Title
Towards Network Reliability: Harnessing Interference and Dynamics

Author
Mishra, Shaunak

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Towards Network Reliability: Harnessing Interference and Dynamics

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Shaunak Mishra

2016

ABSTRACT OF THE DISSERTATION

Towards Network Reliability: Harnessing Interference and Dynamics

by

Shaunak Mishra

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2016

Professor Suhas N. Diggavi, Chair

In the past few decades, networked systems have revolutionized, among other things, the way we communicate, and the way we control physical processes. However, addressing the unreliability associated with such networked systems continues to be a fundamental challenge. The unreliability in such networked systems can be attributed to a variety of factors including the lack of co-ordination among network components, noisy measurements, and security vulnerabilities leading to malicious elements in the network. In view of the factors mentioned above, this dissertation takes steps towards enhancing network reliability by addressing some fundamental challenges in two contemporary network setups: wireless communication networks and cyber-physical systems (CPS).

In the context of wireless communication networks, we focus on the unreliability stemming from interference, *i.e.*, a scenario where multiple transmitter-receiver pairs sharing the same frequency band interfere with each other. This is a fundamental bottleneck in reliably scaling data rates; an essential requirement for supporting the huge growth in mobile Internet traffic. The temporal nature of interference in such networks depends on the underlying traffic and the resource allocation decisions of neighboring base stations. In practice, due to the bursty nature of data traffic and uncoordinated resource allocations across base stations, the resulting interference at the physical layer tends to be bursty. Though recent advances in network information theory have led to

a fundamental understanding of interference in wireless networks, the channel models usually assume that interference is always present and hence do not capture the impact of bursty interference. Thus, the following question emerges: how can we harness bursty interference and potentially improve the (information theoretic) system capacity? In this dissertation, we investigate the above question in the context of parallel (multicarrier) interference channels, and show positive results, *i.e.*, we demonstrate that harnessing burstiness leads to a non-trivial increase in the system capacity. We develop interference management schemes which leverage feedback from receivers to recover interfered symbols in previous transmissions. For the case when feedback from receivers is not available, we develop opportunistic schemes based on a degraded message set approach to exploit burstiness. In addition, we develop tight outer bounds for a variety of regimes, and hence prove the information theoretic optimality of our interference management schemes in those regimes.

In the context of CPS, we focus on the unreliability arising out of security vulnerabilities. Due to the close interaction between the cyber and physical components, CPS pose unique security challenges. Furthermore, conventional cyber security methods which are oblivious to the underlying physical dynamics leave open the possibility of leveraging such dynamics for addressing security vulnerabilities in CPS. With this motivation, and considering the fact that state estimation is a crucial component in CPS, we study the problem of securing state estimation in linear dynamical systems despite active adversaries. In particular, we focus on two attack scenarios: (i) attacks on software/hardware where state estimation is performed (observer attacks), and (ii) attacks on sensors used for state estimation. To protect against observer attacks we propose an architecture where state estimation is performed across multiple computing nodes (observers), and derive tight bounds on the number of attacked observers which can be tolerated for accurate state estimation. To protect against sensor attacks, we propose a secure state estimation algorithm, and derive (optimal) bounds on the achievable state estimation error given an upper bound on the number of attacked sensors. The proposed state estimator involves Kalman filters operating over subsets of sensors to search for a sensor subset which is reliable for state estimation. As a result of independent interest, we give a coding theoretic view of attack detection and state estimation against sensor attacks in a noiseless dynamical system.

Hence, in a nutshell, this dissertation takes fundamental steps towards enhancing the reliability of wireless communication networks (by harnessing bursty interference), and cyber-physical systems (by leveraging physical dynamics for security).

The dissertation of Shaunak Mishra is approved.

Christina Panagio Fragouli

Paulo Tabuada

Rafail Ostrovsky

Suhas N. Diggavi, Committee Chair

University of California, Los Angeles

2016

To Maa and Bapa.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Guru Brahma guru Vishnu,
guru devo Maheshwara,
guru sakshat param Brahma,
tasmai shri gurave namah.

The above Sanskrit saying from the ancient Indian text *Skanda Purana* highlights the importance of a teacher (*guru*) and translates to: *guru is the creator, manager and destroyer of the world; sublime prostrations to him*.

I would like to express my heartfelt gratitude and respect to my advisor Professor Suhas Diggavi. For a period spanning almost six years, I've had the fortune of working closely with him. These years have not only sharpened my technical abilities but have also influenced my views towards life in general; I sincerely thank my advisor for his care and guidance. My habit of coming up with random ideas, and then rushing to him for comments is something which continues till date, and I cannot thank him enough for his insightful evaluations stemming from years of experience. I would also like to thank my PhD thesis committee members: Professor Paulo Tabuada, Professor Christina Fragouli and Professor Rafail Ostrovsky. At multiple points during my PhD, I received crucial comments from them which directly shaped my PhD projects. In particular, I had the fortune of collaborating with Professor Tabuada and members of his group on multiple projects. I sincerely thank Professor Tabuada for this fruitful collaboration. In addition, I would like to thank Professor Fragouli for her collaboration during my initial PhD years, and for hosting me as an intern at EPFL during the summers of 2011 and 2012.

During my PhD years, I had the incredible fortune of working with amazingly talented postdoctoral scholars: Vinod Prabhakaran, I-Hsiang Wang, Nikhil Karamchandani and Yasser Shoukry. I would like to thank them for their guidance and support. In particular, I deeply cherish the bonding

I have with Nikhil and Yasser; I cannot imagine surviving tough times during my PhD years without their help. I had the privilege of doing summer internships at industrial research labs (summer of 2013 and 2015), and I am grateful to my internship mentors Raja Bachu and Niranjan Ratnakar (at Qualcomm), and Amir Ingber (at Yahoo! labs) for their guidance. I also take this opportunity to thank members of our lab at UCLA: Can, Jad, Mehrdad, Joyson, Wei and Yair. In particular, my friendship with Can and Jad, and the good times we have had together in our lab is something I will cherish for the rest of my life. The same goes for my friends who made my stay at UCLA memorable: Mihir, Riccha, Neha, Yifan, Saakshita, Shailesh, Pratik, Joseph, Ayan, Sina, Janaki, Iris, Martina, Yahya, Karmoose, Ayca, Kasra, Nikhil, Sumeet, Ashish, Sunav, Nachiket, Sushant, Aayush, Debanjan, Ilya and Vignesh. A special thank you to my friend Gaelen, who has also been my apartment mate for the past four years; however sad the situation may be, his (poor) jokes never fail to cheer me up!

Moving to the United States from India for graduate studies had a sad but inevitable consequence; distance from my family and friends in India. I would like to thank my parents and brother for their unconditional love and support throughout my life. I owe the accomplishments in my life, both professional and personal, to them. I would also like to thank my close friends in India, Anubhav and Srushti, for their love and encouragement during tough times.

Finally, I am thankful to the National Science Foundation[1] and the University of California, Los Angeles for their financial support while working on this dissertation.

[1]This dissertation was supported in part through NSF grants 1136174, 1321120, 1314937 and 1514531.

VITA

2010	B.Tech in Electronics and Electrical Communication Engineering IIT Kharagpur, India.
2011	M.S. in Electrical Engineering UCLA, Los Angeles, California.
Summer 2011, Summer 2012	Summer intern at EPFL Lausanne, Switzerland.
Summer 2013	Summer intern at Qualcomm Bridgewater, New Jersey.
Summer 2015	Summer intern at Yahoo! labs Sunnyvale, California.
2016	Ph.D. candidate in Electrical Engineering UCLA, Los Angeles, California.

PUBLICATIONS

Shaunak Mishra, Yasser Shoukry, Nikhil Karamchandani, Suhas Diggavi and Paulo Tabuada, "Secure state estimation against sensor attacks in the presence of noise," under review for IEEE TCNS Special Issue on Secure Control of Cyber Physical Systems.

Shaunak Mishra, I-Hsiang Wang and Suhas Diggavi, "Harnessing bursty interference in multicarrier systems with feedback," under review for IEEE Transactions on Information Theory.

Mahdi Jafari Siavoshani, Shaunak Mishra, Christina Fragouli and Suhas Diggavi, "Multi-party secret key agreement over state-dependent wireless broadcast channels," under review for IEEE Transactions on Information Forensics & Security.

Shaunak Mishra, Yasser Shoukry, Nikhil Karamchandani, Suhas Diggavi and Paulo Tabuada, "Secure state estimation: optimal guarantees against sensor attacks in the presence of noise," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2015).

Shaunak Mishra, Nikhil Karamchandani, Paulo Tabuada and Suhas Diggavi, "Secure state estimation and control using multiple (insecure) observers," in Proceedings of the IEEE Conference on Decision and Control (CDC 2014).

Shaunak Mishra, I-Hsiang Wang and Suhas Diggavi, "Harnessing bursty interference in multicarrier systems with feedback," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2014).

Shaunak Mishra, I-Hsiang Wang and Suhas Diggavi, "Opportunistic interference management for multicarrier systems," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2013).

Shaunak Mishra, Christina Fragouli, Vinod Prabhakaran and Suhas Diggavi "Using feedback for secrecy over graphs," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2013).

Mahdi Jafari Siavoshani, Shaunak Mishra, Christina Fragouli and Suhas Diggavi, "Group secret key agreement over state-dependent wireless broadcast channels," in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011).

CHAPTER 1

Network reliability challenges: wireless interference management and CPS security

In an effort to address fundamental reliability challenges in wireless communication networks and cyber-physical systems (CPS), this dissertation focuses on two broad topics: (i) interference management in wireless networks, and (ii) CPS security. In this introductory chapter, we give an overview of contemporary challenges associated with wireless interference management, and the contributions of this dissertation towards addressing the same. We then give an overview of the challenges and contributions of this dissertation with respect to CPS security. We end this introductory chapter with a description of the organization of this dissertation.

1.1 Wireless interference management

1.1.1 Challenges

Mobile Internet traffic has witnessed an unprecedented growth in the last few years. Wireless networks are becoming increasingly dense to support this growth, and in such dense wireless networks, interference management continues to be a major challenge in achieving higher data rates. Thus, understanding the capacity of interference limited wireless networks is a topic of great practical interest. In this context, though the capacity characterization of the two user interference channel is still an open question, that of the two user Gaussian interference channel is characterized to within 1 bits/s/Hz [2]. In addition, when there is channel output feedback from the receivers, the capacity region is characterized to within 2 bits/s/Hz [3]. These approximate capacity results

are important milestones in understanding interference networks, and provide valuable insight into the design of optimal interference management schemes in the high SNR regime.

However, in the system model for most works studying the capacity of interference channels (including [2] and [3]), the interfering link is assumed to be always present. In practice, the temporal nature of interference in wireless networks tends to be *bursty*. Such burstiness can be attributed to multiple factors: the bursty nature of data traffic, distributed medium access control mechanisms, and decentralized networking protocols. As an example, consider an OFDM based cellular setup (with full frequency reuse), and two users in neighboring cells which are close to the cell boundary. If the subcarrier allocation decisions of neighbouring base stations are uncoordinated, the two users at the cell boundary may be assigned an overlapping set of subcarriers, leading to interference. However, as the allocation decisions for an user change over time, the resulting interference tends to be bursty. In view of the above observations, considering a system model where interference is always present may be pessimistic; it fails to address the possibility of exploiting bursty interference for improvements in the achievable rate.

1.1.2 Contributions

In this dissertation, we focus on studying the problem of harnessing bursty interference in multicarrier systems. More specifically, we focus on two setups: (i) when output feedback is available from the receivers, and (ii) when there is no feedback from the receivers. Our contributions[1] with respect to the first setup (*i.e.*, with output feedback) are as follows:

- For the multicarrier setup, we develop inner and outer bounds which are non-trivial extensions of single carrier results [1]. We identify regimes where treating subcarriers separately is optimal. For the remaining regimes, we employ coding across subcarriers (helping mechanism) to achieve tight results. Our outer bounds involve a subset entropy inequality by Madiman and Tetali [4].

- Based on our inner and outer bounds, we have a complete capacity region characterization

[1] Joint work with I-Hsiang Wang and Suhas Diggavi.

for the linear deterministic interference channel setup [5]. In the setup with Gaussian noise, we have a tight generalized degrees of freedom (GDoF) characterization and provide outer bounds on the capacity region.

Our contributions[2] with respect to the second setup (*i.e.*, without feedback) are as follows:

- We develop opportunistic achievability schemes (inner bounds) using erasure coding across subcarriers, signal-scale alignment [6, 7] and Han-Kobayashi scheme. These schemes are based on a degraded message set approach which, despite the absence of feedback, provide opportunistic rate increments whenever interference is absent. Furthermore, they guarantee a base rate whenever interference is present.
- We develop outer bounds using techniques inspired by multilevel diversity codes.
- Our inner and outer bounds coincide for several regimes.

Hence, our contributions demonstrate that burstiness of interference can be harnessed for an increase in the system capacity. Furthermore, our multicarrier results are of direct relevance to currently deployed systems like OFDM.

1.2 CPS security

1.2.1 Challenges

CPS essentially consist of a physical process (dynamical system) and a network of sensors, controllers and actuators which realize a feedback loop for managing the underlying dynamical system. The vast majority of critical infrastructure (*e.g.*, smart grid, water supply and transportation systems) is currently managed by such CPS. CPS security is a problem of increasing importance as we discover that such critical infrastructure is quite vulnerable to malicious attacks [8, 9]. For instance, an attacker can physically corrupt sensors of such CPS, and affect the state estimation

[2] Joint work with I-Hsiang Wang and Suhas Diggavi.

quality. Erroneous state estimates could then result in erroneous (and potentially destabilizing) control inputs into the dynamical system which in turn could result in serious physical damage. In addition to such physical attacks which require physical proximity to the sensors, CPS can also be remotely attacked [10]. CPS that are remotely operated, such as Unmanned Aerial Vehicles (UAVs) and parts of the power grid, can be vulnerable to several attack methodologies, since most of these systems rely on complex control algorithms running on networked digital platforms. For example, this could be enabled by hardware malware in the chips used in these platforms that becomes active at the discretion of an attacker [11].

1.2.2 Contributions

In the context of the challenges mentioned above, conventional cyber security methods which are oblivious to the underlying physical dynamics, leave open the possibility of leveraging such dynamics for securing the operation of CPS. With this motivation, in this dissertation we focus on the problem of securing state estimation in CPS against malicious attacks. In particular, we consider two attack scenarios for state estimation in linear dynamical systems: (i) observer attacks, and (ii) sensor attacks.

For the setup with observer attacks, we[3] consider the problem of estimating the state in a private and secure manner despite active adversaries that can attack the software/hardware where state estimation is performed. To combat such threats, we propose an architecture where state estimation is performed across multiple computing nodes (observers). We then show that even when ρ out of a total $3\rho+1$ observers are actively attacked:

- Using a combination of outputs from the observers, the state is still correctly estimated.
- The physical plant is still correctly controlled.
- The adversary can only obtain limited knowledge about the state.

Our approach leverages the underlying linear dynamics, and is inspired by techniques in cryptogra-

[3] Joint work with Nikhil Karamchandani, Paulo Tabuada and Suhas Diggavi.

phy for secure message transmission and information-theoretic secrecy. In addition, our guarantees on the secrecy of the plant's state against corrupting observers are based on the Cramer-Rao lower bound from estimation theory.

For the setup with sensor attacks, we[4] consider the problem of estimating the state of a noisy linear dynamical system when an unknown subset of sensors is arbitrarily corrupted by an adversary. Our contributions in this context are as follows:

- We propose a secure state estimation algorithm, and derive (optimal) bounds on the achievable state estimation error given an upper bound on the number of attacked sensors. The proposed state estimator involves Kalman filters operating over subsets of sensors to search for a sensor subset which is reliable for state estimation.
- We give a coding theoretic view of attack detection and state estimation against sensor attacks in a noiseless dynamical system.

1.3 Organization of this dissertation

This dissertation has three parts. The first part covers the problem of harnessing bursty interference; Chapter 2 deals with the problem of harnessing bursty interference with feedback, and Chapter 3 deals with the problem of opportunistic (bursty) interference management in the absence of feedback from the receivers. The second part deals with secure state estimation, with Chapter 4 dealing with observer attacks and Chapter 5 dealing with sensor attacks in the presence of noise. Chapter 6 deals with the conclusions of this dissertation and directions for future research. Finally, the third part covers the Appendix with detailed proofs associated with results in the preceding chapters.

[4] Joint work with Yasser Shoukry, Nikhil Karamchandani, Paulo Tabuada and Suhas Diggavi.

Part I

Harnessing Bursty Interference

CHAPTER 2

Harnessing Bursty Interference in Multicarrier Systems with Feedback

2.1 Introduction

In this chapter[1], we consider the problem of harnessing bursty interference in multicarrier systems when output feedback is available from the receivers. Feedback from receivers is not only a resource which is available in practice, but is also known to provide an unbounded gain in capacity for the (non-bursty) Gaussian interference channel [3]. To study benefits of feedback, [1] considered a single carrier setup with bursty interference and output feedback from the receivers. In [1], bursty interference is modeled using a Bernoulli random state (instantiated i.i.d. over time), and a complete capacity characterization is given for the linear deterministic setup [5]. In addition, schemes in [1] also employ coding across several instantiations of bursty interference. Results in [1] show that, depending on the regime of interference and the level of burstiness, there are significant gains in the capacity compared to the non-bursty setup. In this chapter, we study the multicarrier version of [1], *i.e.*, a setup with output feedback in multicarrier systems with bursty interference. Our motivation stems from the positive results in [1], and the widespread usage of multicarrier systems like OFDM. Since [1] developed optimal single carrier schemes, a natural question arises in the multicarrier version: is it always optimal to treat each subcarrier *separately* and just copy the optimal scheme in [1] on each subcarrier? As the following example illustrates, such a separation may not be always optimal.

[1] Joint work with I-Hsiang Wang and Suhas Diggavi.

Toy example Consider two parallel symmetric 2-user linear deterministic interference channels (LDICs) [5] as shown in Figure 2.1(a) (details of the linear deterministic channel model are described in Section 2.2). The first subcarrier has one direct link ($n_1 = 1$) and one interfering link ($k_1 = 1$, hence $\alpha_1 = \frac{k_1}{n_1} = 1$) and the second subcarrier has one direct link and three interfering links ($\alpha_2 = \frac{k_2}{n_2} = 3$). Causal output feedback is available from the receivers to the respective transmitters. Bernoulli random states $S_1[t]$ and $S_2[t]$ indicate the presence of interference in the first and second subcarrier respectively and are instantiated i.i.d. (over time) from an arbitrary joint distribution $\mathbb{P}_{S_1S_2}$. For this example, we assume the expectation of both the states to be $p = \frac{1}{2}$. Our goal here

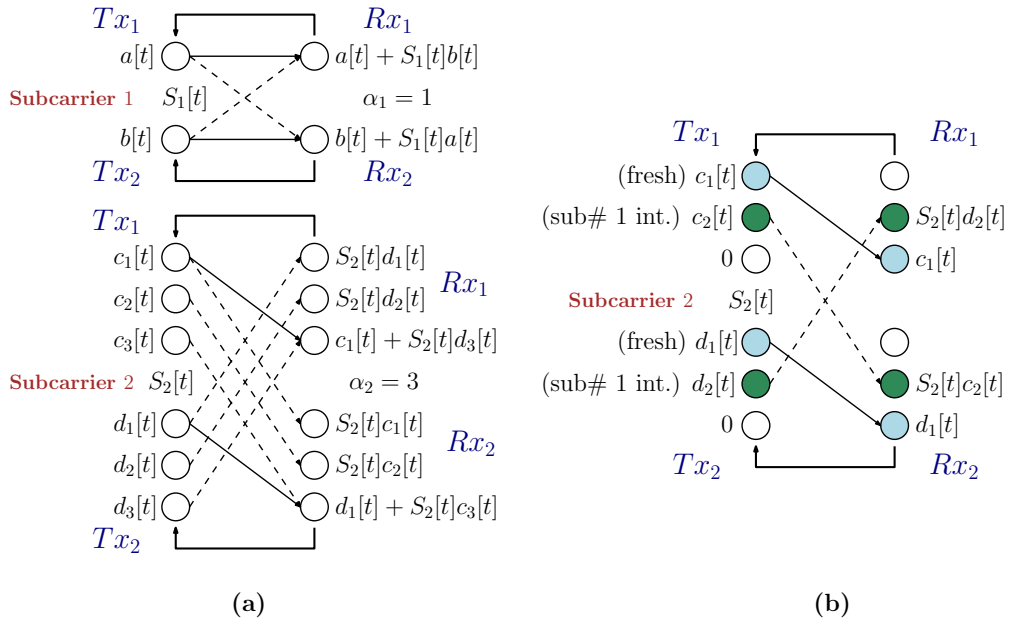


Figure 2.1: Toy example with bursty interference in 2 subcarriers: (a) setup with subcarrier 1 ($n_1 = 1$, $k_1 = 1$) and subcarrier 2 ($n_2 = 1$, $k_2 = 3$) and marginal probability of interference in each subcarrier is $p = \frac{1}{2}$, (b) for achieving symmetric capacity, fresh symbols ($c_1[t]$ and $d_1[t]$) are sent in the top most level of both subcarriers, and the middle level is used to recover interfered symbols (in the previous block) of subcarrier 1.

is to find the maximum achievable symmetric rate. Using the optimal single carrier schemes in [1], we can achieve symmetric rate 0.667 from the first subcarrier and symmetric rate 1.25 from the second subcarrier. Summing these rates, we can achieve a total symmetric rate 1.917. Now,

we will show that rate 2.0 is achievable by coding *across* the subcarriers rather than treating the subcarriers separately. We use a block based pipelined scheme (block length N_B) as follows. The transmitters always send fresh symbols in the first subcarrier (a-symbols for Rx_1 and b-symbols for Rx_2 as shown in Figure 2.1). In the first subcarrier, for sufficiently large N_B, with high probability (w.h.p.) only pN_B a-symbols in a block get interfered at Rx_1 (and pN_B b-symbols at Rx_2). At the end of a block, due to feedback from Rx_1, Tx_1 knows exactly which of its transmitted a-symbols caused interference at Rx_2 (since the same state variable $S_1[t]$ holds for both the receivers). For the next block, Tx_1 creates N_B linear combinations of these pN_B a-symbols (which caused interference at Rx_2 in the previous block) and sends these N_B linear combinations as $c_2[t]$ (in the second subcarrier as shown in Figure 2.1 (b)) over the next N_B time slots. Due to bursty interference, w.h.p. only pN_B of these linear combinations appear at Rx_2; but this is sufficient to decode pN_B a-symbols constituting the linear combinations. Using these a-symbols Rx_2 can now recover all the interfered b-symbols in the previous block and hence achieve rate 1 from the first subcarrier (same for Rx_1 due to symmetry). For the remaining levels in the second subcarrier, the following is done: lowest levels are not used ($c_3[t] = d_3[t] = 0$), and the transmitters send fresh symbols in the highest level (as shown in Figure 2.1 (b)) which appear interference free at the receivers (as the lowest levels are not used). This leads to an additional rate 1 from the second subcarrier. Adding rates from the two subcarriers, we achieve symmetric rate 2. This is in fact the symmetric capacity; an easy consequence of the outer bounds developed in this chapter.

