam of rays travels to the receiver through different paths leading to the multiple arrivals of the signal in different time intervals at the receiver side. The impulse response of a UWA channel is affected by the signal's refraction and reflection properties which determine the number of the strong paths, the strengths and delays of the multipath. Fig. 3.1 shows a shallow water multipath propagation and its corresponding impulse response. Strictly speaking, there are an infinite number of echoes, however, many of them experience multiple reflections and can be neglected, leaving only a few strong paths. Typical multipath delay for UWA channel is on the order of 10 ms which is much larger than for terrestrial radio channels. This motivates multicarrier transmission schemes that can cope with multipath fading [12].

3.1.3 Temporal Variability

Two sources of UWA channel's temporal variability are the inherent random changes in the propagation medium and the relative motion between transmitter and receiver, as shown in Fig. 3.2. On the one hand, inherent changes may occur over a long period of time (e.g., seasonal temperature variation), which do not affect the instantaneous level of a signal, or over a short period of time which results in fast time-varying communication channels. Unlike the case of a terrestrial radio channel, there are many sources of short-term temporal variability, e.g., surface waves, air bubbles, and fish shoals [116], and that is why there is no standard UWA channel model [12]. Among these sources, surface waves induce signal scattering and the Doppler spread due to the random displacements of the reflection points, and thus lead to significant changes on impulse response [12, 116].

On the other hand, the relative motion between transmitter and receiver adds to the channel's variability through the Doppler effect which is more severe in UWA channels due to the low speed of sound. Besides, underwater instruments are subject to drifting with waves, currents, tides, resulting in Doppler spread with the magnitude almost as large as the Doppler spread induced by some intentional motions. Motions that influence the Doppler effect can be generally categorized into three types: surface waves, drifting, and intentional motions [114]. It is noteworthy that the Doppler spreads may be time-varying which brings more challenges to UWA channel modeling.

3.2 UWA Channel Modeling

Due to the high cost of system deployment and testing in underwater environments, there is an increasing need for an accurate acoustic channel modeling. In this section, the development of UWA channel modeling is reviewed to gain a better understanding

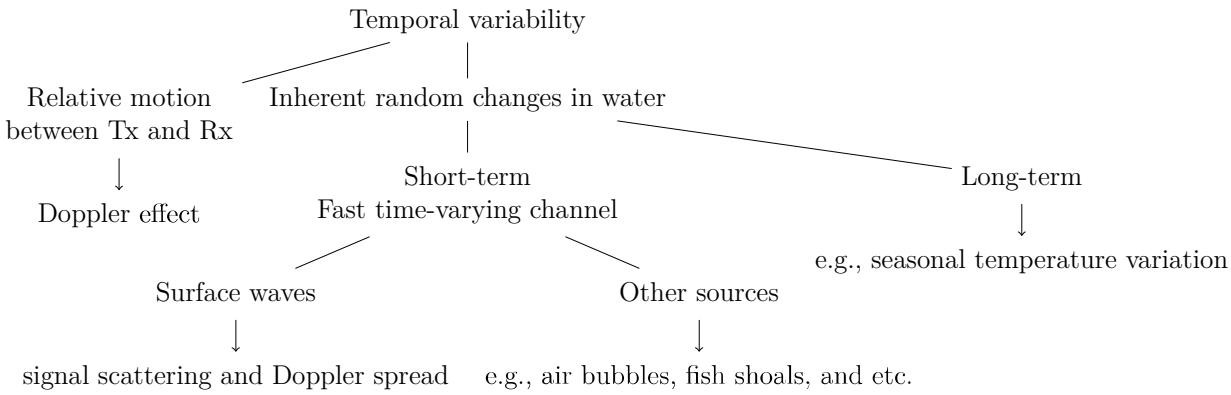


Figure 3.2: Time-variability in UWA channels.

on strengths and limitations of the existing models. Specifically, a survey of three popular channel simulators, which are referred to as BELLHOP [1, 117], the time variable acoustic propagation model (TV-APM) [118], and a recently-proposed SACM [114] are presented.

3.2.1 The Development of UWA Channel Modeling

As mentioned, the UWA propagation is characterized by frequency-dependent attenuation, time-varying multipath, and low-speed of sound. In the late 1980s, during which there was a lack of accurate knowledge of UWA channels, UWA channels were modeled as time-invariant systems within a certain period of time. Based on this assumption, various receiving techniques, e.g., matched filter, that were used in conventional digital communication systems were applied in early UWA communication systems [17, 67]. However, the time-invariant model is incapable of tracking channel time variations and thus, is impractical in the case of highly dynamic environment. To accommodate the Doppler effect, many researchers, in the late 1990s, started modeling UWA channels with the path-uniform Doppler, that is assuming that the same Doppler speed is applied to all paths. Therefore, the common Doppler scaling factor can be compensated by resampling and frequency shifting at the receiver side [119]. However, the UWA channels

are inherently wideband due to the sufficiently large ratio of the emitted signal bandwidth to the carrier frequency. Consequently, with the development of OFDM-UWA systems, large Doppler spread destroys the orthogonality of the subcarriers and induces intercarrier interference (ICI). Many methods were proposed to address this problem by taking consideration of non-uniform Doppler in channel modeling. In [9], non-uniform Doppler shifts were considered for a path-based channel model. The Doppler spread was treated to have two components, i.e., the path-uniform Doppler spread which can be compensated through resampling, and the residual Doppler, which can vary for different paths. Recently, advanced modeling approaches have been under investigation for better performance. For example, a higher-order polynomial approximation was employed to describe the amplitude and delay variations for different paths at the cost of increased complexity [120].

In addition to the different approaches in the aforementioned UWA channel models, each model can be applied to specific scenarios. Time-invariant channel model is usually sufficient for communication in calm water with less relative movement between the transmitter and the receiver, and the coherence time is remarkably larger than the symbol duration. Whereas, the model with a path-uniform Doppler suits the scenario for calm water with a strongly reflected multipath structure occurring when the transmitter and receiver are far apart so that incidence angles of acoustic rays are large. For the model which considers non-uniform Doppler shifts, reliable UWA communications can be achieved in rough water environment with multipath if the Doppler can be effectively compensated at the receiver. Throughout the development of UWA channel modeling, continuous efforts are made to achieve a trade-off between the model accuracy and computational complexity. Among many simulation tools, three representative software are presented in the following subsections.

Table 3.1: Input environmental data for BELLHOP[1] in [1]

Data Type	Description
Sound speed profile	Collected through conductivity-temperature-depth sensors (CTD) measurements over 1999-2001
Top boundary	Assumed vacuum
Bottom boundary	Acoustic-Elastic half space with the range-dependent bathymetry
Bathymetry	Collected through the use of sonar scans which can be found from some oceanic data source managed by research organizations, e.g., the National Geophysical Data Center (NGDC)
Water depth	99.59 m
Transmitter depth	78.96 m
Receiver depth	49.37 m
Distance between the transmitter and receiver	3.9 km

[1] Partial of the input data are collected off the coast of Newfoundland (46 29.5N 48 29.4W).

3.2.2 BELLHOP Ray Tracing

BELLHOP is a well-known ray tracing program for computing acoustic pressure fields in ocean environment, given a sound speed profile or variable absorbing boundaries [117]. It can produce a variety of outputs such as transmission loss, eigenrays, arrivals and received time-series. In [1], BELLHOP runs on a collection of regional environmental data to output multipath arrivals. These arrivals are then used to calculate the channel impulse responses and transfer functions. The set of data collected to input to BELLHOP includes location-specified temperature and salinity profile of the ocean column, a description of the surface and bottom properties, e.g., the bathymetry data, as well as the noise signature affected by wind, ship activity, or biological activity. Table 3.1 describes the reported environmental data as the input to the BELLHOP program, and by using the actual environmental data, a model is provided which shows the same trend with the empirical Thorp's spreading loss approximation, when an ideal ocean condition is considered. However, the model does not take into account random channel time variation.

3.2.3 The Time Variable Acoustic Propagation Model (TV-APM)

The simulator in [118] generates time variable simulated acoustic channel responses between moving sources and receivers. The main feature of the simulated model is to take into account the Doppler effect induced by the relative motion between transmitter and receiver as well as the effects of that motion propagating through the acoustic channel. Essentially, it loads the configuration data and iteratively calls the BELLHOP program to simulate dynamic signal propagation. Each call relies on the corresponding geometry of the transmitter-receiver. After each iteration, it checks the incoherence between the channels in the current and previous iterations. If the incoherence is above a given threshold, a new set of calculations is performed. Iterations will terminate if the incoherence between the new and the old set of calculations is below the threshold or if the maximum number of iterations is exceeded. However, the complexity can be high when Doppler spreads are affected not necessarily by a constant relative motion between the transmitter and receiver. That is to say the Doppler spread caused by environmental variations, e.g., surface waves, drifting, may result in a larger number of iterations. Recently, there has been a growing need for developing statistical UWA channel models for computationally efficiency [114].

3.2.4 A Statistical Modeling of a Class of UWA Channels

The recently-proposed model in [114] offers a statistical characterization that incorporates acoustic propagation as well as the effects of random local displacement. Firstly, random displacements which are treated on two scales: large-scale and small-scale, as shown in Fig. 3.3. For the large-scale random displacement, it usually spans on the order of many wavelengths, while for the small-scale case, the displacement is usually

on one or a few wavelengths. Since the nominal channel geometry can be considered as a time-invariant system expressed in (3.3), it can be determined using ray tracing, given a specified sound speed profile and typical signal frequency. However, the changes in the surface height or seabed terrain result in the deviation of the path length from the nominal ones, as $l_p = \bar{l}_p + \Delta_{l_p}$, where the deviation of the pth path Δ_{l_p} is random, and $\bar{l}_p$ is the length of the nominal path. Thus, approximations of the path gain and delay are calculated to capture the large-scale uncertainty based on the assumption that Δ_{l_p} is exponentially distributed. Moreover, small-scale scattering is considered to be a major contributor to signal random fluctuations. To model scattering in a UWA channel, each multiplath is split into a number of micropaths associated with the intrapath gain $h_{p,i}$ and the intrapath delay $\tau_{p,i}$, where i is the micropath index of the pth path. More specifically, $\tau_{p,i} = \tau_p + \Delta_{\tau_{p,i}}$ is treated as random to account for random displacement of scattering points. Specially, the model assumes that $\Delta_{\tau_{p,i}}$ follows zero-mean Gaussian distribution with variance $\sigma^2_{\Delta_p}$ and a small-scale coefficient is defined as

$$\gamma_p(f) = \frac{1}{h_p} \sum_i h_{p,i} e^{-j2\pi f \Delta_{\tau_{p,i}}}, \tag{3.5}$$

where h_p is the pth multipath gain. In general, $\gamma_p(f)$ is complex Gaussian with mean $\bar{\gamma}_p(f)$ and variance $2{\sigma_p}^2(f)$, provided the variances of the real and imaginary of $\gamma_p(f)$ are both ${\sigma_p}^2(f)$. Additionally, motion-induced Doppler shifts are addressed in the proposed channel model. Given v_p is the path associated relative velocity and c is the speed of sound, the Doppler scaling $a_p = v_p/c$ is influenced by three types of motion, i.e., surface wave, drifting, and intentional transmitter/receiver motion. Thus, the overall transfer function can be expressed as

$$H(f,t) = \bar{H}_0(f) \sum_p h_p \gamma_p(f) e^{j2\pi a_p f t}, \tag{3.6}$$

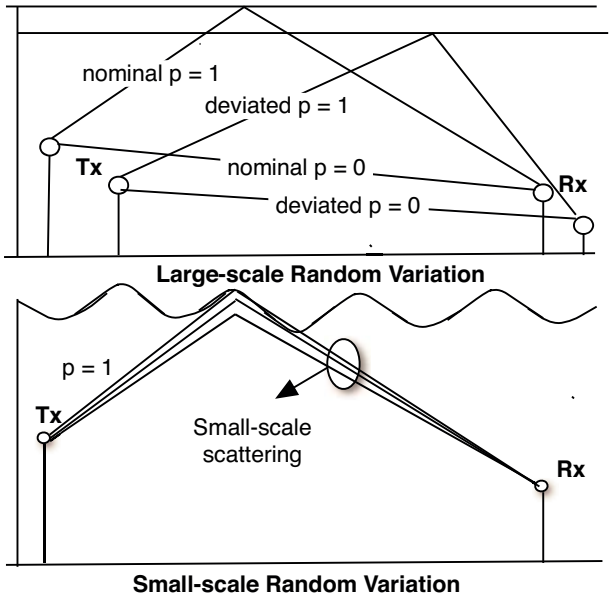


Figure 3.3: Two types of random phenomena.

where $\bar{H}_0(f)$ is the reference path transfer function. The SACM addresses the larger-scale channel gain, multipath, surface scattering, and motion-induce Doppler based on the assumption of the Gaussian-distributed intrapath delays of the scattered paths. It can be considered as a starting point for statistical UWA channel modeling, and can be extended to incorporate more complex physical properties such as spatial correlations, surface curvature, and effects of breaking waves. A comparison of the three channel simulators is shown in Table 3.2.

Table 3.2: Summary of the three UWA channel simulators

Models	**BELLHOP**	**TV-APM**	**SACM**
Description	Ray tracing program for computing acoustic pressure field in ocean environment, given environmental data	Generates time variable simulated acoustic channel responses between moving Tx and Rx	A statistical characterization that incorporates acoustic propagation as well as the effects of random local displacement
Pros	Widely accepted	Takes into account the Doppler effect	Computational efficient
Cons	Does not consider random variations	Low computational efficiency	Characterize the shallow-water channels, recently published and has not been widely accepted

3.3 Simulated UWA Channel

In this section, the simulated CIR and correlation functions of the fading coefficients, using collected environmental data described in Table 3.1, are presented. BELLHOP was used to obtain the large-scale gain of the acoustic channel with the specific system geometry and sound speed profile. For each small-scale realization, the small-scale coefficients $\gamma_p(f)$ were generated as described in (3.5). To allow for motion-induced Doppler, the term associated with Doppler scaling factor is included and assumed to be constant over a short time interval.[12] The carrier frequency and bandwidth are 13 and 10 kHz, respectively. B_{Δ_p}, which is set to 0.1 Hz, is the 3 dB width of the power spectrum density of $\Delta_{\tau_{p,i}}(t)$ [114]. The signal is transmitted for 3 minutes. Fig. 3.4 shows the simulated CIR which have a sparse structure formed by multipath arrivals that are resolved in delay. Given that there is no intentional motion between the transmitter and receiver, it can be observed that the multipath delay varies in time even if there is no intentional transmitter-receiver motion. This is due to the unintentional motion-induced Doppler effects, i.e., drifting or surface wave induced Doppler shifting. Moreover, it is noteworthy that the direct path p_0 may reach the receiver after the indirect path, e.g., the bottom reflected path p_b, which travels at higher speed. This results in a non-minimum phase channel response [13].

Next, the channel's small-scale analysis can be made through assessing correlation functions in both the frequency and time domains, which are defined as follows,

$$C_{\gamma_p}(\Delta_f, \Delta_t) = \begin{cases} E[\gamma_p(f+\Delta_f) - \bar{\gamma}_p(f+\Delta_f)][\gamma_p(f) - \bar{\gamma}_p(f)]^* & \Delta_t = 0 \\ E[\gamma_p(f, t+\Delta_t) - \bar{\gamma}_p(f)][\gamma_p(f,t) - \bar{\gamma}_p(f)]^* & \Delta_f = 0 \end{cases} \tag{3.7}$$

The Gaussian distributed delays $\Delta_{\tau_{p,i}}(t)$ obeys the first-order autoregressive process (AR-1). Fig. 3.5 and Fig. 3.6 illustrate the frequency and time correlation for the de-

[12]Focusing on small-scale phenomenon, it is reasonable to assume the constant velocity for any motion within a short time period (e.g., several ms)

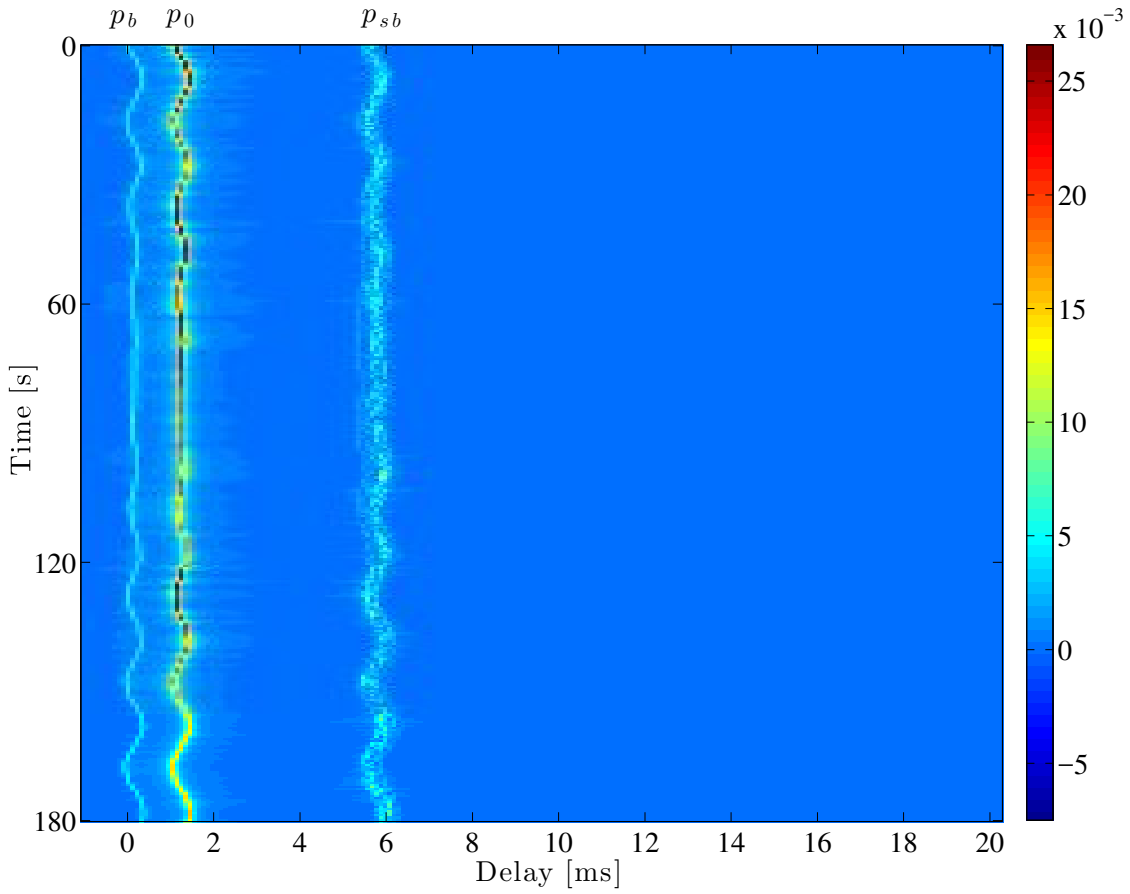


Figure 3.4: Time evolution of the magnitude baseband impulse response.

ployed system frequency and several values of the standard deviation σ_{Δ_p}, respectively. It is noted that the large intrapath delay variance σ_{Δ_p} represents the large deviation of the micropath from the nominal path, and thus leads to lower coherence of small-scale coefficients in frequency and time. Therefore, utilizing these correlation properties leads to a computationally efficient channel model simulation.

3.4 UWA Channel Sparsity

A good channel model represents a precise description of the physical properties of the channel, and thus may involve many coefficients. This will cause the problem

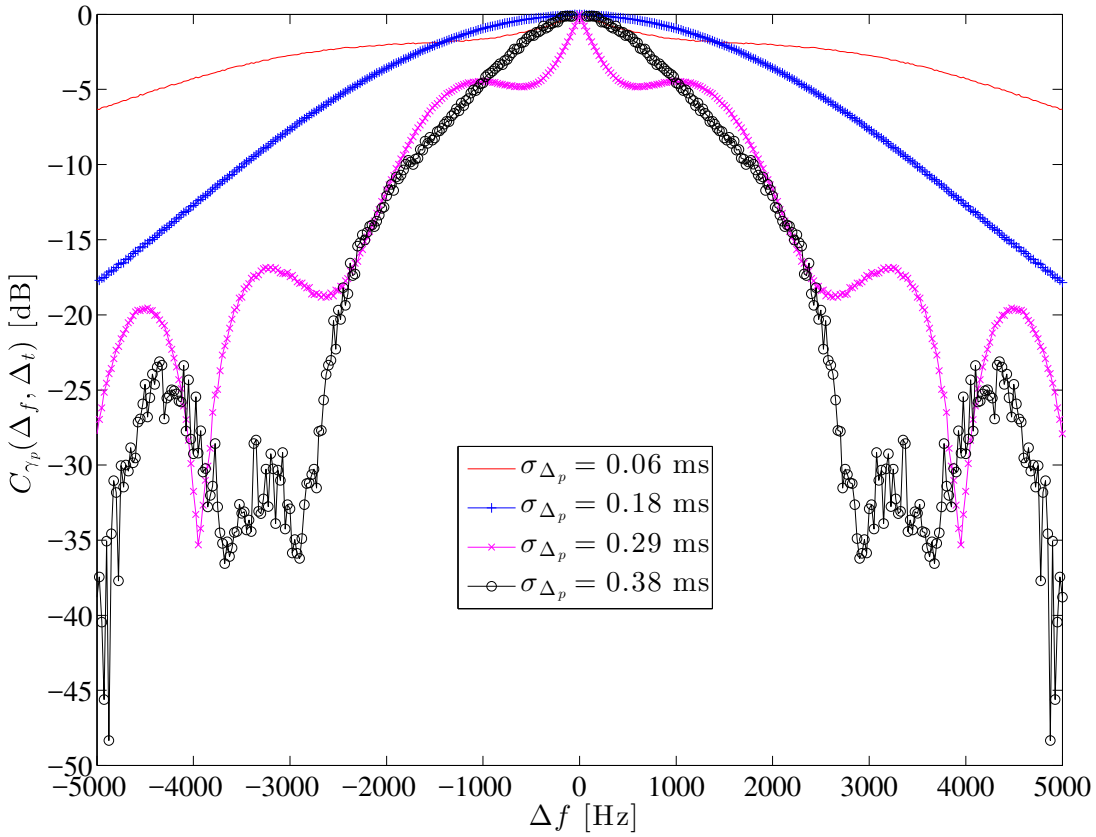


Figure 3.5: Small-scale frequency correlation function for several values of the standard deviation σ_{Δ_p}. Δ_t is set to zero, while Δ_f is varied.

of high computational complexity which requires excessively sophisticated receiving techniques. To achieve a trade-off between the accuracy and the complexity of a model, one can combine the channel sparsity with the modeling task. The aforementioned time-invariant channel model and uniform Doppler channel model can utilize the sparsity over delay domain. The uniform Doppler channel model assumes the constant speed for all paths, therefore a specific Doppler for different frequencies needs to be estimated and compensated. Although higher complexity is involved in the receiver, the system performance is improved. As more coefficients are involved in non-uniform Doppler channel model, it is more beneficial to exploit the existing channel sparsity. Given

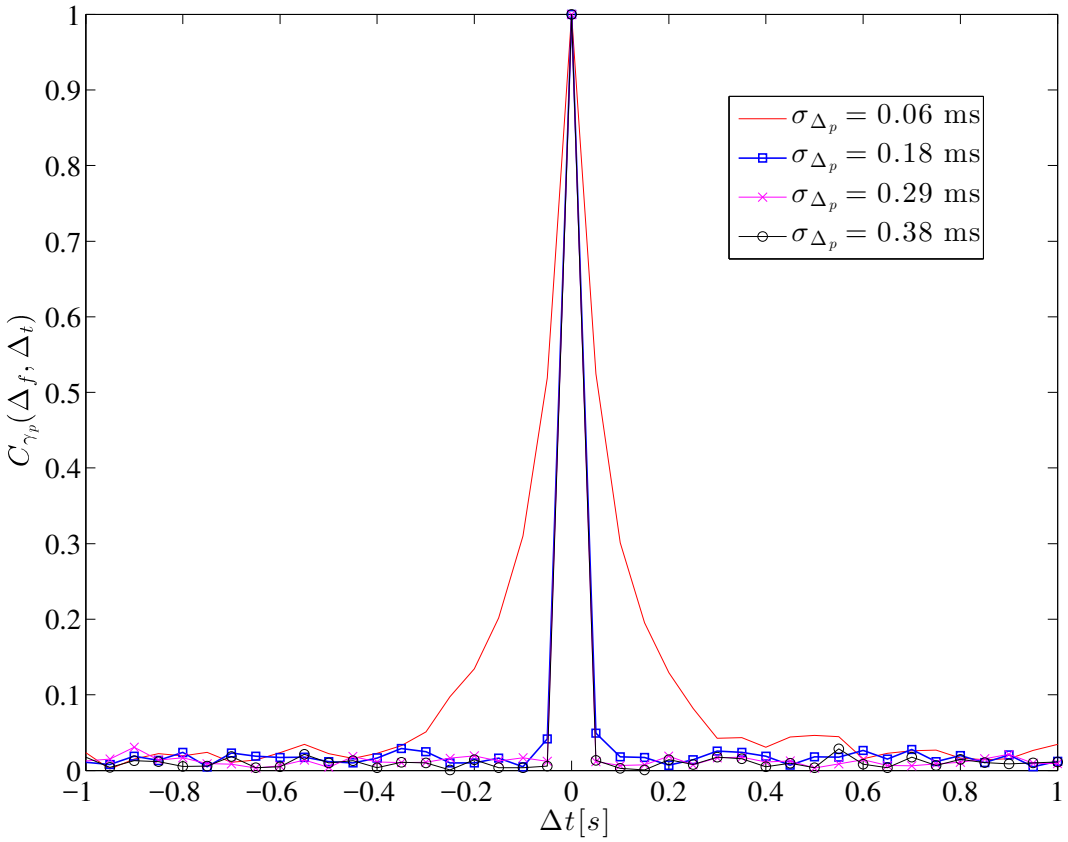


Figure 3.6: Small-scale time correlation function for several values of the standard deviation σ_{Δ_p}, Δ_f is set to zero, while Δ_t is varied.

that only significant paths need to be considered while the remaining paths can be neglected due to overly low energy, the sparsity can be utilized on path level over the delay and Doppler domain. An example of the sparse representation of a UWA channel in Delay and Doppler domains is shown in Fig. 3.7. Each square represents a resolution bin associated with a channel coefficient, and the dots represent the multiple paths contributing to each channel coefficient. A channel is considered to be sparse when the number of significant paths is much smaller than the total number of grids. In the following chapters, we propose a novel sparse signal reconstruction algorithm and applied it in both CS-based time-invariant and time-varying (with non-uniform

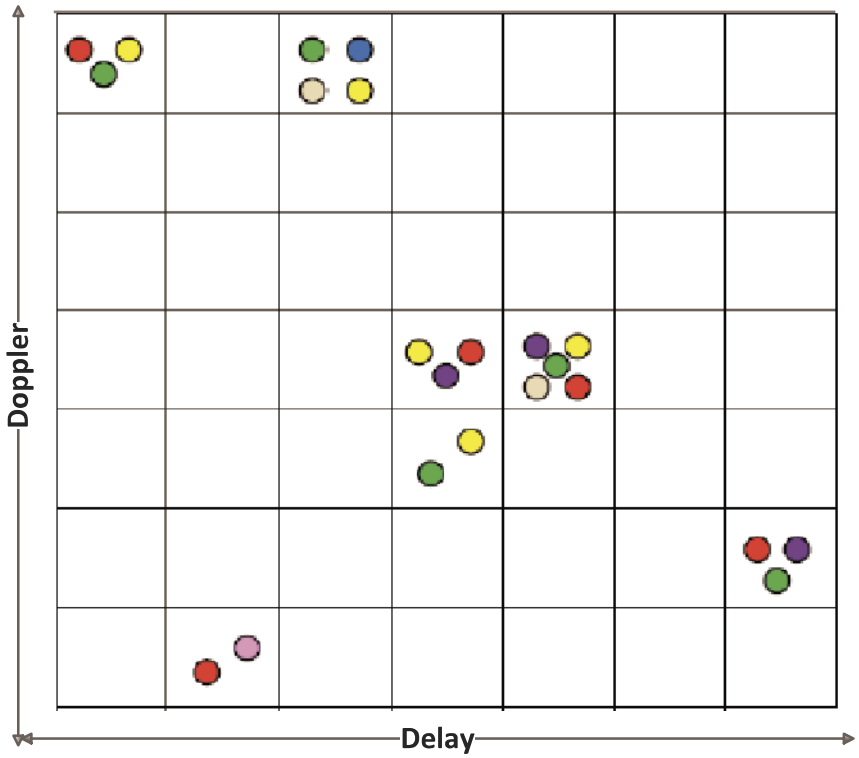


Figure 3.7: Sparse representation of a UWA channel in Doppler and delay domains.

Doppler) channel estimation.

3.5 Conclusion

In this chapter, acoustic propagation in water is briefly described. Based on the properties which are studied, an overview of various ray-theory-based UWA channel models is provided, in chronological order, along with a survey of the three representative channel simulation software. Through the comparison, one can expect to gain a better understanding on the pros and cons of each modeling method and its applicable scope. Furthermore, the simulated channel impulse response based on a recently-reported statistical model using a collection of the real environment data is shown. The reported statistical model takes into account physical propagation and random channel varia-

tions on large and small scales, and describes the scattering-induced intrapath delays using Gaussian distribution. The correlation properties in time and frequency are investigated for different values of the intrapath delay variance. Through observation of the simulated channel, it is worth noting that the sparsity can be incorporated to channel modeling and estimation to achieve a trade-off between the system error performance and complexity.

Chapter 4

CS-Based Time-Invariant Channel Estimation for OFDM System

Recently, studies have suggested that many multipath channels, besides the UWA channels, tend to exhibit a sparse structure in the sense that the majority of CIR taps end up being either zero or negligible [35]. A few examples include: a) in the North American HDTV broadcasting standard, there are only a few significant dominant echoes over a typical delay spread [121]; b) channels of broadband wireless systems in hilly environment also exhibit a sparse CIR [122]. Unlike the traditional channel estimation methods, methods which exploit the sparsity of the wireless channels, often reduce the required number of pilots and lead to a higher spectral and energy efficiency.

This chapter presents a novel greedy algorithm for CS-based time-invariant sparse channel estimation in OFDM systems, which is referred to as the AS-SaMP. Compared with other state-of-the-art MP-type algorithms, the proposed algorithm possesses the feature of not requiring *a priori* knowledge of the sparsity level, and moreover, adjusts the step size adaptively to approach the true sparsity level. This makes it a promising candidate for many practical applications where the sparsity level of a signal is not

available. Additionally, theoretical performance analysis and a symbolic computational complexity of the proposed algorithm are provided. Numerical simulation results show that a good performance is attained by the AS-SaMP algorithm without increasing the complexity significantly when compared with the other considered recovery algorithms.

4.1 System Model

We consider an N-subcarrier OFDM system in which P subcarriers are used as pilots. The symbols transmitted on the kth subcarrier, $X(k)$, $0 \leq k \leq N-1$, are assumed to be independent and identically distributed random variables drawn from a phase-shift keying (PSK) or quadrature amplitude modulation (QAM) signal constellation. Assume that the discrete multipath channel has the impulse response

$$h(n) = \sum_{p=0}^{N_p-1} \eta_p(n)\delta(n - \tau_p(n)), \tag{4.1}$$

where N_p is the number of paths, and $\eta_p(n)$ and $\tau_p(n)$ are the amplitude gain and the delay associated with the pth path, respectively. Although $\eta_p(n)$ and $\tau_p(n)$ are time-varying, the condition $\eta_p(n) \approx \eta_p$ can be true in one OFDM symbol period [29]. Moreover, by considering a relative stationary transmitter and receiver, the assumption $\tau_p(n) \approx \tau_p$ can be made [123]. Then the vector of received signal after the DFT is expressed as

$$\mathbf{Y} = \mathbf{XH} + \mathbf{W} = \mathbf{XDh} + \mathbf{W}, \tag{4.2}$$

where $\mathbf{X}$ is an $N \times N$ diagonal matrix with the elements $X(k)$, $0 \leq k \leq N-1$, on the main diagonal, $\mathbf{Y} = [Y(0), Y(1), ..., Y(N-1)]^T$, $\mathbf{H} = [H(0), H(1), ..., H(N-1)]^T$, and $\mathbf{W} = [W(0), W(1), ..., W(N-1)]^T$ are the frequency response vectors of the received symbol, channel and additive white Gaussian noise (AWGN), respectively, and $\mathbf{h} =$

$[h(0), h(1), ..., h(L-1)]^T$, with L as the number of taps. The (m, n) element of $\mathbf{D}$ is given by $[\mathbf{D}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$, where $0 \leq m \leq N-1$ and $0 \leq n \leq L-1$. After extracting the pilot subcarriers, there is

$$\mathbf{Y_p} = \mathbf{X_p D_p h} + \mathbf{W_p} = \mathbf{Ah} + \mathbf{W_p}, \tag{4.3}$$

where $\mathbf{Y_p} = \mathbf{SY}$, $\mathbf{X_p} = \mathbf{SXS}^T$, $\mathbf{D_p} = \mathbf{SD}$, $\mathbf{W_p} = \mathbf{SW}$, and $\mathbf{S}$ is a $P \times N$ matrix for selected pilot subcarriers. In addition, $\mathbf{A} = \mathbf{X_p D_p}$ is a $P \times L$ matrix, referred to as the measurement matrix according to Section 2.1.2.

The goal of CS-based channel estimation is to estimate $\mathbf{h}$ from the received pilot $\mathbf{Y_p}$, given the measurement matrix $\mathbf{A}$. In (4.3), from which $\mathbf{h}$ needs to be estimated, the jth index of $\mathbf{h}$ corresponds to the jth path delay $\tau(j)$, and the corresponding value represents the gain of that path. The actual delay of the channel may not coincide with the assumed delay points; this is known as the off-grid problem [124][13]. Clearly, finer-grained delay points lead to a better approximation of the continuous delay, thus improving the estimation quality [28]. In this chapter, we use a sampling rate of $N\Delta f$, where Δf is the subcarrier spacing, and assume that τ_p is multiple integer of the sampling time, as it is commonly used in the literature [3, 9, 28].

[13]It is worth noting that the off-grid problem is common to all CS-based algorithms.