The above example demonstrates a *helping* mechanism; the second subcarrier *helped* the first subcarrier in recovering interfered symbols in a pipelined fashion. In this chapter, we generalize this idea for an arbitrary collection of subcarriers with the following constraint: interference states across subcarriers are drawn from an arbitrary joint distribution (instantiated i.i.d. over time) and the marginal probability of interference is the same for each subcarrier. The main idea behind the generalization is to use specific *levels* in very strongly interfered subcarriers to recover interfered signals for strongly and weakly interfered subcarriers in a pipelined fashion as shown in the toy example. Another aspect captured by the toy example is the importance of burstiness; subcarriers in the above example are separable (due to our results and [1]) when interference is always present.

Hence, the proposed helping mechanism owes its relevance to bursty interference.

The remainder of this chapter is organized as follows. Section 2.2 deals with the notation and setup. Section 2.3 summarizes the main results of this chapter. This is followed by Section 2.4 on inner bounds for the linear deterministic setup, and Section 2.5 on the GDoF characterization for the setup with Gaussian noise. Section 2.6 deals with the outer bounds.

2.2 Notation and Setup

We consider a system with two base stations (transmitters) Tx_1 and Tx_2, and two users (receivers) Rx_1 and Rx_2. For $i \in \{1,2\}$, Tx_i has message $W^{(i)}$ for Rx_i. There are M parallel channels from Tx_i to Rx_i (subcarriers indexed by $j \in \{1,2,\dots M\}$). In this chapter, we consider two setups for the subcarrier channel: the first one is based on the linear deterministic model [5, 7] (LD setup[2]), and the second one is based on the Gaussian interference channel (GN setup). The subcarrier channel model for both the setups, followed by the bursty interference model and rate requirements are described below.

Subcarrier channel model In the LD setup, each subcarrier is modeled by a 2-user (symmetric) LDIC [5, 7] with a bursty interfering link (explained below) and feedback from respective receivers. At discrete time index $t \in \{1,2,\dots N\}$, the transmitted signal in subcarrier j of Tx_i is $\mathbf{x}_j^{(i)}[t] \in \mathbb{F}^{q_j}$ where $\mathbb{F}$ is a finite field. The received signal in subcarrier j at Rx_i is given by:

$$\mathbf{y}_j^{(i)}[t] = \mathbf{G}_j^{q_j-n_j}\mathbf{x}_j^{(i)}[t] + S_j[t]\mathbf{G}_j^{q_j-k_j}\mathbf{x}_j^{(i')}[t], \tag{2.1}$$

where $\mathbf{G}_j$ is a $q_j \times q_j$ shift matrix in the terminology of deterministic channel models [5], $S_j[t]$ is a Bernoulli random variable (details in bursty interference model below) determining the presence of interference in subcarrier j at time index t, $\mathbf{x}_j^{(i')}[t]$ denotes the transmitted signal on subcarrier j of user $i' \neq i$, and parameters n_j and k_j represent the direct and interfering link strengths [5] in

[2]Unlike its name suggests, the LD setup in our chapter is not purely deterministic; there is a stochastic aspect related to the burstiness of interference. However, since the term deterministic in the interference channel literature is usually ascribed to the absence of receiver noise, we stick to this convention.

subcarrier j. Figure 2.2 shows the channel model for subcarrier j in the LD setup. Without loss of generality, we assume $q_j = \max(n_j, k_j)$ and let $\alpha_j = \frac{k_j}{n_j}$ denote the normalized strength of the interfering signal in subcarrier j. For every time instant, it is convenient to consider a subcarrier as indexed levels of bit pipes [5]; each bit pipe carries a symbol from $\mathbb{F}$. Note that, if we assume $S_j[t] = 1$ in (2.1), then the subcarrier channel model is precisely the usual 2-user LDIC [5, 7] where interference is assumed to be always present.

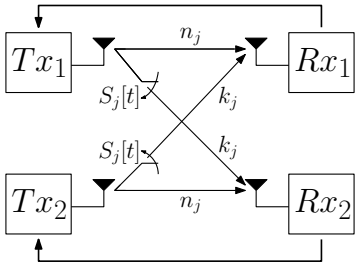


Figure 2.2: Bursty interference channel (with feedback) for subcarrier j in LD setup: n_j and k_j represent direct and interfering link strengths. Presence of interference at time index t is determined by Bernoulli random variable $S_j[t]$.

In the GN setup, at discrete time index $t \in \{1, 2, \ldots N\}$, the transmitted signal in subcarrier j of Tx_i is $x_j^{(i)}[t] \in \mathbb{C}$, such that $\frac{1}{N}\sum_{t=1}^{N} |x_j^{(i)}[t]|^2 \leq 1$. The received signal in subcarrier j at Rx_i is given by:

$$y_j^{(i)}[t] = g_{D,j} x_j^{(i)}[t] + S_j[t] g_{I,j} x_j^{(i')}[t] + z_j^{(i)}[t], \tag{2.2}$$

where $g_{D,j},\ g_{I,j} \in \mathbb{C}$ denote the direct and interfering channel gains, and $z_j^{(i)}[t] \sim \mathscr{CN}(0,1)$ is Gaussian noise. As in the LD setup, $S_j[t]$ is the interference state. In both LD and GN setups, Tx_i receives causal feedback from Rx_i (feedback consists of the received signal and the interference state).

Bursty interference model We consider the same interference statistics for both LD and GN setups. As described above, the presence of interference in subcarrier j at time index t is given by a Bernoulli random variable $S_j[t]$ (takes values in $\{0,1\}$). The M Bernoulli random variables

$\{S_1[t], S_2[t], \ldots S_M[t]\}$ have a joint probability distribution $\mathbb{P}(S_1[t] = s_1, S_2[t] = s_2, \ldots S_M[t] = s_M) = \mathbb{P}(S_1 = s_1, S_2 = s_2, \ldots S_M = s_M)$ instantiated i.i.d. over time. In this chapter, we restrict the analysis to joint distributions with the same marginal probabilities for every $S_j[t]$, *i.e.*, $\forall j$, $\mathbb{E}(S_j[t]) = p$. The transmitters are assumed to know the above statistics, but are limited to causal information on the interference realizations in the subcarriers (through feedback).

Achievable rates We consider the same rate requirements for both LD and GN setups. Base station Tx_i intends to send message $W^{(i)}$ to Rx_i over N time slots (time index $t \in \{1, 2, \ldots N\}$). Rate $R^{(i)}$ (corresponding to $W^{(i)}$) is considered achievable if the probability of decoding error is vanishingly small as $N \to \infty$.

2.3 Main Results

Theorem 1 (LD setup capacity) *The capacity region for* $\left(R^{(1)}, R^{(2)}\right)$ *in the LD setup is given by the following rate inequalities:*

$$R^{(i)} \leq p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p, \tag{2.3}$$

$$R^{(i)} + pR^{(i')} \leq p\Delta + \sum_{j=1}^{M} n_j(1+p), \tag{2.4}$$

$$R^{(i)} + R^{(i')} \leq p\Delta + 2\sum_{j=1}^{M} n_j, \tag{2.5}$$

where $i, i' \in \{1,2\}$ *and* $i \neq i'$*, and*

$$\begin{aligned}\Delta &= \sum_{j=1}^{M} \max(n_j, k_j) + (n_j - k_j)^+ - 2n_j \\ &= \sum_{j:\alpha_j > 2} (k_j - 2n_j) - \sum_{j:\alpha_j \leq 1} k_j - \sum_{j:1<\alpha_j \leq 2} (2n_j - k_j).\end{aligned} \tag{2.6}$$

Remark 1 *The value of* Δ*, as defined above, plays an important role in our achievability schemes. As we describe later in Section 2.4 on inner bounds,* $\Delta > 0$ *implies that there are enough levels in subcarriers with* $\alpha_j > 2$ *(very strong interference) to recover the interfered signals for subcarriers with* $\alpha_j \leq 1$ *(weak interference) and* $1 < \alpha_j \leq 2$ *(strong interference). Also, as shown in*

Figure 2.3, the shape of the capacity region depends on the value of Δ. The details of the rate

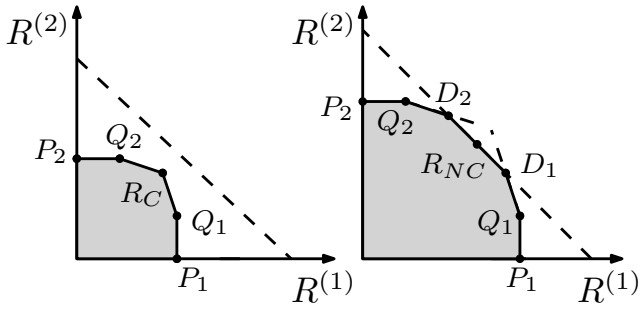


Figure 2.3: Capacity region (LD setup) when $\Delta < 0$ and $\Delta > 0$. The dashed line representing inequality (2.5) is active only when $\Delta > 0$. Symmetric capacity (C_{sym}) for $\Delta < 0$ and $\Delta \geq 0$ is given by $R_C = \frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j$ and $R_{NC} = \frac{p}{2}\Delta + \sum_{j=1}^{M} n_j$ respectively.

tuples $\left(R^{(1)}, R^{(2)}\right)$ marked in Figure 2.3 are listed below:

- P_1 : $\left(p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p, 0\right)$
- Q_1 : $\left(p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p, \sum_{j=1}^{M} (n_j - k_j)^+\right)$
- D_1 : $\left(p\Delta + \sum_{j=1}^{M} n_j, \sum_{j=1}^{M} n_j\right)$
- $R_C \equiv (R_C, R_C)$: $\left(\frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j, \frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j\right)$
- $R_{NC} \equiv (R_{NC}, R_{NC})$: $\left(\frac{p}{2}\Delta + \sum_{j=1}^{M} n_j, \frac{p}{2}\Delta + \sum_{j=1}^{M} n_j\right)$
- D_2 : $\left(\sum_{j=1}^{M} n_j, p\Delta + \sum_{j=1}^{M} n_j\right)$
- Q_2 : $\left(\sum_{j=1}^{M} (n_j - k_j)^+, p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p\right)$
- P_2 : $\left(0, p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p\right)$.

Remark 2 *For the case when $\Delta < 0$, the symmetric capacity is given by R_C as shown in Figure 2.3 (a). The subscript C in R_C stands for causal, and is related to the fact that R_C is derived from outer bounds which consider causal knowledge of bursty interference realizations. In a similar spirit, for symmetric capacity R_{NC} when $\Delta > 0$, the subscript NC stands for non-causal. We describe the proofs for these outer bounds in Section 2.6.*

Corollary 1 (separability in LD setup) *In the LD setup, for achieving symmetric capacity, treating subcarriers separately is optimal when:*

1. $0 < p < 1$ *and all* $\alpha_j \leq 2$,
2. $0 < p < 1$ *and all* $\alpha_j \geq 2$,
3. $p \in \{1, 0\}$ *(degenerate non-bursty case).*

For the remaining cases, coding across subcarriers achieves symmetric capacity.

Theorem 2 (GN setup outer bounds) *The following rate inequalities are outer bounds on achievable* $(R^{(1)}, R^{(2)})$ *in the GN setup:*

$$R^{(i)} \leq \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + |g_{D,j}|^2 + |g_{I,j}|^2\right), \tag{2.7}$$

$$R^{(i)} + pR^{(i')} \leq p\Delta_G + (1+p) \sum_{j=1}^{M} \log\left(1 + |g_{D,j}|^2\right), \tag{2.8}$$

$$R^{(i)} + R^{(i')} \leq p\Delta_G + 2 \sum_{j=1}^{M} \log\left(1 + |g_{D,j}|^2\right), \tag{2.9}$$

where $i,\ i' \in \{1,2\}$ *and* $i \neq i'$, *and* $\Delta_G = \sum_{j=1}^{M} \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) + \log\left(1 + \frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right) - 2\log\left(1 + |g_{D,j}|^2\right)$.

Theorem 3 (GN setup GDoF) *In the GN setup, assuming* $g_{D,j} = \sqrt{SNR}$, $g_{I,j} = \sqrt{INR_j}$ *and* $INR_j = SNR^{\beta_j}$ *(rational* β_j*),*

$$\begin{aligned} GDoF\left(\beta_1, \ldots \beta_M\right) &= \limsup_{SNR \to \infty} \frac{C_{sym}\left(SNR, \beta_1, \ldots \beta_M\right)}{M \log(SNR)} \\ &= 1 + \min\left(\frac{\frac{p}{2}\Delta_{GDoF}}{M}, \frac{\frac{p}{1+p}\Delta_{GDoF}}{M}\right), \end{aligned} \tag{2.10}$$

where C_{sym} *denotes the symmetric capacity and* $\Delta_{GDoF} = \sum_{j=1}^{M} (\max\left(1, \beta_j\right) + (1-\beta_j)^+ - 2)$.

Corollary 2 (separability in GN setup) *Similar to the separability in LD setup (Corollary 1), in the GN setup, treating subcarriers separately is GDoF optimal when all* $\beta_j \leq 2$ *or all* $\beta_j \geq 2$.

2.4 Inner bounds: LD setup

In this section, we focus on schemes for achieving the symmetric capacity in the LD setup (see Appendix A.0.6 and A.0.7 for achievability of remaining corner points in Figure 2.3). In Section 2.4.1, we briefly review the single carrier schemes in [1] and describe a *bursty relaying* technique (used in our multicarrier schemes). In Section 2.4.2, we mention the cases where treating subcarriers separately is optimal (*i.e.,* simply copying the optimal single carrier scheme [1] on each subcarrier leads to the symmetric capacity). For the remaining cases, we propose multicarrier schemes (covered in Sections 2.4.3 and 2.4.4), which employ a helping mechanism where some *helper* levels in subcarriers with $\alpha_j > 2$ are used to recover interfered signals in subcarriers with $\alpha_j < 2$. For $\Delta \geq 0$ (Section 2.4.3), the helping mechanism is optimal; whereas for $\Delta < 0$ (Section 2.4.4) the helping mechanism is run in parallel with the single carrier schemes [1] to achieve symmetric capacity.

2.4.1 Single carrier symmetric capacity and bursty relaying

The single carrier version of our setup (*i.e.,* $M = 1$) was studied in [1]. For notational consistency, we use $j = 1$ (subcarrier index) in stating the results from [1]. We simply restate below the schemes in [1] for the regimes $\alpha_1 \leq 1$ and $1 < \alpha_1 \leq 2$; but for the regime $\alpha_1 > 2$ we mention a slightly different scheme that makes describing our multicarrier schemes in Sections 2.4.3 and 2.4.4 more convenient.

Regime $\alpha_1 \leq 1$ For this regime, the symmetric capacity is $n_1 - \frac{p}{1+p}k_1$. To achieve this, a two phase scheme (same for Tx_1 and Tx_2) is used as briefly described below[3] (see [1] for details):

- Phase F: Transmitters in phase F at time index t send fresh symbols on all n_1 levels. If there is no interference at time index t (occurs w.p. $1-p$), all n_1 symbols can be decoded at the intended receiver and both transmitters stay in phase F for time index $t+1$. If there is interference (occurs w.p. p), only the bottom k_1 symbols get interfered at a receiver and the transmitters transition to phase R for time index $t+1$.

[3]The scheme for $\alpha_1 = 1$ has slight variation from this scheme. For details, see [1].

- Phase R: Transmitters send the past interference (obtained from receiver feedback) on the top k_1 levels and fresh symbols on the remaining $(n_1 - k_1)$ levels. Both transmitters transition to phase F for the next time index after phase R.

Figure 2.4 shows the underlying Markov chain for this scheme. Figure 2.5 shows the scheme for

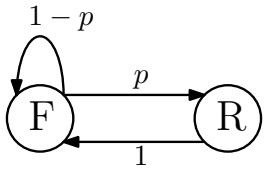


Figure 2.4: Underlying Markov chain for the single carrier schemes in [1] for $\alpha_1 \leq 2$.

the setup where $n_1 = 3$ and $k_1 = 2$.

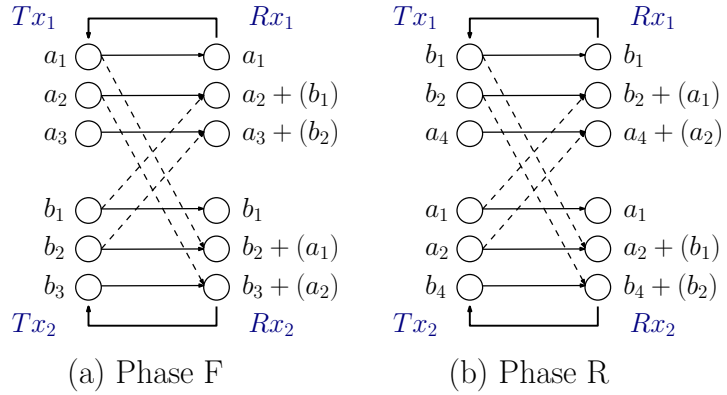


Figure 2.5: Single carrier scheme [1] for $n_1 = 3$ and $k_1 = 2$. Symbols in $(\cdot)$ appear only when interference is present.

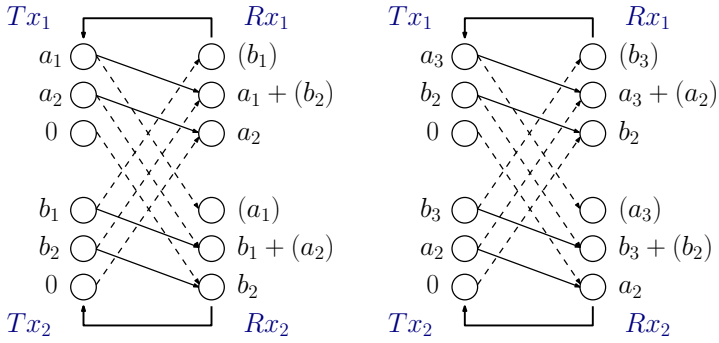
Regime $1 < \alpha_1 \leq 2$ For this regime, the symmetric capacity is $\frac{1-p}{1+p} n_1 + \frac{p}{1+p} k_1 = n_1 - \frac{p}{1+p}(2n_1 - k_1)$. To achieve this, a two phase scheme is used as briefly described below (see [1] for details):

- Phase F: Transmitters in phase F at time index t send fresh symbols on the top n_1 levels and the bottom $k_1 - n_1$ levels are not used. If there is no interference at time index t (occurs w.p. $1-p$), all n_1 symbols can be decoded at the intended receiver and both transmitters stay in

phase F. If there is interference at time index t (occurs w.p. p), only $2n_1 - k_1$ symbols get interfered at a receiver and the transmitters transition to phase R for time index $t+1$.

- Phase R: Transmitters send fresh symbols in the top $k_1 - n_1$ levels. In the next $2n_1 - k_1$ levels (below the top $k_1 - n_1$ levels), the $2n_1 - k_1$ interfering symbols (obtained through receiver feedback) from the previous time index are sent. The remaining $k_1 - n_1$ levels in the bottom are not used. Both transmitters transition to phase F for the next time index after phase R.

The underlying Markov chain in this scheme is same as the one in Figure 2.4. Figure 2.6 shows the scheme for the setup where $n_1 = 2$ and $k_1 = 3$.



(a) Phase F (b) Phase R

Figure 2.6: Single carrier scheme [1] for $n_1 = 2$ and $k_1 = 3$. Symbols in $(\cdot)$ appear only when interference is present.

Regime $\alpha_1 > 2$ (bursty relaying) For this regime, the symmetric capacity is $n_1 + \frac{p}{2}(k_1 - 2n_1)$. In [1], this is achieved using a Markov chain based scheme similar to the ones described above. To help describe the multicarrier schemes, we develop a block version of the scheme in [1] as follows. In each block of duration N_B, transmitters send fresh symbols on the top n_1 levels and never use the bottom n_1 levels. Since the bottom n_1 levels are never used, the fresh symbols from the top n_1 levels are always received interference free. This realizes rate n_1. From the $k_1 - 2n_1$ levels in the middle (below the top n_1 levels), we realize an additional rate $\frac{p}{2}(k_1 - 2n_1)$ over two blocks as follows. For the first block, Tx_i creates $N_B(k_1 - 2n_1)$ linear combinations from $pN_B(k_1 - 2n_1)$

fresh symbols and sends these linear combinations in the middle $k_1 - 2n_1$ levels. For large enough N_B, w.h.p. $Rx_{i'}$ receives $pN_B(k_1 - 2n_1)$ such linear combinations. $Rx_{i'}$ decodes the constituent fresh symbols from these linear combinations and sends them to $Tx_{i'}$ (through feedback). $Tx_{i'}$ now creates $N_B(k_1 - 2n_1)$ new linear combinations from these symbols and sends them in the $(k_1 - 2n_1)$ middle levels during the next block. W.h.p. Rx_i receives $pN_B(k_1 - 2n_1)$ such linear combinations and decodes all the constituent symbols. This leads to an additive rate of $\frac{pN_B(k_1-2n_1)}{2N_B} = \frac{p}{2}(k_1 - 2n_1)$ at Rx_i (and similarly at $Rx_{i'}$). In the remainder of this chapter, we refer to this technique (for middle levels in subcarriers with $\alpha_j > 2$) as *bursty relaying* since Tx_i-Rx_i pair effectively acts as a relay for $Tx_{i'}$-$Rx_{i'}$ and vice versa. Figure 2.7 illustrates this technique of bursty relaying. Adding the rate from bursty relaying in $(k_1 - 2n_1)$ middle levels and rate n_1 from the top n_1 levels, we achieve rate $n_1 + \frac{p}{2}(k_1 - 2n_1)$.

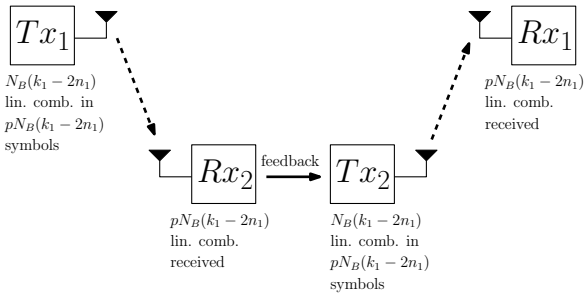


Figure 2.7: Bursty relaying using $k_1 - 2n_1$ middle levels (below the top n_1 levels) when $\alpha_1 > 2$. As shown, Rx_2 receives $pN_B(k_1 - 2n_1)$ linear combinations in $pN_B(k_1 - 2n_1)$ symbols during a block of duration N_B. It decodes and sends the constituent symbols to Tx_2 which again creates $N_B(k_1 - 2n_1)$ linear combinations from these symbols. In the next block, Rx_1 receives $pN_B(k_1 - 2n_1)$ linear combinations from Tx_2 and decodes the constituent symbols.

2.4.2 Multicarrier separability

Using outer bounds (2.4) and (2.5) for LD setup and achievability rates for the single carrier schemes in [1], the following can be easily verified (see Appendix A.0.5 for verification details):

- For $p \in \{0,1\}$, *i.e.,* when interference is either never present or always present, the symmetric capacity can be achieved by treating the subcarriers separately.

- For $0 < p < 1$, when all subcarriers have $\alpha_j \leq 2$, the symmetric capacity can be achieved by treating the subcarriers separately.

- For $0 < p < 1$, when all subcarriers have $\alpha_j \geq 2$, the symmetric capacity can be achieved by treating the subcarriers separately.

Hence, the subcarriers are *separable* in the above cases. When we have subcarriers with $\alpha_j \leq 2$ as well as subcarriers with $\alpha_j > 2$ (and $0 < p < 1$), we employ coding across subcarriers (through a helping mechanism described in the next subsection) to achieve symmetric capacity; we assume such a *mixed* collection of subcarriers in describing our multicarrier schemes in Sections 2.4.3 and 2.4.4.

2.4.3 Achieving symmetric capacity when $\Delta \geq 0$

As defined in (2.6), $\Delta = \sum_{j:\alpha_j>2} (k_j - 2n_j) - \sum_{j:\alpha_j\leq 1} k_j - \sum_{j:1<\alpha_j\leq 2} (2n_j - k_j)$, and when $\Delta \geq 0$, $C_{sym} = R_{NC} = \frac{p}{2}\Delta + \sum_{j=1}^{M} n_j$. We will now describe the achievability of R_{NC} using a block based scheme. In each block of duration N_B, fresh symbols are sent in the following levels (same for both transmitters by symmetry):

- All n_j levels for subcarriers with $\alpha_j \leq 1$.

- Top n_j levels for subcarriers with $\alpha_j > 1$.

In addition, the following levels are not used:

- Bottom $k_j - n_j$ levels of subcarriers with $1 < \alpha_j \leq 2$.

- Bottom n_j levels of subcarriers with $\alpha_j > 2$.

Because of the above choices, in every block (for large enough N_B):

- In subcarriers with $\alpha_j \leq 1$, w.h.p. pN_Bk_j fresh symbols get interfered.
- In subcarriers with $1 < \alpha_j \leq 2$, w.h.p. $pN_B(2n_j - k_j)$ fresh symbols get interfered.
- In subcarriers with $\alpha_j > 2$, the top n_j fresh symbols are always received interference free.

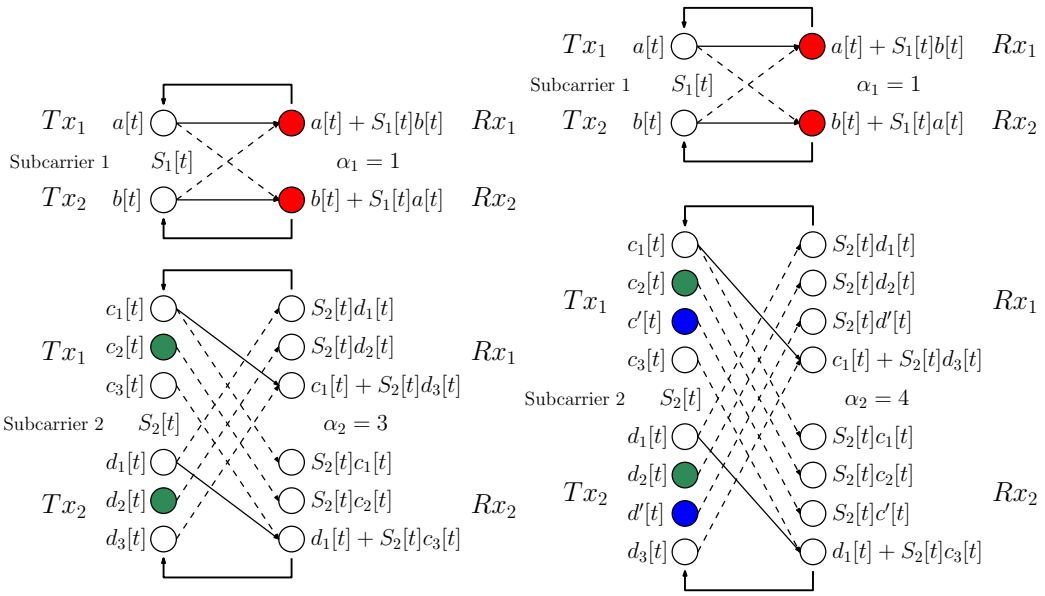
In total, each receiver needs to recover $pN_B(\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j)$ interfered symbols in each block. This recovery is done in a pipelined fashion in the next block using a *helping mechanism* described below.

Helping mechanism We will use the term *helper levels* for the middle $k_j - 2n_j$ levels (below the top n_j levels) in subcarriers with $\alpha_j > 2$; hence $\sum_{j:\alpha_j>2} k_j - 2n_j$ helper levels in total. After each block, due to feedback from Rx_i, Tx_i knows exactly which of its transmitted symbols caused interference at $Rx_{i'}$. The number of such symbols, as described above, is w.h.p. equal to $pN_B(\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j)$. Tx_i now creates
$N_B(\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j)$ linear combinations of these symbols and sends the linear combinations on any $(\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j)$ of the helper levels in the subsequent block. W.h.p. $pN_B(\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j)$ of such linear combinations are received at $Rx_{i'}$. This is sufficient to recover all the interfered symbols at $Rx_{i'}$ in the previous block.

As all the interfered symbols in a block are recovered using the above mechanism, we realize rate $\sum_{j=1}^{M} n_j$. If $\Delta > 0$, some of the helper levels are still available; precisely $(\sum_{j:\alpha_j>2} k_j - 2n_j) - (\sum_{j:\alpha_j \leq 1} k_j + \sum_{j:1<\alpha_j \leq 2} 2n_j - k_j) = \Delta$ of them. We realize an additional rate of $\frac{p}{2}\Delta$ from such leftover helper levels using the bursty relaying scheme described in Section 2.4.1. Adding the rate from the leftover helper levels to $\sum_{j=1}^{M} n_j$, we achieve the symmetric capacity $\frac{p}{2}\Delta + \sum_{j=1}^{M} n_j$.

Illustrative examples The toy example in Section 2.1 considered two subcarriers with $n_1 = 1$, $k_1 = 1$, $n_2 = 1$ and $k_2 = 3$ (and $p = \frac{1}{2}$). As illustrated in the toy example, the middle level in the second subcarrier helped in recovering interfered symbols in the first subcarrier. With reference to our achievability scheme for $\Delta \geq 0$, the middle level in the second subcarrier is a helper level (green level in Figure 2.8 (a)) whereas the (only) level in the first subcarrier is a helped level (red

level in Figure 2.8 (a)). Since there is only one helped level and one helper level, $\Delta = 1-1 = 0$ and $C_{sym} = 2$. To illustrate the idea behind our achievability scheme for $\Delta > 0$, we slightly modify



(a) $n_1 = 1$, $k_1 = 1$, $n_2 = 1$ and $k_2 = 3$. **(b)** $n_1 = 1$, $k_1 = 1$, $n_2 = 1$ and $k_2 = 4$.

Figure 2.8: Toy example and its modification: (a) original toy example, and (b) Example 1.

the toy example as described below.

Example 1 ($n_1 = 1, k_1 = 1, n_2 = 1$ **and** $k_2 = 4$) *Compared to the original toy example, we have modified only the second subcarrier such that it has one extra middle level (blue level in Figure 2.8 (b)). For this case,* $\Delta = 2-1 = 1$ *and* $C_{sym} = 2 + \frac{1}{4}$. *The helping mechanism is used as in the original toy example to achieve rate* 2. *Additional rate* $\frac{1}{4}$ *is achieved using the bursty relaying technique for the extra middle level in the second subcarrier (blue level in Figure 2.8 (b)).*

2.4.4 Achieving symmetric capacity when $\Delta < 0$

When $\Delta < 0$, $C_{sym} = R_C = \frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j$. Before we proceed to the details, we give a high level idea of the scheme as follows. Simply copying the scheme for $\Delta \geq 0$ in Section 2.4.3 does not work for this case since there are not enough helper levels ($\sum_{j:\alpha_j>2} k_j - 2n_j$) compared to the number of levels facing interference ($\sum_{j:\alpha_j\leq 1} k_j + \sum_{j:1<\alpha_j\leq 2} 2n_j - k_j$). The trick in this case is

to *help as much as possible*. For each subcarrier with $\alpha_j < 2$, we select h_j *helped* levels; these levels face interference and the interfered symbols are recovered using the helping mechanism described in Section 2.4.3. The total number of helped levels $\sum_{j:\alpha_j<2} h_j$ equals the number of helper levels ($\sum_{j:\alpha_j>2} k_j - 2n_j$). For the remaining interfered levels in subcarriers with $\alpha_j < 2$, we run the optimal single carrier scheme [1] (with slight modifications depending on the value of α_j as discussed later in Sections 2.4.4.1, 2.4.4.2 and 2.4.4.3) in parallel with the helping mechanism. Adding the rates from the single carrier schemes and the helping mechanism, we achieve the symmetric capacity. This high level idea can also be illustrated by rewriting $R_C = \frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j$ as shown below:

$$\begin{aligned}
&\frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j \\
&= \left(\sum_{j:\alpha_j\geq 2} n_j\right) + \left(\sum_{j:\alpha_j<2} h_j\right) + \left(\sum_{j:\alpha_j\leq 1} (n_j - h_j) - \frac{p}{1+p}(k_j - h_j)\right) \\
&\quad + \left(\sum_{j:1<\alpha_j<2} (n_j - h_j) - \frac{p}{1+p}(2(n_j - h_j) - (k_j - h_j))\right) \\
&= \left(\sum_{j:\alpha_j\geq 2} n_j\right) + \left(\sum_{j:\alpha_j<2} h_j\right) + \left(\sum_{j:\alpha_j\leq 1} \tilde{n}_j - \frac{p}{1+p}\tilde{k}_j\right) + \left(\sum_{j:1<\alpha_j<2} \tilde{n}_j - \frac{p}{1+p}(2\tilde{n}_j - \tilde{k}_j)\right),
\end{aligned} \tag{2.11}$$

where $\sum_{j:\alpha_j<2} h_j = \sum_{j:\alpha_j>2} k_j - 2n_j$ is the total number of helped levels, and for subcarriers (with $\alpha_j < 2$) being helped the effective direct and interfering link strengths are $\tilde{n}_j = n_j - h_j$ and $\tilde{k}_j = k_j - h_j$. The last two terms in (2.11) come from the optimal single carrier schemes for $\alpha_j < 2$ (that run in parallel with the helping mechanism).

We now describe the achievability of R_C in detail. In subcarriers with $\alpha_j \geq 2$, the transmitters always send fresh symbols in the top n_j levels and never use the bottom n_j levels. This realizes rate $\sum_{j:\alpha_j\geq 2} n_j$. For each subcarrier with $\alpha_j < 2$, we assign a non-negative integral value h_j with the following constraints: (a) $h_j \leq k_j$ for $\alpha_j \leq 1$ and $h_j \leq 2n_j - k_j$ for $1 < \alpha_j < 2$, (b) $\sum_{j:\alpha_j<2} h_j = \sum_{j:\alpha_j>2} k_j - 2n_j$. Simply put, h_j denotes the number of helped levels in a subcarrier and the total number of such levels equals the number of helper levels available in subcarriers with $\alpha_j > 2$. Having fixed h_j for each subcarrier with $\alpha_j < 2$, we now describe the modifications needed in the

optimal single carrier scheme [1] for parallel execution with the helping mechanism.

2.4.4.1 Modification for $\alpha_j < 1$

The bottom h_j levels (of the direct link) are selected as helped levels as shown in Figure 2.9 (a) and interfered symbols in these levels are recovered using the helping mechanism described in Section 2.4.3. For the modified single carrier scheme, phase F remains the same as in [1] and the modification is only in Phase R. For illustration purposes consider that in phase F for a subcarrier with $\alpha_j < 1$, Tx_1 sends fresh symbols $[a_1\ a_2\ \ldots a_{n_j}]$ (as shown in Figure 2.9 (a)) and Tx_2 sends fresh symbols $[b_1\ b_2\ \ldots b_{n_j}]$. If there is no interference, all the fresh symbols are received and the transmitters stay in phase F. If there is interference, the transmitters transition to phase R. In the scheme in [1], all k_j interfering symbols were sent on the top k_j levels in phase R; in the modified scheme the transmitters just send the top $\tilde{k}_j = k_j - h_j$ interfering symbols in the top $\tilde{k}_j$ levels as shown in Figure 2.9 (a). In the remaining levels, fresh symbols are sent (starred symbols in Figure 2.9 (a)). Ignoring the bottom h_j levels, the resulting system of linear equations at the receivers is exactly the same as in [1] with direct link strength $\tilde{n}_j$ and interfering link strength $\tilde{k}_j$. Thus at end of phase R, Rx_1 is able to decode $\{a_{\tilde{n}_j-\tilde{k}_j+1}, a_{\tilde{n}_j-\tilde{k}_j+2}, \ldots a_{\tilde{n}_j}\}$ (interfered symbols in phase F) and $\{a^*_{\tilde{n}_j+1}, a^*_{\tilde{n}_j+2}, \ldots a^*_{2\tilde{n}_j-\tilde{k}_j}\}$ (fresh symbols in phase R). To decode interfered symbols in the helped levels, the helping mechanism is used (which collects all interfered symbols in helped levels during a block of duration N_B and enables their recovery in the next block). So effectively, the rate obtained from a subcarrier with $\alpha_j \leq 1$ is $h_j + (\tilde{n}_j - \frac{p}{1+p}\tilde{k}_j)$.

2.4.4.2 Modification for $\alpha_j = 1$

The case $k_j = n_j$ is just an aggregated version of the simple case $k_j = n_j = 1$. For this simple case, either $h_j = 0$ or $h_j = 1$. If $h_j = 1$, we use the helping mechanism to recover the interfered symbols. If $h_j = 0$, there are no helped levels and we simply use the scheme for $\alpha_j = 1$ in [1].

2.4.4.3 Modification for $1 < \alpha_j < 2$

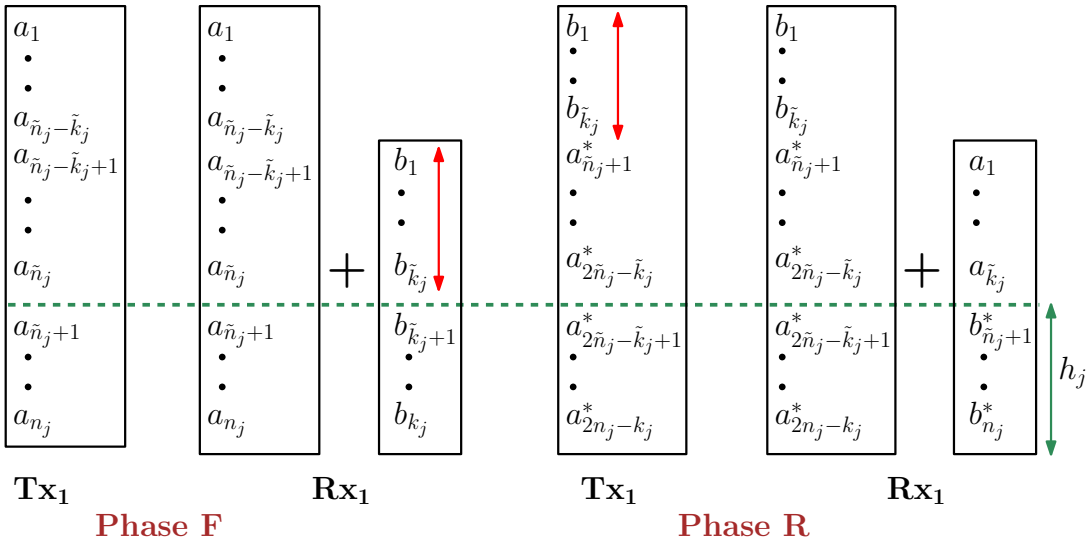
The top h_j levels (of the direct link at the receiver) are selected as helped levels as shown in Figure 2.9 (b). Again, phase F remains the same as in [1] and the modification is only for phase R. For illustration purposes, consider that in phase F for a subcarrier with $1 < \alpha_j < 2$, Tx_1 sends fresh symbols $[a_{\tilde{n}_j+1}\ a_{\tilde{n}_j+2}\ \ldots a_{n_j}\ a_1\ a_2 \ldots a_{\tilde{n}_j}]$ on the top n_j levels[4] (as shown in Figure 2.9 (b)). Similarly, Tx_2 sends fresh symbols $[b_{\tilde{n}_j+1}\ b_{\tilde{n}_j+2}\ \ldots b_{n_j}\ b_1\ b_2 \ldots b_{\tilde{n}_j}]$ on the top n_j levels. The bottom $k_j - n_j$ levels are not used. If there is no interference, all the fresh symbols are received and the transmitters stay in phase F. If there is interference, the transmitters transition to phase R. In phase R of the scheme in [1], the bottom $k_j - n_j$ levels were not used and the $2n_j - k_j$ interfering symbols in phase F were sent on the $2n_j - k_j$ levels above the unused levels. In the modified scheme, the transmitters send only $2n_j - k_j - h_j = 2\tilde{n}_j - \tilde{k}_j$ interfering symbols (from phase F) on the $2\tilde{n}_j - \tilde{k}_j$ levels above the $k_j - n_j$ unused levels in the bottom. These interfering symbols correspond to the $2\tilde{n}_j - \tilde{k}_j$ levels below the top h_j levels in the direct link at the receiver as shown in Figure 2.9 (b). In the remaining levels, fresh symbols are sent (starred symbols in Figure 2.9 (b)). Ignoring the h_j helped levels, the resulting system of linear equations at the receivers is exactly the same as in [1] with direct link strength $\tilde{n}_j$ and interfering link strength $\tilde{k}_j$. Thus at end of phase R, Rx_1 is able to decode $\{a_1, a_2, \ldots a_{2\tilde{n}_j - \tilde{k}_j}\}$ (interfered symbols in phase F) and $\{a^*_{\tilde{n}_j+1}, a^*_{\tilde{n}_j+2}, \ldots a^*_{\tilde{k}_j}\}$ (fresh symbols in phase R). To decode interfered symbols in the helped levels, the helping mechanism is used (which collects all interfered symbols in helped levels during a block of duration N_B and enables their recovery in the next block). So effectively, the rate obtained from a subcarrier with $1 < \alpha_j < 2$ is $h_j + (\tilde{n}_j - \frac{p}{1+p}(2\tilde{n}_j - \tilde{k}_j))$.

Taking into account the above modifications and adding the rates across subcarriers we achieve rate R_C. To give an illustrative example of our achievability scheme for $\Delta < 0$, we slightly modify the toy example in Section 2.1 as described below.