4.2 The AS-SaMP Algorithm with Application to Sparse Channel Estimation

4.2.1 Comparison of Reconstruction Algorithms in the Literature

This section focuses on the reconstruction algorithms of the MP family, that identify the support set[14] of the target signal iteratively. At each iteration, one or more columns of the measurement matrix that are most correlated with the current residual are selected (this is referred to as the maximum correlation test), and the residual is updated by projecting the measurements onto the linear space spanned by the selected columns. The reconstruction algorithms considered here are the OMP, CoSaMP, and SaMP. Fig. 4.1 depicts the corresponding flow charts, where $\mathbf{C}^i$, $\mathbf{F}^i$, $\mathbf{r}^i$, and K denote the candidate support set, the final support set, the residual vector in the ith iteration, and the sparsity level of the target signal, respectively. As seen from Fig. 4.1, in the OMP algorithm, $\mathbf{F}^i$ is expanded by adding coordinates successively and it uses only one maximum correlation test to add one coordinate to $\mathbf{F}^i$.

Meanwhile, the CoSaMP algorithm refines a fixed-size $\mathbf{F}^i$ by selecting coordinates from a set of candidates $\mathbf{C}^i$. It uses a *preliminary correlation test* and a *final correlation test*, which are simply referred to as *preliminary test* and *final test*, to add one or more coordinates to $\mathbf{F}^i$. The *final test* removes the wrong coordinates added in the *preliminary test*, which is referred to as *backtracking*, and therefore improves the accuracy of the estimation [63]. However, since most natural signals are compressible rather than strictly sparse, the sparsity level K for these signals could not be well-defined. It is shown that the reconstruction accuracy can be significantly degraded as we either

[14]For a vector $\beta = [\beta_1, \beta_2, ..., \beta_L] \in \mathbb{R}^L$, the support set is defined as $\{i|\beta_i \neq 0\}$ [52].

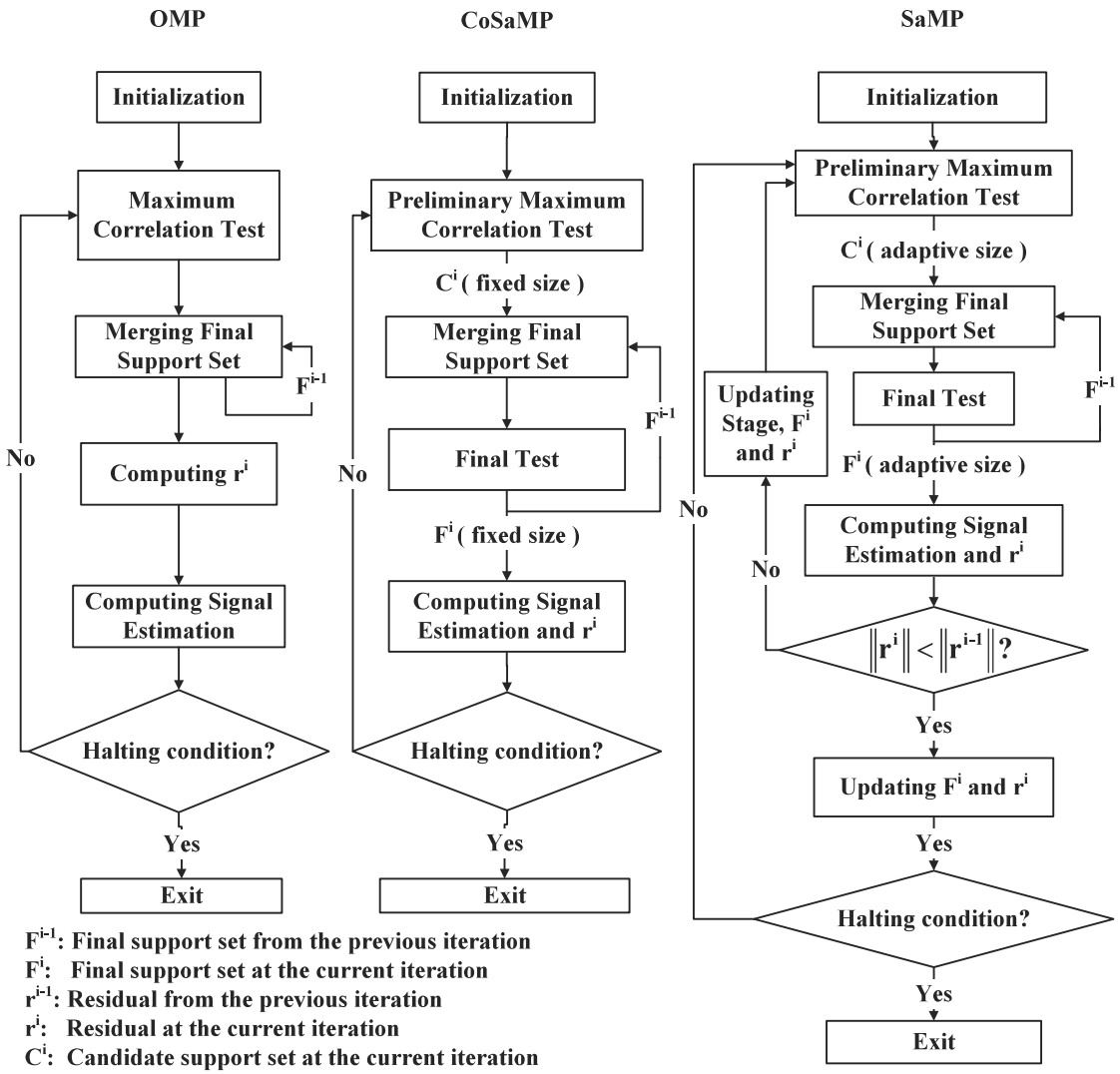


Figure 4.1: Flow charts of the OMP, CoSaMP, and SaMP algorithms.

underestimate or *overestimate* K [64]. Unlike the CoSaMP algorithm, SaMP does not require *a priori* knowledge of K. It adopts a stage-wise approach to identify $\mathbf{F}^i$ through the backtracking strategy. The size of $\mathbf{F}^i$ stays the same among iterations in each stage; however, when it moves to the next stage, the size of $\mathbf{F}^i$ is increased by a fixed step size s to search for more coordinates of the recovered signal which correspond to the least residual. This process continues until the residual of the recovered signal falls below a predetermined threshold.

Although SaMP guarantees exact recovery after a finite number of iterations (see proof in [64]), it leaves an open question about the choice of the step size s to achieve the trade-off between accuracy of estimation and complexity. This motivates us to address the problem of adaptively adjusting the step size between consecutive stages. Recently, a variable step size algorithm has been proposed in [85]. However, the increment of the step size is based on a particular relationship between the number of the measurements and the sparsity level,[15] which is not always valid in our application as the sparsity level of the channel can vary in time. Therefore, we propose a novel AS-SaMP algorithm which can adaptively adjust the step size to approach the true sparsity level in the next section.

4.2.2 The AS-SaMP Algorithm

Since a smaller step size s in the SaMP algorithm leads to a better estimation accuracy while the complexity increases, and a larger s degrades the accuracy of the estimation while the complexity decreases, an adaptively adjusted s may lead to a better trade-off between the accuracy of estimation and the complexity of the algorithm. Specifically, an adaptively adjusted s means that the change of s depends on how far the current reconstruction state, e.g., current reconstructed signal energy or its estimated sparsity level, is from the state of the true signal. Because the sparse elements with large values are reconstructed in the initial stages of the algorithm, the energy difference of the reconstructed signal between consecutive stages is reduced at a declining rate as the number of stages increases. In other words, the energy of the reconstructed signal tends to be stable when the estimated sparsity level is close to the true sparsity level K. Following this property, the AS-SaMP algorithm begins with a larger step size (the

[15] According to [85], the step size is changed based on a practical rule i.e., if ones want to recover the signal, the sample number should be four times the signal sparsity level.

Algorithm 1 AS-SaMP

Input: Received signal at pilot subcarriers $\mathbf{Y_p}$, measurement matrix $\mathbf{A}$, tolerance ϵ, threshold Γ, initial step size s_I;

Initialize $\hat{\mathbf{h}} = [0, 0..., 0]^T$, $\hat{\mathbf{h}}_{old} = [0, 0..., 0]^T$, $\mathbf{r}_{temp} = [0, 0..., 0]^T$, indices set $\mathbf{B}^0 = \emptyset$, candidate support set $\mathbf{C}^0 = \emptyset$, residual $\mathbf{r}^0 = \mathbf{Y_p}$, size of final support set $L_F = s = s_I$, final support set $\mathbf{F}^0 = \emptyset$, iteration index $i = 1$
while $(\|\mathbf{r}^{i-1}\| > \epsilon)$ **do**
  Calculate signal $\mathbf{SP} = |\mathbf{A^H r}^{i-1}|$
  Select indices set $\mathbf{B}^i$ in $\mathbf{A}$ corresponding to the L_F largest elements in $\mathbf{SP}$ {Preliminary test}
  Merge chosen indices and final support set from previous iteration into candidate support set $\mathbf{C}^i = \mathbf{B}^i \cup \mathbf{F}^{i-1}$
  Refine candidate set to final set $\mathbf{F}^i$ by selecting indices corresponding to the L_F largest elements of $|\mathbf{A}^{\dagger}_{\mathbf{C}^i}\mathbf{Y_p}|$ {Final test}
  Solve least-square problem $\hat{\mathbf{h}}(\mathbf{F}^i) = \mathbf{A}^{\dagger}_{\mathbf{F}^i}\mathbf{Y_p}$
  Calculate current residual $\mathbf{r}_{temp} = \mathbf{Y_p} - \mathbf{A}_{\mathbf{F}^i}\mathbf{A}^{\dagger}_{\mathbf{F}^i}\mathbf{Y_p}$
  if $(\|\mathbf{r}_{temp}\| < \epsilon)$ **then**
    $\mathbf{r}^i = \mathbf{r}^{i-1}$
    Break
  else if $(\|\mathbf{r}_{temp}\| \geq \|\mathbf{r}^{i-1}\|)$ **then**
    if $(\| \hat{\mathbf{h}}(\mathbf{F}^i) \| - \| \hat{\mathbf{h}}_{old} \| < \Gamma)$ **then**
      $s = \lceil s/2 \rceil$, $L_F = L_F + s$, $\hat{\mathbf{h}}_{old} = \hat{\mathbf{h}}(\mathbf{F}^i)$, $\mathbf{r}^i = \mathbf{r}^{i-1}$, $i = i + 1$ {Fine tuning}
    else
      $L_F = L_F + s$, $\hat{\mathbf{h}}_{old} = \hat{\mathbf{h}}(\mathbf{F}^i)$, $\mathbf{r}^i = \mathbf{r}^{i-1}$, $i = i + 1$ {Fast approaching}
    end if
  else
    $\mathbf{F}^{i-1} = \mathbf{F}^i$, $\mathbf{r}^i = \mathbf{r}_{temp}$, $i = i + 1$
  end if
end while
return $\hat{\mathbf{h}}$

Output: Estimation of baseband channel impulse response $\hat{\mathbf{h}}$.

initial step size is denoted as s_I) to expedite the convergence. Then, the step size is adaptively decreased to provide fine tuning in later stages, as the change rate of the reconstructed signal's energy decreases. Consequently, an additional threshold Γ is used to specify the beginning of the fine tuning.

The pseudocode for the AS-SaMP algorithm is shown as **Algorithm 1**. The algorithm is stage-wise with a variable size of $\mathbf{F}^i$ in different stages. During each stage, it adopts two correlation tests iteratively, i.e., candidate and final tests, to search for a certain number of coordinates corresponding to the largest correlation values between the signal residual and the columns of the measurement matrix. Then, the algorithm moves to the next stage until the recovered signal with the least residual is found. As opposed to SaMP, the proposed algorithm incorporates two threshold values into the halting criterion: tolerance ϵ and Γ. Therefore, AS-SaMP halts when the residual's norm is smaller than ϵ, in which ϵ is set to be the noise energy. Meanwhile, s is decreased when the energy difference of the reconstructed signal falls below Γ, whose value is chosen based on empirical observations. Starting with a sufficiently large initial step size ($s_I \leq K$), the proposed algorithm quickly approaches the target signal. However, when the difference in the energy of the reconstructed signals becomes smaller than the preset Γ, the step size is reduced (by a factor of two) to avoid *overestimation* of the K-sparse target signal. This *overestimation* can significantly degrade the accuracy of the algorithm [64]. For practical applications, the values of Γ and ϵ need to be selected; some details are presented as follows.

The threshold Γ: In the AS-SaMP algorithm, the step size is reduced when the energy of the reconstructed signal changes by an amount less than Γ. Clearly, $\Gamma \leq \|\mathbf{h}\|$ in order to provide fine tuning. The larger Γ is, the earlier a fine tuning starts and thus requires a larger number of iterations. On the other hand, a smaller Γ leads to fewer iterations. However, a higher estimation quality can be achieved with an earlier initiation of fine

tuning.

The tolerance ϵ: As in the SaMP algorithm, AS-SaMP halts when the norm of the residual channel falls below a preset threshold ϵ. According to [64], $\epsilon = 0$ when the measurements are noiseless, while ϵ can be chosen as the norm of the noise vector when the measurements are noisy.

In the next section, a theoretical guarantee of exact recovery of AS-SaMP, in both noiseless and noisy cases, is provided with the corresponding proofs shown in Appendix A.

4.3 Theoretical Performance Analysis of AS-SaMP

The recovery performance of the proposed AS-SaMP algorithm is based on the theoretical performance of SaMP and subspace pursuit (SP) [125]; therefore, the proofs which follow the format in [64, 125] are developed for two cases: exact recovery from noiseless measurements and approximate recovery from noisy measurements.

4.3.1 Reconstruction Performance for Noiseless Measurements

Before stating **Theorem 1** for the exact recovery of the AS-SaMP algorithm, we need two results summarized in the lemmas below.

Lemma 1. Given an arbitrary K-sparse signal $\mathbf{h}$ and the corresponding measurement $\mathbf{Y_p} = \mathbf{Ah}$. Let the total number of stages decided by AS-SaMP be J and $s_i, i \in \{1, 2, ..., J\}$ be the step size of the ith stage. If $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_J} < 0.06$ [41], where $K_J = \sum_{i=1}^{J} s_i$ is the estimated sparsity level, the last stage of AS-SaMP is equivalent to SaMP algorithm with estimated sparsity level K_J, except with possibly different contents in the final support set and the observation residual vector.

Lemma 2. AS-SaMP guarantees the convergence of the recovery process. The upper-bounded number of iterations that AS-SaMP involves is

$$-\log(\frac{|h_{\min}|}{\parallel \mathbf{h} \parallel})(\frac{-1}{\log(C_{K_1})}+\frac{-1}{\log(C_{K_2})}+...+\frac{-1}{\log(C_{K_J})})+J, \tag{4.4}$$

where $h_{\min}$ is the non-zero element with the minimum magnitude and $C_{K_i}=\frac{2\delta_{3K_i}(1+\delta_{3K_i})}{(1-\delta_{3K_i})^3}$, $i=1,2,...,J$, δ_{3K_i} is the RIP parameter in the ith stage, and K_i is the size of final support set in the ith stage.

Proof of **Lemma 1** and **Lemma 2** are deferred to Appendix A.1. Next, the upper-bounded number of iterations for AS-SaMP is compared with that for SaMP in the Corollary below.

Corollary 1. Provided that $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_{s-AS-SaMP}}<0.06$ and $\delta_{3K_{s-SaMP}}<0.06$, where $K_{s-AS-SaMP}$ and K_{s-SaMP} are the estimated sparsity level for AS-SaMP and SaMP, respectively, the upper-bounded number of iterations for AS-SaMP is smaller than that for SaMP.

The proof of the **Corollary 1** is postponed to Appendix A.2. Now based on the lemmas above, a sufficient condition for exact reconstruction is drawn in the following theorem.

Theorem 1. (*Exact recovery from noiseless measurements*): Let $K_{s-AS-SaMP}=s_I J$, where $s_I=s_1$ is the initial step size and J is the total number of stages of the AS-SaMP algorithm. If the sensing matrix $\mathbf{A}$ satisfies the RIP with the parameter $\delta_{3K_{s-AS-SaMP}}<0.06$, the AS-SaMP algorithm is guaranteed to exactly recover $\mathbf{h}$ from $\mathbf{Y_p}$ via a finite number of iterations.

Proof. Based on **Lemma 1** and **Lemma 2**, when the RIP condition is satisfied, because the last stage is equivalent to SaMP with estimated sparsity level $K_{s-AS-SaMP}$, the AS-

SaMP algorithm guarantees exact recovery of the target signal after this stage, and it takes finite number of iterations to reach $K_{s-AS-SaMP}$. ■

Remark 1. From the **Lemma 1**, a sufficient condition which is required for $\mathbf{A}$ to guarantee an exact recovery is $\delta_{K_J} < 0.06$, where $K_J = \sum_{i=1}^{J} s_i$ and s_i is the step size in the ith stage. As $s_I \geq s_2 \geq \cdots \geq s_J$, we have $K_J \leq s_I J$. Therefore, a more restrictive requirement of the RIP parameter of $\mathbf{A}$ will be $\delta_{3s_I J} < 0.06$ which is $\delta_{3K_{s-AS-SaMP}} < 0.06$. The sufficient condition for SaMP is more restrictive than SP algorithm as the estimated sparsity level $K_{s-SaMP} = s\lceil K/s \rceil$ where s is the fixed step size in SaMP, is always at least as large as the true sparsity level K [64]. Similarly, to compare the restrictiveness of the condition of the AS-SaMP, the values of K_{s-SaMP}, $K_{s-AS-SaMP}$ and K needs to be compared. As $\lceil K/J \rceil \leq s_I \leq \frac{s\lceil K/s \rceil}{J}$, so $K \leq s_I J \leq s\lceil K/s \rceil$ and thus $K \leq K_{s-AS-SaMP} \leq K_{s-SaMP}$.

Furthermore, because of monotonicity of δ_{3K}, $\delta_{3K_{s-AS-SaMP}} \leq \delta_{3K_{s-SaMP}}$, as a result, if $\delta_{3K_{s-SaMP}} < 0.06$, $\delta_{3K_{s-AS-SaMP}} < 0.06$ holds, which means that the requirement of $\mathbf{A}$ for AS-SaMP is less restrictive than that for SaMP. Moreover, as $\mathbf{A}$ is a $P \times N$ partial DFT matrix in our application, and indices of the P pilots are randomly chosen, $\mathbf{A}$ satisfies the RIP with an overwhelming probability provided that

$$K \leq C_1 \frac{P}{(\log N)^6}, \tag{4.5}$$

where C_1 depends only on the RIP parameter (by overwhelming probability, it means that the probability is at least $1 - N^{-\frac{1}{C_1}}$) [40] and K is the sparsity level of the target CIR. In fact, (4.5) expresses the minimum number of pilots ($P \geq \frac{K(\log N)^6}{C_1}$) required such that a random subset of $\mathbf{A}$ with average cardinality $3K_{s-AS-SaMP}$ satisfies the RIP with high probability. Specifically, for $P \geq 8K$, the recovery rate is above 90% [40].

4.3.2 Reconstruction Performance for Noisy Measurements

The second part of theoretical performance analysis is to investigate the approximate recovery from inaccurate measurements of the AS-SaMP algorithm. Two types of inaccurate measurements are considered: one is subject to noise perturbation and the other one is subject to approximately sparse signal whose non-significant elements are comparatively small (but not zero) and noise.

Theorem 2. (*Approximate recovery from noisy measurements*): Consider $\mathbf{h} \in \mathbb{R}^N$ as a K-sparse signal, $\mathbf{Y_p} = \mathbf{Ah} + \mathbf{W_p} \in \mathbb{R}^P$ as the noisy measurement vectors and $\mathbf{W_p}$ as a noise vector generated from a Gaussian distribution with zero mean and variance σ^2. If the measurement matrix $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_{s-AS-SaMP}} < 0.03$, the signal approximation $\hat{\mathbf{h}}$ satisfies:

$$\begin{aligned} \| \mathbf{h} - \hat{\mathbf{h}} \| &\leq \frac{1 + \delta_{3K_{s-AS-SaMP}}}{\delta_{3K_{s-AS-SaMP}}(1 - \delta_{3K_{s-AS-SaMP}})} \| \mathbf{W_p} \| \\ &= \frac{1 + \delta_{3K_{s-AS-SaMP}}}{\delta_{3K_{s-AS-SaMP}}(1 - \delta_{3K_{s-AS-SaMP}})} \sigma \end{aligned} \tag{4.6}$$

Corollary 2. (*Approximate recovery from signal and noise perturbations*): Consider $\mathbf{h} \in \mathbb{R}^N$ as a compressible K-sparse signal. Let $\mathbf{h}_K$ represent the K most significant entries. The signal $\mathbf{h}$ is compressibly sparse if $\mathbf{h} - \mathbf{h}_K \neq \mathbf{0}$. With the same assumption of **Theorem 2**, if $\mathbf{A}$ satisfies the RIP with parameter $\delta_{6K_{s-AS-SaMP}} < 0.03$, the reconstruction distortion of the AS-SaMP algorithm is written as below:

$$\| \mathbf{h} - \hat{\mathbf{h}} \| \leq \frac{1 + \delta_{6K_{s-AS-SaMP}}}{\delta_{6K_{s-AS-SaMP}}(1 - \delta_{6K_{s-AS-SaMP}})} (\sigma + \sqrt{\frac{1 + \delta_{6K_{s-AS-SaMP}}}{K}} \| \mathbf{h} - \mathbf{h}_K \|_1) \tag{4.7}$$

The proofs of the **Theorem 2** and **Corollary 2** are similar to the corresponding theorem and corollary in [125] as the AS-SaMP is equivalent to SaMP with the estimated sparsity $K_{s-AS-SaMP}$ at the last stage except for the different contents of candidate and

final support set which does not affect stability of AS-SaMP under both signal and noise perturbations.

4.4 Computational Complexity of the Existing Algorithms in the Literature

In this section, the computational complexity of the existing algorithms in the literature and the AS-SaMP algorithm is compared in terms of the number of operations, which equals the product of the number of operations per iteration and the number of iterations. Generally, each algorithm performs six major steps during each iteration[16]: forming signal proxy, identifying the largest indices, merging the set of indices, approximating the signal on the merged set of indices by least-squares, pruning to obtain next approximation, and updating the residual which is the part of the signal that has not been approximated [52, 63]. Assume that each step involves the standard technique[17]; the dimension of the measurement matrix is $P \times L$ and the sparsity level is K. According to [63], the operation counts per iteration for the corresponding steps are $O(LP)$, $O(L)$, $O(K)$, $O(K^2P)$,$O(K)$, and $O(KP)$, where $O(\cdot)$ represents the big-O notation.[18] Among the steps of each algorithm, the LS estimation dominates the contribution to the complexity unless L is much larger than K^2. In addition, the complexity of the algorithms also depends on the number of iterations. Since the number of stages of

[16]The OMP algorithm involves five major steps during each iteration: forming signal proxy, identifying the largest indices, merging the set of indices, approximating the signal on the merged set of indices by least-squares, and updating the residual [84].

[17]Fast implementation exists for each step of all the algorithms, e.g., in the LS estimation step, iterative methods such as Richardson's iteration [63] or Cholesky decompositions [126] can be used for efficiency. However, we adopt the standard technique in each step for all the algorithms.

[18]As a tight bound and a lower bound (represented by Ω and Θ-notations, repectively) of each step in the algorithm may vary from one iteration to another, and the overall complexity is dominated by the step which has the largest complexity, the big-O notation can be used to show a valid upper bound of the algorithm's complexity.

SaMP is upper bounded by $\lceil K/s \rceil$ [64], and during each stage a portion of coordinates in the true support set are identified and refined via up to K iterations, an upper bound of the number of iterations is $\lceil K/s \rceil K$. On the other hand, due to the fast convergence of the AS-SaMP algorithm, fewer stages are required to provide the same quality of estimates, and as the computational complexity of each step is the same, the AS-SaMP algorithm is less complex when compared with the SaMP algorithm. Hence, an upper bound of the number of iterations of AS-SaMP is also $\lceil K/s \rceil K$. Note that the upper bounds obtained for SaMP and AS-SaMP are quite loose, as the number of iterations which varies from a stage to another is likely to be equal to or smaller than $s(\ll K)$ for most of the stages. Thus, we present an improved upper-bounded number of iterations for the AS-SaMP algorithm in Table 4.1 (see **Lemma 2** in Section 4.3.1 for the proof) and we show that the upper-bounded number of iterations is smaller than that for the SaMP algorithm in **Corollary 1**, Section 4.3.1. Moreover, according to [84], the number of iterations which the OMP algorithm involves is upper bounded by the number of the element in the target signal with the amplitudes larger than $\frac{2b}{1-(2K-1)\mu}$, where b is the upper bound of the ℓ_2 norm of the observation noise and μ is the mutual coherence of the measurement matrix (see the definition in Eq. (5.1), Section 5.1). For CoSaMP, the number of iterations is upper bounded by $\log(\|\mathbf{h}\|/\epsilon_1)$ in which ϵ_1 is a precision parameter of the reconstructed signal [63]. A summary of the number of iterations of the considered algorithms is provided in Table 4.1.

4.5 Simulation Results

In this section, simulation results are presented to illustrate the performance of the proposed algorithm for sparse channel estimation in OFDM communication systems. In order to evaluate the performance, MSE is adopted as the metric to quantize the

Table 4.1: The number of iterations

Methods	
OMP†	$card(\mathbf{h})$ with $\lvert h_{\min}\rvert > \frac{2b}{1-(2K-1)\mu}$
CoSaMP	$\log(\lVert\mathbf{h}\rVert/\epsilon_1)$
SaMP‡	$\leq [-J\log(\frac{\lvert h_{\min}\rvert}{\lVert\mathbf{h}\rVert})\frac{-1}{\log(C_{K_{s-SaMP}})}+J]PL$
AS-SaMP‡	$\leq [-J\log(\frac{\lvert h_{\min}\rvert}{\lVert\mathbf{h}\rVert})\frac{-1}{\log(C_{K_{s-AS-SaMP}})}+J]PL$

† $h_{\min}$ is the non-zero element with the minimum magnitude in $\mathbf{h}$.
‡ $\mathbf{h}$ is the target signal and J is the total number of stages. $C_{K_{s-SaMP}}$ and $C_{K_{s-AS-SaMP}}$ are RIP-related parameters for SaMP and AS-SaMP, respectively (see Section 4.3.1 for a detailed explanation).

channel estimation error. One definition of MSE is

$$MSE = E[\sum_{m=1}^{N} |H(m) - \hat{H}(m)|^2]. \tag{4.8}$$

4.5.1 Simulation Setup

A block diagram of the simulation setup is shown in Fig. 4.2. We consider UWA channel estimation for a coded OFDM transmission[19] with $N = 1024$ subcarriers and bandwidth of $B = 9.8$ kHz, leading to a subcarrier spacing of $\Delta f = 9.5$ Hz. The CP duration equals 26 ms, which corresponds to the length of CP $N_{CP} = 256$. Unless otherwise mentioned, the number of pilots is $P = 256$. The data symbols are drawn independently from a 16-QAM constellation and are coded using a $(1024, 512)$ binary low-density parity-check (LDPC) code. We consider the channel model described in (4.1) with $N_p = 15$ multipaths, in which the inter-arrival times are exponentially distributed with a mean of 1 ms, i.e., $E[\tau_{j+1}-\tau_j] = 1$ ms, $j \in \{0, 1, ..., N_p-1\}$. The amplitudes are Rayleigh distributed with the average power decreasing exponentially with the delay, and the difference between the amplitudes at the beginning and the end of the CP is

[19]The proposed algorithm can also be applied to channel estimation in other communication systems, provided a sparse channel representation.

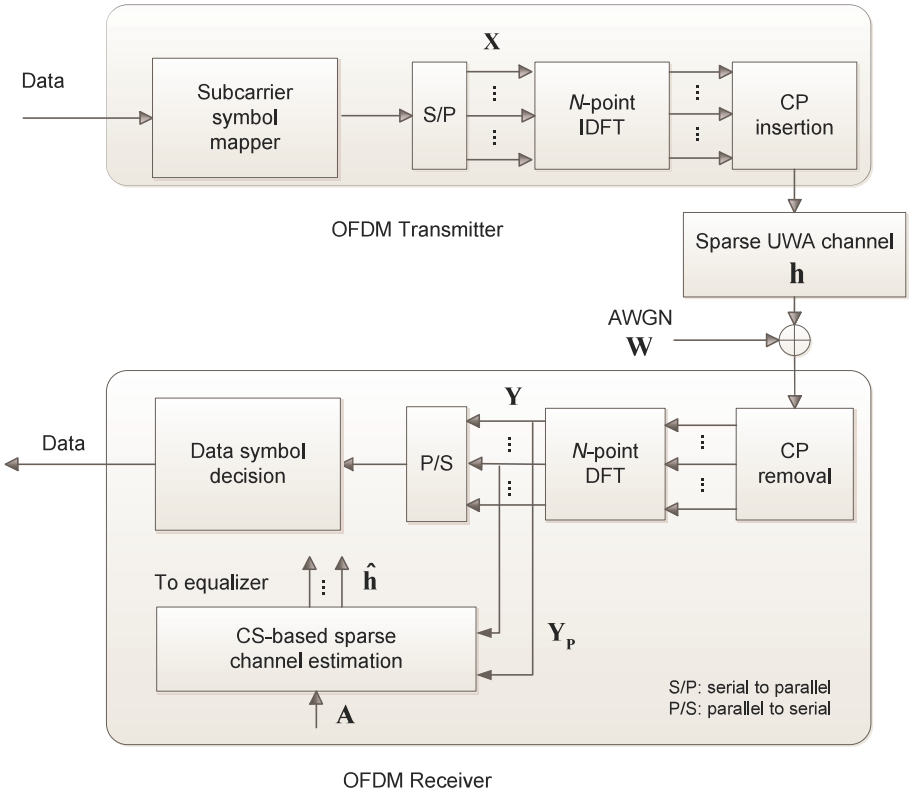


Figure 4.2: Simulation setup.

20 dB. These parameters are assumed to be constant within an OFDM symbol.

In Chapter 4 and 5, the aforementioned OMP-like methods as well as an AMP-like method, namely EM-BG-AMP [83] are considered. The parameters for the considered OMP-like reconstruction algorithms are given in Table 4.2. For EM-BG-AMP, the maximum number of EM[20] iterations is 200 and the convergence tolerance for EM is 10^{-5} [83]. For SaMP and AS-SaMP, since there is a trade-off between the initial step size s and the reconstruction speed, three choices of the s value ($s \leq N_p$) are used, which correspond to a small, medium and large step size. Also, the effect of various choices of Γ (0.01, 0.1, 1 and 10) on the MSE and CPU running time performance of AS-SaMP is evaluated in [26]; the results showed that when $\Gamma > 1$, the MSE of

[20]In EM-BG-AMP, the *a priori* distributions of the target signal and noise are unknown and learned through the EM algorithm [83].

Table 4.2: Parameters of the selected MP-type algorithms

Name	**MaxIter**[‡]	**Sparsity level** K	**Step size s**	**Tolerance** ϵ	**Threshold** Γ	**Ratio** u
OMP	20	not required	not required	norm(Noise)[§]	not required	not required
CoSaMP	20	15	not required	norm(Noise)	not required	not required
SaMP	not required	not required	1, 6, 8	norm(Noise)	not required	not required
VSStAMP	not required	not required	initially 1, 6, 8	norm(Noise)	not required	$\leq \frac{1}{4}$
AS-SaMP	not required	not required	initially 1, 6, 8	norm(Noise)	1	not required

[‡] Maximum iterations.
[§] norm($\mathbf{V}$)=$\sqrt{\sum |\mathbf{V}|^2}$.

the AS-SaMP algorithm starts to saturate. Therefore, the threshold Γ is set to 1. Moreover, among the compared algorithms, only CoSaMP requires the sparsity level as *a priori* information, and the same stopping criterion is used for fairness, i.e., all algorithms stop when the signal residual falls below ϵ. The MSE and BER are used to measure the channel estimation accuracy and the system performance, respectively. The CPU running time[21] is used to provide a rough estimation of the channel estimation computational complexity. Simulations are performed in MATLAB R2014a using a 2.8 GHz Intel Core i7 CPU with 8 GB of memory storage, and we use 10^4 Monte-Carlo trials to average the results. Performance of the proposed reconstruction algorithm is shown next.

4.5.2 Performance of the AS-SaMP Algorithm

First, we compare the AS-SaMP algorithm with two classic recovery algorithms, namely LS and OMP using different numbers of randomly distributed pilots. Fig. 4.3 shows the MSE of these algorithms versus SNR. As the number of pilots increases, MSE decreases for all algorithms. It is worth noting that OMP has, in general, a better MSE performance than LS for the same number of pilots. Similarly, AS-SaMP achieves a better MSE performance than the OMP algorithm. For example, at SNR = 15 dB

[21]The running time is used instead of the number of iterations for comparison, as the complexity in each iteration varies for different algorithms. However, it is worth noting that different hardware configurations may result in different running time measurements.

and $P = 64$, the MSE for LS, OMP and AS-SaMP algorithms are 9×10^{-2}, 1.2×10^{-2} and 3.5×10^{-3}, respectively. In other words, for the same level of MSE performance, the AS-SaMP algorithm uses fewer pilots than the other two algorithms.

Next, Figs. 4.4, 4.5 and 4.6 plot the MSE, BER and the CPU running time for all the algorithms, respectively. To get an idea of the potential MSE gain achieved by exploiting the sparsity into channel estimation, we compare the MSE performance of the considered algorithms with the MSE lower bound of an ideal channel estimator with known indices of the non-zero entries of $\mathbf{h}$ for the OMP-like algorithms[22]. According

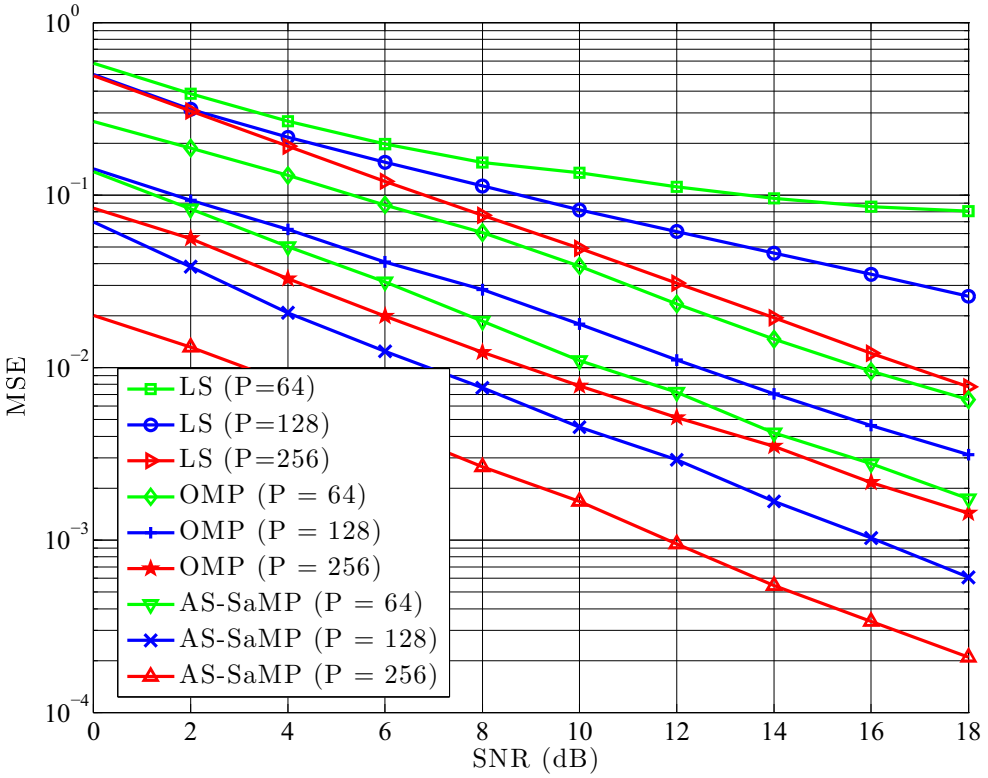


Figure 4.3: MSE performance of the LS, OMP and AS-SaMP algorithms with various number of pilots.