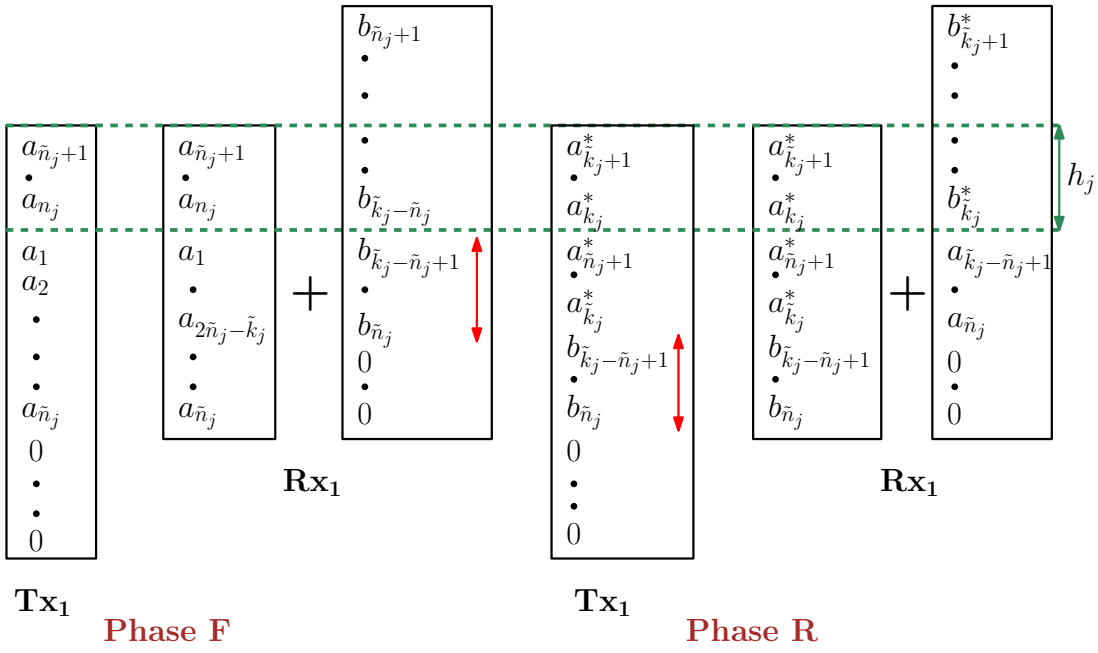
Example 2 ($n_1 = 2$, $k_1 = 2$, $n_2 = 1$ and $k_2 = 3$) *Compared to the original toy example as shown in Figure 2.10 (a), we have modified only the first subcarrier. For this case, there are two levels in*

[4]This particular labeling of the symbols is just for convenience in describing the modification in phase R.

the first subcarrier which face may interference but there is only one helper level (green level in Figure 2.10 (b)) available in the second subcarrier. Hence $\Delta = 1-2 = -1$ *and* $C_{sym} = 2+\frac{2}{3}$*. We help the bottom level in the first subcarrier (as we did in the original toy example) and by simply copying the scheme in the original toy example we achieve rate* 2*. For the top level in the first subcarrier (gray level in Figure 2.10 (b)), we use the optimal single carrier scheme for* $\alpha_1 = 1$ *[1] and achieve additional rate* $\frac{2}{3}$*. In this example, it is easy to see that the helping mechanism and the single carrier scheme can be executed in parallel.*

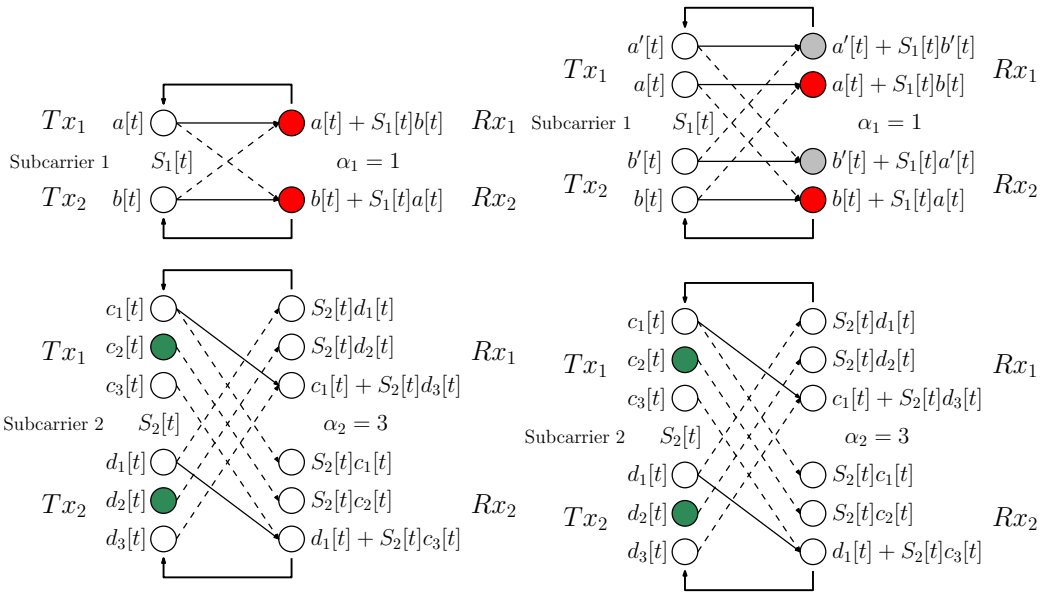


(a) $\alpha_j < 1$



(b) $1 < \alpha_j < 2$

Figure 2.9: Modified single carrier schemes for $\alpha_j < 1$ and $1 < \alpha_j < 2$ which run in parallel with the helping mechanism when $\Delta < 0$. Because of h_j helped levels, the effective direct and interfering link strengths are $\tilde{n}_j = n_j - h_j$ and $\tilde{k}_j = k_j - h_j$. The bidirectional red arrows indicate the interfering symbols (from phase F) sent in phase R of the modified scheme.



(a) $n_1 = 1$, $k_1 = 1$, $n_2 = 1$ and $k_2 = 3$.

(b) $n_1 = 2$, $k_1 = 2$, $n_2 = 1$ and $k_2 = 3$.

Figure 2.10: Toy example and its modification: (a) original toy example, and (b) Example 2.

2.5 GDoF: GN setup

In this section, we describe outer bounds followed by inner bounds (in Sections 2.5.1 and 2.5.2) on the GDoF for GN setup. As mentioned in Section 2.3, for the GDoF analysis we assume $g_{D,j} = \sqrt{SNR}$, $g_{I,j} = \sqrt{INR_j}$ and $INR_j = SNR^{\beta_j}$. We assume a rational β_j to simplify the achievability schemes (described in Sections 2.5.1 and 2.5.2). With the above assumptions, the GDoF for GN setup is defined as follows:

$$GDoF\left(\beta_1,\beta_2,\dots\beta_M\right) = \limsup_{SNR\to\infty} \frac{C_{sym}\left(SNR,\beta_1,\beta_2,\dots\beta_M\right)}{M\log(SNR)},$$

where C_{sym} is the symmetric capacity. From outer bounds (2.8) and (2.9) for the GN setup (proved in Section 2.6), we have bounds on C_{sym} as follows:

$$\begin{aligned} C_{sym} &\le \min\left(\frac{p}{2}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right),\ \frac{p}{1+p}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right)\right) \\ &= \begin{cases} \frac{p}{2}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right) & \text{if } \Delta_G \ge 0, \\ \frac{p}{1+p}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right) & \text{if } \Delta_G < 0, \end{cases} \end{aligned} \tag{2.12}$$

where $\Delta_G = \sum_{j=1}^{M}\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right) + \log\left(1+\frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right) - 2\log\left(1+|g_{D,j}|^2\right)$. Using (2.12), the following outer bound on GDoF holds:

$$\begin{aligned} &GDoF(\beta_1,\dots,\beta_M) \\ &\le \min\left(\lim_{SNR\to\infty}\frac{\frac{p}{2}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right)}{M\log(SNR)},\ \lim_{SNR\to\infty}\frac{\frac{p}{1+p}\Delta_G + \sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right)}{M\log(SNR)}\right) \\ &= \min\left(\frac{\frac{p}{2}\Delta_{GDoF}}{M}+1,\ \frac{\frac{p}{1+p}\Delta_{GDoF}}{M}+1\right), \end{aligned} \tag{2.13}$$

where Δ_{GDoF} is derived from Δ_G as shown below:

$$\begin{aligned} \Delta_{GDoF} &= \lim_{SNR\to\infty}\frac{\Delta_G}{\log\left(SNR\right)} \\ &= \sum_{j=1}^{M}\left(\max\left(1,\beta_j\right) + \left(1-\beta_j\right)^+ - 2\right) \\ &= \left(\sum_{j:\beta_j>2}\beta_j - 2\right) - \left(\sum_{j:\beta_j\le 1}\beta_j\right) - \left(\sum_{j:1<\beta_j\le 2} 2-\beta_j\right). \end{aligned} \tag{2.14}$$

In the remainder of this section, we describe achievability schemes (inner bounds) which achieve outer bound (2.13). The schemes for the GDoF setting mimic the achievability schemes for symmetric capacity in the LD setup by using techniques from [12]. Hence, the scheme for $\Delta_{GDoF} \geq 0$ (Section 2.5.1) in the GDoF setting mimics the scheme for $\Delta \geq 0$ in LD setup and the scheme for $\Delta_{GDoF} < 0$ (Section 2.5.2) mimics the scheme for $\Delta < 0$ in LD setup.

2.5.1 GDoF inner bound when $\Delta_{GDoF} \geq 0$

We use a block based scheme (block size N_B) which mimics the scheme for $\Delta \geq 0$ in Section 2.4.3 for LD setup. For convenience in describing our scheme, we will work with the following *real* channel (the achievable rate for the complex channel in GN setup is just twice the achievable rate for this channel):

$$y_j^{(i)}[t] = \sqrt{SNR}\, x_j^{(i)}[t] + \big(S_j[t]\big)\sqrt{INR_j}\, x_j^{(i')}[t] + z_j^{(i)}[t], \tag{2.15}$$

where $x_j^{(i)}[t]$, $x_j^{(i')}[t] \in \mathbb{R}$, $\frac{1}{N}\sum_{t=1}^{N} |x_j^{(i)}[t]|^2 \leq 1$ and $z_j^{(i)}[t] \sim \mathcal{N}(0,1)$. Similar to the analysis in [12], we consider

$$SNR = Q^{2m}, \tag{2.16}$$

where Q and m are positive integers. Furthermore, m is such that $\forall j \in \{1,2,\ldots M\}$, $m\beta_j$ is an integer (always possible since all β_j are rational). By letting m grow to infinity, we get a sequence of SNRs that approach infinity. Using (2.16), the received signal in (2.15) can be rewritten as follows:

$$y_j^{(i)}[t] = Q^m x_j^{(i)}[t] + \big(S_j[t]\big)\, Q^{m\beta_j} x_j^{(i')}[t] + z_j^{(i)}[t]. \tag{2.17}$$

Following [12], we will express positive real signals in Q-ary representation using Q-ary digits $0,1,\ldots Q-1$ (which we will refer to as "qits", similar to [12]). To mimic the achievability scheme for $\Delta \geq 0$ in LD setup (Section 2.4.3), we use the following structure for the input signals (we drop the time index for convenience):

- For j with $\beta_j > 2$,

$$x_j^{(i)} = \left[0 \,.\, x_{j,m\beta_j}^{(i)} \, x_{j,m\beta_j-1}^{(i)} \, \ldots x_{j,1}^{(i)}\right]_Q, \tag{2.18}$$

where $x_{j,1}^{(i)} = x_{j,2}^{(i)} = \ldots = x_{j,m}^{(i)} = 0$ and for the remaining $r \in \{1,2,\ldots m\beta_j\} - \{1,2,\ldots m\}$, $x_{j,r}^{(i)} \in \{1,2,\ldots Q-2\}$.

- For j with $\beta_j \leq 1$,

$$x_j^{(i)} = \left[0 \,.\, x_{j,m}^{(i)} \, x_{j,m-1}^{(i)} \, \ldots x_{j,1}^{(i)}\right]_Q, \tag{2.19}$$

where $x_{j,r}^{(i)} \in \{1,2,\ldots \lfloor \frac{Q-1}{2} \rfloor - 1\}$ for $r \in \{1,2,\ldots m\}$.

- For j with $1 < \beta_j \leq 2$,

$$x_j^{(i)} = \left[0 \,.\, x_{j,m\beta_j}^{(i)} \, x_{j,m\beta_j-1}^{(i)} \, \ldots x_{j,1}^{(i)}\right]_Q, \tag{2.20}$$

where $x_{j,1}^{(i)} = x_{j,2}^{(i)} = \ldots = x_{j,m(\beta_j-1)}^{(i)} = 0$ and for the remaining $r \in \{1,2,\ldots m\beta_j\} - \{1,2,\ldots m(\beta_j - 1)\}$,
$x_{j,r}^{(i)} \in \{1,2,\ldots \lfloor \frac{Q-1}{2} \rfloor - 1\}$.

The structure (*i.e.,* non-zero qits) used is same as in the scheme for LD setup (Section 2.4.3). The restrictions on the values taken by non-zero qits arises from techniques in [12] (these simplify the analysis by preventing carry overs when signals interfere, see [12] for details). In the absence of noise, it is easy to see the similarities between the LD setup and above setup; qits in a signals are similar to *levels* in the LD setup. The following example makes this similarity more precise for the case of subcarriers with $\beta_j > 2$.

Example 3 *In a subcarrier with $\beta_j > 2$, the received signal at Rx_i after interference (in the absence of noise) is as follows:*

$$\left[x_{j,m\beta_j}^{(i)} \, x_{j,m\beta_j-1}^{(i)} \, \ldots x_{j,m\beta_j-m+1}^{(i)} \,.\, x_{j,m\beta_j-m}^{(i)} \ldots x_{j,1}^{(i)}\right]_Q + \left[x_{j,m\beta_j}^{(i')} \, x_{j,m\beta_j-1}^{(i')} \, \ldots \, x_{j,m+1}^{(i')} \, 0 \, 0 \, \ldots 0 \,.\, 0 \, 0\right]_Q. \tag{2.21}$$

*Clearly, the top m qits of the direct signal (*i.e., $x^{(i)}_{j,m\beta_j}\, x^{(i)}_{j,m\beta_j-1}\, \ldots x^{(i)}_{j,m\beta_j-m+1}$*) are interference free in the above scenario and by doing a modulo Q^m operation at the receiver, one can completely recover the direct signal. Even in the presence of noise, due to bounded variance of the noise, the higher qits can be decoded with negligible probability of error (as $m \to \infty$).*

Having shown the similarity between LD setup and the above setup in the absence of noise, we now describe the rates that we can achieve from the subcarriers in the GDoF setting.

$\beta_j \leq 1$ In this case, over a block only $(pN_B)m\beta_j$ qits in the direct signal are interfered. Assuming we are able to recover all (except $o(m)$) interfering qits (using the helping mechanism described for $\beta_j > 2$ below), we can achieve the following rate:

$$m \log_Q \left(\lfloor \frac{Q-1}{2} \rfloor - 1 \right) + o(m).$$

The above rate follows directly from the analysis in [12].

$1 < \beta_j \leq 2$ In this case, over a block only $(pN_B)m(2-\beta_j)$ qits in the direct signal are interfered. Assuming we are able to recover all (except $o(m)$) interfering qits (using the helping mechanism described for $\beta_j > 2$ below), we can achieve the following rate:

$$m \log_Q \left(\lfloor \frac{Q-1}{2} \rfloor - 1 \right) + o(m).$$

$\beta_j > 2$ The top m qits in the subcarriers with $\beta_j > 2$ are always received interference free. So from them we can achieve rate:

$$m \log_Q (Q-2) + o(m).$$

We now describe the helping mechanism for the GDoF setting. For removing the interfering qits for subcarriers with $\beta_j < 2$ in the previous block, we need to use $\sum_{j:1<\beta_j\leq 2} m(2-\beta_j) + \sum_{j:\beta_j\leq 1} m\beta_j$ *helper* qits in subcarriers with $\beta_j > 2$; these are the middle $m(\beta_j - 2)$ qits below the top m qits. Since $\Delta_{GDoF} \geq 0$, we have sufficient number of such helper qits to recover all interfering qits in subcarriers with $\beta_j < 2$. The helping mechanism is same as described for the LD setup (with minor

changes for the Q-ary setup). From the leftover helper qits, we can achieve an additional rate using the bursty relaying technique. Summing the rates for all subcarriers we have the following inner bound (a factor of $\frac{1}{2}$ is included to account for the complex channel):

$$\begin{aligned}
&\frac{1}{2}C_{sym}(SNR,\beta_1,\dots\beta_M)\\
&\geq \left(m\sum_{j:\beta_j\leq 2}\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)+o(m)\right)+\left(m\sum_{j:\beta_j>2}\log_Q(Q-2)+o(m)\right)\\
&\quad+\left(\frac{p}{2}m\left(\sum_{j:\beta_j>2}(\beta_j-2)-\sum_{j:\beta_j\leq 1}\beta_j-\sum_{j:1<\beta_j\leq 2}(2-\beta_j)\right)\log_Q(Q-2)+o(m)\right). \qquad (2.22)
\end{aligned}$$

Hence,

$$\begin{aligned}
GDoF(\beta_1,\dots\beta_M) &= \limsup_{m\to\infty}\frac{C_{sym}(SNR,\beta_1,\dots\beta_M)}{M\log_Q(Q^{2m})}\\
&\overset{(a)}{\geq}\frac{\frac{p}{2}\left(\left(\sum_{j:\beta_j>2}\beta_j-2\right)-\left(\sum_{j:\beta_j\leq 1}\beta_j\right)-\left(\sum_{j:1<\beta_j\leq 2}2-\beta_j\right)\right)}{M}+1\\
&=\frac{\frac{p}{2}\Delta_{GDoF}}{M}+1,
\end{aligned}$$

where (a) follows from large enough Q. Since the inner bound on GDoF matches the outer bound, we have a tight result when $\Delta_{GDoF}\geq 0$.

2.5.2 GDoF inner bound when $\Delta_{GDoF}<0$

As in the case of $\Delta_{GDoF}\geq 0$ in Section 2.5.1, we focus on the real channel in (2.17) for our achievability scheme. The scheme for this case mimics the achievability of symmetric capacity in LD setup for $\Delta<0$ by using the techniques from [12]. Since we have already illustrated the usage of techniques from [12] (for the case $\Delta_{GDoF}\geq 0$) in mimicking the LD setup schemes for the GDoF setting, we will briefly sketch the inner bound for $\Delta_{GDoF}<0$.

Following the strategy of *helping as much possible* for the case $\Delta<0$ in LD setup, we use the middle $m(\beta_j-2)$ qits (below the top m qits) in subcarriers with $\beta_j>2$ as helper qits. All the helper qits are used to recover interference in helped qits in subcarriers with $\beta_j<2$ (each subcarrier with $\beta_j<2$ has h_j helped qits and $\sum_{j:\beta_j<2}h_j=\sum_{j:\beta_j>2}m(\beta_j-2)$). So we get the following rates from

subcarriers:

- For j with $\beta_j \geq 2$: $m\log_Q(Q-2)+o(m)$.
- For j with $1<\beta_j<2$: $\left(h_j+\frac{1-p}{1+p}(m-h_j)+\frac{p}{1+p}(m\beta_j-h_j)\right)\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)+o(m)$.
- For j with $\beta_j \leq 1$: $\left(h_j+(m-h_j)-\frac{p}{1+p}(m\beta_j-h_j)\right)\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)+o(m)$.

It should be noted that due to noise, some of the interfering qits in phase F (of the single carrier scheme executed in parallel with the helping mechanism) may not be decoded correctly at Tx_i (after feedback) and this may affect the recovery of qits in phase R. However, it can be shown that such an *error propagation* leads to $o(m)$ reduction (compared to the case without noise) in the achievable rate for a subcarrier. Combining the rates from all subcarriers, we have the following bound (factor of 2 included for the complex channel):

$$\begin{aligned}
&C_{sym}(SNR,\beta_1,\dots\beta_M)\\
&\geq 2\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)\times\\
&\Bigg(\sum_{j:\beta_j<2}h_j+\sum_{j:\beta_j\geq 2}m+\sum_{j:\beta_j\leq 1}(m-h_j)-\frac{p}{1+p}(m\beta_j-h_j)\\
&\qquad+\sum_{j:1<\beta_j<2}\frac{1-p}{1+p}(m-h_j)+\frac{p}{1+p}(m\beta_j-h_j)\Bigg)+o(m)\\
&\overset{(a)}{=}2\left(\frac{p}{1+p}m\Delta_{GDoF}+\sum_{j=1}^{M}m\right)\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)+o(m),
\end{aligned}\tag{2.23}$$

where (a) follows from $\sum_{j:\beta_j<2}h_j=\sum_{j:\beta_j>2}m(\beta_j-2)$. Now, we have the following bound on the GDoF:

$$\begin{aligned}
GDoF(\beta_1,\beta_2\dots\beta_M)&=\limsup_{m\to\infty}\frac{C_{sym}(SNR,\beta_1,\dots\beta_M)}{M\log_Q(Q^{2m})}\\
&\geq\lim_{m\to\infty}\frac{\left(\frac{p}{1+p}m\Delta_{GDoF}+\sum_{j=1}^{M}m\right)\log_Q\left(\lfloor\frac{Q-1}{2}\rfloor-1\right)+o(m)}{mM}\\
&\overset{(a)}{=}\frac{\frac{p}{1+p}\Delta_{GDoF}}{M}+1,
\end{aligned}\tag{2.24}$$

where (a) follows from large enough Q. The above inner bound matches outer bound (2.13) when $\Delta_{GDoF}<0$ and this completes the GDoF characterization.

2.6 Outer bounds: LD and GN setups

In this section, we focus on proofs of outer bounds in the LD and GN setups. We refer to outer bounds (2.4) and (2.8) as causal outer bounds as they account for the causal knowledge of subcarrier interference states at the transmitter. For proving these causal outer bounds, we use a subset entropy inequality by Madiman and Tetali which we describe in Section 2.6.1, prior to the proofs. Then we introduce some additional notation in Section 2.6.2 followed by outer bound proofs for the LD setup (Section 2.6.3) and GN setup (Section 2.6.4).

2.6.1 Madiman-Tetali subset inequality

We now describe a subset entropy inequality by Madiman and Tetali [4]. Consider a hypergraph $(U, \mathscr{E})$ where U is a finite ground set and $\mathscr{E}$ is a collection of subsets of U. A function $\mathscr{G} : \mathscr{E} \to \mathbb{R}^+$ is called a fractional partition of $(U, \mathscr{E})$ if it satisfies the following condition $\forall\, j \in U$:

$$\sum_{E \in \mathscr{E} : j \in E} \mathscr{G}(E) = 1. \tag{2.25}$$

With the above definition, the subset entropy inequality can now be stated as follows:

$$\sum_{E \in \mathscr{E}} \mathscr{G}(E) H(X_E) \geq H(X_U), \tag{2.26}$$

where $\mathscr{G}$ is a fractional partition and the above inequality holds for any collection of jointly distributed random variables X_U. The differential entropy version of the above inequality has the same form [4]. To use these inequalities in our setups, we first choose a suitable fractional partition as explained below. For $\mathbf{s} \in \{0,1\}^M$, let $\mathbf{S}[t] = (S_1[t], S_2[t], \ldots S_M[t]) = \mathbf{s}$ denote the collection of interference states of all the M subcarriers at time index t. As specified in Section 2.2, the occurrence of $\mathbf{S}[t] = \mathbf{s}$ is governed by the joint probability distribution $\mathbb{P}(\mathbf{S}[t] = \mathbf{s})$. To define a fractional partition, we consider the ground set $U = \{1, 2, \ldots M\}$ (*i.e.,* the index set of subcarriers) and view $\mathbf{s} \in \{0,1\}^M$ as a collection of M indicator functions for representing any subset of U. The power set of U (excluding subsets $\mathbf{s}$ such that $\mathbb{P}(\mathbf{S}[t] = \mathbf{s}) = 0$) is chosen as set $\mathscr{E}$. Now, we define a

fractional partition $\mathscr{G} : \mathscr{E} \to \mathbb{R}^+$ as follows:

$$\mathscr{G}(E) = \frac{\mathbb{P}(\mathbf{S}[t] = \mathbf{s}_E)}{p}, \tag{2.27}$$

where $E \in \mathscr{E}$ and $\mathbf{s}_E$ denotes the joint state where only the subcarriers whose index is in set E face interference. The fractional partition condition holds as follows:

$$\sum_{E \in \mathscr{E}: j \in E} \frac{\mathbb{P}(\mathbf{S}[t] = \mathbf{s}_E)}{p} = \frac{\mathbb{E}\left(S_j[t]\right)}{p} = 1. \tag{2.28}$$

In Section 2.6.3.1, we demonstrate the application of inequality (2.26), in conjunction with the fractional partition defined in (2.27), for proving outer bound (2.4). Similarly, in Section 2.6.4.1, for proving outer bound (2.8) we use the differential entropy version [4] of inequality (2.26) with the same fractional partition.