[22]The ideal channel estimation is formed through $\hat{\mathbf{h}} = \mathbf{A_T}^{\dagger}\mathbf{Y_P}$, where $\mathbf{A_T}$ is the submatrix obtained by extracting N_p columns of $\mathbf{A}$ corresponding to the known indices set $\mathbf{T}$. Thus, the lower bound of the MSE for the matching pursuit algorithms is $\frac{{N_p}^2\sigma^2}{trace(\mathbf{A_T}^H\mathbf{A_T})}$ [35], where σ^2 is the noise power and $trace(\mathbf{x})$ is defined to be the sum of the elements on the main diagonal of $\mathbf{x}$.

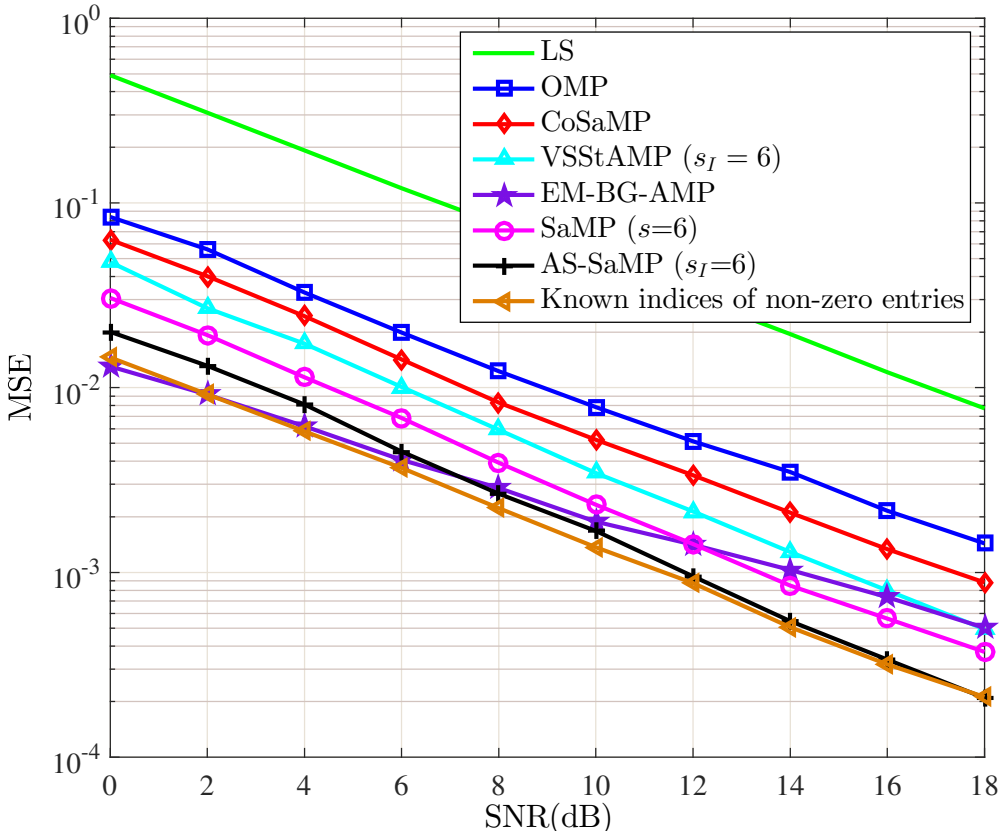


Figure 4.4: MSE performance of the LS, OMP, CoSaMP, SaMP, VSStAMP, AS-SaMP and EM-BG-AMP algorithms. The lower bound for known indices of the non-zero entries of **h** is also included.

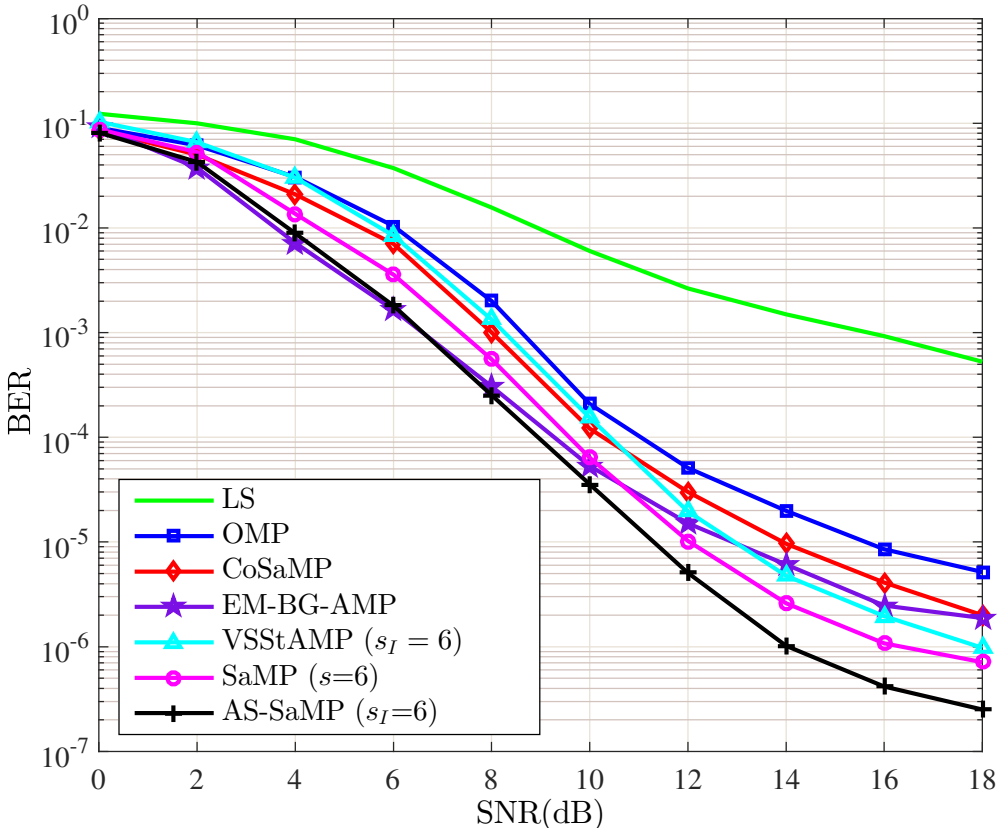


Figure 4.5: BER performance of the LS, OMP, CoSaMP, SaMP, VSStAMP, AS-SaMP, and EM-BG-AMP algorithms.

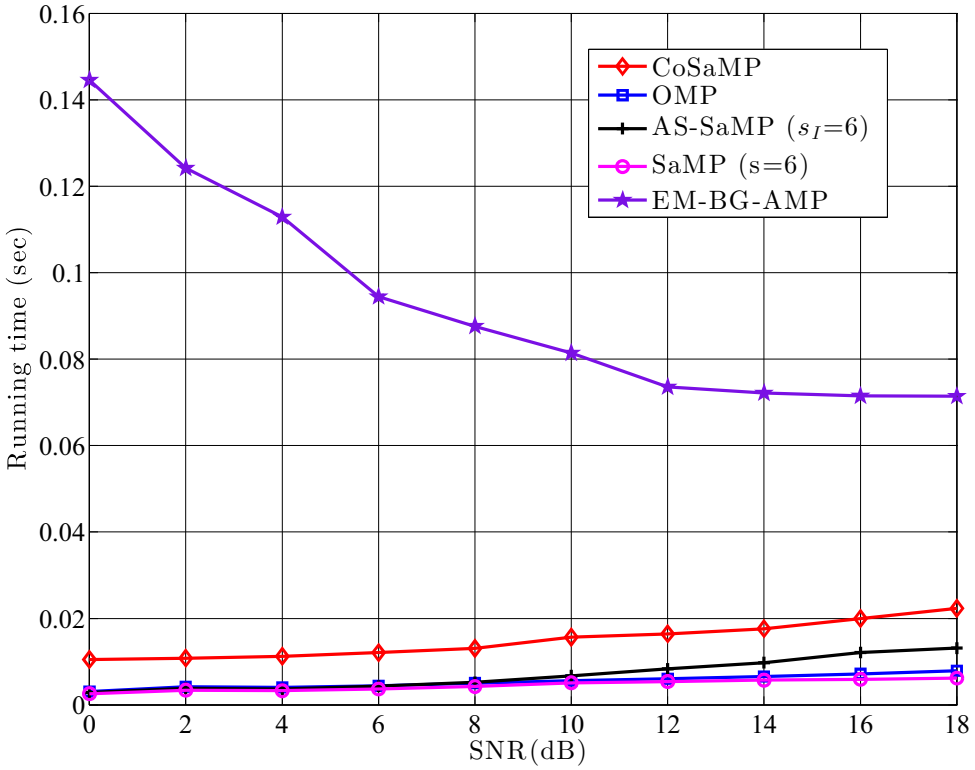


Figure 4.6: CPU running time of the OMP, CoSaMP, SaMP, AS-SaMP, and EM-BG-AMP algorithms.

to results in Figs. 4.4 and 4.5, the CS-based channel estimators give better MSE and BER performance than the conventional LS estimator. In other words, the channel estimators based on the VSStAMP, SaMP and AS-SaMP algorithms outperform those based on the OMP and CoSaMP algorithms in the sense that the former algorithms offer the same performance even if fewer pilots were used. It can also be seen that, although EM-BG-AMP slightly outperforms the other algorithms for lower SNRs (SNR < 8 dB), the proposed algorithm outperforms the other algorithms for higher SNRs (SNR ≥ 8 dB). Moreover, given the same sparsity level, the MSE of AS-SaMP is closer to the aforementioned lower bound than that of the other algorithms at higher SNRs. As shown in Fig. 4.6, the CPU running time of EM-BG-AMP is significantly larger than those of the other algorithms due to the iterative statistical parameters learning process via EM, and at lower SNRs, EM-BG-AMP requires a larger number of iterations for EM. Furthermore, it is noted that the complexity of the CoSaMP algorithm is higher than that of other algorithms; this can be explained through Table 4.3 in which the running time of each step per iteration for CoSaMP is higher than that for other algorithms. This can be further explained as the size of the support set for CoSaMP is larger $(2K)$ than the support set size for the other algorithms. Additionally, the complexity of the AS-SaMP algorithm is higher than that of the SaMP; this is because the step size is reduced during the fine tuning stages given the same initial step size in AS-SaMP. Moreover, from Table 4.3, it is noteworthy that the LS approximation step dominates the contribution to the total running time per iteration among the steps of all algorithms.

Figs. 4.7 and 4.8 depict the MSE performance and computational complexity of the AS-SaMP and SaMP algorithms with different step sizes, respectively. As can be seen, for a medium or large initial step size ($s = s_I = 6$ or $s = s_I = 8$), AS-SaMP outperforms SaMP with a small increase in complexity, while for $s = s_I = 1$, the same performance

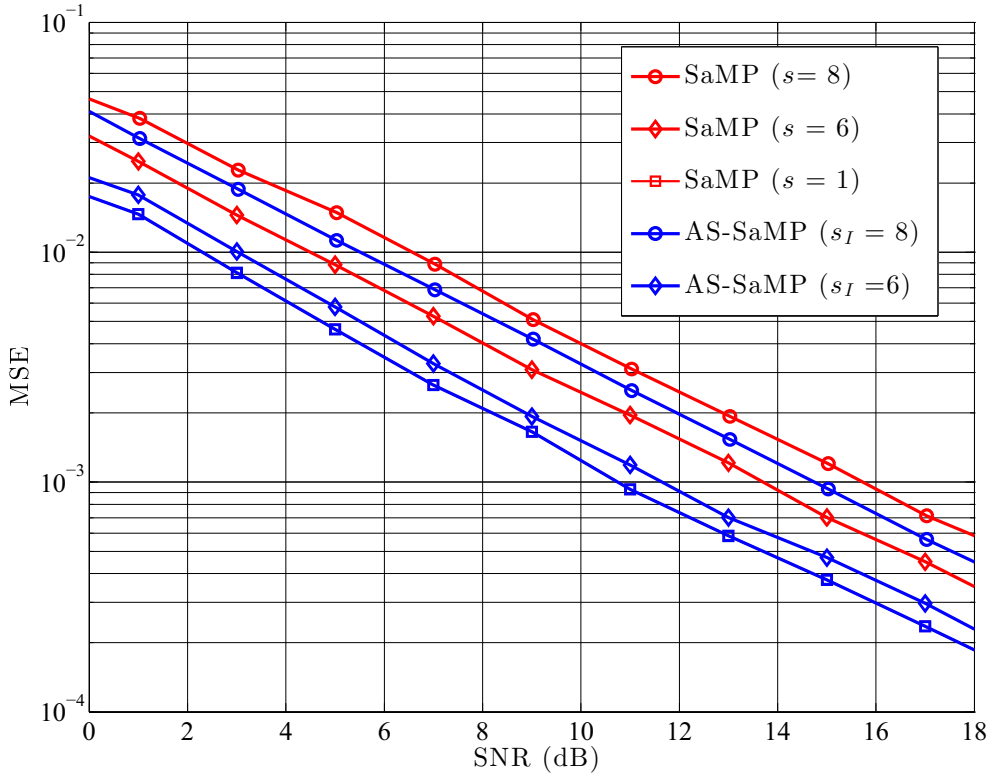


Figure 4.7: MSE performance of the SaMP and AS-SaMP algorithms with different step sizes.

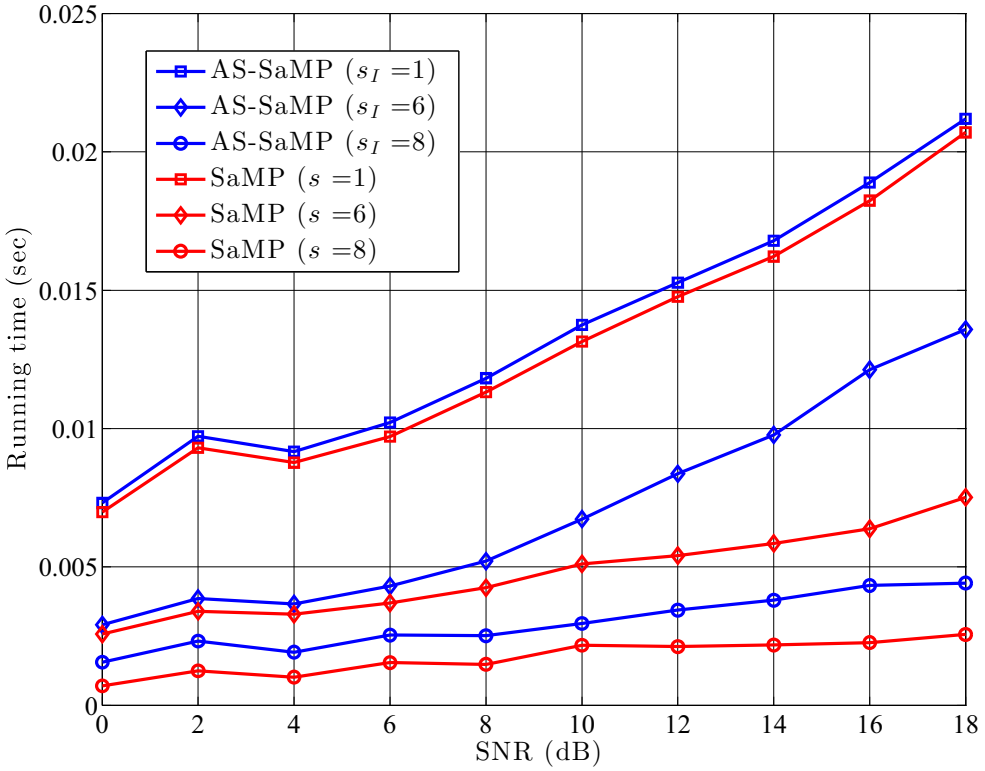


Figure 4.8: CPU running time of the SaMP and AS-SaMP algorithms with different step sizes.

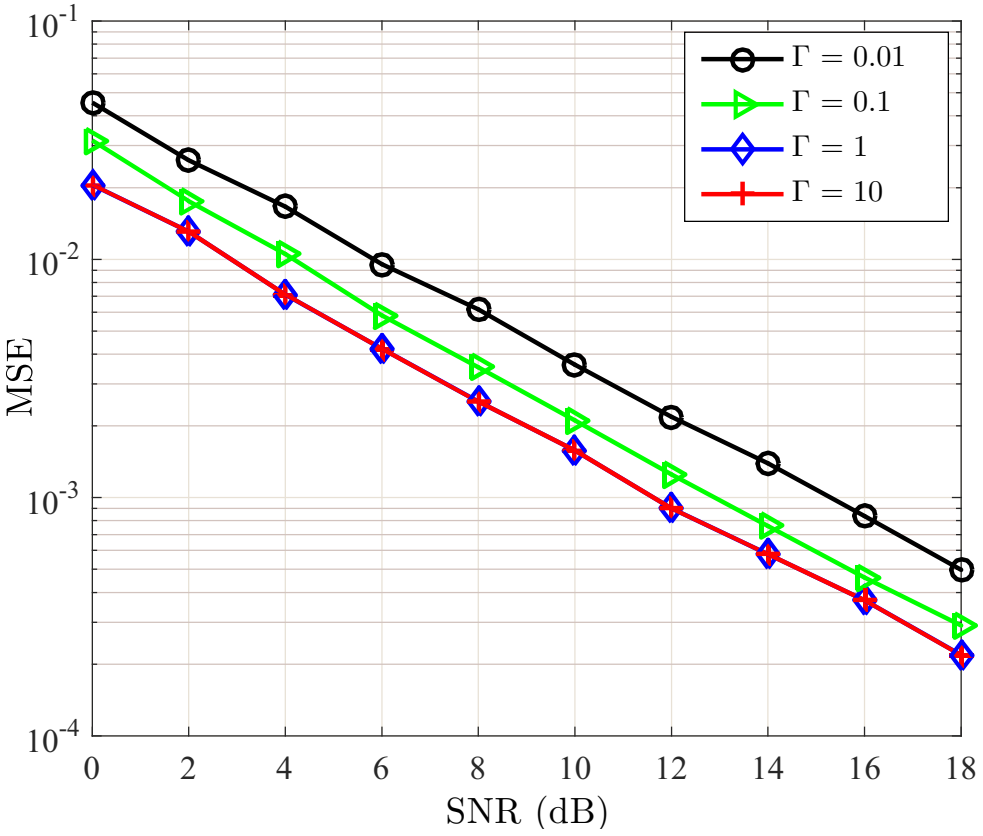


Figure 4.9: MSE performance of AS-SaMP for various Γ values, and with the initial step size $s = 6$.

Table 4.3: CPU running time (sec) of each step per iteration for CS-based algorithms, SNR = 10 dB and $P = 256$.

Step	**OMP**	**CoSaMP**	**SaMP** ($s = 6$)	**AS-SaMP** ($s_I = 6$)
Form proxy	4.21×10^{-5}	8.28×10^{-5}	5.11×10^{-5}	5.06×10^{-5}
Identification	7.23×10^{-6}	2.36×10^{-5}	1.80×10^{-5}	1.84×10^{-5}
Support merger	1.14×10^{-5}	1.80×10^{-5}	7.80×10^{-6}	8.03×10^{-6}
LS approximation	3.32×10^{-4}	4.4×10^{-3}	7.9×10^{-4}	8.46×10^{-4}
Pruning	NA*	9.15×10^{-5}	7.99×10^{-6}	1.53×10^{-5}
Residual update	1.56×10^{-5}	4.80×10^{-5}	3.58×10^{-5}	5.50×10^{-5}

* NA: not applicable.

Table 4.4: CPU running time (sec) for AS-SaMP with various Γ at SNR = 15 dB and the initial step size $s = 6$.

$\Gamma = 0.01$	$\Gamma = 0.1$	$\Gamma = 1$	$\Gamma = 10$
0.008	0.01	0.062	0.079

is achieved using a slightly larger CPU running time for AS-SaMP. This can be easily explained, as when $s = s_I = 1$, AS-SaMP becomes equivalent to SaMP except an additional criterion for changing stages. Note that SaMP and AS-SaMP require $s \leq N_p$ and $s_I \leq N_p$, respectively, to avoid overestimation. In general, the AS-SaMP algorithm is more accurate without significantly increasing the complexity of the estimation.

Finally, the impact of the threshold Γ on the MSE and CPU running time is investigated for the proposed algorithm. Fig. 4.9 and Table 4.4 show the MSE and CPU running time for various Γ values for the initial step size, $s = 6$, respectively. As seen, a larger Γ leads to a smaller MSE and a relatively high complexity; on the other hand, a smaller Γ leads to a larger MSE and a relatively low complexity. This is consistent with the discussion in Section 4.2.2. Moreover, it is noted that the MSE curves of AS-SaMP with $\Gamma = 1$ and $\Gamma = 10$ overlaps. This can be explained as when $\Gamma \geq 1$, fine tuning starts when the algorithm moves from the first stage to the second, and the performance starts to saturate.

4.6 Conclusion

In this chapter, an adaptive step size SaMP algorithm, namely AS-SaMP, is proposed and analyzed for application to sparse channel estimation in OFDM systems. As the name suggests, it features an adaptively adjusted step size without requiring *a priori* information of the estimated channel sparsity. Simulation results demonstrate that the proposed algorithm provides an improved MSE and BER performance when compared with the conventional LS and other CS-based reconstruction algorithms, without significantly increasing the computational complexity. Therefore, it can be concluded that the AS-SaMP algorithm offers a better trade-off between performance and complexity. In the subsequent chapter, it is demonstrated that the proposed algorithm can be adopted in sparse time-varying channel estimation.

Chapter 5

Near-Optimal Pilot Placement Scheme in OFDM Systems

In this chapter, the pilot placement scheme based on the MIP for sparse channel estimation in OFDM systems is investigated. Although a brute-force search guarantees the optimal pilot placement, it is prohibitive to examine all possibilities due to high computational complexity. It is known that by minimizing the mutual coherence of the measurement matrix when the signal is sparse on the unitary DFT matrix, the optimal set of pilot locations is a CDS. Based on this, an efficient near-optimal pilot placement scheme is proposed in cases where CDS does not exist. Simulation results demonstrate the effectiveness of the proposed schemes, and moreover, show that the previously proposed AS-SaMP estimation algorithm, with the new pilot placement scheme, achieves a better trade-off between the performance—in terms of MSE and BER—and complexity, when compared to the other aforementioned estimation algorithms. More importantly, the proposed estimation algorithm and the pilot placement scheme can also be applied to channel estimation in other communication systems, provided a sparse channel estimation and a DFT-based measurement matrix.

5.1 Problem Statement

According to the CS theory, an accurate recovery of a sparse signal relies on the RIP of the measurement matrix. However, the RIP evaluation for a particular matrix is an NP-hard problem [42]. An alternative property which evaluates if a measurement matrix can preserve well the information of the sparse signal in the measurements is the mutual coherence of the measurement matrix [40–42]. According to Eq. (4.3), the measurement matrix is the product of the transmitted pilots and the DFT submatrix and is determined by both the symbols on the pilot subcarriers and the set of pilot location indices, which is also referred to as the pilot placement/arrangement. This chapter focuses on the pilot placement by assuming that the same symbol is transmitted on all the pilot subcarriers. As mentioned before, a common property to theoretically evaluate the reliability of a certain measurement matrix is the MIP. The mutual coherence of a $P \times L$ measurement matrix $\mathbf{A}$ is defined as the maximum absolute correlation between any two normalized columns, which is

$$\mu(\mathbf{A}) = \max_{1 \le i,j \le L, i \ne j} \frac{| < \boldsymbol{a}_i \cdot \boldsymbol{a}_j > |}{\|\boldsymbol{a}_i\| \cdot \|\boldsymbol{a}_j\|}. \tag{5.1}$$

Given equal-power pilots, and substituting $\mathbf{A}$ with $\mathbf{X_p D_p}$, (5.1) becomes

$$\mu(\mathbf{A}) = \max_{1 \le i,j \le L, i \ne j} |X(p_c)|^2 \frac{| < \mathbf{d}_{\mathbf{p}i} \cdot \mathbf{d}_{\mathbf{p}j} > |}{\|\mathbf{d}_{\mathbf{p}i}\| \cdot \|\mathbf{d}_{\mathbf{p}j}\|}, \tag{5.2}$$

where $c = 1, 2, ...P$ and $\mathbf{d}_{\mathbf{p}_i}$ denotes the ith column of $\mathbf{D_p}$ with the jth element given by $\frac{1}{\sqrt{N}} e^{-j2\pi i p_j / N}$, $j = 1, 2, ..., P$. We aim to find the set of pilot location indices $\boldsymbol{\Omega} = \{p_1, p_2, ..., p_{\mathrm{P}}\}$ which minimizes $\mu(\mathbf{A})$. Although the optimal pilot pattern can be obtained through exhaustive searching over all possible patterns, it is computationally prohibitive to form the search space for a large number of subcarriers and pilots. Sev-

eral methods have been suggested to search the suitable solutions iteratively [86,90–92]; however, the complexity of these methods is potentially high because the search space grows rapidly (even exponentially) as the numbers of subcarriers and pilots increase. In [93], the authors identify a collection of matrices formed by deterministic selection of rows of Fourier matrices which satisfy RIP. In particular, the selected indices of the rows correspond to the integer outputs of certain polynomial functions [93]. Here, a novel pilot placement scheme is proposed, which aims to provide a near-optimal solution without suffering from the fast-growing complexity. Next, an analysis of optimal pilot pattern based on CDS is presented.

5.2 Analysis of Optimal Pilot Pattern Based on CDS

Assume that the measurement matrix $\mathbf{A}$ is composed by P rows of $\mathbf{D}$ where the (m, n) element of $\mathbf{D}$ is given by $[\mathbf{D}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$ $(0 \leq m \leq N-1$ and $0 \leq n \leq L-1)$, and the indices set of the selected rows is $\mathbf{\Omega}$. All the pilot symbols are equal-powered to be $E_P = |X(p_c)|^2$, $c = 1, 2, ...P$. According to [90–92], if $\mathbf{\Omega}$ is a CDS with parameters, then the mutual coherence of the resulted measurement matrix, $\mu(\mathbf{A})$, is minimized. Specifically, recall the definition of a CDS [127],

Definition 2. Cyclic Difference Set (CDS): Let $\mathbf{\Xi}$ be a finite additive Abelian group of order N. The P-element subset $\mathbf{Q}$ is called a (N, P, λ) CDS in $\mathbf{\Xi}$ if the list of difference $(q(i) - q(j))$ modulo N, $i, j = 1, 2, ...P$, and $i \neq j$, represents each nonidentity elements in $\mathbf{\Xi}$ exactly λ times.

For example, the $(7, 3, 1)$ CDS is $\mathbf{Q} = \{1, 2, 4\}$, which satisfies that any integer between 1 and 6 will occur and repeat exactly once in the set $\{q(i) - q(j)$ (modulo $7) | 1 \leq i \neq j \leq 3\}$.

It should be noted that we have refined the analysis presented in [90] and [91], so that it

is applicable in the general case. As in (5.2), $| < \mathbf{d}_{\mathbf{p}_i} \cdot \mathbf{d}_{\mathbf{p}_j} > |$ only depends on $\Delta_i = i - j$ and $\| \mathbf{d}_{\mathbf{p}_i} \| = \| \mathbf{d}_{\mathbf{p}_j} \| = 1$, designing the optimal pilot pattern can be formulated as:

$$\begin{aligned} \mathbf{\Omega}_{opt} &= \arg\min_{\mathbf{\Omega}} \max_{1 \le \Delta_i \le L-1} E_P | < \mathbf{d}_{\mathbf{p}_i} \cdot \mathbf{d}_{\mathbf{p}_{i+\Delta_i}} > | \\ &= \arg\min \max_{1 \le \Delta_i \le L-1} E_P | \sum_{r=1}^{P} \omega^{p_r \cdot \Delta_i} |. \end{aligned} \tag{5.3}$$

where $\omega = e^{-j\frac{2\pi}{N}}$. To maximize $|\sum_{r=1}^{P} \omega^{p_r \cdot \Delta_i}|$ is equivalent to maximize $|\sum_{r=1}^{P} \omega^{p_r \cdot \Delta_i}|^2 = \sum_{m=1}^{P} \omega^{p_m \cdot \Delta_i} \cdot \sum_{n=1}^{P} \omega^{-p_n \cdot \Delta_i}$, and E_P is a constant, therefore (5.3) can be re-written as:

$$\mathbf{\Omega}_{opt} = \arg\min_{\mathbf{\Omega}} \max_{1 \le \Delta_i \le L-1} \sum_{m=1}^{P} \sum_{n=1}^{P} \omega^{(p_m - p_n)\Delta_i}. \tag{5.4}$$

It is worth noting that $\sum_{m=1}^{P} \sum_{n=1}^{P} \omega^{(p_m - p_n)\Delta_i}$ in above equation is a complex number generally. Thus, to make it applicable in the general case, a revision is made as follows:

$$\begin{aligned} \mathbf{\Omega}_{opt} &= \arg\min_{\mathbf{\Omega}} \max_{1 \le \Delta_i \le L-1} \underbrace{P}_{m=n} + \underbrace{\sum_{m=1}^{P} \sum_{n=1}^{P} Re[\omega^{|p_m - p_n|\Delta_i}]}_{m \ne n} \\ &= \arg\min_{\mathbf{\Omega}} \max_{1 \le \Delta_i \le L-1} \underbrace{P}_{m=n} + \underbrace{\sum_{m=1}^{P} \sum_{n=1}^{P} \cos(\frac{2\pi}{N} |p_m - p_n| \Delta_i)}_{m \ne n}. \end{aligned} \tag{5.5}$$

In above equation, the mutual coherence of the resulting matrix depends not only on the space between two columns but also on the space between two pilots. Define a set $\mathbf{G} = \{(p_m - p_n) \text{ modulo } N | 1 \le m, n \le P, m \ne n\}$ which contains N different numbers $g = 1, 2, ...N-1$ and each number repeats λ_g times. Equation (5.5) can be re-written as:

$$\mathbf{\Omega}_{opt} = \arg\min_{\mathbf{\Omega}} \max_{1 \le \Delta_i \le L-1} E_P(\underbrace{P}_{m=n} + \underbrace{\sum_{g=1}^{N-1} \lambda_g \cos(\frac{2\pi}{N} g \Delta_i)}_{m \ne n}). \tag{5.6}$$

The problem in above equation is equivalent to the problem of finding the optimal pilot patter which minimizes the maximum value of $\sum_{g=1}^{N-1} \lambda_g \cos(\frac{2\pi}{N} g\Delta_i), 1 \leq \Delta_i \leq L-1$. To show that the patterns based on CDS are optimal, $\lambda_1 = \lambda_2 = ... = \lambda_{N-1}$ needs to be satisfied. Because

$$\max_{1 \leq \Delta_i \leq L-1} \sum_{g=1}^{N-1} \lambda_g \cos(\frac{2\pi}{N} g\Delta_i) \geq \frac{\sum_{\Delta_i=1}^{L-1} (\sum_{g=1}^{N-1} \lambda_g \cos(\frac{2\pi}{N} g\Delta_i))}{L-1}, \tag{5.7}$$

equality happens when $\lambda_1 = \lambda_2 = ... = \lambda_{N-1} = \frac{P^2-P}{N-1}$, and thus, the minimum value of the mutual coherence is

$$\begin{aligned} \mu(\mathbf{A})_{min} &= E_P \sqrt{P + \frac{P^2-P}{N-1} \frac{\sum_{\Delta_i=1}^{L-1} (\sum_{g=1}^{N-1} \cos(\frac{2\pi}{N} g\Delta_i))}{L-1}}, \\ &= E_P \sqrt{\frac{PN-P^2}{N-1}}. \end{aligned} \tag{5.8}$$

5.3 Proposed Pilot Placement Scheme Based on the Concatenated CDS

Previously, we have shown that for an N-subcarrier OFDM system using P subcarriers as pilots, the pilot pattern according to the CDS (N, P, $\frac{P^2-P}{N-1}$) is optimal. However, CDS exists only for some specific N and P. In this section, a pilot placement scheme based on the concatenated CDS is proposed for the cases where there is no CDS. According to **Definition 2**, for an existing (N, P, λ) CDS of order P, assuming that $\mathbf{G}$ is the set of cyclic differences of any two elements of the CDS, every non-identity element in the set $\mathbf{G}$ of order N has exactly the same number of repetitions, λ [127]. In other words, if we denote the number of repetitions of the different elements of $\mathbf{G}$ as $\boldsymbol{\lambda_G} = \{\lambda_g | g = 1, 2, ...N-1\}$, then $\lambda_1 = \lambda_2 = ...\lambda_{N-1} = \lambda$ which also means that the variance of $\boldsymbol{\lambda_G}$ is zero. Moreover, it is noticed that a pilot pattern with a smaller

variance of $\boldsymbol{\lambda_G}$ is likely to give a smaller mutual coherence of the resulting measurement matrix, and thus, more accurate estimates. Consider an OFDM system with $N = 1024$, in which $P = 256$ identical pilot symbols are randomly scattered, and the number of taps of the sparse CIR is $L = 256$. To show that as the variance of $\boldsymbol{\lambda_G}$ increases, it is likely that so does the mutual coherence of $\mathbf{A}$ and the MSE of estimates, the Spearman's rank correlation[23] is adopted to measure the strength of a monotonic relationship (i.e., values of elements in a vector either increase or decrease with every increase in an associated vector) between paired vectors [128], and is defined as below.

Definition 3. The Spearman's rank correlation: For two vectors of size V, $\mathbf{A} = [A(1), A(2), ..., A(V)]$ and $\mathbf{B} = [B(1), B(2), ..., B(V)]$, the Spearsman's correlation is calculated from

$$r_s(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^{V}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{V}(a_i - \bar{a})^2 \sum_{i=1}^{V}(b_i - \bar{b})^2}},$$

where a_i and b_i are the positions in the ascending order (ranks) of $A(i)$ and $B(i)$, respectively. $\bar{a}$ and $\bar{b}$ are the means of a_i and b_i, $i = 1, 2, ..., V$.

Table 5.1 shows the Spearman's rank correlation between any pair of the following four vectors: the variance of $\boldsymbol{\lambda_G}$, the mutual coherence $\mu(\mathbf{A})$, and the MSE for both the OMP and AS-SaMP algorithms, obtained based on 10^4 pilot patterns; 10^3 OFDM symbols and 10 dB SNR were considered. From Table 5.1, it can be seen that a smaller variance of $\boldsymbol{\lambda_G}$ tend to correspond smaller $\mu(\mathbf{A})$ and the average MSEs of OMP and AS-SaMP.

It is noted that by concatenating a CDS the variance of the number of repetitions tends to be small. A concatenated CDS is shown in the following example. For the $(7, 3, 1)$ CDS, a concatenated CDS is obtained through $\{1, 2, 4, (1\times7+1), (1\times7+2), (1\times$

[23]The Spearman's rank correlation can take values from 1 to -1, with $1\,(-1)$ indicating that two vectors can be described using a monotonic increase (decrease) function, and 0 meaning that there is no tendency for one vector to either increase (decrease) when the other increases [128].

Algorithm 2 Pilot Placement Based on Concatenated CDS with an Iterative Tail Search

Input: An existing (v, u, a) CDS $\mathbf{C}$ for concatenation, the number of total subcarriers N, the number of pilot subcarriers P, the partial DFT matrix $\mathbf{D}$ of which the (m, n) element is $\frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$, where $0 \leq m \leq N-1$, $0 \leq n \leq L-1$, and L is the number of taps of the CIR;

1: Initialize $\mathbf{\Omega}_c^0 = \emptyset$, $\mathbf{\Omega}_{temp} = \emptyset$
2: **for** i from 1 to $\lfloor \frac{N}{v} \rfloor$ **do**
3: $\quad \mathbf{\Omega}_c^i = \{\mathbf{\Omega}_c^{i-1} \bigcup [\mathbf{C} + (i-1) \times v]\}$
4: **end for**
5: $P_r = P - u \times \lfloor \frac{N}{v} \rfloor$, $\mathbf{\Omega} = \mathbf{\Omega}_c$
6: **for** j from 1 to P_r **do**
7: $\quad \mathbf{\Omega}_{temp} = \mathbf{\Omega}$
8: $\quad$ Form all $P_r - j + 1$ possible subsets of size j by adding an element to $\mathbf{\Omega}_{temp}$: $\mathbf{\Omega} = \{\mathbf{\Omega}_{temp} \bigcup k \in \{P_r + 1, P_r + 2, ..., N\} \backslash \mathbf{\Omega}_{temp}\}$
9: $\quad$ Form the matrix $\mathbf{A}$ by selecting rows of $\mathbf{D}$ for each j-element sets generated from the previous step, and the indices set of the selected rows is $\mathbf{\Omega}$
10: $\quad$ For all $(P_r - j + 1)$ of $\mathbf{A}$ matrices generated from the previous step, calculate the corresponding mutual coherence, and choose the set with the minimum mutual coherence
11: **end for**
12: **return** $\mathbf{\Omega}$

Output: The pilot indices set $\mathbf{\Omega}$

$7+4), (2\times7+1), (2\times7+2), (2\times7+4), ..., (i\times7+1), (i\times7+2), (i\times7+4), ...\}$, where $i \in \mathbb{Z}^+$ and $i \geq 1$. From these observations, we propose a pilot placement scheme based on the concatenated CDS for pairs of (P, N) where CDS does not exist. First, a CDS needs to be chosen for concatenation according to the ratio of the number of pilots to the number of subcarriers, i.e., P/N. Specifically, the existing CDS with the parameters (v, u, a), in which u/v is the closest to P/N, is selected. For instance, to select indices for 256 pilots from 1024 positions, the (133,33,8) CDS is used. After concatenating

Table 5.1: Spearman's rank correlations

	$\text{var}(\boldsymbol{\lambda_G})^{\S}$	$\mu(\mathbf{A})$	The average MSE of OMP	The average MSE of AS-SaMP
$\text{var}(\boldsymbol{\lambda_G})$	1	0.7475	0.7467	0.7527
$\mu(\mathbf{A})$	0.7475	1	0.7481	0.7534
The average MSE of OMP	0.7467	0.7481	1	—
The average MSE of AS-SaMP	0.7527	0.7534	—	1

§ The variance of $\boldsymbol{\lambda_G}$ is $\frac{1}{N-1}\sum_{g=1}^{N-1}(\lambda_g)^2 - (\frac{1}{N-1}\sum_{g=1}^{N-1}\lambda_g)^2$.

the selected CDS, an iterative procedure is adopted to find the rest of pilot positions which minimize the mutual coherence of the resulting measurement matrix. We refer to this as to the iterative tail search; the pseudocode for the proposed scheme is shown as **Algorithm 2**. It is worth noting that the size of the search space is greatly reduced after concatenation, and hence, the proposed method converges significantly faster when compared to the iterative methods in [90–93].

5.4 Simulation Results

In this section, we adopt the same simulation set-up which can be found in Section 4.5. Here, five pilot placement schemes, i.e., random, the procedure in [90], the stochastic sequential search (SSS) in [92], the RIP-based scheme in [93], and our proposed scheme, are considered. Two polynomials are used to generate the pilot indices sets for the RIP-based scheme, i.e., $f_1(n) = 10n + n^2$ and $f_2(n) = 10n + n^2 + n^3$ [93]. Equal power pilots are assumed for all scenarios. With the random scheme, the pilots are selected randomly among all subcarriers and 10^4 trials are generated for averaging the results. With the proposed placement scheme, the pilots are arranged based on the concatenated CDS with an iterative tail search, referred to as C-CDS with TS. Because the selection of the existing CDS depends on the ratio P/N, the (273,17,1) and (133,33,8) CDS are chosen for the cases of $P = 64$ and $P = 256$, respectively.

5.4.1 MSE Performance of the AS-SaMP for Different Pilot Placement Schemes

Fig. 5.1 shows the MSE performance of the OMP and AS-SaMP algorithms with the random and the proposed pilot placement scheme, given $P = 64$. For the randomly placed pilots, error bars are used to indicate the standard deviations of the MSE; this was calculated based on 10^4 indices sets. It can be seen that the proposed method provides a superior channel estimation performance when compared to the random placements, as a reduced mutual coherence μ is obtained. For instance, at SNR = 10 dB, the average MSEs of the random scheme for OMP and AS-SaMP are approximately

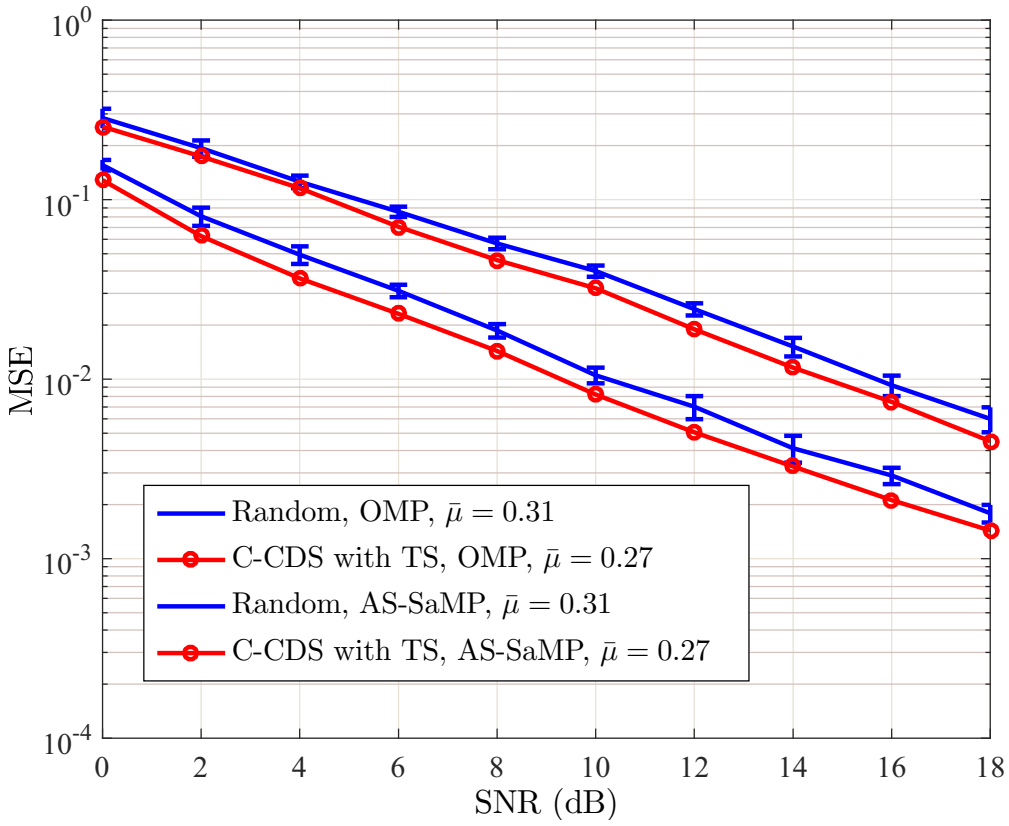


Figure 5.1: MSE performance of the OMP and AS-SaMP algorithms with random and the proposed pilot placement, for $P = 64$. Solid lines are used for OMP and dashed lines for AS-SaMP.

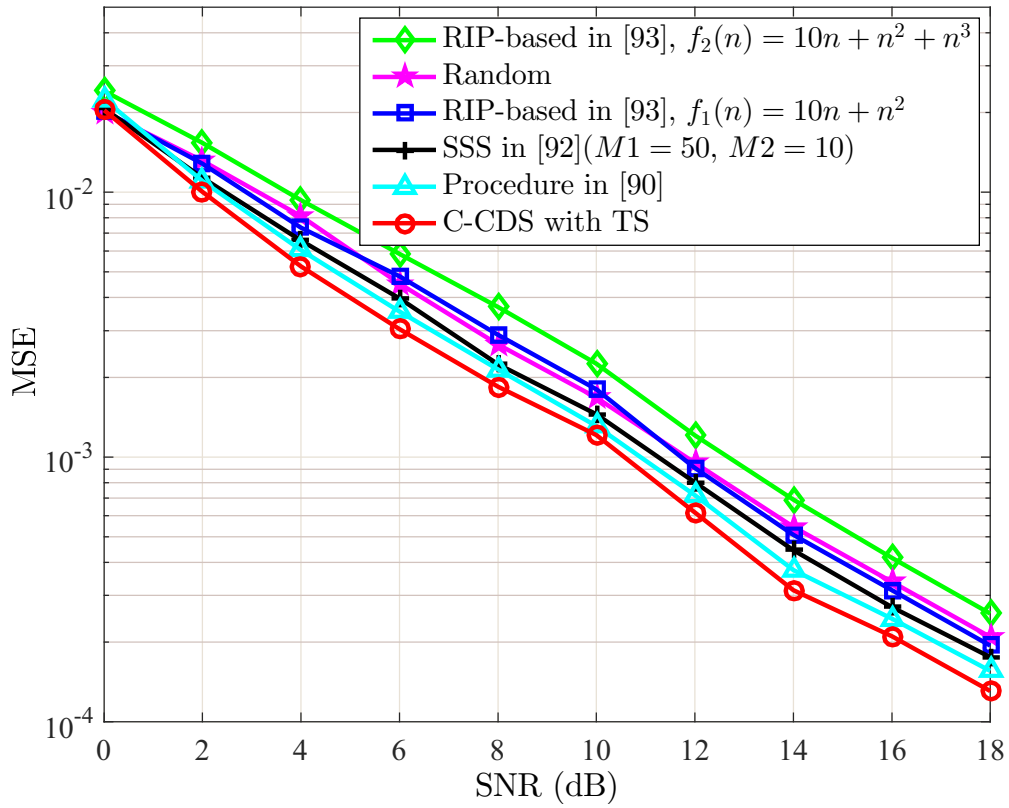


Figure 5.2: MSE performance of the AS-SaMP algorithm for different pilot placements, for $P = 256$.

4×10^{-2} and 1×10^{-2} , respectively, while the MSEs of the proposed scheme for OMP and AS-SaMP are approximately 3×10^{-2} and 8×10^{-3}, respectively. It should be noted that for each SNR, the MSE with the proposed method is smaller than the mean of the MSE minus the standard deviation with the randomly placed pilots. More specifically, it equals approximately the mean minus twice the standard deviation; as such, the proposed method provides a better MSE performance than most of the random pilot arrangements. Also, AS-SaMP achieves a better MSE performance when compared with OMP for both pilot placement schemes.

Fig. 5.2 shows the MSE of the AS-SaMP algorithm for the five previously mentioned pilot placement schemes for $P = 256$. Among them, the proposed C-CDS with

TS has the best performance. Moreover, since the C-CDS pilot arrangement is deterministic, and the iterative search is only conducted for the tail, the searching space is significantly reduced. Therefore, the number of iterations of the proposed method, which is proportional to its computational complexity, is significantly lower than that of the procedure in [90]. An example is provided as follows. When $N = 1024$ and $P = 256$, the procedure in [90] requires $(2N - P + 1)P/2 = 229,504$ iterations, and the SSS in [92] requires $M1 \times M2 \times P = 128,000$ iterations, where $M1$ and $M2$ are the number of the outer and inner loop, respectively. For the proposed scheme, if a (133,33,8) CDS is used for concatenation, there are $1024 - 133 \times \lfloor \frac{1024}{133} \rfloor = 93$ subcarriers at the tail. To search the rest of 25 ($256 - 33 \times \lfloor \frac{1024}{133} \rfloor$) pilot indices which minimize μ, $(2 \times 93 - 25 + 1) \times 25/2 = 2025$ iterations are required.

5.4.2 BER Performance of the AS-SaMP for Different Pilot Placement Schemes

Finally, the BER performance of the overall OFDM system is assessed for different pilot placement schemes, with results shown in Figs. 5.3 and 5.4. In Fig. 5.3, the AS-SaMP algorithm is considered for the five pilot placement schemes. Clearly, AS-SaMP with the proposed C-CDS with TS is slightly better than the other pilot schemes. Fig. 5.4 compares the BER performance of the OMP, CoSaMP, SaMP and AS-SaMP algorithms with the random and proposed pilot arrangements. In general, AS-SaMP with the proposed pilot allocation scheme provides the best BER performance among all the estimation algorithms with the considered pilot placement schemes.

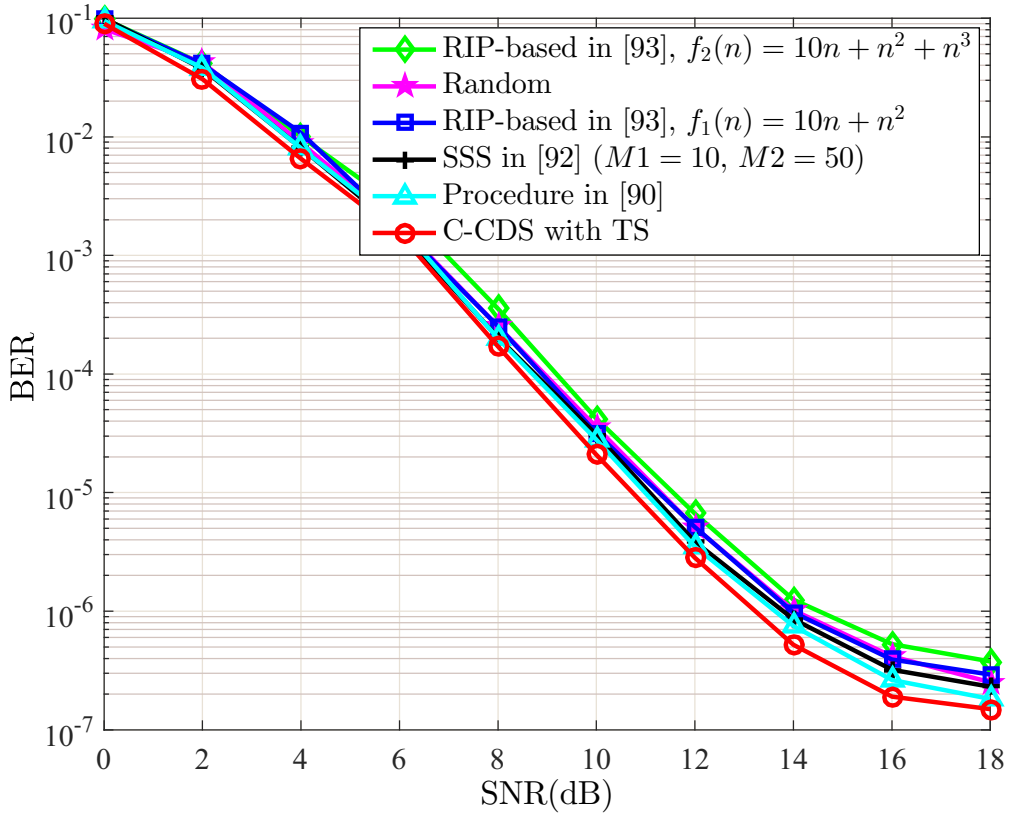


Figure 5.3: BER performance of the AS-SaMP algorithm for different pilot placements, for $P = 256$

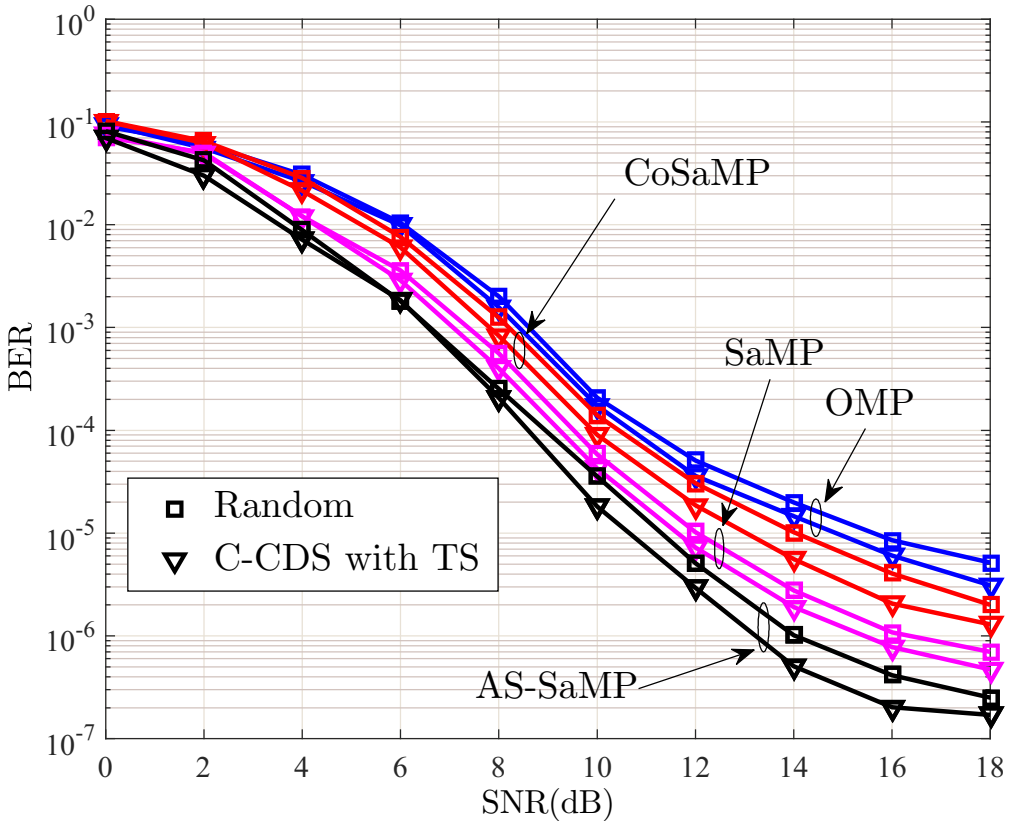


Figure 5.4: BER performance of the OMP, CoSaMP, SaMP, and AS-SaMP algorithms for random and the proposed pilot placement, for $P = 256$. Solid lines are used for OMP, dot lines for CoSaMP, dashed-dot lines for SaMP, and dashed lines for AS-SaMP.

5.5 Conclusion

This chapter presents an analysis of MIP-based optimal pilot placement for sparse channel estimation in OFDM systems. Moreover, a near-optimal pilot placement scheme is proposed, which is based on the concatenated CDS with an iterative tail search (C-CDS with TS). Because the search space of the proposed method is significantly reduced, its complexity is much lower than the iterative procedures in the literature. We have shown the AS-SaMP algorithm with the new pilot placement scheme provides a better MSE performance for the channel estimate, as well as the system BER when compared with existing schemes in the literature, without significantly increasing the computation complexity.

Chapter 6

CS-Based Sparse Estimation and Prediction for Time-Varying Underwater Acoustic Channels

The estimation and prediction of time-varying sparse acoustic communication channels are investigated in this chapter. First, given sufficiently large transmission bandwidth, the channel impulse response of the time-varying UWA channel can have sparse structures, formed by multipath arrivals which can be resolved in the delay and Doppler domains [43, 46, 129]. Such representation is referred to as DDSF [130], and is considered a first-order approximation to the fast time-varying channels in which each channel component is associated with a Doppler shift that is assumed constant within a specific period of time [129]. The aforementioned sparse channel estimation techniques based on MP algorithms can be used to efficiently identify the dominant components in the channel's DDSF representation.

Next, we consider several channel estimation-based equalizers in the literature to investigate the impact of channel estimation on overall system performance. In particular,

the residual prediction error decision feedback equalizer (RPE-DFE) in [15], which improves the robustness of the traditional DFE with respect to channel estimation errors, is presented.

Furthermore, an adaptive channel prediction scheme which does not require any statistical *a prior* knowledge of channels and noise is proposed to provide future CSI. The proposed scheme operates on the channel's DDSF, and can reduce the computational complexity when compared with the channel prediction in time domain. To further improve the prediction accuracy, past detected symbols are fed back to assist the proposed predictor with an up-to-date channel estimate. The performance of the proposed channel estimation and prediction approaches are demonstrated through numerous simulation results. Note that the notations used in this chapter are independent with those used in Chapters 4 and 5 for the time-invariant channels.

6.1 Sparse Estimation of the Channel Delay-Doppler Spreading Function

6.1.1 Problem Statement

Considering the sampled complex baseband representation for all signals and CIRs, the received signal is sampled at the symbol rate, and can be expressed as

$$\begin{aligned} y(n) &= \sum_{k=0}^{K-1} h(k,n)x(n-k+1) + w(n), \\ &= \mathbf{h}^T(n)\mathbf{x}(n) + w(n), \end{aligned} \tag{6.1}$$

where the transmitted symbol $x(n)$, whose elements are assumed to be independent and identically distributed random variables drawn from a linear digital signal constellation.

$w(n)$ is the AWGN, and $w(n) \sim \mathcal{N}(0,\, \sigma_w^2)$, where σ_w^2 is the variance of the noise. $\mathbf{x}(n)$ is the column vector at time n with its kth element corresponding to $x(n-k+1)$, $0 \leq k \leq K-1$, and K is the number of sampled delay taps. The acoustic channel is modelled as a discrete time-varying system having a K-by-1 impulse response $\mathbf{h}(n)$. The kth element of $\mathbf{h}(n)$ is given by $h(k,n)$ which represents the response of the system at time n to a unit impulse input at time $(n-k)$. It is known that the linear time-varying channel can be described through its DDSF, which is defined as the Fourier transform of its impulse response with respect to time domain [46]. Let $u(l,k)$ denote the sampled DDSF, where $0 \leq l \leq L-1$, and L is the number of sample Doppler points. Thus, $h(k,n)$ can be written as

$$h(k,n) = \sum_{l=0}^{L-1} u(l,k) e^{j2\pi\nu_l n\Delta t}, \tag{6.2}$$

where $\nu_l = \nu_{min} + l\Delta\nu$ is the lth sampled Doppler frequency, ν_{min} is the minimum Doppler frequency, $\Delta\nu$ and Δt are the sampling intervals in Doppler and time domains, respectively. Substituting $h(k,n)$ in (6.1) with (6.2) yields

$$y(n) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} u(l,k) e^{j2\pi\nu_l n\Delta t} x(n-k+1) + w(n). \tag{6.3}$$

Assuming that $u(l,k)$ remains constant over a received symbol block of which the length is M, the input-output relationship can be written as

$$\mathbf{y}(n) = \mathbf{\Phi}(n)\mathbf{u} + \mathbf{w}(n), \tag{6.4}$$

where

$$\mathbf{y}(n) = [y(n)\ y(n+1)\ ...\ y(n+M-1)]^T,$$

$$\mathbf{w}(n) = [w(n)\ w(n+1)\ ...\ w(n+M-1)]^T,$$

$$\mathbf{u} = [u(0,0)\ u(0,1)\ ...\ u(0,K-1)\ ...\ u(L-1,K-1)]^T.$$

$$\mathbf{\Phi}(n) =$$

$$\begin{bmatrix} x(n)e^{j2\pi\nu_1 n\Delta t} & \cdots & x(n-K+1)e^{j2\pi\nu_1(n-K+1)\Delta t} & \cdots & x(n-K+1)e^{j2\pi\nu_{L-1}(n-K+1)\Delta t} \\ x(n+1)e^{j2\pi\nu_1(n+1)\Delta t} & \cdots & x(n-K+2)e^{j2\pi\nu_1(n-K+2)\Delta t} & \cdots & x(n-K+2)e^{j2\pi\nu_{L-1}(n-K+2)\Delta t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(n+M-1)e^{j2\pi\nu_1(n+M-1)\Delta t} & \cdots & x(n-K+M)e^{j2\pi\nu_1(n-K+M)\Delta t} & \cdots & x(n-K+M)e^{j2\pi\nu_{L-1}(n-K+M)\Delta t} \end{bmatrix}.$$

Let N_p represent the number of dominant components in $\mathbf{u}$,[24] and N_u denote the length of $\mathbf{u}$. Clearly, we have $N_u = K \times L$. $\mathbf{\Phi}(n)$ is referred to as the measurement matrix in the CS theory. In essence, (6.4) represents the received symbols as the weighted sum of delayed and Doppler shifted transmitted sequences; the goal is to estimate $\mathbf{u}$, given $\mathbf{y}(n)$ and $\mathbf{\Phi}(n)$. In particular, $N_p \ll M < N_u$ and thus, the previously mentioned CS-based algorithms can be used to estimate the significant entries in $\mathbf{u}$. Then, the CIR can be calculated using the estimated $\mathbf{u}$ according to (6.2).

6.1.2 Key Parameter Selection

The set of parameters that need to be specified for channel estimation includes the block length M, the Delay and Doppler range, and the number of Delay and Doppler samples, K and L, respectively. Several factors should be considered in selecting M. First, the assumption considered in (6.4) requires $\mathbf{u}$ to remain approximately constant over M received symbols. The validity of this assumption depends on both the value

[24] In the CS theory, N_p is referred to as the sparsity level of the target signal [42].

of M and the fluctuation rate of $\mathbf{u}$.[25] Estimation using an excessively long block data would introduce significant errors due to the Doppler effect [43]. For example, given a tap with a Doppler shift $\nu = 6.25$ Hz and the sampling frequency $f_s = 5000$ Hz, the phase of the tap will change by $\frac{\pi}{2}$ during a time period of 1000 samples. Second, from the perspective of compressed sensing, a large M ($\gg N_p$) will lead to a measurement matrix $\mathbf{\Phi}(n)$ with better orthogonality among its columns. This property is desired in order to obtain a satisfactory estimation accuracy [42, 43]. Therefore, the value of M, which leads to a tradeoff between the accuracy of the estimation and the validity of the aforementioned assumption, needs to be determined.

Furthermore, to calculate the number of Doppler samples L, the Doppler shift range and the Doppler sample interval need to be chosen. The Doppler shift range is determined in order to cover the most significant part of Doppler, and the Doppler sample interval, $\Delta\nu$, is decided such that the columns in the resulting $\mathbf{\Phi}(n)$ possess near-orthogonality, i.e., the correlation between any two normalized columns is sufficiently low.[26] Similarly, the number of Delay samples (also known as the channel length), K, should be chosen to cover the maximum delay spread of a multipath channel.

6.1.3 Computational Complexity of the Considered Estimation Algorithms

Another important performance metric is the complexity of the channel estimation algorithms. Here, several existing MP-based algorithms , i.e., OMP, CoSaMP, ORLSMP, SaMP, and the proposed AS-SaMP algorithms [30] are considered. In Chapter 4, a

[25] The channel fluctuation rate can be reflected from channel coherence time, which highly depends on the propagation environment and transmitted signal frequencies. Typical channel coherence time spans from hundreds of miliseconds to tens of minutes; this requires a receiver to be capable predicting the future CSI since an estimation of CSI based on the training symbols may be outdated when channel varies rapidly.

[26] A lower bound of $\Delta\nu$, i.e. $\Delta\nu > 1.4/(\frac{\pi M}{f_s})$, is derived in [43] to guarantee that the correlation between any two normalized columns in $\mathbf{\Phi}(\mathbf{n})$ is below $1/\sqrt{2}$.

complexity analysis of OMP, CoSaMP, SaMP and AS-SaMP was provided; thus, this section only focuses on the computational complexity of the ORLSMP algorithm, and its comparison with that of the other algorithms.

On the one hand, for CoSaMP, SaMP and AS-SaMP, computation mainly consists of six steps in each iteration[27], i.e., forming the signal proxy, identifying the largest indices, merging the set of indices, approximating the signal on the merged set of indices by least squares, pruning to obtain the next approximation, and updating the residual. It is assumed that each step involves a standard technique, and the dimension of the measurement matrix is M by N_u, and N_p is the sparsity level of the vector to be estimated. The operation counts per iteration for the corresponding steps are $O(MN_u)$, $O(N_u)$, $O(N_p)$, $O(N_p^2M)$, $O(N_p)$ and $O(N_pM)$. Among the steps, the amount of the computations involved in each iteration is mainly dominated by the first and the fourth steps, which is $O(MN_u + N_p^2M)$ (see the analysis in Section 4.4).

On the other hand, for the ORLSMP algorithm, the LS criterion is adopted to select a new column onto which the projection of the residual vector together with all the previously selected columns is maximum. This is different from the previously mentioned MP-based algorithms in the sense that the new columns are selected to maximize the rank-one projection of the residual vector. Therefore, the operation count per iteration for the ORLSMP algorithm is in the order of $O(MN_p^2 + N_uMN_p^2 + N_uN_p)$, given $N_u > M \gg N_p$ (see Appendix B for a description of the ORLSMP algorithm and a list of operation count of each step). Therefore, the amount of computations involved per iteration in the ORLSMP algorithm is increased over that in the other MP-based algorithms.

[27]As mentioned in Section 4.4, there are five steps per iteration in the OMP algorithm.

6.2 Channel Estimation Based Equalization

As a vital component of receiver in practical communication systems, a previously reported equalizer in [15] which takes channel estimation errors into consideration, is presented in this section for the completeness.

6.2.1 System Model

Two common equalization techniques are considered, i.e., linear equalizer (LE) and DFE, as shown in Fig. 6.1. Each consists of a forward finite impulse response (FIR) filter and, in the case of DFE, a feedback FIR filter which filters and feeds back the detected symbols $\hat{x}(n)$. The filter coefficients of the LE are denoted as $\mathbf{h}_{lin}$, and the feedforward and feedback filter coefficients in the DFE are denoted as $\mathbf{h}_{ff}$ and $\mathbf{h}_{fb}$, respectively. Here we focus on training-based channel estimation methods, with partitioned transmitted symbols as shown in Fig. 6.2. For the LE, after passing through the filter, the soft decision estimate of the transmitted signal can be represented as

$$r(n) = \mathbf{h}_{lin}^T \mathbf{y}'(n), \tag{6.5}$$

where $\mathbf{y}'(n) = \begin{bmatrix} y(n-L_f+1) & \cdots & y(n-1) & y(n) \end{bmatrix}^T$ is a vector of received symbols at time n, and L_f is the length of the filter coefficients $\mathbf{h}_{lin}$.