2.6.2 Additional notation

For notational convenience, we use indicator functions $\mathbb{I}_{j \notin \mathbf{s}}$ and $\mathbb{I}_{j \in \mathbf{s}}$ to denote the absence and presence of interference in subcarrier j when the joint state realization across M subcarriers is $\mathbf{S}[t] = \mathbf{s} \in \{0,1\}^M$. Also, in the proofs we use $\sum_{\mathbf{s}}$ to denote $\sum_{\mathbf{s} \in \{0,1\}^M}$. The additional notation used for LD setup proofs is listed below in Table 2.1.

Table 2.1: Notation used in LD setup proofs.

$\mathbf{S}_{1:t}$	$\triangleq$	$(\mathbf{S}[1], \mathbf{S}[2], \ldots, \mathbf{S}[t])$
$\mathbf{Y}^{(i)}[t]$	$\triangleq$	$\left(\mathbf{y}_1^{(i)}[t], \mathbf{y}_2^{(i)}[t], \ldots, \mathbf{y}_M^{(i)}[t]\right)$, *i.e.*, received signal (across M subcarriers) for Rx_i at time index t
$\mathbf{Y}_{\mathbf{s}}^{(i)}[t]$	$\triangleq$	received signal (across M subcarriers) for Rx_i at time t when $\mathbf{S}[t] = \mathbf{s}$
$\mathbf{Y}_{1:t}^{(i)}$	$\triangleq$	$\left(\mathbf{Y}^{(i)}[1], \mathbf{Y}^{(i)}[2], \ldots, \mathbf{Y}^{(i)}[t]\right)$
$\mathbf{V}_{\mathbf{s}}^{(i)}[t]$	$\triangleq$	interfering signals (across M subcarriers) for Rx_i when $\mathbf{S}[t] = \mathbf{s}$
$\mathbf{V}_{1:t}^{(i)}$	$\triangleq$	$\left(\mathbf{V}_{\mathbf{S}[1]}^{(i)}[1], \mathbf{V}_{\mathbf{S}[2]}^{(i)}[2], \ldots, \mathbf{V}_{\mathbf{S}[t]}^{(i)}[t]\right)$

$\tilde{\mathbf{V}}^{(i)}[t] \triangleq$ interfering signals (across M subcarriers) at Rx_i when all its subcarriers face interference at time index t; equivalent to $\mathbf{V}_{\mathbf{s}}^{(i)}[t]$ with $\mathbf{s} = \{1, 1, \ldots, 1\}$

$\hat{\mathbf{X}}^{(i)}[t] \triangleq$ received signal (across M subcarriers) at Rx_i when all its subcarriers are interference free at time index t; equivalent to $\mathbf{Y}_{\mathbf{s}}^{(i)}[t]$ with $\mathbf{s} = \{0, 0, \ldots, 0\}$

The notation used for GN setup proofs is listed below in Table 2.2. Some notation is common to both LD and GN setup proofs, and hence we have some repetitions from Table 2.1 in Table 2.2 for easier lookup.

Table 2.2: Notation used in GN setup proofs.

$\mathbf{S}_{1:t}$	$\triangleq$	$(\mathbf{S}[1], \mathbf{S}[2], \ldots, \mathbf{S}[t])$
$\mathbf{Y}^{(i)}[t]$	$\triangleq$	$\left(y_1^{(i)}[t], y_2^{(i)}[t], \ldots, y_M^{(i)}[t]\right)$, *i.e.*, received signal (across M subcarriers) for Rx_i at time index t
$\mathbf{Y}_{\mathbf{s}}^{(i)}[t]$	$\triangleq$	received signal (across M subcarriers) for Rx_i at time t when $\mathbf{S}[t] = \mathbf{s}$; the difference between $\mathbf{Y}^{(i)}[t]$ and $\mathbf{Y}_{\mathbf{s}}^{(i)}[t]$ is that the state at time t is specified as $\mathbf{s}$ in the latter
$\mathbf{Y}_{1:t}^{(i)}$	$\triangleq$	$\left(\mathbf{Y}^{(i)}[1], \mathbf{Y}^{(i)}[2], , \ldots, \mathbf{Y}^{(i)}[t]\right)$
$\mathbf{Z}^{(i)}[t]$	$\triangleq$	$\left(z_1^{(i)}[t], z_2^{(i)}[t], \ldots, z_M^{(i)}[t]\right)$, *i.e.*, receiver noise (across M subcarriers) for Rx_i at time index t
$\mathbf{Z}_{\mathbf{s}}^{(i)}[t]$	$\triangleq$	receiver noise in interfered subcarriers for Rx_i at time index t when $\mathbf{S}[t] = \mathbf{s}$

$\mathbf{Z}_{\mathbf{s}^c}^{(i)}[t] \triangleq$ receiver noise in interference free subcarriers for Rx_i at time index t when $\mathbf{S}[t] = \mathbf{s}$

$\mathbf{V}_{\mathbf{s}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t] \triangleq \left(S_1[t]g_{I,1}x_1^{(i')}[t] + z_1^{(i)}[t],\ S_2[t]g_{I,2}x_2^{(i')}[t] + z_2^{(i)}[t], \ldots, S_M[t]g_{I,M}x_M^{(i')}[t] + z_M^{(i)}[t]\right)$, *i.e.*, interfering signal (if present) plus noise, across M subcarriers, for Rx_i at time index t when $\mathbf{S}[t] = \mathbf{s}$

$\mathbf{V}_{\mathbf{s}}^{(i)}[t] \uplus \mathbf{Z}_{\mathbf{s}}^{(i)}[t] \triangleq$ interfering signal plus noise in interfered subcarriers for Rx_i at time index t when $\mathbf{S}[t] = \mathbf{s}$; this does not include the noise terms for subcarriers which do not face interference at time t (unlike $\mathbf{V}_{\mathbf{s}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t]$)

$\mathbf{V}_{1:t}^{(i)} \uplus \mathbf{Z}_{1:t}^{(i)} \triangleq \left(\mathbf{V}_{\mathbf{S}[1]}^{(i)}[1] \uplus \mathbf{Z}^{(i)}[1],\ \mathbf{V}_{\mathbf{S}[2]}^{(i)}[2] \uplus \mathbf{Z}^{(i)}[2], \ldots, \mathbf{V}_{\mathbf{S}[t]}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t]\right)$

$\tilde{\mathbf{V}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t] \triangleq$ interfering signal plus noise (across M subcarriers) at Rx_i when all its subcarriers face interference at time index t; equivalent to $\mathbf{V}_{\mathbf{s}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t]$ with $\mathbf{s} = \{1, 1, \ldots, 1\}$

$\hat{\mathbf{X}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t] \triangleq$ received signal (across M subcarriers) at Rx_i when all its subcarriers are interference free at time index t; equivalent to $\mathbf{Y}_{\mathbf{s}}^{(i)}[t]$ with $\mathbf{s} = \{0, 0, \ldots, 0\}$

$\hat{\mathbf{X}}_{1:t}^{(i)} \uplus \mathbf{Z}_{1:t}^{(i)} \triangleq \left(\hat{\mathbf{X}}^{(i)}[1] \uplus \mathbf{Z}^{(i)}[1],\ \hat{\mathbf{X}}^{(i)}[2] \uplus \mathbf{Z}^{(i)}[2], \ldots, \hat{\mathbf{X}}^{(i)}[t] \uplus \mathbf{Z}^{(i)}[t]\right)$

2.6.3 Outer bounds: LD setup

For the LD setup, outer bounds (2.3) and (2.5) are straightforward (multicarrier) extensions of the outer bounds in the single carrier setup [1] (their proof is described in Appendix A.0.1 and Appendix A.0.2 respectively). Our main contribution in terms of outer bound techniques lies in proving outer bound (2.4) (and its corresponding version (2.8) for the GN setup). We focus on the proof of outer bound (2.4) in Section 2.6.3.1.

2.6.3.1 Proof of outer bound (2.4)

For proving outer bound (2.4), we first obtain a bound on $R^{(1)}$ followed by a bound on $R^{(2)}$. Finally, to obtain the bound on $R^{(1)} + pR^{(2)}$, we add the bounds on $R^{(1)}$ and $R^{(2)}$ accordingly, removing intermediate *interference* terms in the process. In particular, to facilitate the removal of such interference terms in the bound for $R^{(1)} + pR^{(2)}$, we use the subset entropy inequality (2.26) while bounding $R^{(1)}$. In addition, we also account for the causal knowledge of the state sequence $\mathbf{S}_{1:N}$ in the proof. We describe the proof details below.

Using Fano's inequality for Rx_1, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
&NR^{(1)} - N\varepsilon \\
&\leq I\left(W^{(1)}; \mathbf{Y}^{(1)}_{1:N}, \mathbf{S}_{1:N}\right) \\
&= \sum_{t=1}^{N} I\left(W^{(1)}; \mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t]\right) \\
&= \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{Y}^{(1)}_{\mathbf{s}}[t] \middle| \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
&\quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{Y}^{(1)}_{\mathbf{s}}[t] \middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
&\overset{(a)}{\leq} \sum_{t=1}^{N} \sum_{\mathbf{s}} \left[\mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \sum_{j=1}^{M} \left\{n_j \mathbb{I}_{j \notin \mathbf{s}} + \max(n_j, k_j) \mathbb{I}_{j \in \mathbf{s}}\right\}\right] \\
&\quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t] \middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
&= \sum_{t=1}^{N} \sum_{j=1}^{M} \sum_{\mathbf{s}} \left[n_j \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \mathbb{I}_{j \notin \mathbf{s}} + \max(n_j, k_j) \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \mathbb{I}_{j \in \mathbf{s}}\right] \\
&\quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t] \middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
&\overset{(b)}{=} N \sum_{j=1}^{M} (1-p) n_j + p \max(n_j, k_j) \\
&\quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t] \middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
&\overset{(c)}{=} N \sum_{j=1}^{M} (1-p) n_j + p \max(n_j, k_j) - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t] \middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right)
\end{aligned}
$$

$$\overset{(d)}{\leq} N\sum_{j=1}^{M}(1-p)n_j + p\max(n_j,k_j) - p\sum_{t=1}^{N} H\left(\tilde{\mathbf{V}}^{(1)}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right)$$

$$\overset{(e)}{=} N\sum_{j=1}^{M} n_j + (\max(n_j,k_j) - n_j)p - p\sum_{t=1}^{N} H\left(\tilde{\mathbf{V}}^{(1)}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right), \tag{2.29}$$

where (a) follows from upper bounding the entropy of output in an interference free subcarrier by n_j (for an interfered subcarrier it is upper bounded by $\max\left(n_j,k_j\right)$). In addition, for step (a), we use the fact that $H\left(\mathbf{Y}^{(1)}_{\mathbf{s}}[t]\middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t]=\mathbf{s}\right) = H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t]\middle| W^{(1)}, \mathbf{Y}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}, \mathbf{S}[t]=\mathbf{s}\right)$ (*i.e.*, with the given conditioning, the direct signal can be determined and only the interfering signal contributes to the output uncertainty). Step (b) above follows from the fact that for a fixed subcarrier j, $\sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t]=\mathbf{s})\,\mathbb{I}_{j\in\mathbf{s}} = p$ (*i.e.*, the marginal probability of interference for subcarrier j), and (c) follows from the observation that the conditioning term $\mathbf{S}[t]=\mathbf{s}$ can be dropped because it is implicit in $\mathbf{V}^{(1)}_{\mathbf{s}}[t]$. In addition, for step (c) we have used the observation that given $W^{(1)}$, $\mathbf{Y}^{(1)}_{1:t-1}$ and $\mathbf{S}_{1:t-1}$, we can determine the direct signal till time $t-1$ and hence the interfering signal $\mathbf{V}^{(1)}_{1:t-1}$; also, given $W^{(1)}$, $\mathbf{V}^{(1)}_{1:t-1}$ and $\mathbf{S}_{1:t-1}$, we can determine $\mathbf{Y}^{(1)}_{1:t-1}$. Step (d) above follows from using the Madiman-Tetali subset inequality (2.26) for the fractional partition defined in (2.27). More precisely, in the above context, we can rewrite the Madiman-Tetali subset inequality (2.26) as

$$\sum_{\mathbf{s}} \frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p} H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right) \geq H\left(\tilde{\mathbf{V}}^{(1)}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right),$$

where $\frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p}$ corresponds to the fractional partition defined in (2.27) and the term involving entropy of subsets, *i.e.*, $\sum_{\mathbf{s}} \frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p} H\left(\mathbf{V}^{(1)}_{\mathbf{s}}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right)$ is bounded by the joint entropy term, *i.e.*, $H\left(\tilde{\mathbf{V}}^{(1)}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right)$. Finally, step (e) simply follows by rewriting $(1-p)n_j + p\max(n_j,k_j)$ as $n_j + (\max(n_j,k_j)-n_j)p$.

Note that step (d) above is a crucial step in the proof since the resulting *interference* term (*i.e.*, the joint entropy term $H\left(\tilde{\mathbf{V}}^{(1)}[t]\middle| W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right)$) in the bound for $R^{(1)}$ can be removed from the final outer bound using a similar interference term in the bound for $R^{(2)}$ (described below).

Using Fano's inequality for Rx_2, for any $\varepsilon > 0$, there exists a large enough N such that:

$$NR^{(2)} - N\varepsilon$$

$$
\begin{aligned}
&\leq I\left(W^{(2)};\mathbf{Y}_{1:N}^{(1)},\mathbf{Y}_{1:N}^{(2)},\mathbf{S}_{1:N},W^{(1)}\right)\\
&= I\left(W^{(2)};\mathbf{Y}_{1:N}^{(1)},\mathbf{Y}_{1:N}^{(2)},\mathbf{S}_{1:N}\middle|W^{(1)}\right)\\
&= \sum_{t=1}^{N} I\left(W^{(2)};\mathbf{Y}^{(2)}[t],\mathbf{Y}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]\right)\\
&= \sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\left(H\left(\mathbf{Y}_{\mathbf{s}}^{(2)}[t],\mathbf{Y}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\right.\\
&\quad\left.-H\left(\mathbf{Y}_{\mathbf{s}}^{(2)}[t],\mathbf{Y}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},W^{(2)},\mathbf{S}[t]=\mathbf{s}\right)\right)\\
&\overset{(a)}{=} \sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)H\left(\mathbf{Y}_{\mathbf{s}}^{(2)}[t],\mathbf{Y}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(b)}{=} \sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)H\left(\hat{\mathbf{X}}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(c)}{\leq} \sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(d)}{=} \sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(e)}{=} \sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)\\
&\overset{(f)}{\leq} \sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)\\
&= \sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{V}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right), \qquad (2.30)
\end{aligned}
$$

where (a) follows from the observation that $\mathbf{Y}_{\mathbf{s}}^{(2)}[t]$ and $\mathbf{Y}_{\mathbf{s}}^{(1)}[t]$ can be determined given $\mathbf{Y}_{1:t-1}^{(2)}$, $\mathbf{Y}_{1:t-1}^{(1)}$, $\mathbf{S}_{1:t-1}$, $W^{(1)}$, $W^{(2)}$ and $\mathbf{S}[t]=\mathbf{s}$;
hence, $H\left(\mathbf{Y}_{\mathbf{s}}^{(2)}[t],\mathbf{Y}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},W^{(2)},\mathbf{S}[t]=\mathbf{s}\right)=0$. Step (b) follows from the observation that given $\mathbf{Y}_{1:t-1}^{(2)}$, $\mathbf{Y}_{1:t-1}^{(1)}$, $\mathbf{S}_{1:t-1}$, $W^{(1)}$, and $\mathbf{S}[t]=\mathbf{s}$, we can determine the direct signal for Rx_1 and the interfering signal for Rx_2. Hence, the remaining output uncertainty stems from the interfering signal in Rx_1 and the direct signal in Rx_2. Step (c) follows by introducing additional interfering signals for Rx_1, and upper bounding
$H\left(\hat{\mathbf{X}}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)$ by
$H\left(\hat{\mathbf{X}}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\middle|\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)$. Step (d) follows from the observation that

$\hat{\mathbf{X}}^{(2)}[t]$ and $\tilde{\mathbf{V}}^{(1)}[t]$ do not depend on $\mathbf{S}[t]$, and step (e) simply follows from $\sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t]=\mathbf{s})=1$. Step (f) follows from removing the conditioning on $\mathbf{Y}^{(2)}_{1:t-1}$.

Using inequalities (2.29) and (2.30),

$$
\begin{aligned}
& NR^{(1)} - N\varepsilon + pNR^{(2)} - pN\varepsilon \\
& \leq N\sum_{j=1}^{M} n_j + \left(\max(n_j,k_j)-n_j\right)p + p\sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t]\middle|\tilde{\mathbf{V}}^{(1)}[t], W^{(1)}, \mathbf{V}^{(1)}_{1:t-1}, \mathbf{S}_{1:t-1}\right) \qquad (2.31)\\
& \leq N\sum_{j=1}^{M} n_j + \left(\max(n_j,k_j)-n_j\right)p + p\sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t]\middle|\tilde{\mathbf{V}}^{(1)}[t]\right) \\
& \leq N\sum_{j=1}^{M} n_j + \left(\max(n_j,k_j)-n_j\right)p + p\sum_{t=1}^{N}\sum_{j=1}^{M}(n_j-k_j)^+ \\
& = Np\Delta + N\sum_{j=1}^{M}(1+p)n_j, \qquad (2.32)
\end{aligned}
$$

where $\Delta = \sum_{j=1}^{M} \max(n_j,k_j) + (n_j-k_j)^+ - 2n_j$. As mentioned before, the joint entropy term in (2.29) is effectively removed using (2.30) in (2.31). The bound on $pR^{(1)}+R^{(2)}$ follows by symmetry, and this completes the proof of outer bound (2.4).

The above proof demonstrates a connection between subset entropy inequalities and bursty interference in multicarrier systems. Intuitively, the connection stems from the presence of subsets of interfered subcarriers in the outer bound; this leads to entropies of subsets of interference terms, which is bounded by the joint entropy as demonstrated in the proof of (2.4).

2.6.4 Outer bounds: GN setup

The outer bound proofs for the GN setup resemble the outer bound proofs for the LD setup, with a few modifications required for the presence of additive Gaussian noise. Like in the LD setup, our main contribution in terms of outer bound techniques lies in proving (2.8). We describe the proof of (2.8) in Section 2.6.4.1. The proof of outer bounds (2.7) and (2.9) is described in Appendix A.0.3 and Appendix A.0.4 respectively.

2.6.4.1 Proof of outer bound (2.8)

The structure of the proof for outer bound (2.8) is similar to that for proving (2.4). We first obtain a bound on $R^{(1)}$ followed by a bound on $R^{(2)}$. For the bound on $R^{(1)} + pR^{(2)}$, we add the bounds on $R^{(1)}$ and $R^{(2)}$ accordingly, removing intermediate interference terms in the process. To facilitate the removal of intermediate interference terms in the bound for $R^{(1)} + pR^{(2)}$, we use the differential entropy version of the subset entropy inequality (2.26) while bounding $R^{(1)}$. In addition, we also account for the causal knowledge of the state sequence $\mathbf{S}_{1:N}$ in the proof. The proof details are described below.

Using Fano's inequality for Rx_1, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
& NR^{(1)} - N\varepsilon \\
& \leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}\right) \\
& = \sum_{t=1}^{N} I\left(W^{(1)}; \mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t]\right) \\
& = \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
& \quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{Y}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
& \leq \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t] = \mathbf{s}\right) \\
& \quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{Y}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
& \overset{(a)}{\leq} N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) + NM \log(\pi e) \\
& \quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{Y}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right) \\
& \overset{(b)}{=} N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) + NM \log(\pi e) \\
& \quad - \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, \mathbf{S}[t] = \mathbf{s}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(c)}{=} N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+NM\log(\pi e)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]\,\middle|\,W^{(1)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},\mathbf{S}[t]=\mathbf{s}\right)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\mathbf{Z}_{\mathbf{s}^c}^{(1)}[t]\,\middle|\,\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t],W^{(1)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(d)}{=} N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+NM\log(\pi e)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]\,\middle|\,W^{(1)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},\mathbf{S}[t]=\mathbf{s}\right)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\sum_{j=1}^{M}\mathbb{I}_{j\notin\mathbf{s}}\log(\pi e)\\
&=N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+NM\log(\pi e)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]\,\middle|\,W^{(1)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1}\right)-NM(1-p)\log(\pi e)\\
&\overset{(e)}{\leq} N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+NM\log(\pi e)\\
&\quad-p\sum_{t=1}^{N}h\left(\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,W^{(1)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1}\right)-NM(1-p)\log(\pi e)\\
&\overset{(f)}{=} N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+NM\log(\pi e)\\
&\quad-p\sum_{t=1}^{N}h\left(\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,W^{(1)},\mathbf{V}_{1:t-1}^{(1)}\biguplus\mathbf{Z}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1}\right)-NM(1-p)\log(\pi e),
\end{aligned}
\tag{2.33}
$$

where (a) follows from the proof of (A.4) (see Appendix A.0.3). Step (b) follows from the observation that given $W^{(1)}$, $\mathbf{Y}_{1:t-1}^{(1)}$, $\mathbf{S}_{1:t-1}$ and $\mathbf{S}[t]=\mathbf{s}$, the direct signal in Rx_1 can be determined, and the remaining uncertainty in the output stems from the interfering signal and noise; as defined in Table 2.2 the term $\mathbf{V}_{\mathbf{s}}^{(1)}[t]\uplus\mathbf{Z}^{(1)}[t]$ in step (b) denotes the interfering signal (if present) plus the noise across the M subcarriers. Step (c) follows from splitting the term $\mathbf{V}_{\mathbf{s}}^{(1)}[t]\uplus\mathbf{Z}^{(1)}[t]$ on the basis of interfered subcarriers (leading to the term $\mathbf{V}_{\mathbf{s}}^{(1)}[t]\uplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]$) and interference free subcarriers (leading to the term $\mathbf{Z}_{\mathbf{s}^c}^{(1)}[t]$). Step (d) follows from the fact that the (differential) entropy corresponding to Gaussian noise in a subcarrier is $\log(\pi e)$. Step (e) follows by using the differ-

ential entropy version of the Madiman-Tetali subset inequality (2.26) for the fractional partition defined in (2.27). More precisely, in the above context, we can rewrite the Madiman-Tetali subset (differential) entropy inequality as

$$\sum_{\mathbf{s}} \frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p} h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}_{\mathbf{s}}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}\right) \geq h\left(\tilde{\mathbf{V}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}\right), \tag{2.34}$$

where $\frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p}$ corresponds to the fractional partition defined in (2.27) and $h\left(\tilde{\mathbf{V}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}\right)$ is the corresponding joint entropy term (we defined $\tilde{\mathbf{V}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t]$ in Table 2.2). As shown above, the term involving entropy of subsets, *i.e.*, $\sum_{\mathbf{s}} \frac{\mathbb{P}(\mathbf{S}[t]=\mathbf{s})}{p} h\left(\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}_{\mathbf{s}}^{(1)}[t] \middle| W^{(1)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}\right)$ is bounded by the corresponding joint entropy term. Finally, step (f) follows from the observation that using $W^{(1)}$, $\mathbf{Y}_{1:t-1}^{(1)}$ and $\mathbf{S}_{1:t-1}$ we can determine the direct signal at Rx_1 till time $t-1$ and determine $\mathbf{V}_{1:t-1}^{(1)} \biguplus \mathbf{Z}_{1:t-1}^{(1)}$; at the same time, using $\mathbf{V}_{1:t-1}^{(1)} \biguplus \mathbf{Z}_{1:t-1}^{(1)}$, $W^{(1)}$ and $\mathbf{S}_{1:t-1}$ we can determine $\mathbf{Y}_{1:t-1}^{(1)}$.