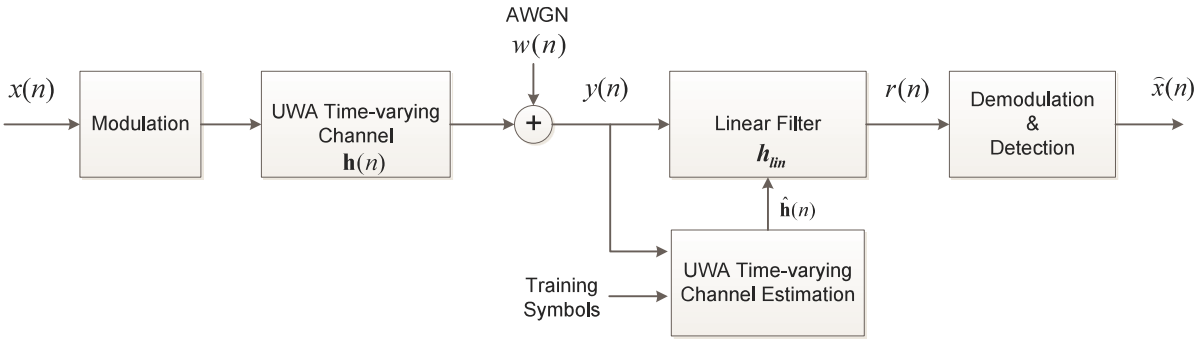
In the case of the DFE,

$$r(n) = \mathbf{h}_{ff}^T \mathbf{y}'(n) + \mathbf{h}_{fb}^T \hat{\mathbf{x}}_{fb}(n), \tag{6.6}$$

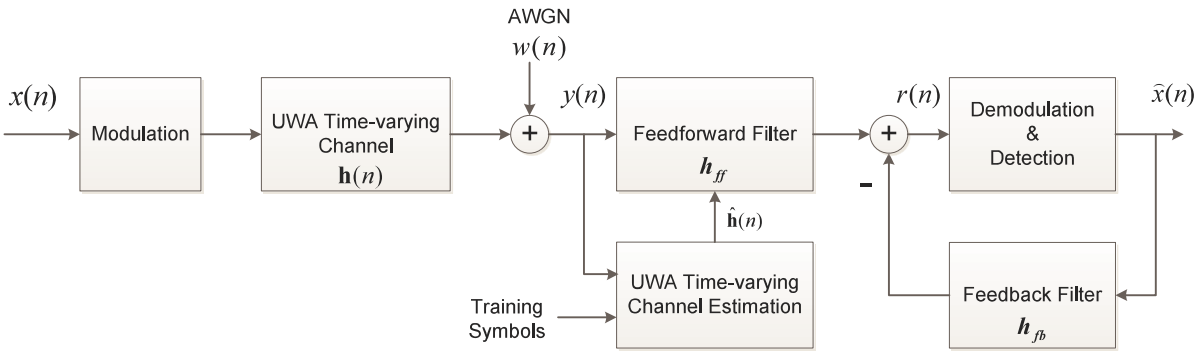
where $\hat{\mathbf{x}}_{fb}(n) = \begin{bmatrix} \hat{x}(n-L_b) & \cdots & \hat{x}(n-1) \end{bmatrix}^T$ is the detected symbol vector, and L_f and L_b are the lengths of $\mathbf{h}_{ff}$ and $\mathbf{h}_{fb}$, respectively. According to (6.1), $\mathbf{y}'(n)$ is expressed as

$$\mathbf{y}'(n) = \underbrace{\begin{bmatrix} h(n-L_f+1,K-1) & \cdots & h(n-L_f+1,0) & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & h(n,K-1) & \cdots & h(n,0) \end{bmatrix}}_{\mathbf{H}^T(n)} \underbrace{\begin{bmatrix} x(n-L_f-K+2) \\ \vdots \\ x(n-1) \\ x(n) \end{bmatrix}}_{\mathbf{x}'(n)} + \underbrace{\begin{bmatrix} w(n-L_f+1) \\ \vdots \\ w(n-1) \\ w(n) \end{bmatrix}}_{\mathbf{w}'(n)}, \tag{6.7}$$

where the subscripts of $h(n,k)$ and $\mathbf{H}^T(n)$ were dropped as (6.7) applies to both linear and DFE. $\mathbf{H}^T(n)$ is the channel matrix with the ith row composed by a properly positioned $\mathbf{h}^T(n-L_f+i)$ along with padded zeros. Representing $\mathbf{H}^T(n)$ by its columns



(a) Linear Equalizer



(b) Decision Feedback Equalizer

Figure 6.1: Block diagrams of linear (top) and decision feedback equalizers (bottom).

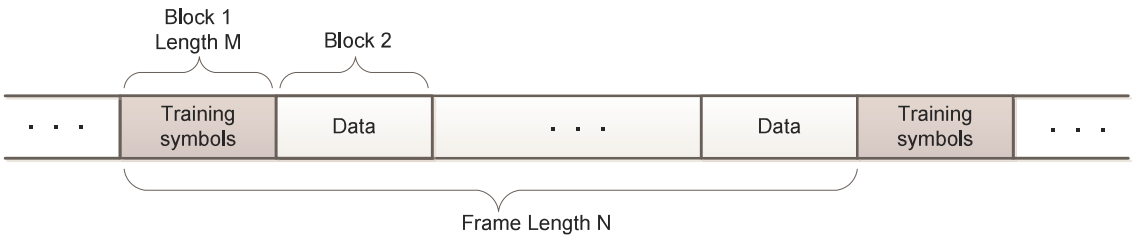


Figure 6.2: The transmitted symbols are partitioned into frames, each containing N transmitted symbols and consisting of multiple blocks of a length M. The first block is used as training symbols for the purpose of channel estimation.

yields

$$\mathbf{H}^T(n) = \begin{bmatrix} \mathbf{c}_{(L_f+K-2)} & \cdots & \mathbf{c}_1 & \mathbf{c}_0 \end{bmatrix}, \tag{6.8}$$

where each column $\mathbf{c}_i$ specifies the contribution from the corresponding transmitted data $x(n-i)$. Moreover, these columns can be partitioned into three parts [15], i.e., $\mathbf{c}_0$, $\mathbf{c}_{fb} = \begin{bmatrix} \mathbf{c}_{L_b} & \cdots & \mathbf{c}_1 \end{bmatrix}$ and $\mathbf{c}_r$ which includes all the remaining columns. In the same manner, the transmitted data vector $\mathbf{x}'(n)$ can be partitioned into three parts: the desired data symbol $x(n)$, the past transmitted symbol vector $\mathbf{x}_{fb}(n) = \begin{bmatrix} x(n-L_b), & \cdots, & x(n-1) \end{bmatrix}$, and the remaining data $\mathbf{x}_r(n)$. Finally, (6.7) can be rewritten as

$$\mathbf{y}'(n) = \mathbf{c}_0 x(n) + \mathbf{c}_{fb}\mathbf{x}_{fb}(n) + \underbrace{\mathbf{c}_r\mathbf{x}_r(n) + \mathbf{w}'(n)}_{\substack{\text{Effective} \\ \text{observation noise}}}. \tag{6.9}$$

The first term in (6.9) represents the received signal corresponding to the desired data symbol, the second term expresses the received portion contributed by the data spanned by the feedback filter, and the last two terms represent the combined signal due to the other data symbols with the noise vector, which is also referred to as the effective observation noise [15]. It is assumed that the data symbols form a zero-mean sequence with variance one, and is uncorrelated with the noise sequence $\mathbf{w}'(n)$. Given perfect

estimate of the noise statistics, the channel estimation based equalizers are expressed as

$$\begin{aligned}\mathbf{h}_{lin} &= \frac{(\hat{\mathbf{Q}} + \hat{\mathbf{c}}_{fb}\hat{\mathbf{c}}_{fb}^H)^{-1}\hat{\mathbf{c}}_0}{1 + \hat{\mathbf{c}}_0(\hat{\mathbf{Q}} + \hat{\mathbf{c}}_{fb}\hat{\mathbf{c}}_{fb}^H)^{-1}\hat{\mathbf{c}}_0} \\ \mathbf{h}_{ff} &= \frac{\hat{\mathbf{Q}}^{-1}\hat{\mathbf{c}}_0}{1 + \hat{\mathbf{c}}_0^H\hat{\mathbf{Q}}^{-1}\hat{\mathbf{c}}_0}, \quad \mathbf{h}_{fb} = \hat{\mathbf{c}}_{fb}^H\mathbf{h}_{ff},\end{aligned} \tag{6.10}$$

where $\hat{\mathbf{Q}} = \mathbf{R}_{\mathbf{w}} + \hat{\mathbf{c}}_r\hat{\mathbf{c}}_r^H$ is the effective noise correlation matrix with $\mathbf{R}_{\mathbf{w}}$ as the covariance matrix of the noise which is independent of the CIR [15]. The filter coefficients in (6.10) are calcuated to minimize the MSE of the soft decision error which is defined as $E[|\epsilon_s(n)|^2]$, where the soft decision error is denoted as $\epsilon_s(n) = r(n) - x(n)$ [15].

6.2.2 Residual Prediction Error based Decision Feedback Equalization

It is worth noting that both the channel estimates and errors in the estimates have an impact on $\epsilon_s(n)$. The performance degradation of the equalizers due to the channel estimation errors results in the need for an equalizer which takes account of channel estimation errors. A DFE based on RPE has been proposed in [15] to alleviate this problem, where the RPE is defined as

$$\mathbf{e}(n) = \mathbf{y}'(n) - \hat{\mathbf{H}}^T(n)\mathbf{x}'(n) = [\mathbf{H}(n) - \hat{\mathbf{H}}^T(n)]\mathbf{x}'(n) + \mathbf{w}'(n). \tag{6.11}$$

Therefore, the RPE can be directly calculated from an output of the equalizer, and can be used to estimate the effective noise correlation matrix $\mathbf{R}_{\mathbf{e}}(n)$. A running average (with a length of L_e) of the correlation matrix of the RPE can be calculated as

$$\mathbf{R}_e(n) = \sum_{m=n-L_e}^{n} \lambda_e^{n-m}\mathbf{e}(m)\mathbf{e}(m)^H, \tag{6.12}$$

with a forgetting factor denoted as λ_e. Then, $\hat{\mathbf{Q}}$ is updated by replacing $\mathbf{R_w}$ with $\mathbf{R_e}(n)$, and the resulting $\hat{\mathbf{Q}}$ is used to calculate the filter weights in (6.10). The equalizer using this method is referred to as RPE-DFE [15]. Although the use of RPE does not completely eliminate the sensitivity to the channel estimation errors, it significantly improves the robustness of the equalizer with respect to channel estimation errors.

6.3 Adaptive Sparse Channel Prediction in Delay and Doppler Domain

In fast-varying channels, an outdated channel estimation may cause errors in equalization and thus degrade the system performance. In practice, adaptive prediction has the advantage of not requiring *a priori* knowledge of the channel and noise statistics. Hence, this section proposes an adaptive channel predictor based on the exponential weighted recursive least square algorithm (EWRLS) [15, 131], which operates on the channel's sparse DDSF.

6.3.1 Problem Statement

As previously assumed, the DDSF of the channel remains unchanged within a block of received symbols (of length M). Prediction of the channel's DDSF differs from the existing prediction of the channel's CIR (e.g., [15, 107, 108, 110, 111]) in the sense that it is only conducted for each block rather than every data sample. Therefore, this significantly reduces the complexity of the proposed predictor in comparison with that of the existing CIR predictors. Correspondingly, b and j are used to represent the index of data block and element in channel's DDSF, respectively. The structure of the proposed DDSF channel predictor is shown in Fig. 6.3. Here, b and j are used to represent the index of the data block and the element in the channel's DDSF, respectively. The input

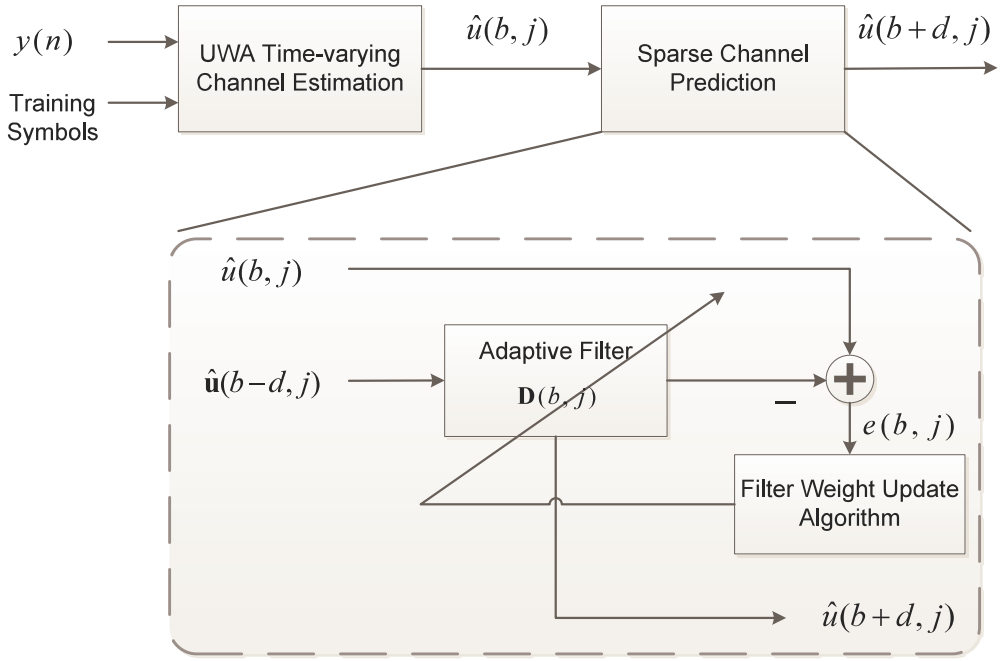


Figure 6.3: The proposed DDSF channel prediction using the EWRLS algorithm.

of the predictor is the current and past estimated channel coefficients in the delay and Doppler domain, i.e., $\hat{u}(b,j)$ and $\hat{\mathbf{u}}(b-d,j) = [\hat{u}(b-d,j), \cdots, \hat{u}(b-d-P+1,j)]^T$, for $j = 0, 1, ..., N_u - 1$ and $b = b_0, ..., N_b$, where $b_0 \geq d$ is the index of the starting block and N_b is the number of data blocks. Here, P and d represent the prediction order and the number of data blocks the channel is predicted ahead (also referred to as prediction horizon [107]), respectively. As the value of d depends on the latency introduced by the receiver processing, e.g., channel estimation, equalization, and detection, it is clear that the prediction accuracy worsens for the prediction obtained with a large value of d [107].

6.3.2 EWRLS Adaptive Predictor

Let us denote $\mathbf{D}(b,j) = [d_0(b,j), d_1(b,j), \cdots, d_{P-1}(b,j)]^T$ as the prediction coefficients. We may further reduce the complexity of the predictor by only considering the significant elements in $\hat{\mathbf{u}}(b,j)$, however, this may cause severe prediction errors due to

the position migration of the significant elements. As such, we consider all elements in $\hat{\mathbf{u}}(b,j)$ and use the EWRLS algorithm to find the $\mathbf{D}(b,j)$ which minimizes the cost function

$$\begin{aligned} J^d(b,j) &= \sum_{q=b_0}^{b} \lambda^{(b-q)}[\hat{u}(q,j) - \sum_{p=0}^{P-1} d_p(b,j)\hat{u}(q-d-p,j)]^2, \\ &= \sum_{q=b_0}^{b} \lambda^{(b-q)}[\hat{u}(q,j) - \mathbf{D}(b,j)^H\hat{\mathbf{u}}(q-d,j)]^2, \end{aligned} \tag{6.13}$$

where $b_0 \geq d$ is the index of the starting block, and λ is the forgetting factor, which is usually determined by $\lambda = 1 - 1/P$ [131]. Let us define the error signal $e^d(b,j) = \hat{u}(b,j) - \sum_{p=0}^{P-1} d_p(b,j)\hat{u}(b-d-p,j)$. Hereinafter, the superscirpts of $J^d(b,j)$ and $e^d(b,j)$ are droped for simplicity, and the adaptive filter weights $\mathbf{D}(b,j)$ are updated through

$$\mathbf{D}(b,j) = \mathbf{D}(b-1,j) + \mathbf{K}(b-d,j)e^*(b,j), \quad b \geq d, \tag{6.14}$$

where x^* denotes the complex conjugate of x. $\mathbf{K}(b,j)$ is referred to as the RLS gain vector [131], and can be calculated from

$$\mathbf{K}(b,j) = \frac{\mathbf{P}(b-1,j)\hat{\mathbf{u}}(b,j)}{\lambda + \hat{\mathbf{u}}^H(b,j)\mathbf{P}(b-1,j)\hat{\mathbf{u}}(b,j)}, \tag{6.15}$$

where $\mathbf{P}(b,j)$ is the inverse correlation matrix of the input signal and can be recursively updated by

$$\mathbf{P}(b,j) = \frac{1}{\lambda}[\mathbf{I} - \mathbf{K}(b,j)\hat{\mathbf{u}}^H(b,j)]\mathbf{P}(b-1,j). \tag{6.16}$$

Finally, the prediction can be calculated from

$$\tilde{u}(b+d,j) = \mathbf{D}(b,j)^H\hat{\mathbf{u}}(b,j), \quad b = b_0, ..., N_b, \quad j = 0, ...KL-1. \tag{6.17}$$

Algorithm 3 EWRLS Filter Weight Update Algorithm

Input: $\hat{\mathbf{u}}(b-d,j)$, $\hat{u}(b,j)$, forgetting factor λ, prediction order P, and δ, which is generally a large positive constant[28];

1: Initialize $\mathbf{D}(0,j) = [1,0,...0]^T$, $\mathbf{P}(0) = \delta \cdot \mathbf{I}$

for b = 0, 1, 2, ..., $N_b - 1$, $j = 0, 1, ..., KL-1$

2: $e(b,j) = \hat{u}(b,j) - \mathbf{D}(b-1,j)^H \hat{\mathbf{u}}(b-d,j)$;

3: $\mathbf{K}(b,j) = \frac{\mathbf{P}(b-1,j)\hat{\mathbf{u}}(b,j)}{\lambda + \hat{\mathbf{u}}^H(b,j)\mathbf{P}(b-1,j)\hat{\mathbf{u}}(b,j)}$;

4: $\mathbf{D}(b,j) = \mathbf{D}(b-1,j) + \mathbf{K}(b-d,j)e^*(b,j)$;

5: $\mathbf{P}(b,j) = \frac{1}{\lambda}[\mathbf{I} - \mathbf{K}(b,j)\hat{\mathbf{u}}^H(b,j)]\mathbf{P}(b-1,j)$;

6: $\tilde{u}(b+d,j) = \mathbf{D}(b,j)^H \hat{\mathbf{u}}(b,j)$

Output: Prediction of the DDSF coefficient of channel $\tilde{u}(b+d,j)$.

A description of the EWRLS filter weight updating algorithm is provided in **Algorithm 3**.

The EWRLS algorithm uses the forgetting factor $\lambda \in (0,1)$ to obtain only a finite memory of input data (also known as the prediction order, P) for tracking time variations of the channel. λ imposes a larger weight to more recent input, and therefore, if the value of λ is large, the algorithm has a relatively long memory length which leads to a decreased prediction error. However, this results in a prohibitively high computational complexity. On the contrary, with a smaller value of λ, the algorithm has a lower complexity, but the prediction error is increased.

[28]Since $\mathbf{P}(b,j)$ is proportional to the covariance matrix of $\mathbf{D}(b,j)$, and the knowledge of $\mathbf{D}(b,j)$ for the initial block is very vague. A high value should be assigned to δ which leads to a high covariance matrix of $\mathbf{D}(b,j)$. The recommended value for δ is $\delta > 100 \cdot \sigma^2_{\hat{\mathbf{u}}}$, where $\sigma^2_{\hat{\mathbf{u}}}$ is the variance of DDSF vector $\hat{\mathbf{u}}(b,j)$ [131]. Given a large number of data blocks, the initial values assigned are not important, since they are forgotten due to the exponential forgetting factor λ.

6.3.3 Computational Complexity of the Adaptive Predictor in Delay-Doppler Domain

The computational complexity per iteration of the **Algorithm 3** in terms of the number of operations for each step is given in Table 6.1. The overall complexity per iteration of the proposed predictor is $O(P^2)$. For the DDSF prediction, there are N_b blocks, and for each block, N_u elements are involved; this yields an overall complexity of $O(N_b N_u P^2)$. For the time domain CIR prediction, the number of iterations is MN_b. Thus, the resulting complexity is $O(MN_b KP^2)$. As seen, the overall complexity of the proposed DDSF predictor is $O(\frac{M}{L})$ times lower than that of the CIR predictor.

Table 6.1: Computational complexity per iteration for Algorithm 3

Step	The number of operations
2	P
3	$P^2 + 2P$
4	P
5	$2P^2$
6	P
Overall complexity per iteration	$O(P^2)$

6.3.4 Decision Feedback Adaptive Channel Predictor

A drawback of the reported predictor is that the input channel estimation becomes outdated quickly for a long frame (when the time duration of a frame is much larger than the channel's coherence time); this can decrease the prediction accuracy and degrade the system performance significantly. Thus, a predictor based on the decision feedback can improve performance over the one based solely on the outdated estimation [106,107,113]. In this section, a receiver structure is proposed, which combines a sparse DDSF channel estimation with a decision feedback adaptive DDSF prediction scheme as illustrated in Fig. 6.4. In Fig. 6.4, $y(n)$, $r(n)$, and $\hat{x}_d(n)$ represent the received symbol, the

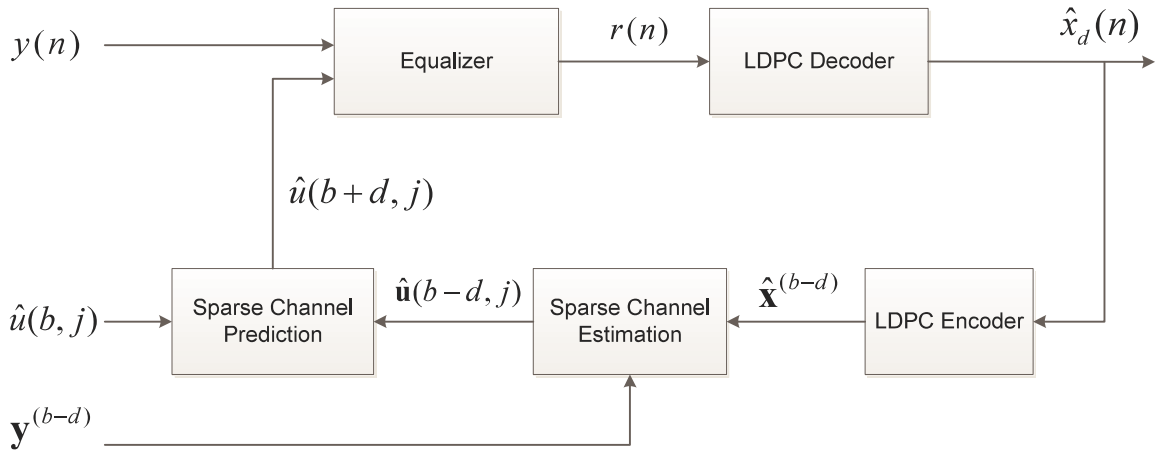


Figure 6.4: Block diagram of the receiver using the decision feedback adaptive channel predictor and sparse channel estimation.

equalized symbol, and the detected data at time n, respectively; $\mathbf{y}^b = [y((b-1)M+1)\ y((b-1)M+2)\ ...\ y(bM)]^T$ represents the received symbol vector for the bth block; $\hat{\mathbf{x}}^b = [\hat{x}((b-1)M-K+2)\ ...\ \hat{x}((b-1)M+1)\ ...\ \hat{x}(bM)]^T$ represents the corresponding detected and re-encoded symbol vector; $\hat{u}(b,j)$ denotes the jth elements in the estimated DDSF for the bth block.

The receiver procedures are illustrated in two modes, i.e., normal and decision feedback. In the normal mode, an up-to-date CSI based on the training block can be directly obtained using the CS-based estimation algorithms, and be used to predict the channel's DDSF for future block. In the decision-feedback mode, the detected bits are re-encoded and input to the channel estimator for an up-to-date CSI. The corrected channel information is processed by the proposed prediction scheme, as shown in Fig. 6.3. In contrast to the channel predictors in [106, 107, 113], the proposed predictor operates in the delay and Doppler domain, and only needs to be performed once per block; therefore, it requires a remarkably low complexity. Moreover, it should be noted that the accuracy of the predictor in decision feedback mode depends on the number of correctly detected bits; this means that error propagation occurs due to the incorrect

detection. Although the error propagation can be neglected when the signal-to-noise ratio (SNR) is above a certain threshold [107],[29] a theoretical analysis of its impact on the prediction accuracy is beyond the scope of this thesis.

6.4 Simulation Results

6.4.1 Simulation Setup

Numerical simulations are conducted to illustrate the performance of the proposed sparse channel estimation and prediction scheme. First, the proposed sparse channel estimation is evaluated in the training scenario, where the perfect knowledge of the transmitted sequence is known. Then, the proposed sparse channel prediction and overall system performance are assessed with a periodic training block shown in Fig. 6.2. It is assumed that the data symbols are modulated using BPSK and are coded using a 1/2 rate binary LDPC code.

6.4.1.1 Experiment Description

The time-varying channels are obtained based on a recently published acoustic channel simulator [114](a detailed description of the channel simulator can be found in Section 3.2.4). The first experiment, called the Surface Processes and Acoustic Communications Experiment (SPACE) was conducted near the coast of Martha's Vineyard in Massachusetts, in the fall of 2008. The carrier frequency was 13 kHz and the transmission rate was 6.5 kilobits per second (kbps). The signal was transmitted for three minutes every two hours. The active three-minute interval of each two-hour period is referred to as one epoch. The experiment lasted for 15 days. The water depth was 10 m, and the transmitter and receiver were fixed at 4 m and 2 m above the sea-floor,

[29] The SNR is defined as $10\log_{10}(\sigma_x^2/\sigma_w^2)$ with σ_x^2 the variance of the transmitted signal.

respectively. Receivers were located southeast of the transmitter at distances of 1000 m.

The second experiment, Mobile Acoustic Communications Experiment (MACE), was conducted in the Atlantic Ocean about 100 miles south of Martha's Vineyard in the summer of 2010. The receiver buoy was suspended at the depth of 40 m and the transmitter was towed at the depth of $50 - 60$ m. During the experiment, the receiver arrays were stationary, while the source array was towed slowly away from the receivers and then towed back, at a speed around 1 m/s. The water depth was approximately 100 m and the distance is set to 500 m. The carrier frequency was 13 kHz and the signals were transmitted continuously at 5 kbps.

The third experiment, known as the Kauai Acomms MURI (KAM), was conducted in July 2011 off the coast of Kauai Island, Hawaii. This system operating frequency was between 8.5 kHz and 17.5 kHz. The transmitter and receiver were deployed approximately mid-way in 100 m water, and were 3 km apart. The signals were modulated onto a carrier of 13 kHz and transmitted at the rate of 6.5 kbps during 9 minute epochs every two hours.

Finally, the fourth experiment, called the Pacific Storm (PS) experiment, was conducted on the submerged portion of San Andreas Fault off the coast of Northern California in September 2010. During this experiment, 3-second-long data packets were repeatedly transmitted every 5 seconds from an AUV to a surface ship. The signals occupied 4 kHz of bandwidth around a center frequency of 10 kHz. The AUV was moving at about 3 m above the bottom, at a depth of approximately 130 m. The transmission distance varied from 200 m to 1 km. In this experiment, the signals were automatically processed by the acoustic modems mounted on the two ends of the link. The received signal strength was recorded once a second. All experiments assume a sinusoidally moving surface and transmitter/receiver drifting. It should be noted that

the relative speed of the AUV in the PS'10 experiment is estimated to be 3 m/s based on the observation from large-scale analysis of the channel, i.e., gain versus distance in [114]. The parameter settings in the numerical simulation, displayed in Table 6.2, are chosen according to the settings used in the SPACE'08, MACE'10, KAM'11 and PS'10 experiments in order to approximate the practical channel measurements.

6.4.1.2 Simulated Channels

First, UWA channel ensembles are obtained using the previsouly proposed SACM simulator [114] with parameter settings in Table 6.2. The nominal geometry maps of the various experiments are plotted in Fig. 6.5. For MACE'10 and PS'10, where there is a relative motion between a specific pair of transmitter and receiver, we assume that the distance between them is initially set to 500 and 200 meters, respectively. Fig. 6.6 shows the ensembles of the simulated channel response using the SPACE'08, MACE'10, KAM'11, and PS'10 experiment settings over a duration of 1 min. Local maxima over the delay axis are visible in the figures, indicating channel taps over which the impulse responses are relatively strong. In all cases, slow varying mean of each path and path spreading, which is a consequence of both micropath dispersion and bandwidth limitation (if the geometry of two paths has similar lengths and the delay resolution of the system is limited due to finite bandwidth, two paths may merge as their arriving times will be too close to be distinguished), are evident. Fig. 6.7 shows the averaged gains versus time for various experiments. A faster variation is noticed in MACE'10 and PS'10 due to the relative motion between the transmitter and receiver. Particularly, in the PS'10 experiment, the transmission range was constantly changing over a longer observation period, resulting in more pronounced gain variations.

Table 6.2: Simulation parameters in four experiments

Parameters	Value			
System Parameters				
Training symbol length M	200, 400, 600, or 800			
Frame length N	3200 or 4800			
Total number of blocks N_b	12000			
Experiment	MACE'10	SPACE'08	KAM'11	PS'10
Transmission symbol rate [symbol per second]	5000	6500	6500	4000
Carrier frequency [kHz]	13	13	13	10
Bandwidth [kHz]	5	9	9	4
Doppler frequency resolution [Hz]	6.7			
Maximum Doppler shift [Hz]	30			
The number of sampled delay points	168			
The number of sampled Doppler points	9			
Deterministic Channel Geometry				
Experiment	MACE'10	SPACE'08	KAM'11	PS'10
The depth of water [m]	100	10	103	130
Transmitter height [m]	45	4	58	3
Receiver height [m]	55	2	59	127
Distance between Tx and Rx [km]	0.5-4	1	3	0.2-1
Relative velocity between Tx and Rx [m/s]	1	0	0	3
Spreading factor	1.5			
Speed of sound in bottom [km/s]	1.3(>1.5 for hard bottom, <1.5 for soft bottom)			
Large-Scale Parameters				
Experiment	MACE'10	SPACE'08	KAM'11	PS'10
Coherence time T_c of the large-scale variation [s]	15	60	60	5
Standard deviation of surface height variation	1.5	1	1.5	1.5
Standard deviation of transmitter height variation	0.5	1	1	0.1
Standard deviation of receiver height variation	1	0.5	1	1.5
Standard deviation of TX and RX distance variation	1	0.5	1.5	0.5
Amplitude of the displacement caused by surface wave A_w [m]	0.05			
Frequency of the displacement caused by surface wave f_w [Hz]	0.01			
Small-Scale Parameters				
Variance of S-S surface variation σ_s^2	1.125			
Variance of S-S bottom variation σ_b^2	0.5625			
Experiment	MACE'10	SPACE'08	KAM'11	PS'10
3-dB width of the p.s.d. of intra-path delays (assumed constant for all paths) B_{δ_p}	2.5×10^{-3}	5×10^{-4}	5×10^{-4}	6×10^{-3}
Number of intra-path (assumed constant for all multipath)	20			
Mean of intra-path amplitudes (assumed constant for all multipath)	0.025			
Variance of intra-path amplitudes (assumed constant for all multipath)	10^{-6}			

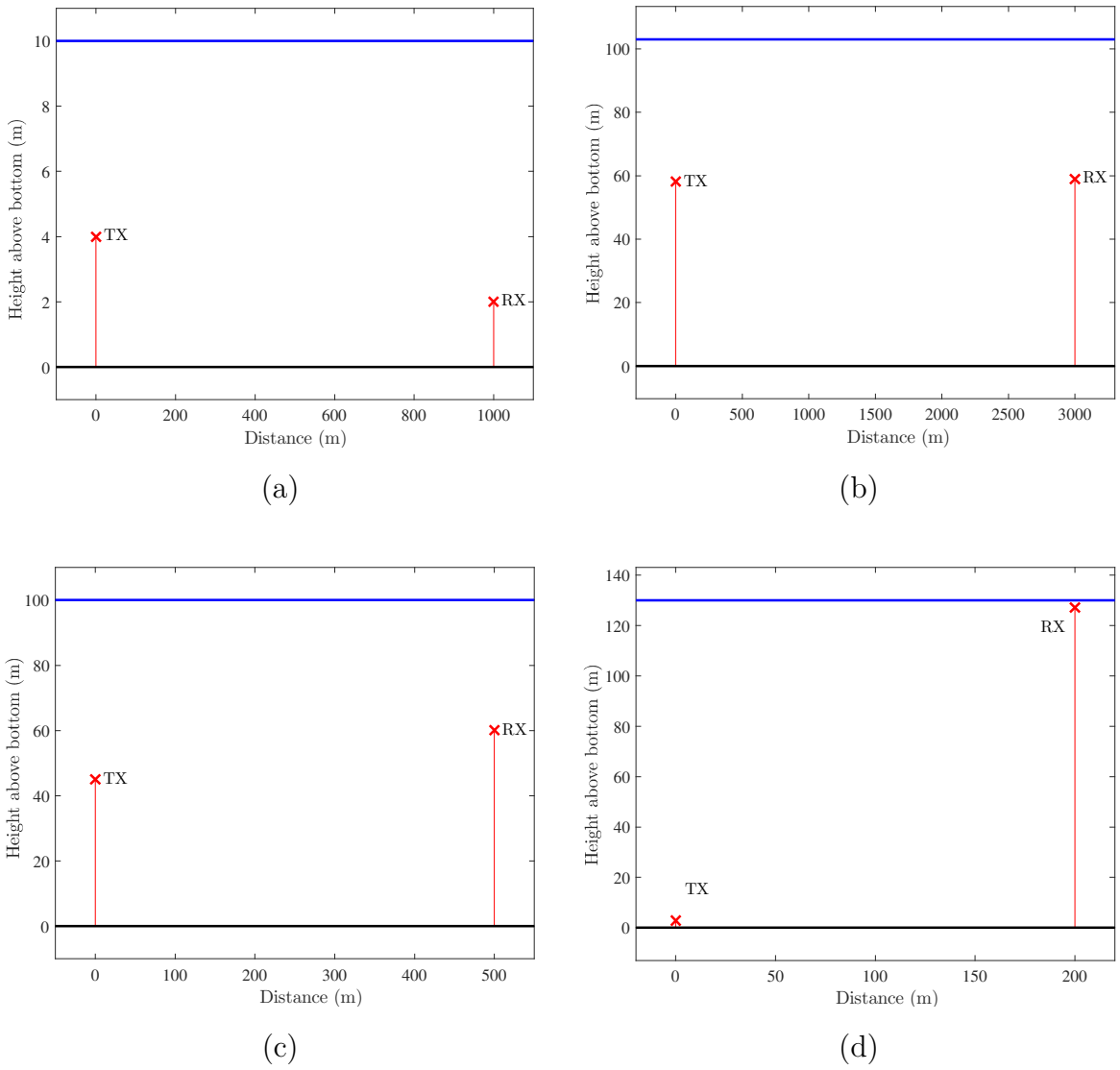


Figure 6.5: Nominal geometry for: (a) SPACE'08; (b) KAM'11; (c) MACE'10; and (d) PS'10. The blue line indicates water surface.

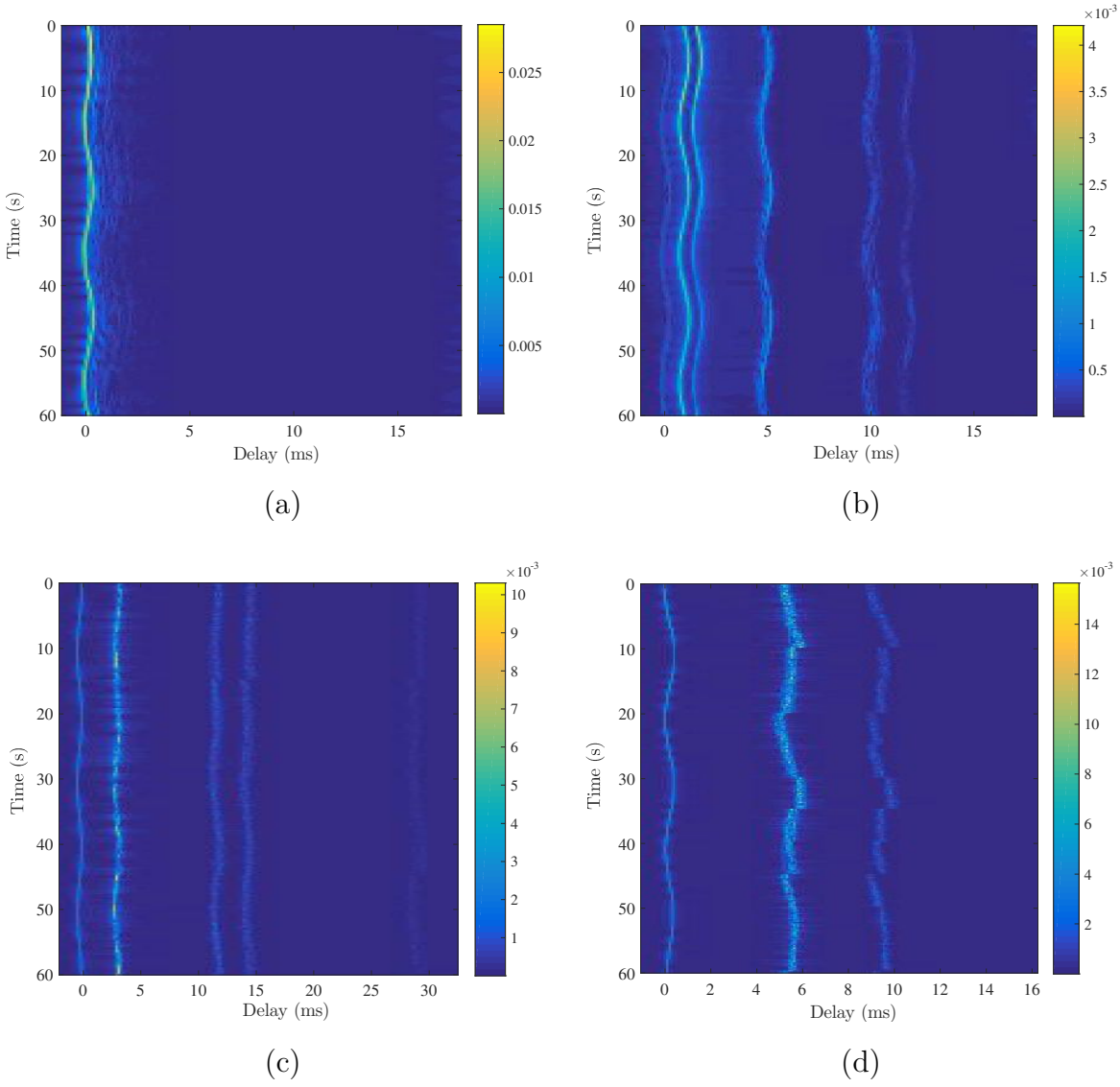


Figure 6.6: Time evolution of the magnitude impulse response for: (a) SPACE'08; (b) MACE'10; (c) KAM'11; and (d) PS'10

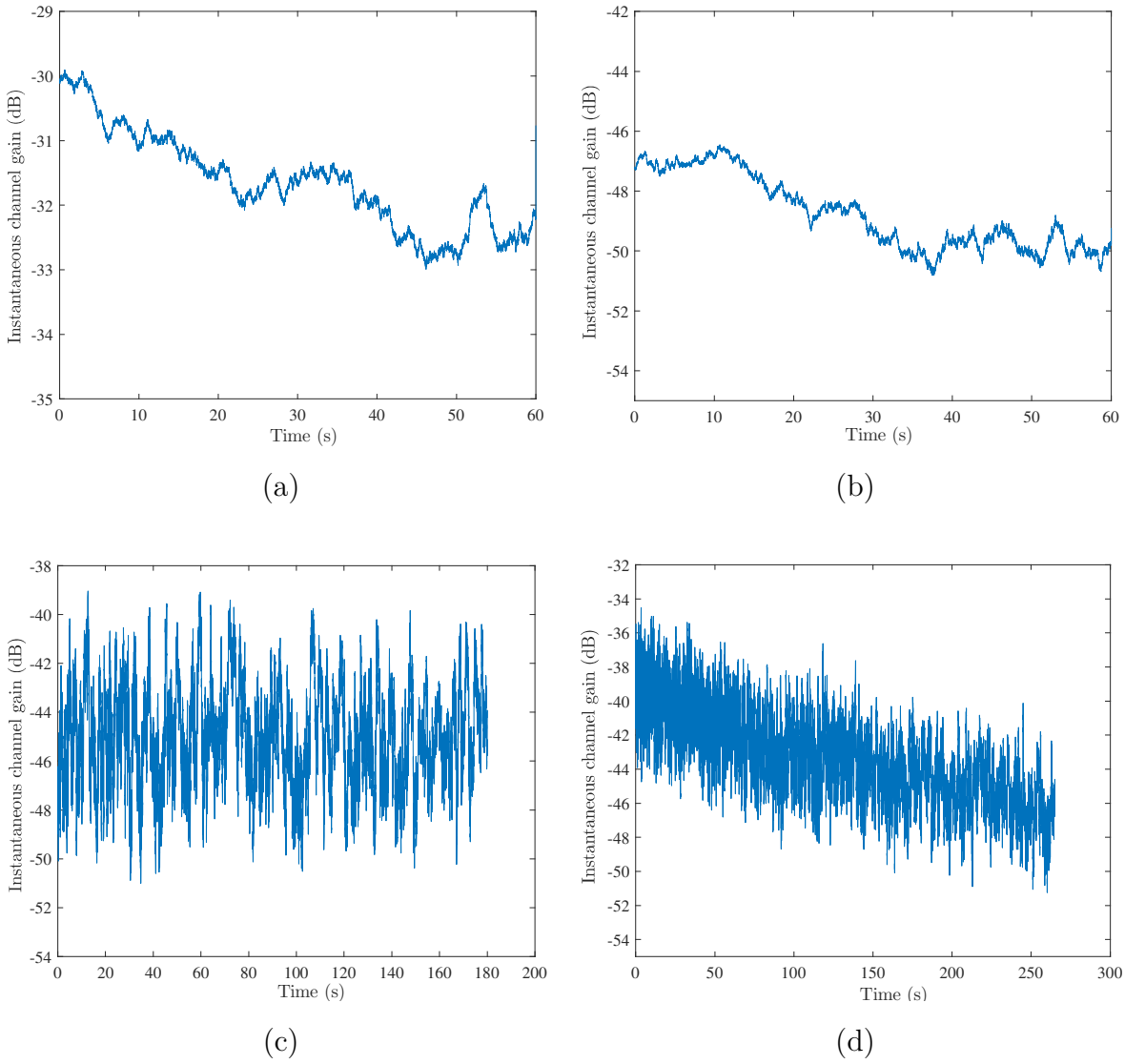


Figure 6.7: Instantaneous gain (averaged over small-scale realizations) versus time: (a) SPACE'08; (b) MACE'10; (c) KAM'11; and (d) PS'10.

6.4.2 Performance of the DDSF Estimation

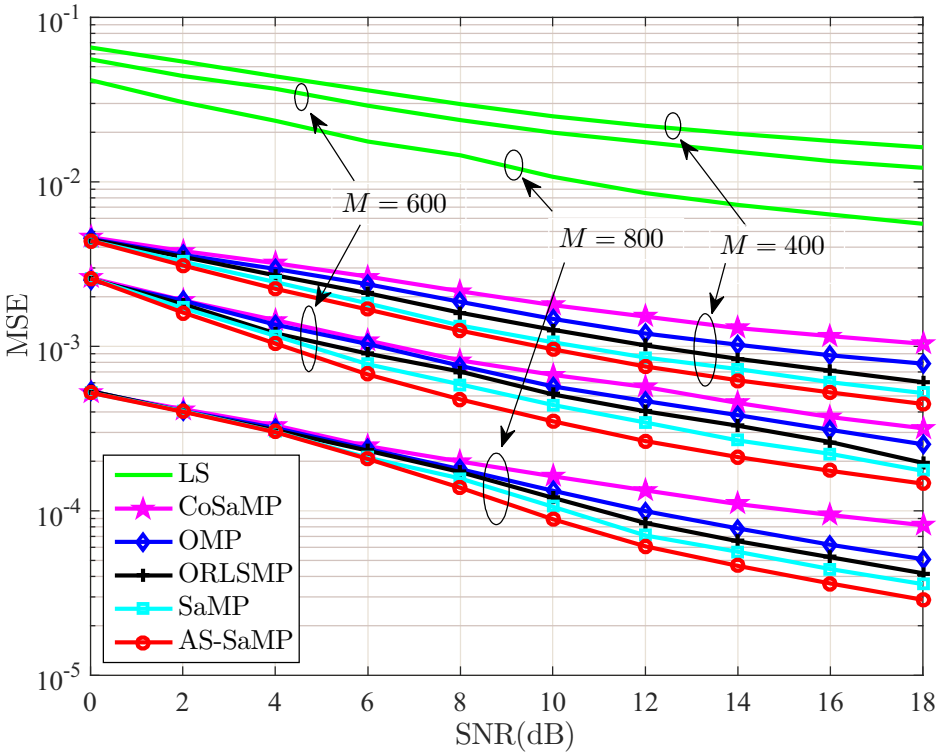
The performance of LS and the five MP-based algorithms, i.e., OMP [52], CoSaMP [63], SaMP [64], ORLSMP [43] and the proposed AS-SaMP [30], are compared in terms of MSE[30] and CPU running time. The initial step size in SaMP and AS-SaMP is set to

[30] $\text{MSE} = \text{E}[\sum_{l=1}^{L}\sum_{k=1}^{K}|u(l,k)-\hat{u}(l,k)|^2]$.

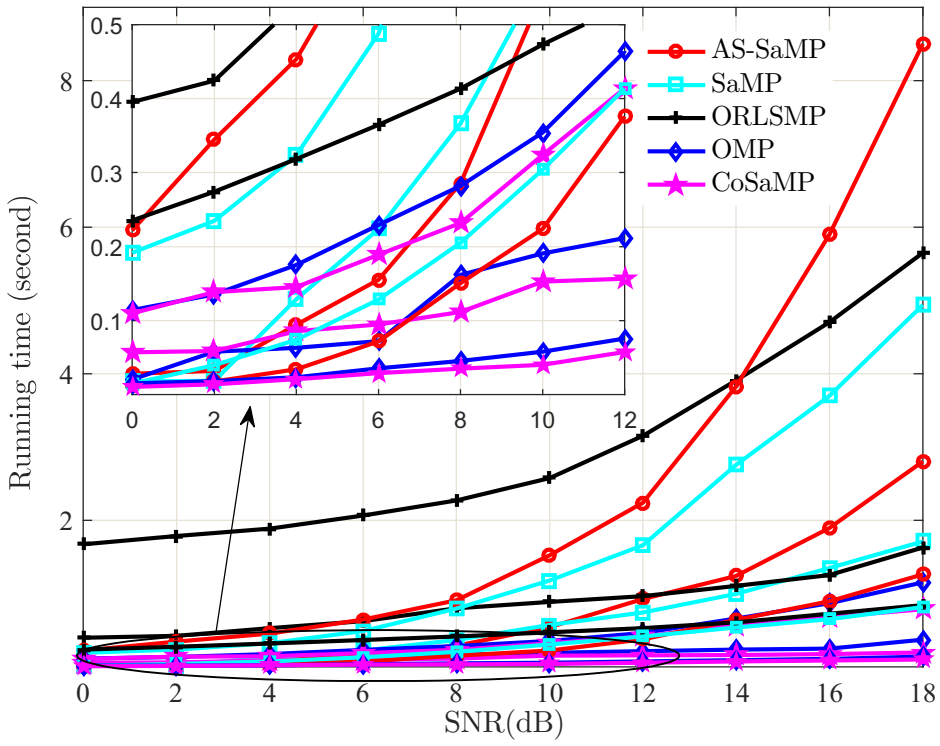
5, and the sparsity level required in CoSaMP is set to 20. Simulations are performed in MATLAB R2015b using a 2.67 GHz Intel Core i7 CPU with 12 GB of memory storage, and 10^3 MonteCarlo trials are used to average the results which correspond to 4.8×10^6 data symbols.

Fig. 6.8 shows the performance of the considered estimation algorithms for various M, and the results are obtained for the MACE'10 channel. In Fig. 6.8 (a), it is clear that the MSE of the CS-based algorithms outperforms the traditional LS algorithm, and AS-SaMP achieves the best MSE performance among all considered algorithms. It should be noted that the MSE of CoSaMP is the worst among the considered CS-based algorithms; this is due to the overestimation which is caused by an inaccurate input of the sparsity level as *a priori* information to the CoSaMP algorithm. Moreover, a better MSE can be achieved by increasing the length of the training symbols for all algorithms. As shown in Fig. 6.8 (b), increasing M results in increased running time for all algorithms. In general, the AS-SaMP algorithm requires longer running time to obtain the estimation at higher SNRs (SNR $>$ 14 dB). This is because an increased number of iterations is involved as SNR increases. However, when compared with SaMP and ORLSMP, it takes equal or less amount of time, with a better quality of estimation for SNR $<$ 10 dB. Furthermore, a higher running time is required for ORLSMP when SNR $<$ 14 dB, given that a similar number of iterations is involved in all considered algorithms. This is because the amount of operation counts per iteration in ORLSMP is higher than that in the other algorithms. In summary, the AS-SaMP algorithm achieves better MSE without using significantly large amount of CPU running time in general.

Fig. 6.9 shows the MSE and running time of the considered algorithms with channels from the other three experiments for $M = 400$ and $M = 800$. From Fig. 6.9 (a) and (c), the MSE of all algorithms for the KAM'11 and SPACE'08 channel are lower than those for channels of the PS'10 and MACE'10 experiment. This is because there is

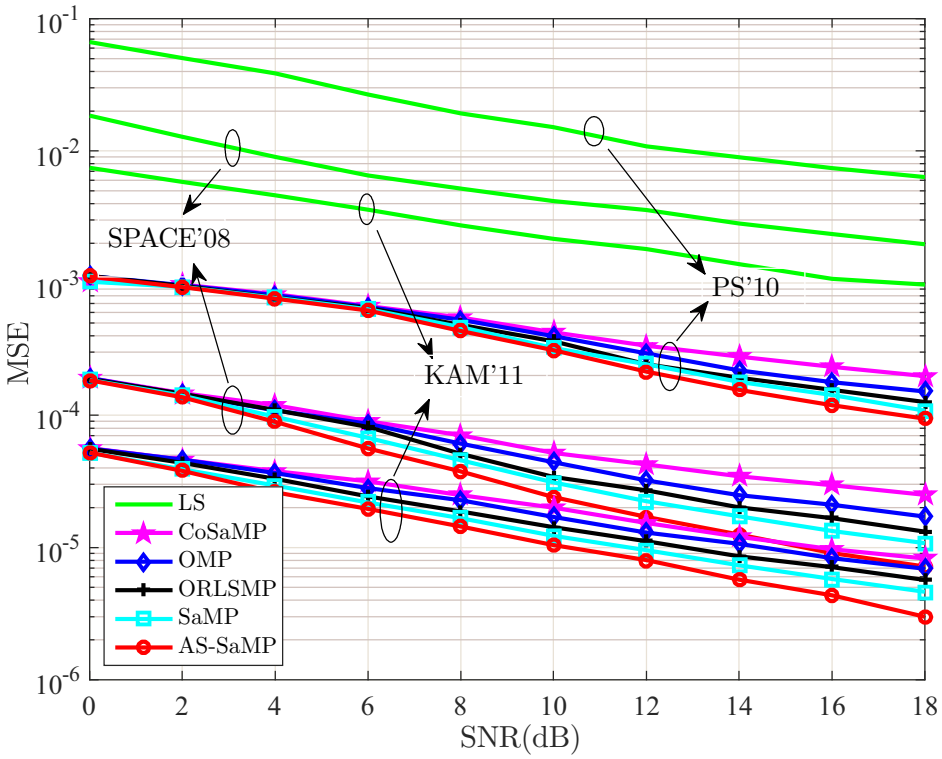


(a)

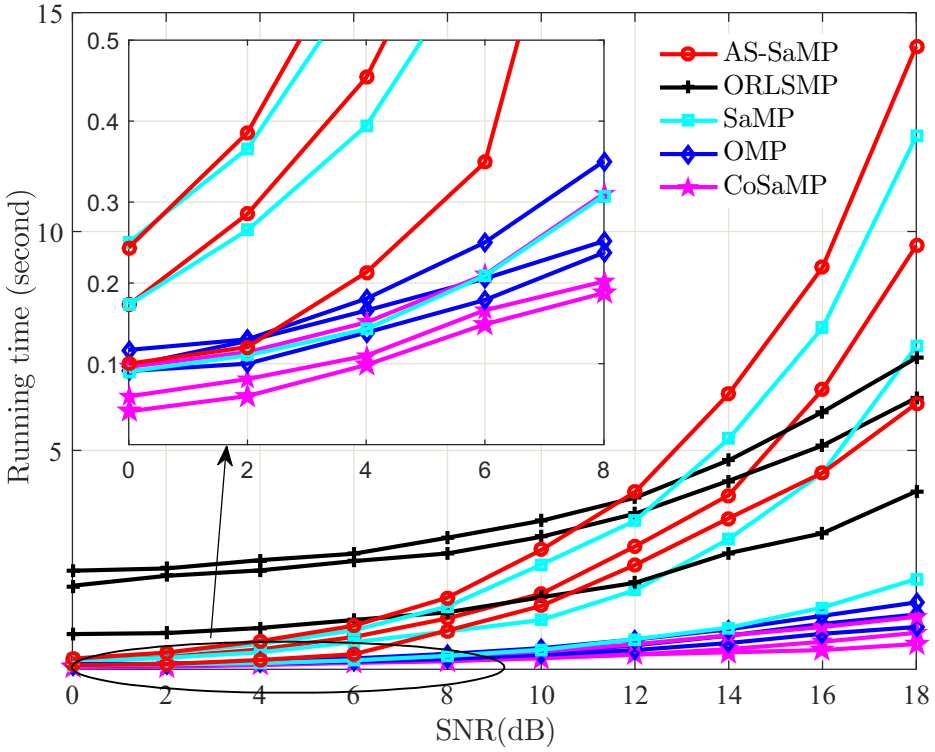


(b)

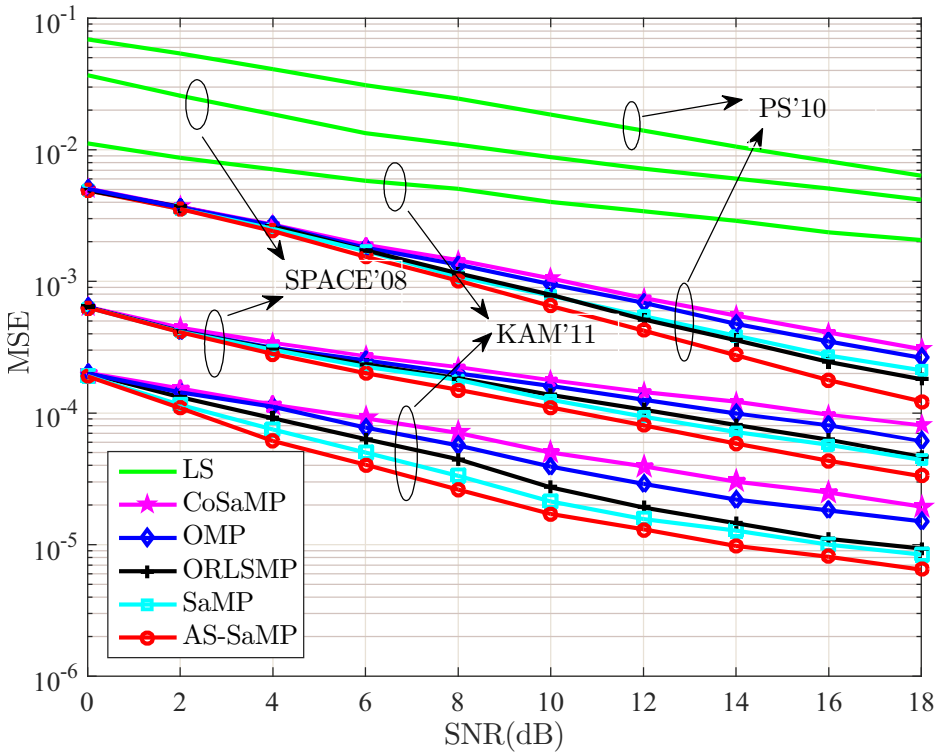
Figure 6.8: Performance of the considered algorithms for $M = 400$, 600, 800, and the experiment MACE'10 (a) MSE; (b) CPU running time. Dot lines are used for $M = 400$, solid lines are for $M = 600$, and dashed lines are for $M = 800$.


MSE
SNR(dB)
SPACE'08
KAM'11
PS'10
LS
CoSaMP
OMP
ORLSMP
SaMP
AS-SaMP

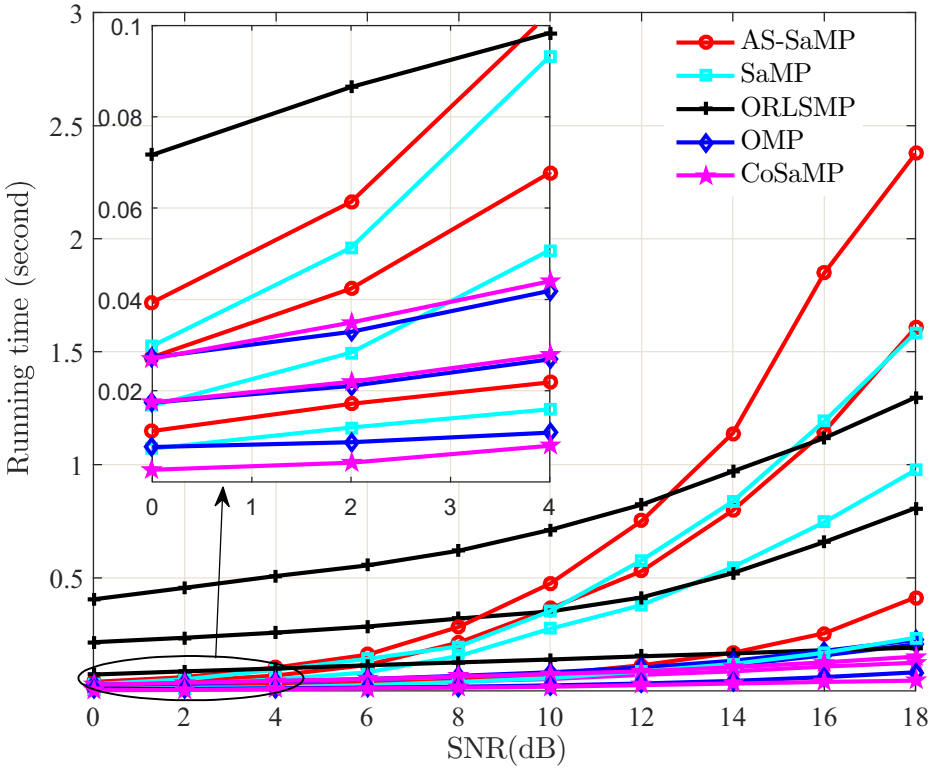
(a)


Running time (second)
SNR(dB)
AS-SaMP
ORLSMP
SaMP
OMP
CoSaMP

(b)



(c)



(d)

Figure 6.9: Performance of the considered algorithms for various channels (a) MSE with $M = 800$; (b) CPU running time with $M = 800$; (c) MSE with $M = 400$; (d) CPU running time with $M = 400$. Solid lines are used for SPACE'08, Dashed lines are for KAM'11, and Dot lines are for PS'10.

no intentional transmitter/receiver motion causing channels to vary slower during a training block period. Meanwhile, the MSE of all algorithms for the PS'10 channel is the worst among all considered channels, since estimation of the fast-varying channel based on a long training block would introduce significant error due to the violation of the underlying assumption, i.e., $\mathbf{u}$ does not change significantly over the M training symbols [43]. In addition, the MP-based estimation methods outperform the traditional LS estimation algorithm, and AS-SaMP outperforms the other considered methods, for a given M. The MSE of CoSaMP is the highest among all the MP-based algorithms due to overestimation. Meanwhile, in Fig. 6.9 (b) and (d), the running time of all algorithms for the KAM'11 channel is higher than that for the channels of the other experiments. This is due to a larger number of multipath which requires more iterations in the considered algorithms. Moreover, the running times of the AS-SaMP and SaMP algorithms are relatively higher than those of the other algorithms for higher SNRs (SNR >12 dB). In contrast, the ORLSMP algorithm takes the longest time for lower SNRs (SNR <12 dB) due to the highest amount of operation counts per iteration, given that a similar number of iterations are involved in all considered algorithms.

Furthermore, it is worth noting that $M = 400$ is considered to be a better choice since considerable savings in running time can be achieved without a significant increase in MSE, especially when the SNR is high. For example, when SNR = 16 dB, the MSE of AS-SaMP for SPACE'08, KAM'11 and PS'10 channels with $M = 400$ are 4.2×10^{-5}, 8×10^{-6} and 1.9×10^{-4}, respectively, while the CPU running times are 1.2 s, 1.8 s and 0.23 s, respectively. However, with the same SNR and channels, the MSEs of AS-SaMP using $M = 800$ are 8×10^{-6}, 4.2×10^{-6} and 1.4×10^{-4}, respectively, while the running times are 6.5 s, 8.4 s, and 4.7 s, respectively. Although the MSEs for SPACE'08, KAM'11, and PS'10 channels increase by 4.25, 0.9 and 0.36 times with $M = 400$, the CPU running times decrease by 4.41, 3.67, and 19.43 times, respectively.

6.4.3 MSE Performance of the DDSF Prediction without Decision Feedback

This section shows the MSE performance of the proposed DDSF predictor without using the detected symbols for the most fast-varying channel, i.e., PS'10. The proposed method uses the EWRLS algorithm for predicting the DDSF of the time-varying channel. It is known that RLS algorithm has a superior tracking ability but its application is limited when complexity is a primary concern. In contrast, the computational complexity of the predictor is considerably reduced as the prediction of channel's DDSF is only needed once every M symbols. Unless otherwise mentioned, the prediction horizon $d = 1$, and the forgetting factor[31] $\lambda = 0.99$, which corresponds to a prediction order $P = 100$. The following two schemes are considered,

- *Scheme 1*: the channel estimation is carried out based on a block of training symbols (with length M), and the estimated channel is used to equalize the rest of the symbols in a data frame (with length N). In other words, no channel prediction method is employed;

- *Scheme 2*: the channel estimation is carried out based on a block of training symbol and the proposed DDSF prediction is used for equalizing the rest of symbols in a data frame.

Fig. 6.10 shows the MSE performance comparison for the above mentioned two scenarios with various CS-based estimation algorithms. A block of training symbols with a length of 400 is employed for channel estimation; this accounts for 8.3% of a data frame whose length N is 4800. It can be clearly seen that the MSE performance is significantly improved for the considered algorithms when the channel prediction is em-

[31]Recall that the prediction order can be determined from the forgetting factor through $P = 1/(1-\lambda)$ [131], during which channel is assumed to be constant.

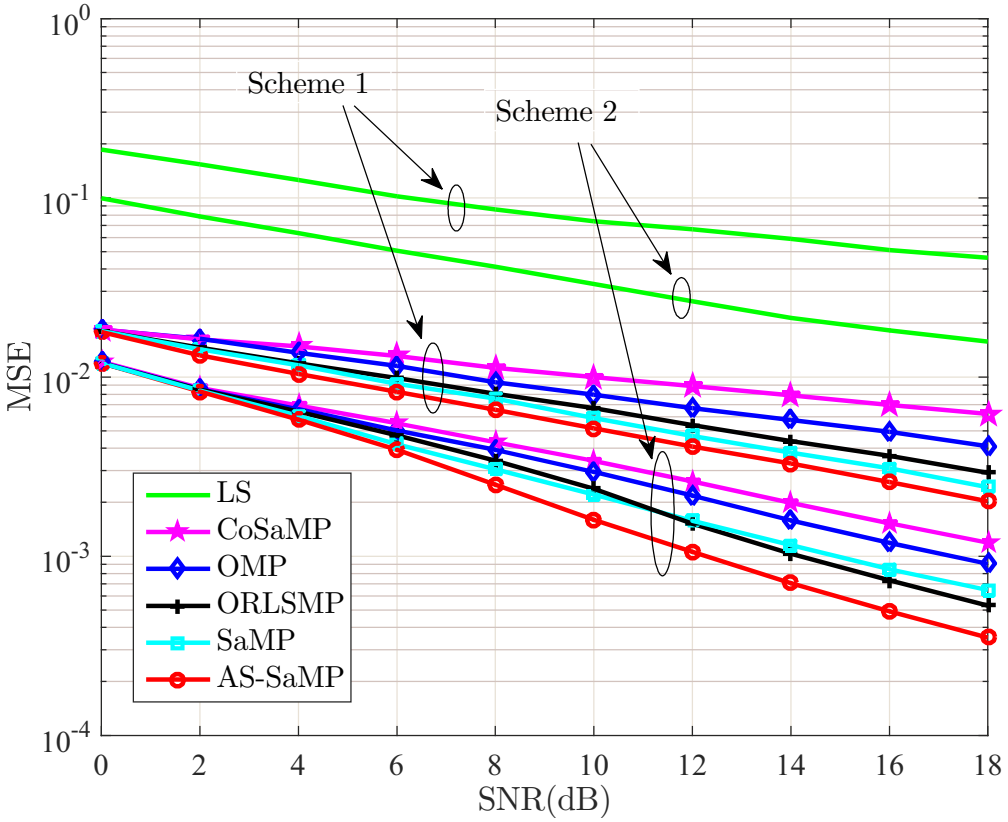


Figure 6.10: MSE performance for two schemes with various channel estimation algorithms, $M = 400$ and $N = 4800$.

ployed (*Scheme 2*). Moreover, among all the considered channel estimation algorithms, the AS-SaMP algorithm gives the best MSE performance in both scenarios.

Next, the impact of M and N on the prediction MSE performance is investigated. Fig. 6.11 shows the MSE results with different M values (200 and 400) for $N = 4800$. It is observed that a better MSE can be achieved by increasing M because of a better quality of the channel estimate. However, the overhead is doubled by increasing M from 200 to 400 for a fixed N. Similarly, given a particular M, e.g.., $M = 400$, the MSE performance of the proposed channel predictor is compared for different values of N, i.e., $N = 3200$ and 4800, as shown in Fig. 6.12. A significant MSE deterioration is observed for all considered estimation algorithms as the less accurate channel prediction

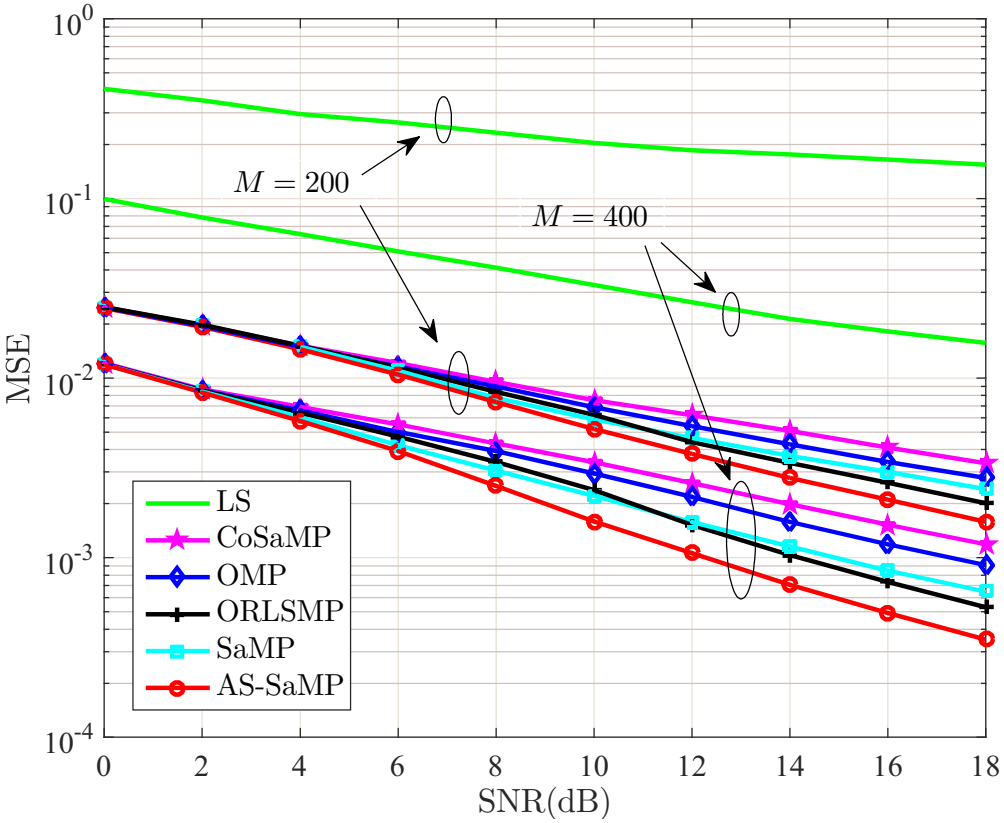


Figure 6.11: The prediction MSE performance for various values of M.