Note that step (e) above is a crucial step in the proof since the resulting *interference* term (*i.e.*, the joint entropy term) is effectively removed from the final outer bound using the bound on $R^{(2)}$ described below.

Using Fano's inequality for Rx_2, for any $\varepsilon > 0$, there exists a large enough N such that:

$$\begin{aligned}
&NR^{(2)} - N\varepsilon \\
&\leq I\left(W^{(2)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
&= I\left(W^{(2)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)}, \mathbf{S}_{1:N} \middle| W^{(1)}\right) \\
&= \sum_{t=1}^{N} I\left(W^{(2)}; \mathbf{Y}^{(2)}[t], \mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, W^{(1)}, \mathbf{S}[t]\right) \\
&= \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t]=\mathbf{s})\, I\left(W^{(2)}; \mathbf{Y}^{(2)}[t], \mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, W^{(1)}, \mathbf{S}[t]=\mathbf{s}\right) \\
&\overset{(a)}{=} \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t]=\mathbf{s}) \times
\end{aligned}$$

$$
\begin{aligned}
& I\left(W^{(2)};\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(b)}{=}\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)I\left(W^{(2)};\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\quad+\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\times\\
&\quad I\left(W^{(2)};\mathbf{Z}_{\mathbf{s}^c}^{(1)}[t]\,\middle|\,\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t],\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(c)}{=}\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)I\left(W^{(2)};\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}_{\mathbf{s}}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(d)}{\leq}\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)I\left(W^{(2)};\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&=\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\quad-\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\times\\
&\quad h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},W^{(2)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\overset{(e)}{=}\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(2)},\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)},\mathbf{S}[t]=\mathbf{s}\right)\\
&\quad-2NM\log(\pi e)\\
&\overset{(f)}{\leq}\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)-2NM\log(\pi e)\\
&\overset{(g)}{=}\sum_{t=1}^{N}h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{Y}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)-2NM\log(\pi e)\\
&=\sum_{t=1}^{N}h\left(\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t],\tilde{\mathbf{V}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]\,\middle|\,\mathbf{V}_{1:t-1}^{(1)}\biguplus\mathbf{Z}_{1:t-1}^{(1)},\mathbf{S}_{1:t-1},W^{(1)}\right)-2NM\log(\pi e),
\end{aligned}
\tag{2.35}
$$

where (a) follows from the observation that given $\mathbf{Y}_{1:t-1}^{(2)}$, $\mathbf{Y}_{1:t-1}^{(1)}$, $\mathbf{S}_{1:t-1}$, $W^{(1)}$ and $\mathbf{S}[t]=\mathbf{s}$, we can determine the direct signal at Rx_1 and the interfering signal at Rx_2. Hence, the remaining uncertainty in the output at Rx_1 corresponds to the interfering signal (if present) plus noise (*i.e.*, the term $\mathbf{V}_{\mathbf{s}}^{(1)}[t]\biguplus\mathbf{Z}^{(1)}[t]$), and the remaining uncertainty in the output at Rx_2 corresponds to the direct signal plus noise (*i.e.*, the term $\hat{\mathbf{X}}^{(2)}[t]\biguplus\mathbf{Z}^{(2)}[t]$). Step (b) follows from splitting the term

$\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t]$ on basis of interfered subcarriers (leading to the term $\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}_{\mathbf{s}}^{(1)}[t]$) and interference free subcarriers (leading to the term $\mathbf{Z}_{\mathbf{s}^c}^{(1)}[t]$). Step (c) follows from the independence of $\mathbf{Z}_{\mathbf{s}^c}^{(1)}[t]$ from $W^{(2)}$ despite the conditioning in step (b). Step (d) follows from introducing extra interference (plus noise) terms in $\mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}_{\mathbf{s}}^{(1)}[t]$ leading to the term $\tilde{\mathbf{V}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t]$. Step (e) follows from the observation that given $\mathbf{Y}_{1:t-1}^{(2)}$, $\mathbf{Y}_{1:t-1}^{(1)}$, $\mathbf{S}_{1:t-1}$, $W^{(1)}$, $W^{(2)}$ and $\mathbf{S}[t] = \mathbf{s}$, the remaining uncertainty in the outputs at Rx_1 and Rx_2 stems from the noise. Step (f) follows by removing the conditioning on $\mathbf{Y}_{1:t-1}^{(2)}$. Finally, step (g) follows from $\sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t] = \mathbf{s}) = 1$.

Using inequalities (2.33) and (2.35),

$$
\begin{aligned}
&NR^{(1)} - N\varepsilon + pNR^{(2)} - pN\varepsilon \\
&\leq N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) + NM \log(\pi e) \\
&\quad - p \sum_{t=1}^{N} h\left(\tilde{\mathbf{V}}^{(1)}[t] \Big| \biguplus \mathbf{Z}^{(1)}[t] \Big| W^{(1)}, \mathbf{V}_{1:t-1}^{(1)} \Big| \biguplus \mathbf{Z}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}\right) - NM(1-p)\log(\pi e) \\
&\quad + p \sum_{t=1}^{N} h\left(\hat{\mathbf{X}}^{(2)}[t] \Big| \biguplus \mathbf{Z}^{(2)}[t], \tilde{\mathbf{V}}^{(1)}[t] \Big| \biguplus \mathbf{Z}^{(1)}[t] \Big| \mathbf{V}_{1:t-1}^{(1)} \Big| \biguplus \mathbf{Z}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, W^{(1)}\right) - 2pNM \log(\pi e) \\
&= N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) \\
&\quad + p \sum_{t=1}^{N} h\left(\hat{\mathbf{X}}^{(2)}[t] \Big| \biguplus \mathbf{Z}^{(2)}[t] \Big| \tilde{\mathbf{V}}^{(1)}[t] \Big| \biguplus \mathbf{Z}^{(1)}[t], \mathbf{V}_{1:t-1}^{(1)} \Big| \biguplus \mathbf{Z}_{1:t-1}^{(1)}, \mathbf{S}_{1:t-1}, W^{(1)}\right) - pNM \log(\pi e)
\end{aligned}
\tag{2.36}
$$

$$
\begin{aligned}
&\leq N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) \\
&\quad + p \sum_{t=1}^{N} \sum_{j=1}^{M} h\left(g_{D,j} x_j^{(2)}[t] + z_j^{(2)}[t] \Big| g_{I,j} x_j^{(2)}[t] + z_j^{(1)}[t]\right) - pNM \log(\pi e) \\
&\leq N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) \\
&\quad + p \sum_{t=1}^{N} \sum_{j=1}^{M} \log\left(\pi e \left(1 + \frac{|g_{D,j}|^2}{1 + |g_{I,j}|^2}\right)\right) - pNM \log(\pi e) \\
&= N \sum_{j=1}^{M} (1-p) \log\left(1 + |g_{D,j}|^2\right) + p \log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right) + p \log\left(1 + \frac{|g_{D,j}|^2}{1 + |g_{I,j}|^2}\right)
\end{aligned}
$$

$$\stackrel{(a)}{=} N\left((1+p)\sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right)+p\Delta_G\right), \tag{2.37}$$

where (a) follows from

$\Delta_G = \sum_{j=1}^{M}\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+\log\left(1+\frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right)-2\log\left(1+|g_{D,j}|^2\right)$. As shown in (2.36), the joint entropy term in (2.33) is effectively removed using (2.35). The bound on $pR^{(1)}+R^{(2)}$ follows by symmetry, and this completes the proof of outer bound (2.8).

CHAPTER 3

Opportunistic Interference Management for Multicarrier Systems

3.1 Introduction

In this chapter[1], we study the problem of harnessing bursty interference (opportunistically) when feedback from the receivers is not available. In particular, we focus on parallel 2-user linear deterministic interference channels. The following toy example, based on parallel linear deterministic channels, captures the intuition behind our problem formulation. Consider 2 transmitters (Tx_1 and Tx_2) and 2 receivers (Rx_1 and Rx_2). For $i \in \{1,2\}$, Tx_i has messages for Rx_i and at discrete time index $t \in \{1,2,\ldots,N\}$, Tx_i can transmit 2 bits $[b_1^i(t)\ b_2^i(t)]$. The 2 bits correspond to 2 subcarriers (parallel channels) allocated to each transmitter-receiver pair. Depending on the interference channel realization (stays constant for $t \in \{1,2,\ldots,N\}$), Rx_i receives one of the three possibilities: $[b_1^i(t)\ b_2^i(t)]$, $[b_1^i(t)+b_1^{i'}(t)\ b_2^i(t)]$ and $[b_1^i(t)\ b_2^i(t)+b_2^{i'}(t)]$ (shown in Figure 3.1), where $i, i' \in \{1,2\}$ and $i' \neq i$. The first possibility corresponds to the interference free case (for Rx_i) and the remaining two possibilities correspond to interference from $Tx_{i'}$ (only one of the subcarriers of Rx_i gets interfered). Hence, there are $3 \times 3 = 9$ distinct possibilities for the pair of received values at Rx_1 and Rx_2 over time duration N. The crucial constraint in this setup is that the transmitters do not know *a priori* the interference channel realization. The channel is used N times (time index $t \in \{1,2,\ldots,N\}$) and we have the following (symmetric) rate requirement: ensure base rate R_1 at a receiver when *any* one of the subcarriers (of the receiver) gets interfered and ensure rate $R_0 + R_1$ at a receiver when both subcarriers (of the receiver) are interference free (*i.e.,* opportunistically deliver incremental

[1] Joint work with I-Hsiang Wang and Suhas Diggavi.

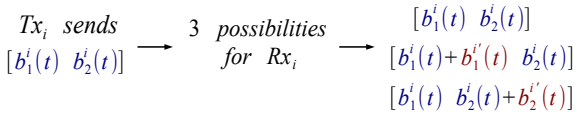


Figure 3.1: Channel realizations for Rx_i in the toy example. The "+" operator denotes modulo 2 addition and indicates the presence of interference. As shown above, interference is not present in all channel realizations for Rx_i (hence bursty); but whenever it is present, it is limited to just 1 out of the 2 transmitted bits.

rate R_0, in addition to R_1, whenever a receiver is interference free). In this setup, we are interested in characterizing the rate region (R_1, R_0) as the performance metric. Clearly, $R_0 \leq 2$ (a maximum of 2 bits per time index can be sent by a transmitter) and corner point $(R_1, R_0) = (0,2)$ is easily achievable. Also, the corner point $(R_1, R_0) = (1,0)$ can be easily achieved by using a repetition code across the 2 subcarriers (*i.e.,* $b_1^1(t) = b_2^1(t)$ and $b_1^2(t) = b_2^2(t)$). The repetition code ensures decodability of the message (of rate R_1) irrespective of which subcarrier gets interfered. Using time sharing between corner points $(0,2)$ and $(1,0)$, we can achieve $2R_1 + R_0 \leq 2$. Intuitively this looks like the best we can do, and indeed it can be shown to be tight using entropy inequalities. The problem pursued in this chapter is a generalization of this example through parallel linear deterministic interference channels (leading to a rate region with more than two non-trivial corner points in most cases).

In [6] and [13], the problem of harnessing bursty interference was studied for a single carrier scenario using a degraded message set approach. This approach guarantees a base rate when the carrier faces interference. In addition to the base rate, an incremental rate is provided whenever the carrier is interference free. In the multicarrier version considered in this chapter, every user (receiver) is allocated M subcarriers (parallel channels) and we extend the degraded message set approach for a rate tuple (R_0, R_L, R_M) as follows: (a) when all M subcarriers of a user get interfered, the user achieves rate R_M (b) when any L out of M subcarriers get interfered, the user achieves rate $R_M + R_L$ and (c) when all M subcarriers are interference free, the user achieves rate $R_M + R_L + R_0$. Thus, the user experiences opportunistic rate increments as the number of interfered subcarriers

decreases. Maintaining low message complexity is the practical idea behind considering the number of interfered subcarriers rather than the specific set of subcarriers interfered. The problem formulation has some similarity with symmetric multilevel diversity coding [14] and our results demonstrate that similar tools (subset entropy inequalities) as in [15] can be used in this context.

The remainder of this chapter is organized as follows. Section 3.2 formalizes the setup and rate requirements. Section 3.3 states the main results. Inner bounds and outer bounds are discussed in Sections 3.4 and 3.5 respectively. We conclude the chapter with a short discussion in Section 3.6.

3.2 Notation and setup

We consider a system with two base stations (transmitters) Tx_1 and Tx_2 and two users (receivers) Rx_1 and Rx_2. For $i \in \{1,2\}$, user Rx_i is allocated M subcarriers $s_1^i, s_2^i, \ldots s_M^i$ by the base station Tx_i. The transmit signals of base stations Tx_1 and Tx_2 are assumed to be independent.

3.2.1 Channel Model

The channel is modeled by a 2-user multicarrier (parallel) linear deterministic interference channel [5] where, similar to [6], interfering links in each subcarrier may or may not be active (unknown to the transmitters). At discrete time index $t \in \{1,2,\ldots N\}$, the transmit signal on subcarrier s_j^i is $\mathbf{x}_j^i(t) \in \mathbb{F}^q$ where $\mathbb{F}$ is a finite field. The received signals on subcarrier s_j^i of Rx_i when s_j^i faces interference from $s_j^{i'}$ (corresponding to user $i' \neq i$) and when it is interference free are described below as (3.1) and (3.2) respectively,

$$\mathbf{y}_j^i(t) = \mathbf{G}^{q-n}\mathbf{x}_j^i(t) + \mathbf{G}^{q-k}\mathbf{x}_j^{i'}(t) \tag{3.1}$$

$$\mathbf{y}_j^i(t) = \mathbf{G}^{q-n}\mathbf{x}_j^i(t) \tag{3.2}$$

where $\mathbf{G}$ is a $q \times q$ shift matrix in the terminology of deterministic channel models [5] and $\mathbf{x}_j^{i'}(t)$ denotes the transmit signal on subcarrier $s_j^{i'}$ for user i'. All operations above are in $\mathbb{F}^q$. Similar to [6], the transmitters are assumed to have prior knowledge of parameters n and k (direct and interfering channel strengths), and the presence (or absence) of interference in a subcarrier is as-

sumed to be constant throughout the channel usage duration. Without loss of generality, we assume $q = \max(n,k)$. Let $\alpha = \frac{k}{n}$ denote the normalized strength of the interfering signal. Since interference free capacity for a single carrier can be achieved when $\alpha \geq 2$ [16], we focus on $0 \leq \alpha \leq 2$. For every time instant, it is convenient to consider a subcarrier as indexed levels of bit pipes. Each bit pipe can carry a symbol from $\mathbb{F}$.

Let $\mathbf{v}_j^i(t) = \mathbf{G}^{q-k}\mathbf{x}_j^{i'}(t)$ denote the interfering signal for Rx_i on subcarrier s_j^i. We use $\mathbf{X}_j^i = [\mathbf{x}_j^i(1)\ \mathbf{x}_j^i(2)...\mathbf{x}_j^i(N)]$ to denote the transmit signals sent during N time slots on s_j^i and $\mathbf{V}_j^i$ is defined similarly from $\mathbf{v}_j^i(t)$. Also, we define $\mathbf{X}_{j_1:j_2}^i = [\mathbf{X}_{j_1}^i \mathbf{X}_{j_1+1}^i, \ldots \mathbf{X}_{j_2}^i]$.

3.2.2 Rate Requirements

The rate requirements for both the users are constrained to be symmetric. For Rx_i, messages (W_0^i, W_L^i, W_M^i) corresponding to rate tuple $(R_0^i, R_L^i, R_M^i) = (R_0, R_L, R_M)$ are encoded in $\mathbf{X}_{1:M}^i$. Based on the number of interfered subcarriers for Rx_i, we have the following requirements for the desired messages:

1. Rx_i decodes W_M^i when all M subcarriers of Rx_i get interfered.
2. Rx_i decodes (W_L^i, W_M^i) when any L out of M subcarriers of Rx_i get interfered.
3. Rx_i decodes (W_0^i, W_L^i, W_M^i) when all M subcarriers of Rx_i are interference free.

A rate tuple is considered achievable if the probability of decoding error is vanishingly small as $N \to \infty$. To simplify our analysis, we consider two setups: $(R_0, R_L, 0)$-setup and $(0, R_L, R_M)$-setup. In the $(R_0, R_L, 0)$-setup, R_M is assumed to be zero and in the $(0, R_L, R_M)$-setup R_0 is assumed to be zero. The rate regions for these two setups are analyzed separately in this chapter.

3.3 Main results

Depending on whether $L \leq \frac{M}{2}$ or $L \geq \frac{M}{2}$, we have different results for $(R_0, R_L, 0)$-setup and $(0, R_L, R_M)$-setup.

3.3.1 Results for $(R_0, R_L, 0)$-setup

3.3.1.1 $L \leq \frac{M}{2}$

We have a tight characterization of capacity in this case.

Theorem 4 *For $L \leq \frac{M}{2}$, the capacity region for $(R_0, R_L, 0)$-setup is as follows.*

$$\begin{aligned} MR_L + (M-L)R_0 &\leq M((M-2L) + L(\max(1,\alpha) \\ &\quad + \max(1-\alpha, 0)))n \qquad (3.3) \\ R_L + R_0 &\leq Mn \qquad (3.4) \end{aligned}$$

3.3.1.2 $L \geq \frac{M}{2}$

In this case, we have a tight characterization in certain regimes.

Theorem 5 *For $L \geq \frac{M}{2}$, consider the following rate inequalities:*

$$\begin{aligned} MR_L &+ (M-L)R_0 \\ &\leq M((M-L)(\max(1-\alpha,0) + \max(1,\alpha)) \\ &\quad + (2L-M)\max(\alpha, 1-\alpha))n \qquad (3.5) \\ R_L + R_0 &\leq Mn \qquad (3.6) \\ 2R_L + R_0 &\leq M(\max(1,\alpha) + \max(1-\alpha,0))n \qquad (3.7) \end{aligned}$$

Inequalities (3.5), (3.6) and (3.7) are inner bounds; (3.5) and (3.6) are outer bounds.

Corollary 3 *We have a tight characterization for the regime $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ in the $(R_0, R_L, 0)$-setup. This follows from the observation that (3.7) is not active in presence of (3.5) and (3.6) for $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ (see Appendix for detailed proof).*

Conjecture 1 *For the $(R_0, R_L, 0)$-setup with $L \geq \frac{M}{2}$, (3.7) is an outer bound.*

If Conjecture 1 holds, we have a tight characterization for $(R_0, R_L, 0)$-setup when $L \geq \frac{M}{2}$.

3.3.2 Results for $(0, R_L, R_M)$-setup

3.3.2.1 $L \leq \frac{M}{2}$

In this case, we have a tight characterization in certain regimes.

Theorem 6 *For $L \leq \frac{M}{2}$, consider the following rate inequalities:*

$$\begin{aligned} R_L + R_M &\leq ((M-2L) + L(\max(1,\alpha) \\ &\quad + \max(1-\alpha, 0)))n \qquad (3.8) \\ R_M &\leq M\max(1-\alpha,\alpha)n \qquad (3.9) \\ MR_L + 2(M-L)R_M &\leq M(M-L)(\max(1,\alpha) \\ &\quad + \max(1-\alpha,0))n \qquad (3.10) \end{aligned}$$

Inequalities (3.8), (3.9) and (3.10) are inner bounds; (3.8) and (3.9) are outer bounds.

Corollary 4 *We have a tight characterization for the regime $\{L \leq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ in the $(0, R_L, R_M)$-setup. This follows from the observation that (3.10) is not active in presence of (3.8) and (3.9) for $\{L \leq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ (see Appendix for detailed proof).*

Conjecture 2 *For the $(0, R_L, R_M)$-setup with $L \leq \frac{M}{2}$, (3.10) is an outer bound.*

If Conjecture 2 holds, we have a tight characterization for $(0, R_L, R_M)$-setup when $L \leq \frac{M}{2}$.

3.3.2.2 $L \geq \frac{M}{2}$

In this case, we have a tight characterization in certain regimes.

Theorem 7 *For $L \geq \frac{M}{2}$, consider the following rate inequalities:*

$$\begin{aligned} R_L + R_M &\leq ((M-L)(\max(1,\alpha) + \max(1-\alpha,0)) \\ &\quad + (2L-M)\max(1-\alpha,\alpha))n \qquad (3.11) \end{aligned}$$

$$R_M \leq M\max(1-\alpha,\alpha)n \tag{3.12}$$

$$R_L + R_M \leq \frac{M}{2}(\max(1,\alpha) + \max(1-\alpha,0))n \tag{3.13}$$

Inequalities (3.11), (3.12) and (3.13) are inner bounds; (3.11) and (3.12) are outer bounds.