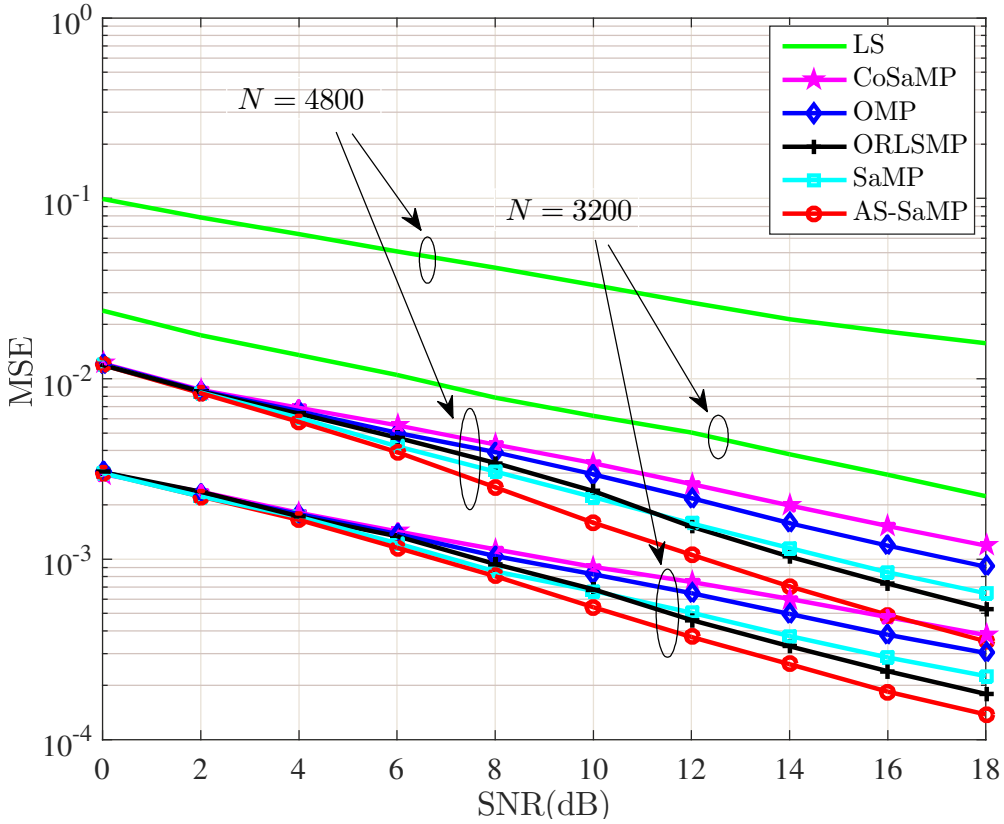


Figure 6.12: The prediction MSE performance for various values of N.

is obtained when N increases, and the channel is more likely to vary within a longer time period. In the following simulation, M and N were chosen as 400 and 4800, respectively, unless indicated otherwise.

Furthermore, the impact of the parameter λ on the performance of the channel prediction is investigated. On the one hand, a large λ leads to a relatively long memory length and thus decreases the prediction error. However, a large λ also causes high computational complexity. On the other hand, a small λ corresponds to a short memory length, at a cost of increased prediction errors. Fig. 6.13 shows a MSE comparison of channel prediction with $\lambda = 0.99$ and 0.95 which correspond to a prediction order $P = 100$ and 20, respectively. Clearly, the MSE of the channel prediction is significantly better with $\lambda = 0.99$, when compared with that of the predictor with $\lambda = 0.95$. How-

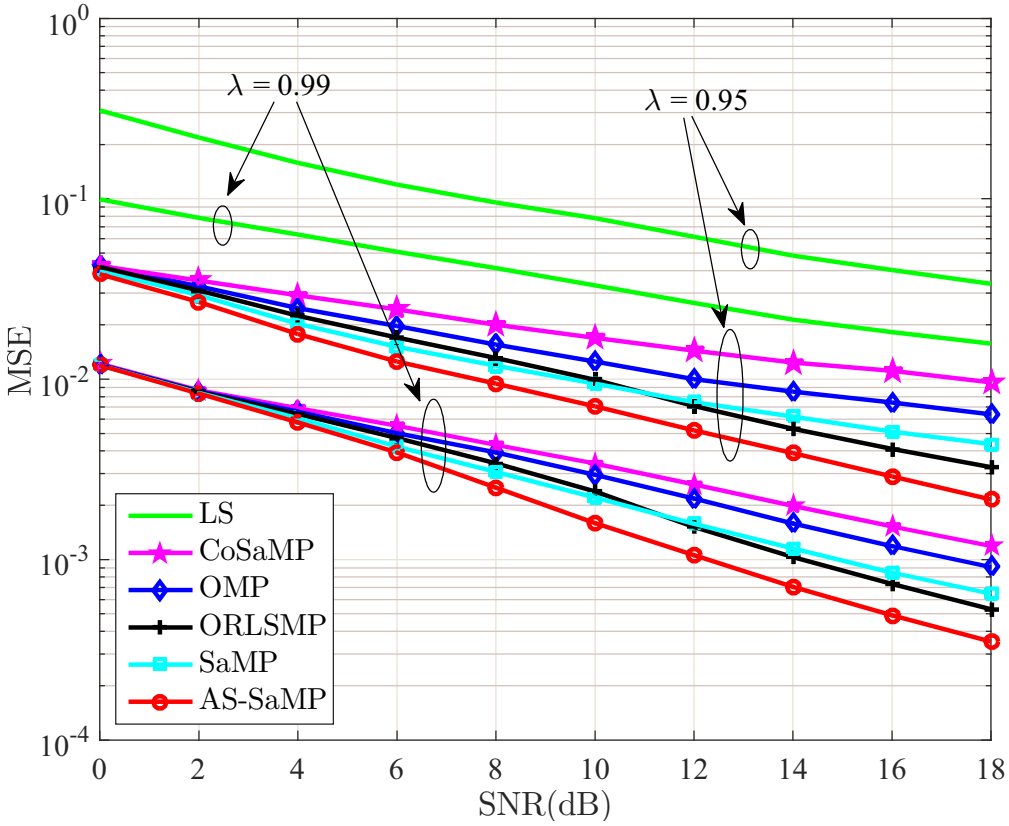


Figure 6.13: The prediction MSE performance for various values of λ.

ever, as expected, the complexity per iteration in the EWRLS algorithm is increased with a larger λ, given that the same number of iterations is involved.

6.4.4 MSE Performance of the DDSF Prediction with Decision Feedback

This section evaluates MSE of the proposed predictor using the previously detected symbols. The performance is shown for the relatively fast-varying PS'10 channels unless mentioned otherwise. For comparison, the following two cases are considered.

- *Case 1*: this case assumes that all feedback signals were correctly detected to exclude effects resulting from error propagation;

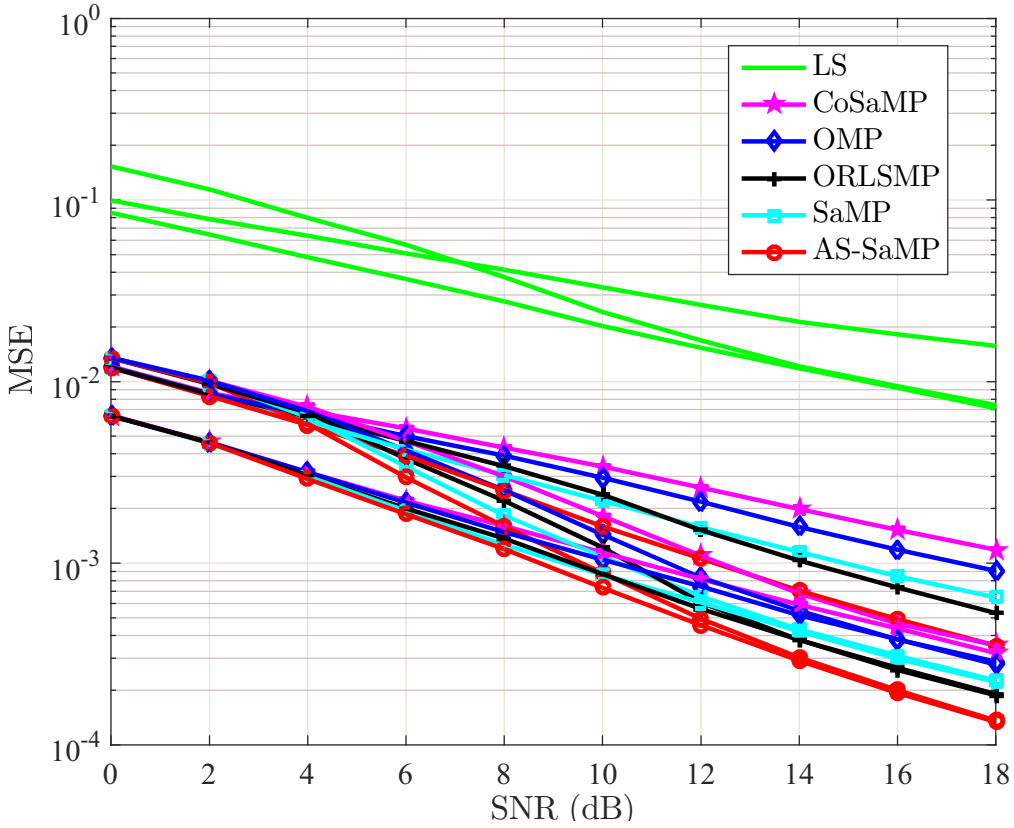


Figure 6.14: MSE performance of channel predictor without/with decision feedback. Solid lines are used for *Case 2*, dashed lines for *Case 1*, and dot lines for the predictor without using decision feedback.

- *Case 2*: this case indicates that feedback signals may contain incorrectly detected symbols, i.e., error propagation can occur which degrades the prediction MSE.

Fig. 6.14 shows the MSE of the channel predictors without and with using decision feedback. One can observe that the prediction accuracy is remarkably improved for both Case 1 and Case 2, when the past detected data were used at the expense of increased computational complexity due to additional channel estimation and re-coding for each block. For instance, considering the AS-SaMP algorithm and Case 1, the prediction MSE with decision feedback decreases by almost 50% at SNR = 6 dB, when compared to that without decision feedback. Moreover, the MSE of the decision feedback predictor for the aforementioned two cases is compared. As expected, the MSE for *Case 2* is

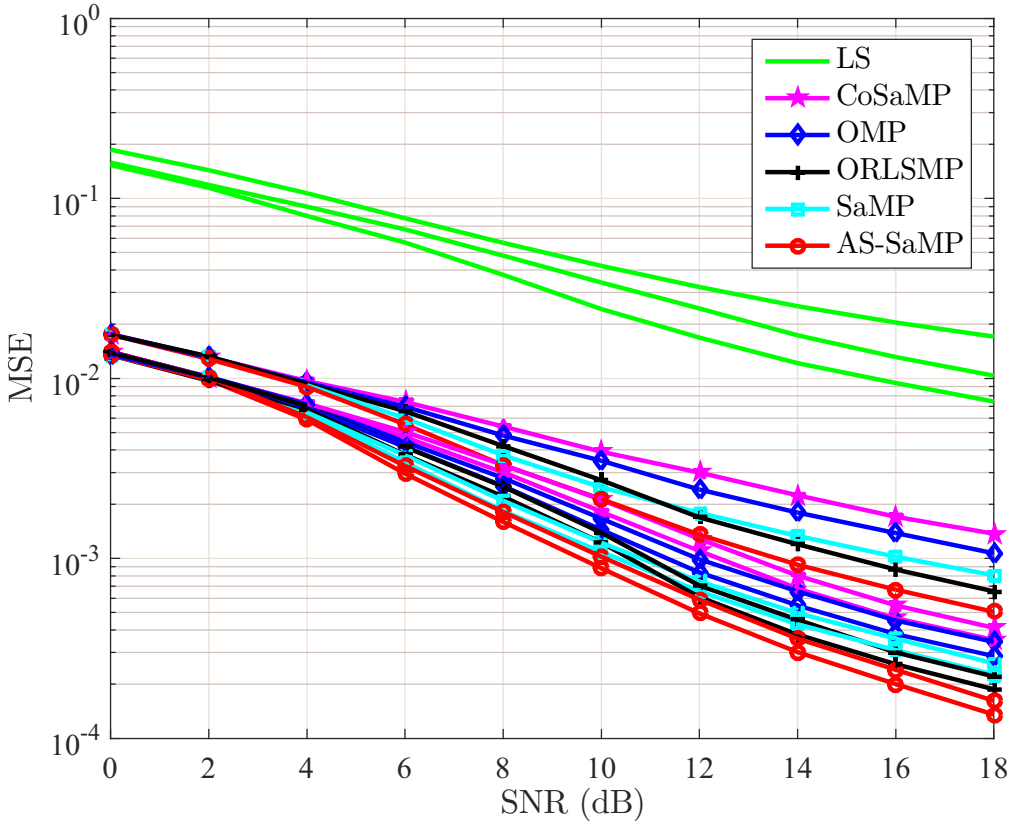


Figure 6.15: MSE performance of the decision feedback channel predictor with various prediction horizon d. Solid lines are used for $d = 1$, dot lines for $d = 2$, and dashed-dot lines for $d = 3$.

higher than that for *Case 1* due to the error propagation of the feedback symbols. It can be seen the impact of error propagation reduces as the SNR increases, and the prediction MSE for *Case 2* is almost the same as that for *Case 1* when SNR $\geq$ 16 dB. It is also worth noting that the MSE curves of the decision feedback predictor for *Case 1* closely approach to those of the channel estimation based on the training sequence[32] as the SNR increases. In other words, the performance of the proposed prediction yields accurate up-to-date CSI for higher SNRs even without training symbols.

Fig. 6.15 illustrates the prediction MSE for different prediction horizons, i.e., d, for

[32]The MSE performance of the training-based channel estimation for PS'10 channels, with $M = 400$ and $N = 4800$, is plotted in Fig. 6.9 (c).

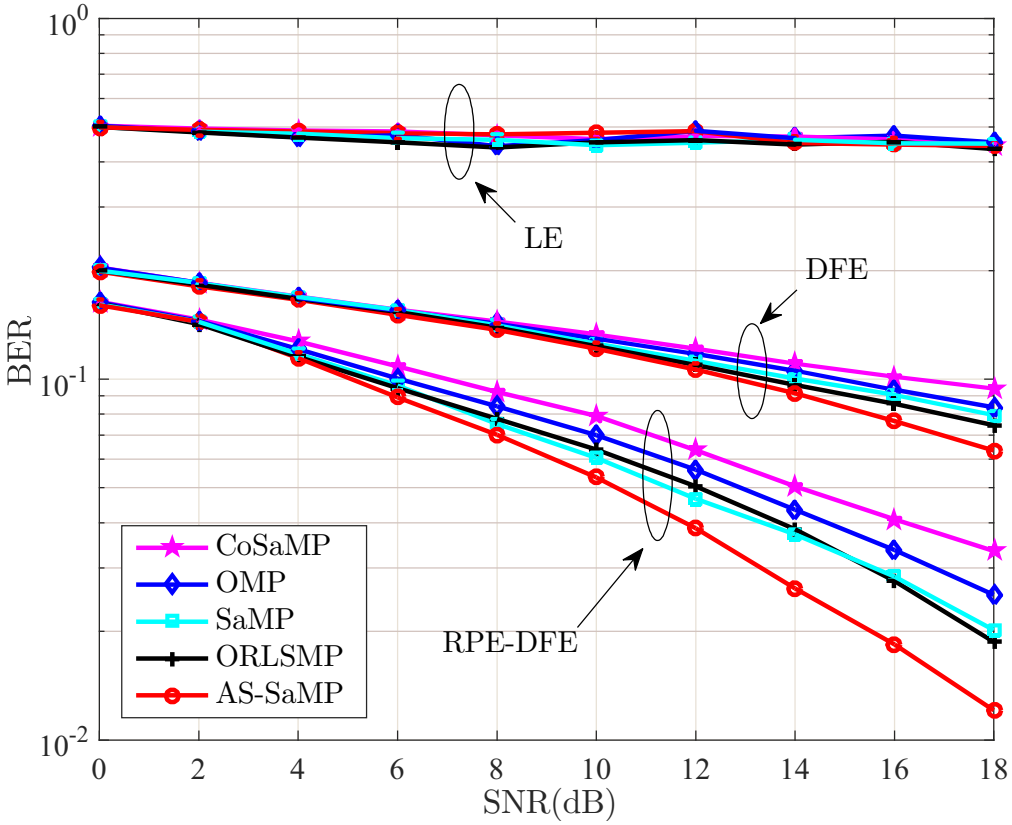


Figure 6.16: Uncoded BER performance for various equalization techniques.

the case *Case 2*. As can be seen, the quality of the channel prediction decreases as d increases. It should be noted that the value of d for an accurate channel prediction depends on how fast the channel can vary. A reliable channel prediction can be achieved even for a large d when then channel varies slowly in time, while the performance of the predictor can be substantially deteriorated even for a significant small d when the channel varies rapidly in time.

6.4.5 System Performance of the Proposed Techniques

Finally, the BER performance of the overall system is evaluated for the various estimation, equalization and prediction methods. An LDPC codes with rate 1/2 was used, and each encoded block size is 400. Fig. 6.16 shows the uncoded BER versus SNR

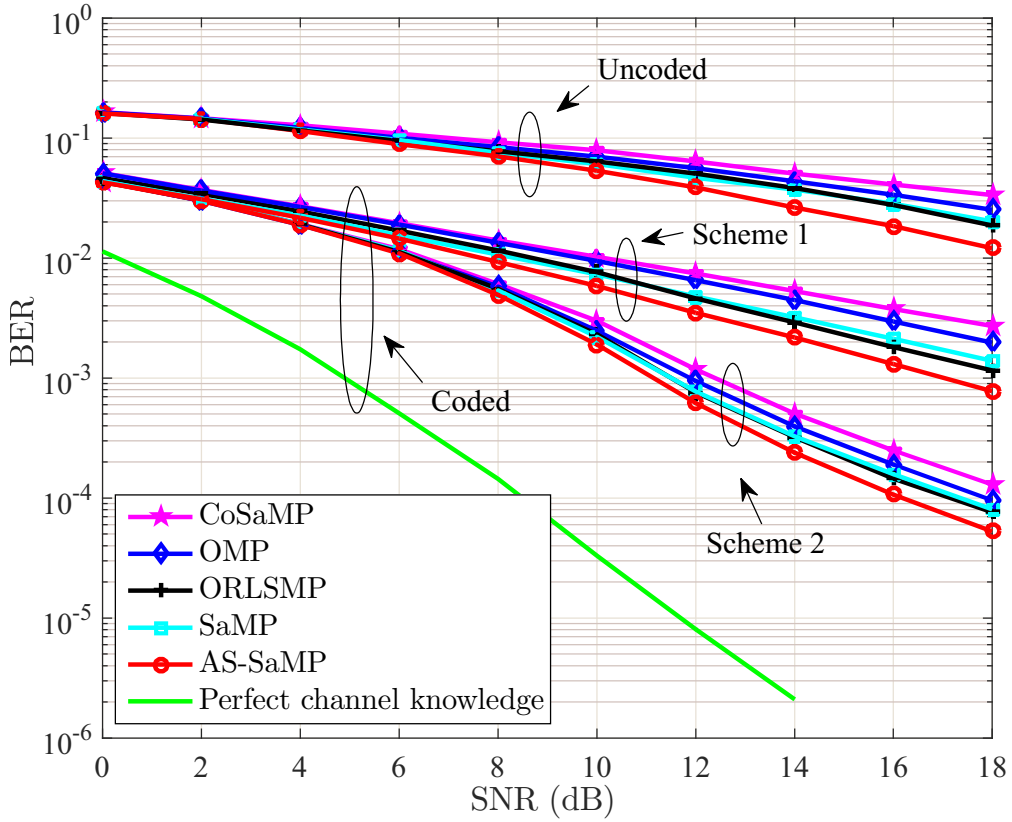


Figure 6.17: BER performance of various receivers.

obtained for various equalization techniques for the PS'10 channels. An effective noise correlation matrix was estimated using Eq. (6.12) and a good result was obtained with $\lambda_e = 0.99$. It is clear that the receiver using RPE-DFE gives the best performance, e.g., more than 8 dB SNR improvement can be observed in comparison with traditional DFE at a BER = 0.06. This is because the RPE-DFE takes into account the statistics of channel estimation errors and noise. However, in marked contrast to the other equalizers, the receiver using LE fails for all the considered estimation algorithms.

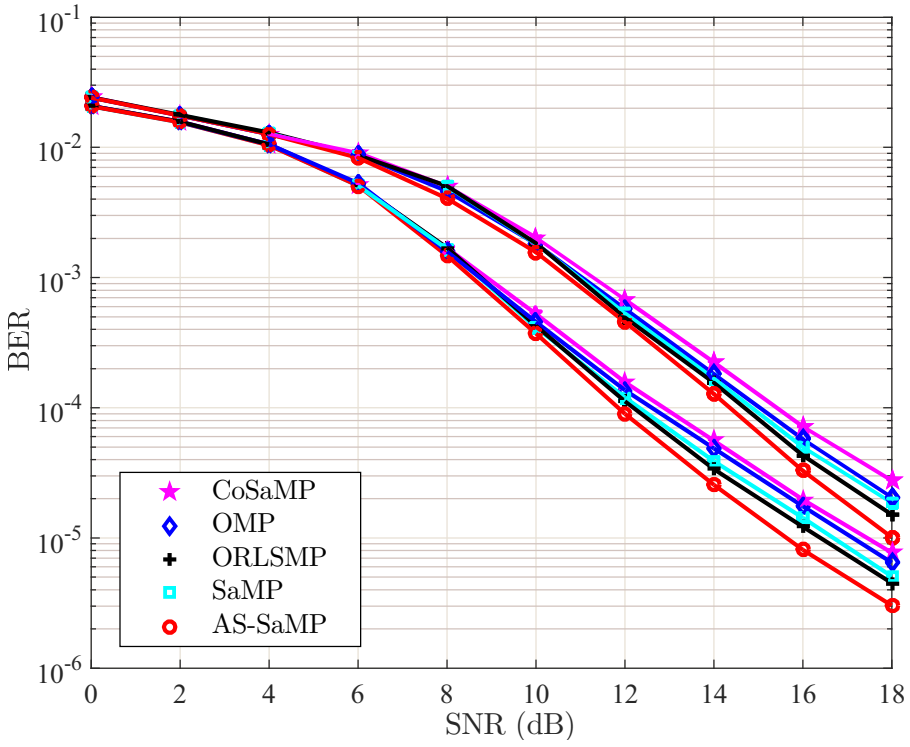
Fig. 6.17 shows the BER performance for the receivers which adopt *Scheme 1* and *Scheme 2*, respectively.[33] For comparison, the performance of a receiver with perfect CSI is also shown. Several observations are made from the figure. First, the BERs

[33] *Scheme 1* and *Scheme 2* were defined in Section 6.4.3

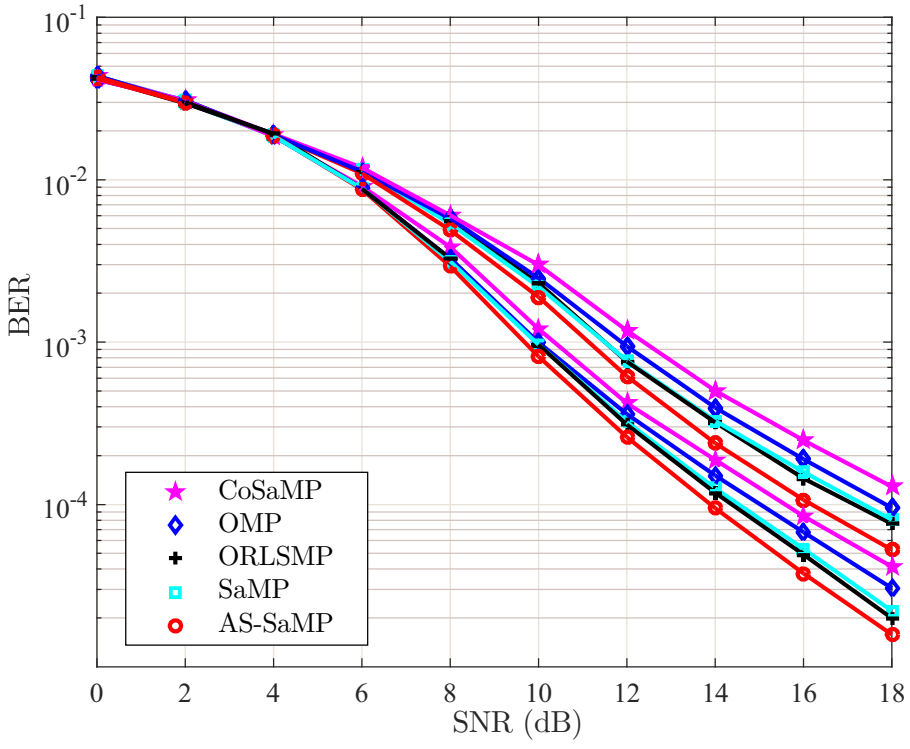
of the receivers using both schemes are significantly decreased for the coded system. Additionally, the BER performance of the receiver which employs channel prediction *Scheme 2* is improved from the one without using channel prediction for all the considered channel estimation algorithms, and such improvements are observed for a wide range of SNRs (SNR $>$ 6 dB). Second, there is at least a 5-dB SNR penalty between the curves for *Scheme 2* and for the receiver with the perfect CSI at a given BER, e.g.,10^{-3}; this indicates that the performance can be further improved with a more accurate channel prediction.

Next the BER performance for systems with channels in the four experiments are shown in Fig. 6.18, where the SPACE'08 and KAM'11 channels are considered as slow-varying channels, while the MACE'10 and PS'10 channels are relatively fast-varying. As seen, the BER results confirm that a better system performance is achieved for the slow-varying channel because of a better quality of the channel estimation and prediction, and the AS-SaMP algorithm outperforms the other considered estimation algorithms for both slow-varying and fast-varying channels. Such improvements are more pronounced at higher SNRs.

Finally, the BER performance of the overall system with the proposed decision feedback channel predictor is evaluated. In Fig. 6.19, at least a 2-dB SNR gain can be observed in the BER curves (for all the considered algorithms) with the decision feedback predictor, when compared with the predictors without using decision feedback. The enhancement is noticeable even at lower SNRs (SNR $<$ 6 dB), and is more prominent at larger SNRs. In addition, the performance with the proposed predictor for *Case 2* is degraded at lower SNRs due to the decision errors. However, the impact of these errors on the performance strongly depends on the specific receiving techniques e.g., channel coding, channel estimation and equalization that are used, and there exists an SNR threshold above which the effect of error propagation can be neglected.



(a)



(b)

Figure 6.18: BER Performance of various receivers with (a) slow-varying channels; (b) fast-varying channels. Solid lines are used for SPACE'08, dashed lines are for KAM'11, dashed-dot lines are for PS'10, and dot lines are for MACE'10.

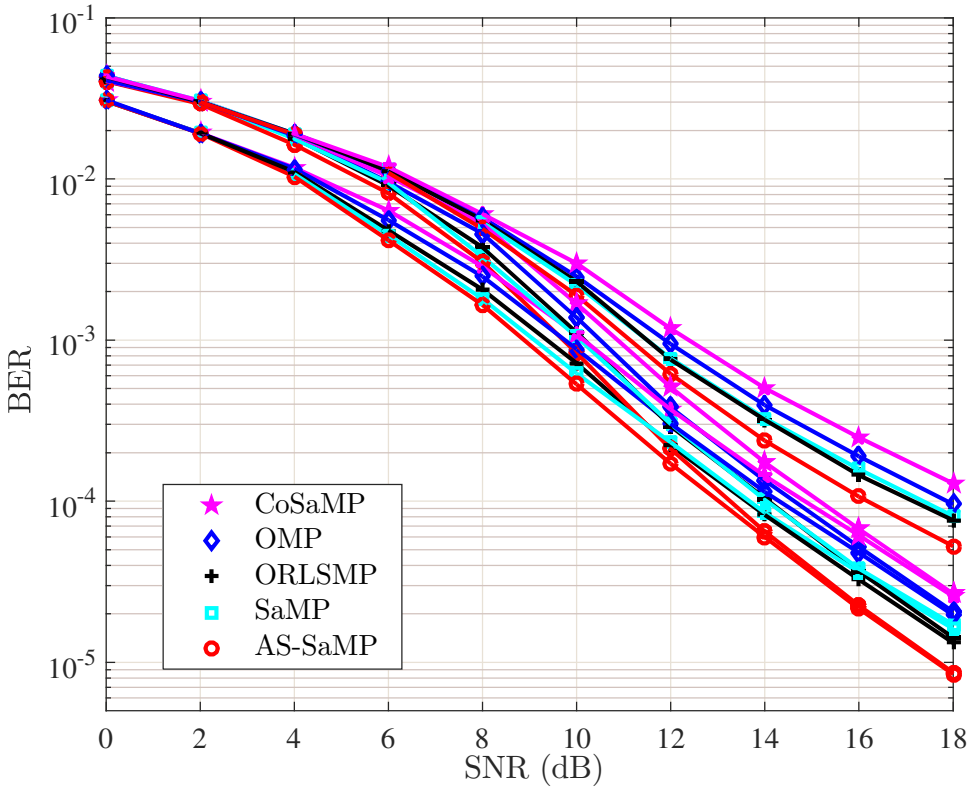


Figure 6.19: BER performance for channel predictor without/with decision feedback for the two cases. Solid lines are used for *Case 2*, dashed lines for *Case 1*, and dot lines for the predictor without using decision feedback.

A theoretical analysis on such an SNR threshold is beyond the scope of this thesis. Furthermore, the BER performance for the proposed predictor with different values of d is compared in Fig. 6.20. The results confirm that a better system performance can be achieved by decreasing the value of d, since a better prediction quality can be obtained for a small d. To summarize, the receiver using the proposed channel estimation and prediction scheme achieves a satisfying system performance with a reasonably low computational complexity.

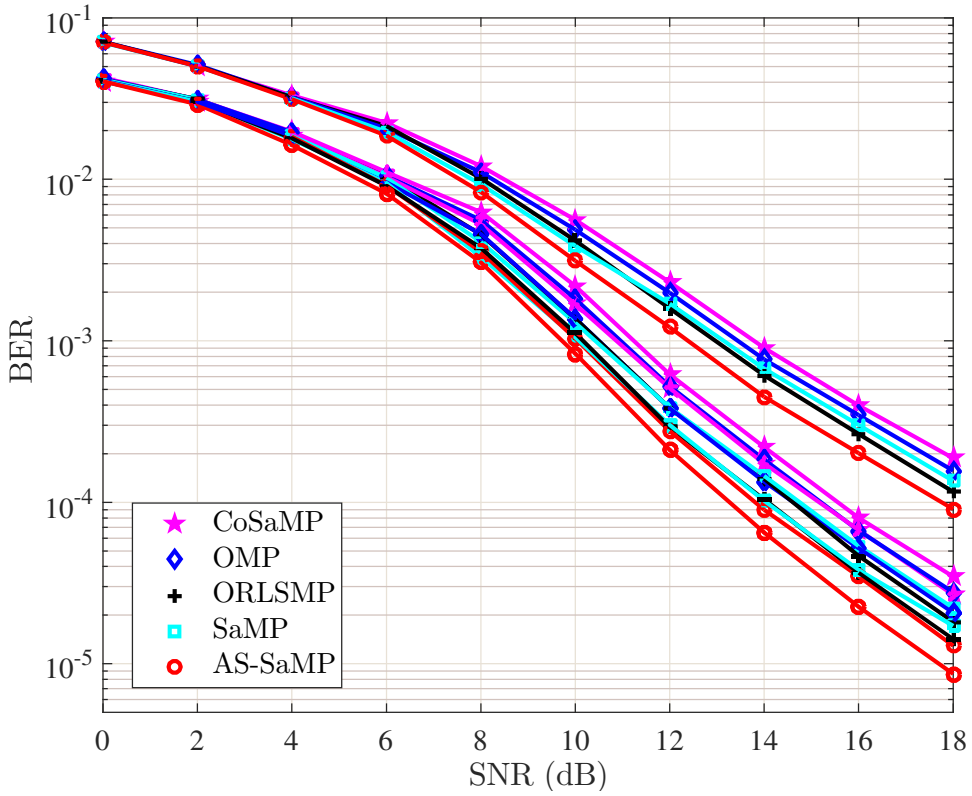


Figure 6.20: BER performance for the decision feedback channel predictor with various prediction horizon d. Solid lines are used for $d = 1$, dot lines for $d = 2$, and dashed-dot lines for $d = 3$.

6.5 Conclusion

This chapter presented a CS-based sparse channel estimation for time-varying UWA communication systems. The DDSF representation is adopted to accommodate the time variability and inherent sparsity of the UWA channels, and the dominant paths on the delay-Doppler plane were identified using various CS-based estimation algorithms. Additionally, channel estimation based equalization techniques were also studied. The robustness of the traditional DFE with respect to channel estimation error was significantly improved by taking into account the effective noise correlation matrix. Moreover, a delay-Doppler domain adaptive channel prediction method was proposed to provide future CSI. The main advantages of the proposed method are its low computational

complexity and the obviation of *a priori* knowledge of the channel and noise statistics. To further increase the prediction accuracy, the proposed channel predictor can operate in decision feedback mode, and a receive structure which combines the sparse channel estimation with the proposed predictor is presented. Numerical results showed that the proposed prediction schemes, with the AS-SaMP estimation algorithm, demonstrate a good system performance over the simulated channels obtained using the parameters obtained from the results of experiments reported in the literature.

Chapter 7

Summary and Future Work

7.1 Summary

In this thesis, sparse channel estimation and prediction algorithms were developed for UWA communication systems. The application of an emerging theory, namely CS, to time-invariant and time-varying sparse channel estimation was studied. Specifically, the contributions of this thesis are summarized as follows:

- A CS-based adaptive step size SaMP algorithm (AS-SaMP) was proposed for sparse channel estimation in UWA-OFDM systems. The AS-SaMP algorithm features an adaptive step size adjustment strategy and possesses the advantage of not requiring *a priori* knowledge of the sparsity of the channel. In addition, a comparative analysis of the existing MP-type reconstruction algorithms was presented, along with a theoretical performance analysis of the proposed algorithm. Both the theoretical analysis and simulation results confirm that the AS-SaMP algorithm can greatly improve the estimation accuracy without introducing significant additional complexity.

- In order to ensure a satisfactory estimation, a near-optimal pilot placement scheme

was proposed, which is based on the concatenated cyclic difference set with an iterative tail search (C-CDS with TS). Because the searching space of the proposed method is significantly reduced, its complexity is much lower than the iterative procedures in the literature. Simulation results showed that the AS-SaMP algorithm with the new pilot placement scheme provides a better MSE performance for the channel estimate, as well as the system BER, without significantly increasing the computational complexity, and thus offers a better trade-off between complexity and performance.

- The DDSF of an UWA time-varying channel, which often has a sparse structure, was adopted and the proposed AS-SaMP algorithm was used to estimate the dominant components in channel's DDSF. It is crucial to obtain accurate future channel estimates based on the fact that an outdated channel estimation can significantly degrade the system performance, and hence, an adaptive channel prediction scheme, which has the advantage of not requiring *a priori* knowledge of the channel and noise statistics, was developed. The proposed predictor employs the EWRLS adaptive algorithm and operates on the estimated DDSF, and therefore has a reduced complexity when compared with the existing CIR predictors.

- An adaptive DDSF predictor which uses the previously detected symbols was developed in order to increase the prediction accuracy. It was shown, through simulation results, that a satisfying performance of the proposed predictor and overall system – in terms of MSE, BER and computational complexity – was achieved using the simulated channels obtained based on experimental settings.

- The proposed technqiues are not limited to time-invariant UWA communication systems, but they can also be applied to time-varying channel estimation in other communication systems – as long as there exist a sparse representation of the

channel, e.g., channels in HDTV broadcasting standard and broadband wireless systems in hilly environment.

7.2 Future Work

The current research in the thesis opens up several interesting fronts for future study:

- Modelling dynamics in UWA channel.

  The statistical acoustic channel model in [114] provides a mathematically rigorous analytical model that takes into account certain physical aspects of acoustic propagation. Meanwhile, an efficient simulation model in which the dynamic behaviour of the small-scale coefficients relies on an autoregressive Gaussian displacement of scattering points. A natural continuation would be to investigate the dynamic properties of the small-scale coefficients from the physics point-of-view and develop explicit dynamic equation in the UWA channel model.

- Extension of proposed algorithms to channels corresponding to different environments.

  This thesis was mainly centred on shallow-water acoustic communications with short transmission range. This means that the considered channels have only nominal and boundary reflected arrivals. However, for channels in various water conditions, e.g., deep sound channel, it will be interesting to investigate how the algorithms developed in this thesis would perform. Additionally, it would be of interest to study the changes needed to be made to increase the algorithm applicability to hybrid environmental conditions.

- Theoretical analysis on impact of the error propagation in the proposed decision feedback based DDSF predictor.

As mentioned in Chapter 6, errors in the previously detected symbols result in system performance degradation. As the simulation results in Chapter 6 show, there exists an SNR threshold above which error propagation effects can be neglected. However, such an SNR threshold relies heavily on the specific modulation and coding scheme, which leaves an open topic for developing an analysis of the error probability as a function of the modulation and coding scheme, under the presence of propagation errors. Such an analysis is important to determine the performance limitations of a receiver which incorporates the proposed predictor.

- Extension towards sparse multiple-input multiple-output (MIMO) channels.

  The work in this thesis focused on exploiting sparsity of single-input single-output (SISO) channels. Since MIMO-OFDM is an appealing solution for high data rate transmissions over UWA channels, it is important to extend the proposed schemes for channel estimation and prediction in MIMO-OFDM systems.

References

[1] P. King, R. Venkatesan, and C. Li, "Modelling a shallow water acoustic communication channel using environmental data for seafloor sensor networks," in *Wireless Communications and Mobile Computing*. Wiley Inter Science, Nov. 2010, no. 11, pp. 1509–1520.

[2] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," *Ad-hoc networks (Elsevier)*, vol. 3, no. 3, pp. 257–279, Mar. 2005.

[3] S. Zhou and Z. Wang, OFDM *for underwater acoustic communications*, First Edition ed. John & Sons, Ltd., 2014.

[4] D. Brady and J. C. Preisig, "Underwater acoustic communications," in *Wireless Communications: Signal Processing Perspectives*, H. V. Poor and G. W. Wornell ed. Prentice Hall, 1998, ch. 8, pp. 330–379.

[5] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," *Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1621–1631, 1993.

[6] W. K. Lam, R. F. Ormondroyd, and J. J. Davies, "A frequency domain adaptive coded decision feedback equalizer for a broadband UWA COFDM system," in *Proceedings of MTS/IEEE Oceans Conference*, vol. 2, Sep. 2008, pp. 794–799.

[7] X. Huang and H.-C. Wu, "Robust and efficient intercarrier interference mitigation for OFDM systems in time-varying fading channels," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2517–2528, Sep. 2007.

[8] L. L. Hanzo, Y. Akhtman, L. Wang, and M. Jiang, MIMO-OFDM *for* LTE,WIFI *and* WIMAX*: Coherent versus non-Coherent and Cooperative Turbo-Transceivers*. John Wiley and IEEE Press, Dec. 2011.

[9] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 2, pp. 198–209, Apr. 2008.

[10] A. F. Molisch, *Wireless Communications*. New York, NY: Wiley, 2005.

[11] "Fcc first report and order: In the matter of revision of part 15 of the commission's rules regarding ultra-wideband transmission systems," FCC 02-48, Tech. Rep., Apr. 2002.

[12] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[13] M. Stojanvic, "Underwater acoustic communications: Design considerations on the physical layer," in *Proceedings of the 5th Annual Conference on Wireless on Demand Network Systems and Services (WONS)*, Jan. 2008, pp. 1–10.

[14] F. Blackmon, E. Sozer, and J. Proakis, "Iterative equalization, decoding techniques for shallow water acoustic channels," in *Proceedings of MTS/IEEE Oceans Conference*, vol. 4, Oct. 2002, pp. 2425–2428.

[15] J. C. Preisig, "Performance analysis of adaptive equlization for coherent acoustic communications in the time-varying ocean environment," *Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 263–278, July 2005.

[16] G. F. Edelmann, H. C. Song, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, "Underwater acoustic communications using time reversal," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 852–864, Oct. 2005.

[17] T. C. Yang, "Correlation-based decision-feedback equalizer for underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 865–880, Oct. 2005.

[18] H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time-reversal communications using adaptive channel equalizers," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 487–496, Apr. 2006.

[19] J. Wu and Y. R. Zeng, "Low complexity soft-input soft-output block decision feedback equalization," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 2, pp. 281–289, Feb. 2008.

[20] J. Tao, Y. R. Zheng, C. Xiao, T. C. Yang, and W.-B. Yang, "Channel estimation, equalization and phase correction for single carrier underwater acoustic communications," in *Proceedings of MTS/IEEE OCEANS Conference*, Kobe, Japan, Apr. 2008, pp. 1–6.

[21] J. Zhang, Y. R. Zheng, and C. Xiao, "Frequency-domain equalization for single carrier MIMO underwater acoustic communications," in *Proceedings of MTS/IEEE OCEANS Conference*, Quebec City, QC, Sep. 2008, pp. 1–6.

[22] Y. R. Zheng, C. Xiao, T. C. Yang, and W. B. Yang, "Frequency-domain channel estimation and equalization for shallow-water acoustic communications," *Elsevier Journal of Physical Communications*, vol. 3, pp. 48–63, Mar. 2010.

[23] C.-J. Wu and D. W. Lin, "Sparse channel estimation for OFDM transmission based on representative subspace fitting," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, May 2005, pp. 495–499.

[24] W. K. Lam and R. F. Ormondroyd, "A coherent COFDM modulation system for a time-varying frequency-selective underwater acoustic channel," in *Proceedings of the 7th International Conference on Electronic Engineering in Oceanography*, Jun. 1997, pp. 198–203.

[25] S.-J. Hwang and P. Schniter, "Efficient multicarrier communication for highly spread underwater acoustic channels," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 9, pp. 1674–1683, Dec. 2008.

[26] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "An adaptive matching pursuit algorithm for sparse channel estimation," in *Proceedings of IEEE on Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, Mar. 2015, pp. 626–630.

[27] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 52–55, Jan. 2005.

[28] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[29] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[30] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Novel compressed sensing-based channel estimation algorithm and near-optimal pilot placement scheme," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2590–2603, Apr. 2016.

[31] H. Arslan and G. E. Bottomley, "Channel estimation in narrowband wireless communication systems," *Wireless Communications and Mobile Computing*, vol. 1, no. 2, pp. 201–219, Apr. 2001.

[32] H. C. Song, P. Roux, W. S. Hodgkiss, W. A. Kuperman, T. Akal, and M. Stevenson, "Multiple-input/multiple-output coherent time reversal communications in a shallow water acoustic channel," *IEEE Journal of Oceanic Engieering*, vol. 32, no. 3, pp. 170–178, Jan. 2006.

[33] A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. B. Porter, and the KausiEx Goup, "Impact of ocean variability on coherent underewater acpustic communications during the Kauai experiment (KauaiEx)," *Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 856–854, 2008.

[34] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcast*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[35] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing:A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[36] M. Stajanovic, "Retrofocusing techniques for high rate acoustic communications," *Journal of Acoustic Society of America*, vol. 117, no. 3, pp. 1173–1185, Mar. 2005.

[37] J. van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in OFDM systems," in *Proceeding of IEEE Vehicular Technology Conference (VTC)*, Chicago, IL, Jul. 1995, pp. 815–819.

[38] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, May 2007.

[39] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[40] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb. 2006.

[41] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[42] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[43] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 929–999, Oct. 2007.

[44] T. Kang and R. A. Iltis, "Iterative carrier frequency offset and channel estimation for underwater acoustic OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 9, pp. 1650–1661, Dec. 2008.

[45] C.-J. Wu and D. W. Lin, "A group matching pursuit algorithm for sparse channel estimation for OFDM transmission," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, May 2006, pp. 429–432.

[46] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications Systems*, vol. CS-11, no. 4, pp. 360–393, Dec. 1963.

[47] D. Eiwen, G. Taubböck, F. Hlawatsch, and H. Feichtinger, "Compressive tracking of doubly selective channels in multicarrier systems based on sequential delay-Doppler sparsity," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2928– 2931.

[48] A. J. Goldsmith and S. G. Chua, "Variable-rate variable power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[49] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 844–855, Jun. 1999.

[50] Y. Zhang, R. Venkatesan, C. Li, and O. A. Dobre, "Compressed sensing-based time-varying channel estimation in UWA-OFDM networks," in *Proceedings*

of International Wireless Communications and Mobile Computing Conference (IWCMC), Aug. 2015, pp. 1520–1525.

[51] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Jul. 2001.

[52] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[53] D. Baron, M. B. Wakin, M. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," *IEEE Signal Processing Magazine*, 2005. [Online]. Available: Available:http://dsp.rice.edu/cs/DCS112005.pdf

[54] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[55] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jul. 2006.

[56] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[57] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 2992–3004, Dec. 2007.

[58] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[59] J. Huang, C. R. Berger, S. Zhou, and J. Huang, "Comparison of basis pursuit algorithms for sparse channel estimation in underwater acoustic OFDM," in *Proceedings of IEEE Oceans Conference-Sydney*, Sydney, Australia, May 2010, pp. 1–6.

[60] Y. Zhang, "User's guide for yall1: Your algorithms for l1 optimization," Tech. Rep. [Online]. Available: downloableathttp://www.caam.rice.edu/?optimization/L1/YALL1/.

[61] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, Nov. 2009, pp. 18 914–18 919.

[62] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Transactions on Signal Processing*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.

[63] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Communications of the ACM: Research Highlights section*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[64] T. T. Do, G. Lu, N. Nam, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2008, pp. 581–587.

[65] A. C. Singer, J. K. Nelson, and S. S. Kozat, "Signal processing for underwater acoustic communications," *IEEE Communication Magazine*, vol. 47, no. 1, pp. 90–96, Jan. 2009.

[66] O.-H. Lee, Y.-J. Son, and K.-M. Kim, "Underwater digital communication using acoustic channel estimation," in *Proceedings of MTS/IEEE Oceans Conference*, vol. 4, Oct. 2002, pp. 2453–2456.

[67] Y. R. Zheng, C. Xiao, T. C. Yang, and W.-B. Yang, "Frequency-domain channel estimation and equalization for single carrier underwater acoustic communications," in *Proceedings of MTS/IEEE Oceans Conference*, Sep. 2007, pp. 1–6.

[68] Y.-F. Cheng and D. Etter, "Analysis of an adaptive technique for modeling sparse systems," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 37, no. 2, pp. 254–264, Feb. 1989.

[69] M. Stojanovic, L. Freitag, and M. Johnson, "Channel-estimation-based adaptive equalization of underwater acoustic signals," in *Proceedings of MTS/IEEE Oceans Conference*, vol. 2, Seattle, WA, Sept. 1999, pp. 985–990.

[70] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proceedings of MTS/IEEE Oceans Conference*, vol. 3, San Diego, CA, Oct. 1995, pp. 1417–1422.

[71] S. Ozen, W. Hillery, M. Zoltowski, S. M. Nereyanuru, and M. Fimoff, "Structured channel estimation based decision feedback equalizers for sparse multipath channels with applications to digital TV receivers," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, vol. 1, Pacific Grove, Nov. 2002, pp. 558–564.

[72] Y. V. Zakharov and V. P. Kodanev, "Multipath-Doppler diversity of OFDM signals in an underwater acoustic channel," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, Jun. 2000, pp. 2941–2944.

[73] B. Li, S. Zhou, M. Stojanovic, and L. Freitag, "Pilot-tone based ZP-OFDM demodulation for an underwater acoustic channel," in *Proceedings of MTS/IEEE Oceans Conference*, Sep. 2006, pp. 1–5.

[74] M. Stojanovic, "Low complexity OFDM detector for underwater acoustic channels," in *Proceedings of MTS/IEEE Oceans Conference*, Sep. 2006, pp. 1–6.

[75] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, pp. 18–48, Jul. 2007.

[76] S. Wu and Y. Bar-Ness, "OFDM channel estimation in the presence of frequency offset and phase noise," in *Proceedings of IEEE International Conference on Communications (ICC)*, vol. 5, May 2003, pp. 3366–3370.

[77] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[78] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapidly dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[79] D. Schafhuber, G. Matz, F. Hlawatsch, and P. Loubaton, "MMSE estimation of time-varying channles for DVB-T systems with strong co-channel interference," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, vol. 3, Toulouse, France, Sept. 2002, pp. 25–28.

[80] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2000.

[81] G. Taubö̈ck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 255–271, Apr. 2010.

[82] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proceedings of IEEE International Symposium on Information Theory*, Saint Petersburg, Russia, Aug. 2011, full version at arXiv:1010.5141.

[83] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proceedings of Asilomar Conference Signals, System and Computing*, Pacific Grove, CA, Nov. 2011, pp. 799–803.

[84] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[85] X. Bia, X. Chen, and Y. Zhang, "Variable step size stagewise adaptive matching pursuit algorithm for image compressed sensing," in *Proceedings of International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Kunming, China, Aug. 2013, pp. 1–4.

[86] X. He, R. Song, and W. Zhu, "Optimal pilot pattern design for compressed sensing-based sparse channel estimation in OFDM systems," *Circuits, Systems, and Signal Processing*, vol. 31, no. 4, pp. 1379–1395, Aug. 2012.

[87] L. Applebaum, W. U. Bajwa, A. R. Calderbank, J. Haupt, and R. Nowak, "Deterministic pilot sequences for sparse channel estimation in OFDM systems," in *Proceedings of International Conference on Digital Signal Processing*, Corfu, Greece, Jul. 2011, pp. 1–7.

[88] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems," *IEEE Signal Processing Letters*, vol. 18, no. 12, pp. 749–752, Dec. 2011.

[89] C. R. Berger, J. Gomes, and J. M. F. Moura, "Study of pilot designs for cyclic-prefix OFDM on time-varying and sparse underwater acoustic channels," in *Proceedings of OES/IEEE Oceans Conference*, Santander, Spain, Jun. 2011, pp. 1–8.

[90] P. Pakrooh, A. Amini, and F. Marvasti, "OFDM pilot allocation for sparse channel estimation," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 59, pp. 1–9, Mar. 2012.

[91] C. Qi and L. Wu, "A study of deterministic pilot allocation for sparse channel estimation in OFDM systems," *IEEE Communications Letters*, vol. 16, no. 5, pp. 742–744, May. 2012.

[92] C. Qi, G. Yue, L. Wu, Y. Huang, and A. Nallanathan, "Pilot design schemes for sparse channel estimation in OFDM systems," *IEEE Transactions on Vehicular Technology*, no. 99, pp. 1–13, Jun. 2014.

[93] J. Haupt, L. Applebaum, and R. Nowak, "On the restricted isometry of deterministically subsampled fourier matrices," in *Proceedings of the Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2010, pp. 1–6.

[94] W.-J. Zeng and W. Xu, "Fast estimation of sparse doubly spread acoustic channels," *Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 303–317, Jan. 2012.

[95] F. Qu, Z. Wang, L. Yang, and Z. Wu, "A journey toward modeling and resolving Doppler in underwater acoustic communications," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 49–55, Feb. 2016.

[96] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 255–271, Apr. 2010.

[97] E. Zamanizadeh, J. Gomes, and J. M. Bioucas-Dias, "Identification and matching of sparse delay-doppler spread functions from high-frequency communications signals," in *Proceedings of MTS/IEEE Oceans Conference*, Seattle, WA, Sept. 2010, pp. 1–10.

[98] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2008, pp. 575–582.

[99] A. D. Hallen, S. Hu, and H. Hallen, "Long range prediction of fading signals: Enabling adaptive transmission for mobile radio channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 62–75, May 2000.

[100] W. Li, "Estimation and tracking of rapidly time-varying broadband acoustic communication channels," Ph.D. dissertation, Technology/Woods Hole Oceanography Institution, MA, USA, 7 2006, ch. 2.

[101] A. W. Fuxjarger and R. A. Iltis, "Adaptive parameter estimation using parallel kalman filtering for spread spectrum code and Doppler tracking," *IEEE Transactions on Communications*, vol. 42, no. 6, pp. 2227–2230, Jun. 1994.

[102] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalzation of fading channels with random coefficients," *Signal Processing*, vol. 53, pp. 211–229, Sept. 1996.

[103] E. J. Sullivan and J. V. Candy, "Space-time array processing: The model based approach," *Journal of Acoustic Society of America*, vol. 102, no. 5, pp. 2809–2820, Nov. 1997.

[104] R. Nadakuditi and J. C. Preisig, "A channel subspace post-filtering approach to adaptive least-square estimation," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1901–1914, Jul. 2004.

[105] S. H. Huang, J. Tsao, T. C. Yang, and S. W. Cheng, "Model-based signal subspace channel tracking for correlated underwater acoustic communication channels," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 2, pp. 343–356, Apr. 2014.

[106] Y. Liu and S. D. Blostein, "Identification of freqeuncy non-selective fading channels using decision feedback and adaptive linear prediction," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1484 – 1492, Mar. 1995.

[107] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 503–602, Mar. 2005.

[108] Z. Wang, S. Zhou, J. C. Preisig, K. Pattipati, and P. Willett, "Clustered adaptation for estimation of time-varying underwater acoustic channels," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3079–3091, Jun. 2012.

[109] A. Radosevic, T. M. Duman, J. G. Proakis, and M. Stojanovic, "Channel prediction for adaptive modulation in underwater acoustic communications," in *Proceedings of OES/IEEE Oceans*, Santander, Spain, Jun. 2011, pp. 1–5.

[110] C. Lv, S. Hou, and W. Mei, "Adaptive prediction of channels with sparse features in OFDM systems," *International Journal of Antennas and Propagation*, vol. 2013, pp. 1–5, Apr. 2013.

[111] N. Lin, H. Sun, E. Cheng, J. Qi, X. Kuai, and J. Yan, "Prediction based sparse channel estimation for underwater acoustic OFDM," *Applied Acoustics*, vol. 96, pp. 94–100, Sept. 2015.

[112] S. O. Haykin, *Adaptive Filter Theory*, 5th ed. New York, NY: Pearson, 2013.

[113] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1412–1417, Jul. 2005.

[114] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 4, pp. 701–717, Oct. 2013.

[115] L. Berkhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics*, 1st ed. New York, NY: Springer, 1982.

[116] P. Gagnon, "Estimation and tracking of rapidly time-varying broadband acoustic communication channels," Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, 9 2012.

[117] M. B. Porter. (http://oalib.jlsresearch.com/Rays/omdex.html) Bellhop code.

[118] A. J. Siva, O. C. Rodriguez, F. Zabel, J. Huillery, and S. M. Jesus, "Underwater acoustic simulations with a time variable acoustic propagation model," in *Proceedings of European Conference on Underwater Acoustics*, Istambul, Turkey, Jul. 2010.

[119] J. Mark, L. Freitag, and M. Stojanovic, "Improved doppler tracking and correlation for underwater acoustic communications," in *Proceedings of IEEE Inter-*

national Conference on Aoustics, Speech, and Signal Processing, vol. 1, Munich, Germany, Apr. 1997, pp. 575–578.

[120] F. Qu, X. Nie, and W. Xu, "A two-stage approach for the estimation of doubly spread acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 1, pp. 131–143, Jan. 2015.

[121] I. . Fevrier, S. Gelfand, and M. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 927–937, Jun. 1999.

[122] S. Ariyavisitakul, N. Sollengerger, and L. Greenstein, "Tap-selectable decision-feedback equalization," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1497–1500, Dec. 1997.

[123] P. Qarabaqia and M. Stojanovic, "Statistical modelling of a shallow water acoustic communication channel," in *Proceedings Underwater Acoustic Measurements Conf.*, Nafplion, Greece, Jun. 2009, pp. 1341–1350.

[124] G. Tang, N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[125] D. Wei and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[126] D. Donoho, V. Stodden, and Y. Tsaig, "About sparselab," Mar. 2007. [Online]. Available: https://sparselab.stanford.edu/

[127] C. J. Colbourn and J. H. Dinitz, "Other combinatorial designs," in *Handbook of Combinatorial Designs*, 2nd ed. CRC Press, 2007, ch. 5, pp. 392–436.

[128] J. Myers and A. D. Well, *Research Design and Statistical Analysis*, 2nd ed. Lawrence Erlbaum, 2003.

[129] W. Li and J. C. Preisig, "Sparse estimation and equalization of rapidly varying wideband acoustic communication channels," in *Proceedings of MTS/IEEE OCEANS Conference*, Boston, MA, May 2006, pp. 1–6.

[130] T. H. Eggn, A. B. Baggeroer, and J. C. Preisig, "Communication over Doppler spread channels, Part I: Channel and receiver presentation," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1, pp. 62–72, Jan. 2000.

[131] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. San Diego, CA: Academic Press, 1979.

Appendices

Appendix A

Theoretical Performance of AS-SaMP: Proof of Lemma 1, 2 and Corollary 1

A.1 Proof of Lemma 1 and 2

Let us recall the **Lemma 1** and **2** from Section 4.3.

Lemma 1. Given an arbitrary K-sparse signal $\mathbf{h}$ and the corresponding measurement $\mathbf{Y_p} = \mathbf{Ah}$. Let the total number of stages decided by AS-SaMP be J and $s_i, i \in \{1, 2, ..., J\}$ be the step size of the ith stage. If $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_J} < 0.06$ [41], where $K_J = \sum_{i=1}^{J} s_i$ is the estimated sparsity level, the last stage of AS-SaMP is equivalent to SaMP algorithm with estimated sparsity level K_J, except possibly different contents in the final support set and the observation residual vector.

Proof. During the last stage of AS-SaMP, the final support set has size K_J. Given the same size of the final support set, both algorithms use the same preliminary and final correlation test, which returns the K_J indices corresponding to the largest absolute

values of $|\mathbf{A}^{\dagger}_{\mathbf{C}^j}\mathbf{Y_p}|$, j denotes the iteration index. The only differences are in the content of the final support set and the observation residual vector. ■

Lemma 2. AS-SaMP guarantees the convergence of the recovery process. The upper-bounded number of iterations that AS-SaMP involves is

$$-\log(\frac{|h_{\min}|}{\parallel \mathbf{h} \parallel})(\frac{-1}{\log(C_{K_1})}+\frac{-1}{\log(C_{K_2})}+...+\frac{-1}{\log(C_{K_J})})+J, \tag{A.1}$$

where $h_{\min}$ is the non-zero element with the minimum magnitude and $C_{K_i}=\frac{2\delta_{3K_i}(1+\delta_{3K_i})}{(1-\delta_{3K_i})^3}$, $i=1,2,...,J$, δ_{3K_i} is the RIP parameter in the ith stage, and K_i is the size of final support set in the ith stage.

Proof. **Lemma 2.1** is introduced which serves as a foundation for the proof of **Lemma 2**.

Lemma 2.1. The energy difference between the signal captured by the final support set from the current iteration and the final support set from the previous iteration, i.e., $\parallel \hat{\mathbf{h}}_{\mathbf{F}^j} \parallel^2 - \parallel \hat{\mathbf{h}}_{\mathbf{F}^{j-1}} \parallel^2$, decreases as the number of iterations increases before the estimated sparsity level reaches the true sparsity level.

Proof of **Lemma 2.1** is postponed to Appendix A.3. Similar to SaMP, AS-SaMP takes a finite number of iterations to approach the sparse estimation. If the algorithm falls into an infinite loop of a certain stage, the final support set will repeat and this is in contradiction to the fact that the energy difference decreases monotonically. Intuitively, AS-SaMP reaches the final estimation with the same estimated sparsity level faster than SaMP because the most significant entries are reconstructed by selecting a larger number of coordinates into the support set during the initial stages. Let the number of iterations required in the ith stage using the AS-SaMP algorithm be n_i^{it}, $i=1,2,...,J$. According to Theorem 6 in [125], for each iteration in a particular stage both SaMP and

AS-SaMP contain two *correlation maximization tests* and the property below holds.

$$\begin{aligned} n_1^{it} &\leq \frac{-\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_1})} + 1, \\ n_2^{it} &\leq \frac{-\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_2})} + 1, ..., \\ n_J^{it} &\leq \frac{-\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_J})} + 1. \end{aligned} \tag{A.2}$$

Let the total number of iterations required be n^{total}, then

$$\begin{aligned} n^{total} &= n_1^{it} + n_2^{it} + ...n_J^{it}, \\ &\leq -\log(\frac{|h_{\min}|}{\| \mathbf{h} \|})(\frac{-1}{\log(C_{K_1})} + ... + \frac{-1}{\log(C_{K_J})}) + J. \end{aligned} \tag{A.3}$$

■

A.2 Proof of Corollary 1

Corollary 1. Provided that $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_{s-AS-SaMP}} < 0.06$ and $\delta_{3K_{s-SaMP}} < 0.06$, where $K_{s-AS-SaMP}$ and K_{s-SaMP} are the estimated sparsity level for AS-SaMP and SaMP, respectively, the upper-bounded number of iterations for AS-SaMP is smaller than that for SaMP.

Proof. Since both the SaMP and AS-SaMP algorithms use the *preliminary* and *final* tests, the upper-bounded number of iterations in **Lemma 2** can also be applied to the SaMP algorithm. Consider the same target signal for both algorithms and according to **Lemma 2** we have

$$n^{total} \leq -\log(\frac{|h_{\min}|}{\| \mathbf{h} \|})(\frac{-1}{\log(C_{K_1})} + ... + \frac{-1}{\log(C_{K_J})}) + J.$$

As $0 < \delta_{3K_1} \leq \delta_{3K_2} ... \leq \delta_{3K_J} < 0.06$, then $0 < \frac{-1}{\log(C_{K_1})} \leq \frac{-1}{\log(C_{K_2})} \leq ... \leq \frac{-1}{\log(C_{K_J})}$. Thus

we have

$$\frac{-1}{\log(C_{K_1})} + \frac{-1}{\log(C_{K_2})} + ... + \frac{-1}{\log(C_{K_J})} \leq \frac{-J}{\log(C_{K_J})},$$

and therefore

$$n^{total} \leq \frac{-J\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_J})} + J.$$

With the same target signal and the total number of stages, the upper bound only depends on C_{K_J}. According to *Remark* 1 in Section 4.3.1, $0 < \delta_{3K_{s-AS-SaMP}} \leq \delta_{3K_{s-SaMP}} < 0.06$, and therefore, $C_{K_{s-AS-SaMP}} \leq C_{K_{s-SaMP}}$. Clearly, $0 < \frac{-1}{\log(C_{K_{s-AS-SaMP}})} \leq \frac{-1}{\log(C_{K_{s-SaMP}})}$, and

$$\frac{-J\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_{s-AS-SaMP}})} + J \leq \frac{-J\log(\frac{|h_{\min}|}{\|\mathbf{h}\|})}{-\log(C_{K_{s-SaMP}})} + J.$$

■

For example, given $\delta_{3K_{s-AS-SaMP}} = 0.01$ which leads to $\log(C_{K_{s-AS-SaMP}}) = -5.59$ and the total number of stages is 5. Suppose $\log(\frac{|h_{\min}|}{\|\mathbf{h}\|}) = -7$, the upper bound of the number of iterations that the AS-SaMP algorithm involves is 11. On the other hand, with the same $\mathbf{h}$ and J, suppose $\delta_{3K_{s-SaMP}} = 0.05$, the upper bound of the number of iterations for the SaMP algorithm is 16.

A.3 Proof of Lemma 2.1

The proof is derived from Theorem 2 in [125] because both the preliminary and final test are correlation maximization tests.

Proof. Provided that the sensing matrix $\mathbf{A}$ satisfies the RIP with parameter $\delta_{3K_J} < 0.06$.

$$\begin{aligned} \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^t} \|^2 &\leq C_{K_i}^2 \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^{t-1}} \|^2, \\ &\leq C_{K_J}^2 \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^{t-1}} \|^2 = \zeta \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^{t-1}} \|^2, \end{aligned} \tag{A.4}$$

where $\hat{\mathbf{h}}_{\overline{\mathbf{F}}^j}$ is the reconstructed signal not captured by $\mathbf{F}^j$ after the jth iteration. (A.4) is based on $0 < C_{K_1} \leq C_{K_2} \leq ... \leq C_{K_J} < 1$, and therefore $\zeta = C_{K_J}^2 < 1$. Thus, the following derivation holds,

$$\begin{aligned} \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^1} \|^2 - \zeta \| \mathbf{h} \|^2 \leq 0 \leq \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^2} \|^2, \\ 0 \leq \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^1} \|^2 - \| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^2} \|^2 \leq \zeta \| \mathbf{h} \|^2, \end{aligned} \tag{A.5}$$

where $\mathbf{h} = \hat{\mathbf{h}}_{\overline{\mathbf{F}}^0}$. As $\| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^1} \|^2 = \| \mathbf{h} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^1} \|^2$ and $\| \hat{\mathbf{h}}_{\overline{\mathbf{F}}^2} \|^2 = \| \mathbf{h} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^2} \|^2$, (A.5) can be written as:

$$\begin{aligned} 0 \leq (\| \mathbf{h} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^1} \|^2) - (\| \mathbf{h} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^2} \|^2) \leq \zeta \| \mathbf{h} \|^2, \\ 0 \leq \| \hat{\mathbf{h}}_{\mathbf{F}^2} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^1} \|^2 \leq \zeta \| \mathbf{h} \|^2 . \end{aligned} \tag{A.6}$$

Similarly, we have

$$\begin{aligned} 0 \leq \| \hat{\mathbf{h}}_{\mathbf{F}^3} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^2} \|^2 \leq \zeta^2 \| \mathbf{h} \|^2, \\ 0 \leq \| \hat{\mathbf{h}}_{\mathbf{F}^4} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^3} \|^2 \leq \zeta^3 \| \mathbf{h} \|^2, \cdots \\ 0 \leq \| \hat{\mathbf{h}}_{\mathbf{F}^t} \|^2 - \| \hat{\mathbf{h}}_{\mathbf{F}^{t-1}} \|^2 \leq \zeta^{t-1} \| \mathbf{h} \|^2 . \end{aligned} \tag{A.7}$$

As $1 > \zeta > \zeta^2 > \zeta^3 > \zeta^4 \cdots$, the energy difference between two consecutive iterations converges to a small positive value which is related to the RIP parameter. ■

Appendix B

Order-Recursive LS-MP Algorithm

The Order-Recursive LS-MP algorithm [43] is described in Algorithm 4, and its operation counts per iteration is summarized in Table B.1.

Table B.1: Computational complexity per iteration for ORLSMP

Step	The number of operations
3	N
5	MN_p^2
6	$N_p^2 + N_p$
7	1
8	N_p
9	NMN_p^2
10	NN_p
Overall complexity per iteration	$O(NMN_p^2 + NN_p + MN_p^2)$

Algorithm 4 Order-Recursive LS-MP Algorithm

Input: $M \times 1$ received signal $\mathbf{y}$, $M \times N$ measurement matrix $\mathbf{\Phi}$ with each column vector $\boldsymbol{\varphi}(j)$ and $j = 0, 1, ..., N-1$, Tolerance T;

Output: $N \times 1$ Estimation channel's DDSF $\mathbf{u}$ with N_p significant elements;

Initialize residual $\mathbf{r}^0 = \mathbf{y}$, Iteration index $t = 2$,
$b^0(j) = |\boldsymbol{\varphi}(j)^H \mathbf{r}^0|$, $\delta^0(j) = \|\boldsymbol{\varphi}(j)\|^2$,
$s^1 = \arg\max_{j=0,...,N-1} \frac{b^0(j)}{\delta^0(j)}$,
Indices set $\mathbf{I}^1 = s^1$, Candidate support set $\mathbf{C}^1 = [\boldsymbol{\varphi}(s^1)]$,
$P^1 = \|\boldsymbol{\varphi}(s^1)\|^{-2}$, $u^1 = P^0 b^0(j)$,
$b^1(j) = b^1(j) - \boldsymbol{\varphi}(j)^H \boldsymbol{\varphi}(s^1) u^1$,
$\mathbf{Q}^1 = \mathbf{I} - \mathbf{C}^1 P^1 \mathbf{C}^{1H}$,
$\delta^1(j) = \boldsymbol{\varphi}(j)^H \mathbf{Q}^1 \boldsymbol{\varphi}(j)$ for $j = 0, ..., N-1, j \notin \mathbf{I}^0$,
$\mathbf{u}^1 = u^1$.

while $(\|\mathbf{r}^t\|_2 > T)$ **do**

$s_t = \arg\max_{j=0,...,N-1, j \notin \mathbf{I}^{t-1}} \frac{|b^(t-1)(j)|^2}{\delta^(t-1)(j)}$

$\mathbf{I}^t = \{\mathbf{I}^{t-1}, s^t\}$, $\mathbf{C}^t = [\mathbf{C}^{t-1}, \boldsymbol{\varphi}(s^t)]$

$$\mathbf{\Gamma}^{t-1} = \begin{bmatrix} -\mathbf{P}^{t-1}(\mathbf{C}^{t-1})^H \boldsymbol{\varphi}(s^t) \\ 1 \end{bmatrix}$$

$$\mathbf{P}^t = \begin{bmatrix} \mathbf{P}^{t-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + (\delta^{t-1}(s^t))^{-1} \mathbf{\Gamma}^{t-1} (\mathbf{\Gamma}^{t-1})^H$$

$u_t = (\delta^{t-1}(s^t))^{-1} b^{t-1}(s^t)$

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}^{t-1} \\ 0 \end{bmatrix} + \mathbf{\Gamma}^{t-1} u_t$$

$b^0(j) - \boldsymbol{\varphi}(j)^H \mathbf{C}^t \mathbf{u}$ for $j = 0, 1, .., N-1, j \notin \mathbf{I}^t$

$\delta^t(j) = \delta^{t-1}(j) - \frac{|\boldsymbol{\varphi}(j)^H \mathbf{C}^t \mathbf{\Gamma}^{t-1}|^2}{\delta^{t-1}(s^t)}$

end while