Corollary 5 *We have a tight characterization for the regime* $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{2}{3}\}$ *in the* $(0,R_L,R_M)$*-setup. This follows from the observation that (3.13) is not active in presence of (3.11) and (3.12) for* $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{2}{3}\}$ *(see Appendix for detailed proof).*

Conjecture 3 *We conjecture that (3.13) is an outer bound for* $(0,R_L,R_M)$*-setup when* $L \geq \frac{M}{2}$.

If Conjecture 3 holds, we have a tight characterization for $(0,R_L,R_M)$-setup when $L \geq \frac{M}{2}$.

3.4 Inner bounds

Figure 3.3 summarizes the inner bounds for different regimes depending on values of α, M and L. The inner bound rate region is obtained from achievable corner points (shown in Figure 3.3) using time-sharing. Achievability schemes for corner points shown in Figure 3.3 can be described as follows.

3.4.1 Achievable corner points (R_L,R_0) in $(R_0,R_L,0)$-setup

- $(0,Mn)$: This appears in cases (1)-(5) in Figure 3.3. It can be achieved by using the top n levels in all the M subcarriers for message W_0^i.
- $(M(1-\alpha)n, M\alpha n)$: This corner point is achievable for $\alpha \leq 1$ and appears in cases (1)-(3) in Figure 3.3. To achieve this, the top $(1-\alpha)n$ levels of each subcarrier are used for W_L^i and the bottom αn levels are used for W_0^i. Since the top $(1-\alpha)n$ levels of a subcarrier are always interference free, using M subcarriers we achieve $(M(1-\alpha)n, M\alpha n)$.

- $((M-L\alpha)n,0)$: This corner point is achievable for $\alpha \leq 1$ and appears in case (1) in Figure 3.3. Since any L out of M subcarriers get interfered, an erasure code[2] (across M subcarriers) can recover symbols at rate $(M-L)\alpha n$ from the bottom αn levels of M subcarriers. Also, an additive rate of $M(1-\alpha)n$ can be obtained by using the top $(1-\alpha)n$ levels of M subcarriers. Adding the contributions from the bottom αn levels and top $(1-\alpha)n$ levels of all M subcarriers, we achieve $R_L = (M-L)\alpha n + M(1-\alpha)n = (M-L\alpha)n$.

- $(M\alpha n, M(2-3\alpha)n)$ *and* $((M\alpha+(M-L)(2-3\alpha))n,0)$: These appear in case (2) in Figure 3.3 and are achievable for $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$ using the following signal-scale alignment technique [7, 6]. The n levels in a subcarrier s_j^i are divided into 4 bands L_1,L_2,L_3 and L_4 as shown in Figure 3.2. For $i \neq i'$, when subcarrier s_j^i faces interference, only L_1 of $s_j^{i'}$ interferes with L_2 and L_3 of s_j^i. Also, only L_2 of $s_j^{i'}$ interferes with L_4 of s_j^i. Given this structure, the trick will be to not transmit any information in band L_2. This keeps L_4 interference free as shown in Figure 3.2. Using L_1 and L_4 of M subcarriers for W_L^i, we achieve $R_L = M\alpha n$. Using only L_3 of M subcarriers for W_0^i we achieve $R_0 = M(2-3\alpha)n$. Hence $(M\alpha n, M(2-3\alpha)n)$ is achievable. For $((M\alpha+(M-L)(2-3\alpha))n,0)$, the same signal-scale alignment trick is used in addition to a rate $\frac{M-L}{M}$ erasure code across M subcarriers for L_3.

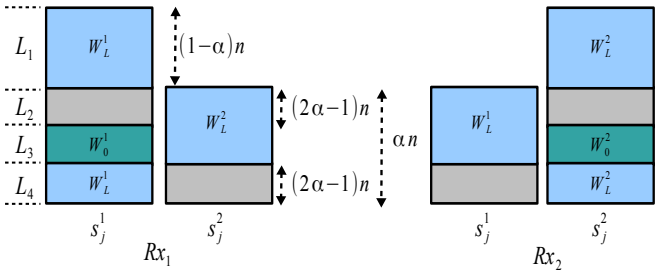


Figure 3.2: Signal-scale alignment technique to achieve $(M\alpha n, M(2-3\alpha)n)$

- $(M(1-\frac{\alpha}{2})n,0)$: This appears in case (3) in Figure 3.3 and is achievable for $\frac{2}{3} \leq \alpha \leq 1$. Han-Kobayashi scheme [17] can achieve rate $(1-\frac{\alpha}{2})n$ for a single interfered subcarrier when $\frac{2}{3} \leq \alpha \leq 1$. This scheme is used for each of the M subcarriers to achieve this corner point.

[2]Interfered levels in the interfered subcarriers are treated as erasures.

- $(M(\alpha-1)n, M(2-\alpha)n)$ *and* $((M-L(2-\alpha))n, 0)$: These are achievable for $1 \leq \alpha \leq 2$. The corner point $(M(\alpha-1)n, M(2-\alpha)n)$ appears in cases (4) and (5) in Figure 3.3 and is achievable using the following signal-scale alignment strategy. The top $(2-\alpha)n$ levels of a subcarrier are used for W_0^i. The next $(\alpha-1)n$ levels are used for W_L^i. This ensures that the levels used for W_L^i are always interference free. Using M subcarriers we achieve, $(M(\alpha-1)n, M(2-\alpha)n)$. To achieve $((M-L(2-\alpha))n, 0)$ (which appears in case (4) in Figure 3.3), a similar scheme is used with a rate $\frac{M-L}{M}$ erasure code (across M subcarriers) for the top $(2-\alpha)n$ levels of a subcarrier.

- $(\frac{M\alpha}{2}n, 0)$: This appears in case (5) in Figure 3.3 and is achievable for $1 \leq \alpha \leq 2$. For the classical two user interference channel (single carrier) with $1 \leq \alpha \leq 2$, rate $\frac{\alpha}{2}n$ is easily achievable. Using this single carrier scheme for M subcarriers, we achieve $(\frac{M\alpha}{2}n, 0)$.

3.4.2 Achievable corner points (R_M, R_L) in $(0, R_L, R_M)$-setup

- $(M(1-\alpha)n, (M-L)\alpha n)$: This appears in cases (6), (7) and (9) in Figure 3.3 and is achievable for $\alpha \leq 1$. Using the top $(1-\alpha)n$ levels of M subcarriers for W_M^i, we achieve $R_M = M(1-\alpha)n$. For W_L^i, a rate $\frac{M-L}{M}$ erasure code is used for the bottom αn levels across M subcarriers to obtain $R_L = (M-L)\alpha n$.

- $(0, (M-L\alpha)n)$: This appears in cases (6), (7) and (9) in Figure 3.3 and is achievable for $0 \leq \alpha \leq 1$. The achievability is same as that of $(R_L, R_0) = ((M-L\alpha)n, 0)$ in the $(R_0, R_L, 0)$-setup.

- $(M\alpha n, (M-L)(2-3\alpha)n)$: This appears in cases (7)-(8) in Figure 3.3 and is achievable for $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$. A signal-scale alignment technique similar to the one in Figure 3.2 is used to achieve $R_M = M\alpha n$. Additionally, a rate $\frac{M-L}{M}$ erasure code across M subcarriers for L_3 is used to achieve $R_L = (M-L)(2-3\alpha)n$.

- $(0, (M\alpha + (M-L)(2-3\alpha))n)$: This appears in case (8) in Figure 3.3 and is achievable for $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$. The achievability is same as that of $(R_L, R_0) = ((M\alpha + (M-L)(2-3\alpha))n, 0)$

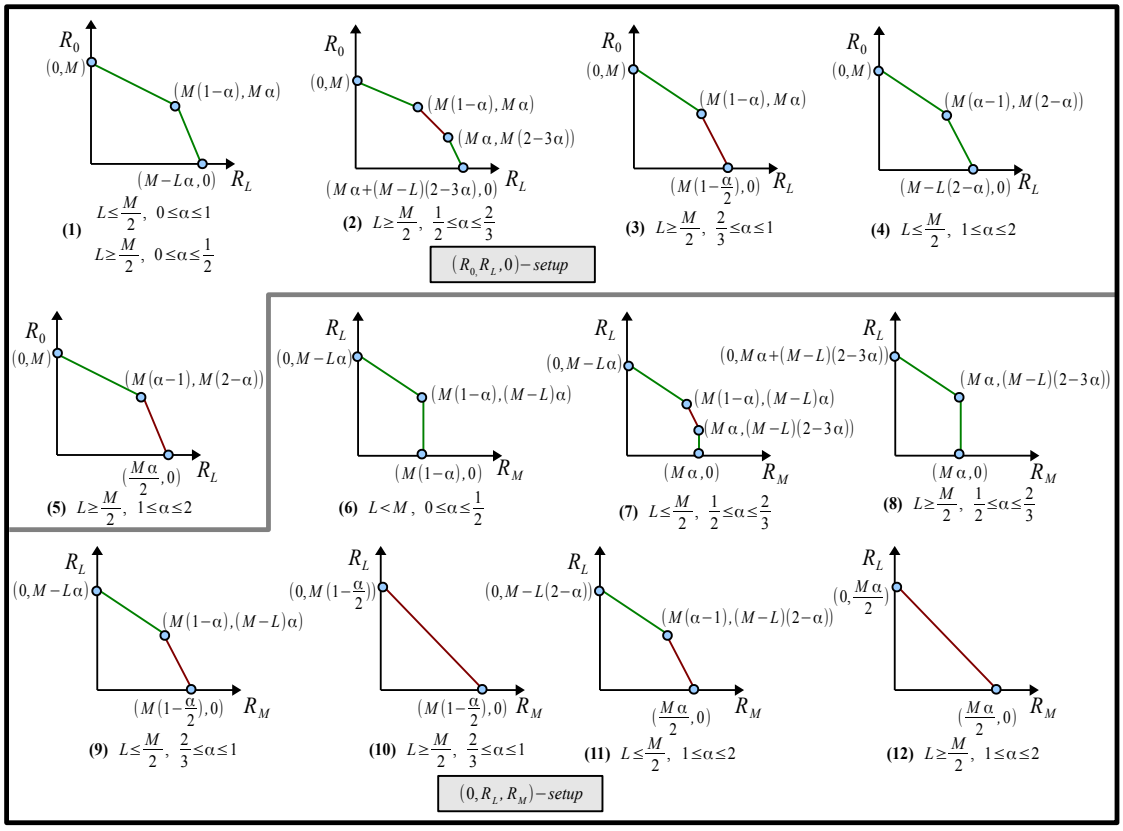


Figure 3.3: Inner bound rate regions for $(0,R_L,R_M)$-setup and $(R_0,R_L,0)$-setup in different regimes. The achievable corner points have been normalized with respect to n and are indicated by blue dots. Lines corresponding to tight outer bounds are colored green and the conjectured outer bounds are colored red.

in the $(R_0,R_L,0)$-setup.

- $(0,M(1-\frac{\alpha}{2})n)$ *and* $(M(1-\frac{\alpha}{2})n,0)$: The corner point $(M(1-\frac{\alpha}{2})n,0)$ appears in cases (9) and (10) while $(0,M(1-\frac{\alpha}{2})n)$ appears in case (10) in Figure 3.3. Both corner points are achievable for $\frac{2}{3}\le\alpha\le1$. To achieve $(0,M(1-\frac{\alpha}{2})n)$, we use the scheme for achieving $(R_L,R_0)=(M(1-\frac{\alpha}{2})n,0)$ in the $(R_0,R_L,0)$-setup (*i.e.*, Han-Kobayashi scheme is used for all the M subcarriers). Also, by using W_M^i instead of W_L^i, the above scheme achieves corner point $(M(1-\frac{\alpha}{2})n,0)$ in the $(0,R_L,R_M)$-setup.

- $(M(\alpha-1)n,(M-L)(2-\alpha)n)$ *and* $(0,(M-L(2-\alpha))n)$: These are achievable for $1 \leq \alpha \leq 2$. The corner point $(M(\alpha-1)n,(M-L)(2-\alpha)n)$ appears in case (11) in Figure 3.3 and is achievable using the following signal-scale alignment strategy. The top $(2-\alpha)n$ levels of a subcarrier are used for W_L^i with a rate $\frac{M-L}{M}$ erasure code across M subcarriers. The next $(\alpha-1)n$ levels are used for W_M^i. This ensures that the levels used for W_M^i are always interference free. Using M subcarriers we achieve, $(M(\alpha-1)n,(M-L)(2-\alpha)n)$. To achieve $(0,M-L(2-\alpha))n,0)$ (which appears in case (11) in Figure 3.3), we use the same scheme as that for $(R_L,R_0)=((M-L(2-\alpha))n,0)$ in the $(R_0,R_L,0)$-setup.

- $(0,\frac{M\alpha}{2}n)$ *and* $(\frac{M\alpha}{2}n,0)$: The corner point $(\frac{M\alpha}{2}n,0)$ appears in cases (11) and (12) while $(0,\frac{M\alpha}{2}n)$ appears in case (12) in Figure 3.3. Both corner points are achievable for $1 \leq \alpha \leq 2$. To achieve $(0,\frac{M\alpha}{2}n)$, we use the scheme for achieving $(R_L,R_0)=(\frac{M\alpha}{2}n,0)$ in the $(R_0,R_L,0)$-setup (case(5) in Figure 3.3). Also, by using W_M^i instead of W_L^i, the above scheme achieves the corner point $(\frac{M\alpha}{2}n,0)$ in the $(0,R_L,R_M)$-setup.

3.5 Outer Bounds

In this section, we first define additional notation for outer bound proofs. This is followed by outer bound proofs for $(R_0,R_L,0)$-setup (which use techniques [15] from multilevel diversity coding) and outer bound proofs for $(0,R_L,R_M)$-setup.

3.5.1 Receiver Configurations

There are $\binom{M}{L}$ ways in which any L out of M subcarriers get interfered. Every such choice is a receiver configuration for a user. We use additional notation for a special set of receiver configurations described below. Consider a circulant matrix $\mathbf{C}_{M,L}$ of dimension M with the first row consisting of $M-L$ consecutive ones followed by L zeros. The other rows are cyclic right shifts of

the first row. As an example, $\mathbf{C}_{3,1}$ is shown below.

$$\mathbf{C}_{3,1} = \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \right)$$

We use $\mathbf{C}_{M,L}$ to list a specific set of receiver configurations in the following manner. Each row corresponds to a receiver configuration with M subcarriers indexed by the columns. In each row, 1 denotes an interference free subcarrier and 0 denotes an interfered subcarrier. Hence, out of $\binom{M}{L}$ choices, $\mathbf{C}_{M,L}$ lists only M receiver configurations. For example, the third row in $\mathbf{C}_{3,1}$ shown above indicates a situation for Rx_i where only subcarrier s_2^i gets interfered. The structure of $C_{M,L}$ corresponds to the choice of receiver configurations we use in some of our outer bound proofs. This structure enables the use of sliding window subset inequality [15] in such proofs.

We now describe additional notation related to receiver configurations of a user. When Rx_i is in receiver configuration indicated by row j of $\mathbf{C}_{M,L}$, we use $\mathscr{Y}^i_{M,L,j}$ to denote the received signal on M subcarriers (over N time slots). In the same spirit, we define $\mathscr{V}^i_{M,L,j}$ as the interfering signal over all M subcarriers for Rx_i in this receiver configuration. The received signal in interference free subcarriers in $\mathscr{Y}^i_{M,L,j}$ is denoted by $\mathscr{X}^i_{M,L,j}$ and the received signal in interfered subcarriers in $\mathscr{Y}^i_{M,L,j}$ is denoted by $\tilde{\mathscr{Y}}^i_{M,L,j}$. When all M subcarriers of Rx_i are interference free, the received signal is denoted by $\mathscr{X}^i_{M,0} = \mathscr{X}^i_{M,0,j}$.

Now, a direct consequence of the sliding window subset inequality [15] in our setting can be stated as follows.

$$\begin{aligned} \sum_{j=1}^{M} H(\mathscr{X}^i_{M,M-1,j}) &\geq \frac{1}{2}\sum_{j=1}^{M} H(\mathscr{X}^i_{M,M-2,j}) \dots \\ \dots &\geq \frac{1}{M}\sum_{j=1}^{M} H(\mathscr{X}^i_{M,0,j}) \end{aligned} \tag{3.14}$$

3.5.2 Outer bounds for $(R_0, R_L, 0)$-setup

3.5.2.1 Proof of outer bound (3.3)

We prove outer bound (3.3) using a careful choice of receiver configurations represented by rows of $\mathbf{C}_{M,L}$. The high level idea is to divide the received signal into interfered and interference free terms followed by the use of (3.14) on the interference free terms. The proof can be described as follows.

Using Fano's inequality for Rx_i $i \in \{1,2\}$, for any $\varepsilon > 0$ there exists a large enough N such that,

$$
\begin{aligned}
& N(MR_L + (M-L)R_0 - (2M-L)\varepsilon) \\
\leq\ & (M-L)I(W_0^i; \mathscr{X}_{M,0}^i | W_L^i) \\
& + \sum_{j=1}^{M} I(W_L^i; \mathscr{X}_{M,L,j}^i \tilde{\mathscr{Y}}_{M,L,j}^i) \\
=\ & (M-L)H(\mathscr{X}_{M,0}^i | W_L^i) - \sum_{j=1}^{M} H(\mathscr{X}_{M,L,j}^i | W_L^i) \\
& + \sum_{j=1}^{M} H(\mathscr{X}_{M,L,j}^i) + \sum_{j=1}^{M} I(W_L^i; \tilde{\mathscr{Y}}_{M,L,j}^i | \mathscr{X}_{M,L,j}^i) \\
\overset{(a)}{\leq}\ & \sum_{j=1}^{M} H(\mathscr{X}_{M,L,j}^i) + \sum_{j=1}^{M} I(W_L^i; \tilde{\mathscr{Y}}_{M,L,j}^i | \mathscr{X}_{M,L,j}^i) \qquad (3.15) \\
\leq\ & \sum_{j=1}^{M} H(\mathscr{X}_{M,L,j}^i) + \sum_{j=1}^{M} H(\tilde{\mathscr{Y}}_{M,L,j}^i) \\
& - \sum_{j=1}^{M} H(\tilde{\mathscr{Y}}_{M,L,j}^i | \mathscr{X}_{M,L,j}^i W_L^i W_0^i) \\
=\ & \sum_{j=1}^{M} H(\mathscr{X}_{M,L,j}^i) \\
& + \sum_{j=1}^{M} H(\tilde{\mathscr{Y}}_{M,L,j}^i) - \sum_{j=1}^{M} H(\mathscr{V}_{M,L,j}^i) \qquad (3.16) \\
\overset{(b)}{\leq}\ & \frac{M-L}{L} \sum_{j=1}^{M} H(\mathscr{X}_{M,M-L,j}^i)
\end{aligned}
$$

$$+\sum_{j=1}^{M} H(\tilde{\mathscr{Y}}^i_{M,L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^i_{M,L,j}) \tag{3.17}$$

(a) follows from (3.14) and (b) follows from $M-L \geq L$ and (3.14). Substituting $i=1$ and $i=2$ in (3.17), we can obtain two inequalities corresponding to different users. On adding these two inequalities,

$$\begin{aligned}
& 2N(MR_L + (M-L)R_0 - (2M-L)\varepsilon) \\
\leq\ & \frac{(M-2L)}{L}\sum_{j=1}^{M} H(\mathscr{X}^1_{M,M-L,j}) + \sum_{j=1}^{M} H(\tilde{\mathscr{Y}}^1_{M,L,j}) \\
& + \frac{(M-2L)}{L}\sum_{j=1}^{M} H(\mathscr{X}^2_{M,M-L,j}) + \sum_{j=1}^{M} H(\tilde{\mathscr{Y}}^2_{M,L,j}) \\
& + (\sum_{j=1}^{M} H(\mathscr{X}^1_{M,M-L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^2_{M,L,j})) \\
& + (\sum_{j=1}^{M} H(\mathscr{X}^2_{M,M-L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^1_{M,L,j})) \\
\overset{(a)}{\leq}\ & 2N(M(M-2L) \\
& + ML\max(1,\alpha) + ML\max(1-\alpha,0))n
\end{aligned} \tag{3.18}$$

where (a) follows from the structure of $\mathbf{C}_{M,L}$.

3.5.2.2 Proof of outer bound (3.5)

The inequality (3.15) in the proof of outer bound (3.3) also holds for $L \geq \frac{M}{2}$. Hence,

$$\begin{aligned}
& N(MR_L + (M-L)R_0 - (2M-L)\varepsilon) \\
\leq\ & \sum_{j=1}^{M} H(\mathscr{X}^i_{M,L,j}) + \sum_{j=1}^{M} I(W^i_L; \tilde{\mathscr{Y}}^i_{M,L,j} | \mathscr{X}^i_{M,L,j}) \\
\leq\ & \sum_{j=1}^{M} H(\mathscr{X}^i_{M,L,j}\tilde{\mathscr{Y}}^i_{M,L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^i_{M,L,j})
\end{aligned} \tag{3.19}$$

Substituting $i=1$ and $i=2$ in (3.19), we can obtain two inequalities corresponding to different users. On adding these two inequalities,

$$2N(MR_L + (M-L)R_0 - (2M-L)\varepsilon)$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{M} H(\mathscr{X}^1_{M,L,j}\tilde{\mathscr{Y}}^1_{M,L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^2_{M,L,j}) \\
&\quad + \sum_{j=1}^{M} H(\mathscr{X}^2_{M,L,j}\tilde{\mathscr{Y}}^2_{M,L,j}) - \sum_{j=1}^{M} H(\mathscr{V}^1_{M,L,j}) \\
&\overset{(a)}{\leq} 2N(M(M-L)(\max(1-\alpha,0)+\max(1,\alpha)) \\
&\quad + M(2L-M)\max(\alpha,1-\alpha))n
\end{aligned} \tag{3.20}
$$

where (a) follows from $L \geq \frac{M}{2}$ and the structure of $\mathbf{C}_{M,L}$.

3.5.3 Outer bounds for $(0, R_L, R_M)$-setup

Outer bounds (3.8), (3.9), (3.11) and (3.12) can be shown by using the El Gamal-Costa injective interference channel bounds [17] as follows.

3.5.3.1 Proof of outer bound (3.8)

For this outer bound proof, we consider two receiver configurations with no interfered subcarriers in common and apply the injective channel bound [17] as shown below.

For any $\varepsilon > 0$ there exists a large enough N such that,

$$
\begin{aligned}
&N(2(R_M+R_L)-2\varepsilon) \\
&\overset{(a)}{\leq} H(\mathscr{Y}^1_{M,L,1}|\mathscr{V}^2_{M,L,L+1}) + H(\mathscr{Y}^2_{M,L,L+1}|\mathscr{V}^1_{M,L,1}) \\
&\leq 2N(L(\max(1,\alpha) \\
&\quad + \max(1-\alpha,0)) + M - 2L)n
\end{aligned} \tag{3.21}
$$

where (a) follows from the injective channel bound [17].

3.5.3.2 Proof of outer bounds (3.9) and (3.12)

Outer bounds (3.9) and (3.12) have the same proof. For the proof, we consider the receiver configuration with all M subcarriers interfered and apply the injective channel bound [17] as shown

below.

For any $\varepsilon > 0$ there exists a large enough N such that,

$$\begin{aligned} & N(2R_M - 2\varepsilon) \\ \overset{(a)}{\leq} \; & H(\mathscr{Y}^1_{M,M,j}|\mathscr{V}^2_{M,M,j}) + H(\mathscr{Y}^2_{M,M,j}|\mathscr{V}^1_{M,M,j}) \\ \leq \; & 2NM\max(1-\alpha,\alpha)n \end{aligned} \tag{3.22}$$

where $\mathscr{Y}^i_{M,M,j}$ corresponds to the receiver configuration with all M subcarriers interfered and (a) follows from the injective channel bound [17].

3.5.3.3 Proof of outer bound (3.11)

For this outer bound proof, we consider two receiver configurations with minimum number of interfered subcarriers (*i.e.,* $2L-M$) in common and apply the injective channel bound [17] as shown below.

For any $\varepsilon > 0$ there exists a large enough N such that,

$$\begin{aligned} & N(2(R_M + R_L) - 2\varepsilon) \\ \overset{(a)}{\leq} \; & H(\mathscr{Y}^1_{M,L,1}|\mathscr{V}^2_{M,L,L+1}) + H(\mathscr{Y}^2_{M,L,L+1}|\mathscr{V}^1_{M,L,1}) \\ \leq \; & 2N((M-L)(\max(1,\alpha) + \max(1-\alpha,0)) \\ & + (2L-M)\max(1-\alpha,\alpha))n \end{aligned} \tag{3.23}$$

where (a) follows from the injective channel bound [17].

3.6 Discussion

It is optimal to treat interference as noise in the regimes where erasure coding across subcarriers leads to tight inner bounds. However, outer bound conjectures on (3.7) and (3.13) (*i.e.,* Conjectures 1 and 3) suggest that this may not be the case for all regimes. For $\alpha = 1$ (and $L > \frac{M}{2}$), both imply $R_L \leq \frac{M}{2}n$; this can be simply achieved by dividing the M subcarriers between the two users. An

erasure coding scheme in this case will lead to $R_L = (M-L)n < \frac{M}{2}n$. Hence, erasure coding across subcarriers may not be optimal in all regimes.

Part II

Secure State Estimation

CHAPTER 4

Secure state estimation and control using multiple (insecure) observers

4.1 Introduction

In this chapter[1], we study the problem of securing state estimation against active adversaries that can attack the software/hardware where state estimation is performed (observer attacks). Against such adversaries, we have two objectives: (i) control the plant correctly despite active attacks, and (ii) prevent the adversary from learning the plant's state. When the state estimates are computed in a single location, an adversary which has access to that location (through hardware or software malware) could use the state estimate for initiating attacks. Therefore, we propose an architecture where state estimation is distributed across several computing nodes (observers), as shown in Figure 4.1 (discussed later in Section 4.2). The challenge is to perform accurate state estimation for controlling the CPS, despite an attacker which has access to a fraction of these observers. In this chapter, we present a solution to this problem and prove that even when ρ out of $3\rho+1$ observers are arbitrarily corrupted, we can still operate the CPS correctly, *i.e.,* we can still control the system as desired and prevent the adversary from learning the state to any desired accuracy.

Our solution is inspired by *secure message transmission* (SMT) [18], a problem studied in cryptography for finite fields. In this problem, a message is securely transmitted between two agents, despite an active adversary who partially controls the communication channels. The main differences in our setup are two-fold: (i) we operate over reals rather than finite fields. This means that it is not possible to give perfect secrecy guarantees and therefore we formulate secrecy as

[1] Joint work with Nikhil Karamchandani, Paulo Tabuada and Suhas Diggavi.

an estimation error guarantee for any adversary. We also give guarantees against a strong active adversary who has complete knowledge of the system parameters and has unbounded power (both transmit and computational). (ii) The SMT problem is posed in a static context, where a given message is to be transmitted. On the other hand, the control and state estimates dynamically change over time in our setup due to the dynamics of a physical plant, and we need to perform these dynamic computations securely. Our techniques are informed by making a connection between our problem and algebraic real error correction (through Reed-Solomon codes [19]) and estimation theory. For simplicity, in this chapter we focus on the case where there is no measurement and actuator noise. However, the ideas can be easily extended for this case, since we ensure that the original state estimate based on the plant output is reconstructed in a distributed and secure manner.

The problem of adversarial attacks in multi-agent networks has been studied in several contexts, for example distributed consensus [20, 21] and function computation [22, 23]. Our goal is not consensus or distributed function computation, but reliable control of a physical plant despite adversarial attacks. Although consensus problems and distributed function computation through linear iterative strategies also involve dynamics, we consider arbitrary linear plants and thus cannot design the dynamics as is possible in these problems. Differential private filtering, studied in [24], consists of a methodology to protect the identity of the nodes contributing information. In our case we seek to protect, not the identity, but the state. In [25] the problem of securing the state of the plant from a passive adversary is studied; in contrast, we allow an active adversary who can also disrupt the legitimate state estimation and control, and our security requirement also differs from their setup.

The remainder of this chapter is organized as follows. Section 4.2 describes the problem setup, system architecture, and notation. Next, we illustrate our key ideas for the case where the adversary attacks a single observer, with Sections 4.3 and 4.4 focusing on a passive and active adversary respectively. In Section 4.5, we extend our results to an active adversary controlling ρ observers and demonstrate how we can operate correctly despite adversarial corruptions when we use at least $3\rho + 1$ observers.

4.2 Notation and Setup

We first describe the model for plant dynamics and then introduce the proposed multiple observer setup. This is followed by the adversary model in the multiple observer setup and the constraints for the plant's operation in the presence of such an adversary.

4.2.1 Plant dynamics

The plant is modeled as a linear time invariant system as shown below:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \tag{4.1}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the plant's state at time t, $\mathbf{u}(t) \in \mathbb{R}^m$ is input to the plant at time t, and $\mathbf{y}(t) \in \mathbb{R}^p$ is the plant's output at time t. For simplicity, in the usual setting (without security constraints), we consider a Luenberger observer [26] for estimating the state of the plant. The Luenberger observer receives the plant's input and output (*i.e.,* $\mathbf{u}(t)$ and $\mathbf{y}(t)$) and uses the following update rule for the state estimate:

$$\hat{\mathbf{x}}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{L}\left(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)\right) \tag{4.2}$$

where $\hat{\mathbf{x}}(t) \in \mathbb{R}^n$ is the observer's state estimate at time t and $\mathbf{L}$ is the observer gain. The state estimate $\hat{\mathbf{x}}(t)$ from the observer is used with the *external* reference command $\mathbf{r}(t) \in \mathbb{R}^m$ (discussed in Section 4.2.5) and a local stabilizing controller with gain matrix K, resulting in the control law:

$$\mathbf{u}(t) = \mathbf{r}(t) + \mathbf{K}\hat{\mathbf{x}}(t). \tag{4.3}$$

In the remainder of this chapter, we will refer to the setup defined by (4.1), (4.2), and (4.3) as the single observer setup.

Throughout the chapter we make the simplifying assumption that the observer estimate $\hat{\mathbf{x}}$ at time $t = 0$ equals the state $\mathbf{x}$ at time $t = 0$. Although counterintuitive, this results in no loss of generality since the secrecy and security guarantees we provide under this assumption extend to the case where $\hat{\mathbf{x}}(0) \neq \mathbf{x}(0)$ (see Appendix A.0.11 for details). Under this assumption, the plant

dynamics can be simplified as follows:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\left(\mathbf{K}\mathbf{x}(t) + \mathbf{r}(t)\right) = \mathbf{A}_{cl}\mathbf{x}(t) + \mathbf{B}\mathbf{r}(t) \tag{4.4}$$

where $\mathbf{A}_{cl} = \mathbf{A} + \mathbf{B}\mathbf{K}$. Without loss of generality, we assume that $\mathbf{x}(0) = \mathbf{0}$ (initial state of the plant). Hence, given a sequence of inputs $\mathbf{r}(0), \mathbf{r}(1), \ldots \mathbf{r}(l-1)$, the sequence of plant states can be written as follows:

$$\begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \vdots \\ \mathbf{x}(l) \end{bmatrix} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{A}_{cl}\mathbf{B} & \mathbf{B} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{A}_{cl}^{l-1}\mathbf{B} & \mathbf{A}_{cl}^{l-2}\mathbf{B} & \ldots & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{r}(0) \\ \mathbf{r}(1) \\ \vdots \\ \mathbf{r}(l-1) \end{bmatrix} = \mathbf{J}_l \mathbf{r}_{0:l-1}. \tag{4.5}$$

As shown above, we use the notation $\mathbf{r}_{t_1:t_2}$ for $\begin{bmatrix} \mathbf{r}^T(t_1) & \mathbf{r}^T(t_1+1) & \ldots & \mathbf{r}^T(t_2) \end{bmatrix}^T$ where T denotes matrix transposition.

4.2.2 Multiple observer setup

In the multiple observer setup, the state observer, as shown in (4.2), is *distributed* among multiple computing nodes. Figure 4.1 shows the multiple observer setup (with d observers). The external

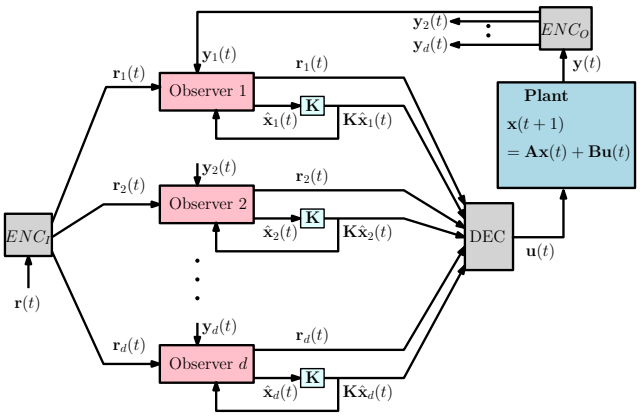


Figure 4.1: A d-observer setup for state estimation.

reference input $\mathbf{r}(t)$ and plant output $\mathbf{y}(t)$ are sent to encoders ENC_I and ENC_O, respectively, as

opposed to being directly sent to the observers. Observer $i \in \{1,2,\ldots,d\}$ receives at time t an encoded version of $\mathbf{r}(t)$, denoted by $\mathbf{r}_i(t)$, from ENC_I and an encoded version of $\mathbf{y}(t)$, denoted by $\mathbf{y}_i(t)$ from ENC_O. In the absence of any adversarial corruptions, the state estimate update rule for observer i is as shown below:

$$\hat{\mathbf{x}}_i(t+1) = \mathbf{A}\hat{\mathbf{x}}_i(t) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_i(t) + \mathbf{r}_i(t)\right) + \mathbf{L}\left(\mathbf{y}_i(t) - \mathbf{C}\hat{\mathbf{x}}_i(t)\right) \tag{4.6}$$

where $\hat{\mathbf{x}}_i(t)$ is the state estimate of observer i at time t. Clearly, the above update rule is similar to (4.2) in the single observer setup; the main difference lies in using $\mathbf{r}_i(t)$ and $\mathbf{y}_i(t)$ instead of $\mathbf{r}(t)$ and $\mathbf{y}(t)$.

In the absence of any adversarial corruptions, the decoder DEC receives $\mathbf{r}_i(t)$ and $\mathbf{K}\hat{\mathbf{x}}_i(t)$ for $i \in \{1,2,\ldots,d\}$ as shown in Figure 4.1. The number d of observers and the design of ENC_I, ENC_O, and DEC is based on the specifications (described in Section 4.2.3) of the adversary who can corrupt a fraction of the observers. We assume that the encoders have access to random number generators and there is no shared randomness between the encoders and the decoder.

4.2.3 Adversary model

We now describe the adversary model in the context of the multiple observer setup described in Section 4.2.2. In this chapter, we consider two types of adversaries: passive and active. The difference between these two types is in the nature of adversarial behavior.

Passive adversary A ρ-passive adversary can tap into any subset of ρ observers in a d-observer setup and access all the inputs to the particular subset of observers. Such adversaries are also referred to as *honest-but-curious* in the cryptography literature since they do not affect the normal operation of a protocol but just try to infer useful information. In the multiple observer setup, the objective of a ρ-passive adversary is to estimate useful information such as the plant's state sequence or the reference input sequence based on inputs to the set of tapped observers.

Active adversary A ρ-active adversary is more powerful than a ρ-passive adversary. It not only has access to all the inputs to the set of affected observers (any ρ observers in a d-observer setup), but can also *inject* errors (of arbitrary magnitude) in the outputs of attacked observers. Furthermore, the adversary can also alter the internal operations (*e.g.,* state estimate update rule) of the attacked observers. Since the outputs from the observers influence the input to the plant, an active adversary can potentially alter the normal operation of the plant.

In both the cases (passive and active), the adversary has unbounded computational power. It also has knowledge of the plant parameters, and the operations done by ENC_I, ENC_O, and DEC. The adversary does not have access to the random number generators in the input encoder (ENC_I) and output encoder (ENC_O); this is essentially the source of secrecy in the multiple observer setup.

4.2.4 Constraints: correctness and secrecy

In a d-observer setup, with initial plant state $\mathbf{x}(0) = \mathbf{0}$ (known to the adversary) and external reference input sequence $\mathbf{r}_{0:l-1}$ (unknown to the adversary) we consider the following constraints:

Correctness The evolution of the plant in the d-observer setup is exactly the same as in the single observer setup; even in the presence of an active adversary which can arbitrarily change the outputs from the set of attacked observers. Formally, for any given input sequence $\mathbf{r}_{0:l-1}$, the plant's state sequence is $\mathbf{x}_{1:l} = \mathbf{J}_l \mathbf{r}_{0:l-1}$ (as shown in (4.5) for the single observer setup with no adversary) despite the attack of an active adversary.

Secrecy An adversary (ρ-active or ρ-passive) having access to the inputs of any ρ observers should have limited knowledge of the external reference input sequence $\mathbf{r}_{0:l-1}$ and plant's state sequence $\mathbf{x}_{1:l}$. Formally, if $\mathbf{E}_{r,0:l-1}$ and $\mathbf{E}_{x,1:l}$ are the error covariance matrices corresponding to the adversary's estimate of $\mathbf{r}_{0:l-1}$ and $\mathbf{x}_{1:l}$, then the following should be satisfied:

$$\mathrm{tr}(\mathbf{E}_{r,0:l-1}) > \phi_r > 0, \quad \mathrm{tr}(\mathbf{E}_{x,1:l}) > \phi_x > 0 \tag{4.7}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace operation, and ϕ_r and ϕ_x are constant design parameters which can be adjusted for any desired level of secrecy. It should be noted that since we assume $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$, $\mathbf{x}(0)$ is known to each observer; but the encoded inputs and encoded outputs are responsible for the observer's uncertainty about $\mathbf{r}_{0:l-1}$ and $\mathbf{x}_{1:l}$.

An important aspect of the d-observer setup is the minimum number d_{min} of observers required to ensure the constraints mentioned above against an adversary (ρ-active or ρ-passive). Clearly, d_{min} depends on ρ, and whether the adversary is active or passive. Using arguments similar to [18], it can be easily shown that $d_{min} \geq \rho + 1$ for a ρ-passive adversary, and $d_{min} \geq 3\rho + 1$ for a ρ-active adversary.

4.2.5 Discussion

The described setup is appropriate for Cyber-Physical Systems that are remotely operated. A case in point are Unmanned Air Vehicles (UAV) where state estimation and the computation of local controllers is performed onboard while the reference input $\mathbf{r}$ is remotely sent by a pilot. Another typical example are SCADA systems where local observers and controllers regulate different physical quantities based on set points that are remotely sent from a central supervisor. In all of these scenarios, we envision attacks on the communication between the operator and the local observers/controllers and between the local observers/controllers and the actuators. We also envision either software or hardware attacks on the observers/local controllers. We exclude from our model attacks on the actuators since an attacker that can command an actuator can immediately cause damage to the plant. Hence, actuators need to be physically hardened to withstand attacks. We also exclude attacks to the operator since in many situations, e.g., UAVs, it is located in a secure facility.

4.3 1-passive adversary

As mentioned in Section 4.2.4, $d_{min} \geq 2$ for a 1-passive adversary. In this section, we show that $d_{min} = 2$ for a 1-passive adversary by designing a 2-observer setup (in Section 4.3.1) and showing

that the correctness and secrecy constraints are satisfied (in Sections 4.3.2 and 4.3.3 respectively).

4.3.1 2-observer setup

The operations of the encoders, observers (indexed by i), and decoder in the 2-observer setup are described below.

Encoder The following operations are done at the input encoder ENC_I which receives $\mathbf{r}(t)$ as input:

$$\mathbf{r}_1(t) = \frac{\mathbf{r}(t)}{2} + \boldsymbol{\theta}(t), \quad \mathbf{r}_2(t) = \frac{\mathbf{r}(t)}{2} - \boldsymbol{\theta}(t) \tag{4.8}$$

where $\boldsymbol{\theta}(t) \in \mathbf{R}^m$ is a random vector drawn from a multivariate Gaussian distribution with zero mean and covariance matrix $\sigma^2\mathbf{I}_m$ ($\mathbf{I}_m$ is the identity matrix of dimension m and σ is a positive real number). In the remainder of this chapter, we use the notation $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. It should be noted that $\boldsymbol{\theta}(t)$ is intentionally generated by the input encoder ENC_I and is i.i.d. (independent and identically distributed) over time. Similarly to ENC_I, the output encoder ENC_O receives $\mathbf{y}(t)$ as input and performs the following operations:

$$\mathbf{y}_1(t) = \frac{\mathbf{y}(t)}{2} + \boldsymbol{\delta}(t), \quad \mathbf{y}_2(t) = \frac{\mathbf{y}(t)}{2} - \boldsymbol{\delta}(t) \tag{4.9}$$

where $\boldsymbol{\delta}(t)$ is intentionally generated random vector $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_p\right)$ and is i.i.d. over time. We justify the use of the Gaussian distribution for $\boldsymbol{\theta}(t)$ and $\boldsymbol{\delta}(t)$ while analyzing the secrecy constraint in Section 4.3.3. Moreover, the random vectors generated by ENC_O and ENC_I are assumed to be independent.

Observer For $i \in \{1,2\}$, observer i receives $\mathbf{r}_i(t)$ and $\mathbf{y}_i(t)$ at time t, and uses update rule (4.6) for its state estimate $\hat{\mathbf{x}}_i(t)$. Recall that we assume that observer i has knowledge of $\mathbf{x}(0)$ and thus sets its initial state estimate as $\hat{\mathbf{x}}_i(0) = \frac{\mathbf{x}(0)}{2}$.

Decoder For $i \in \{1,2\}$, the decoder receives $\mathbf{K}\hat{\mathbf{x}}_i(t)$ and $\mathbf{r}_i(t)$ at time t, and simply adds all its inputs to obtain $\mathbf{u}(t)$ (fed to the plant) as shown below:

$$\mathbf{u}(t) = (\mathbf{K}\hat{\mathbf{x}}_1(t) + \mathbf{r}_1(t)) + (\mathbf{K}\hat{\mathbf{x}}_2(t) + \mathbf{r}_2(t)) \ . \tag{4.10}$$

4.3.2 Correctness

For correctness, given any external reference input sequence $\mathbf{r}_{0:l-1}$, we need the plant's state sequence $\mathbf{x}_{1:l}$ to be exactly as shown in (4.5). We prove the following claim, which is sufficient to show correctness.

Claim 1 *Assuming the operations of ENC_I, ENC_O, DEC, and observers are as described in Section 4.3.1, the following is true for all $t \geq 0$:*

$$\hat{\mathbf{x}}_1(t+1) + \hat{\mathbf{x}}_2(t+1) = \mathbf{A}_{cl}\mathbf{x}(t) + \mathbf{B}\mathbf{r}(t) = \mathbf{x}(t+1). \tag{4.11}$$

Proof 1 *We show this by induction as follows. For the base case $t = 0$:*

$$\begin{aligned}
&\hat{\mathbf{x}}_1(1) + \hat{\mathbf{x}}_2(1) \\
&\overset{(a)}{=} \mathbf{A}\hat{\mathbf{x}}_1(0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_1(0) + \mathbf{r}_1(0)\right) + \mathbf{L}\left(\mathbf{y}_1(0) - \mathbf{C}\hat{\mathbf{x}}_1(0)\right) \\
&\quad + \mathbf{A}\hat{\mathbf{x}}_2(0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_2(0) + \mathbf{r}_2(0)\right) + \mathbf{L}\left(\mathbf{y}_2(0) - \mathbf{C}\hat{\mathbf{x}}_2(0)\right) \\
&\overset{(b)}{=} \mathbf{A}\left(\hat{\mathbf{x}}_1(0) + \hat{\mathbf{x}}_2(0)\right) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_1(0) + \mathbf{K}\hat{\mathbf{x}}_2(0) + \mathbf{r}(0)\right) \\
&\quad + \mathbf{L}\left(\mathbf{y}(0) - \mathbf{C}(\hat{\mathbf{x}}_1(0) + \hat{\mathbf{x}}_2(0))\right) \\
&\overset{(c)}{=} \mathbf{A}\mathbf{x}(0) + \mathbf{B}\left(\mathbf{K}\mathbf{x}(0) + \mathbf{r}(0)\right) + \mathbf{L}\left(\mathbf{y}(0) - \mathbf{C}\mathbf{x}(0)\right) \\
&= \mathbf{A}_{cl}\mathbf{x}(0) + \mathbf{B}\mathbf{r}(0)
\end{aligned} \tag{4.12}$$

where (a) follows from the state estimate update rule (4.6) at the observers, (b) follows from $\mathbf{r}_1(t) + \mathbf{r}_2(t) = \mathbf{r}(t)$ and $\mathbf{y}_1(t) + \mathbf{y}_2(t) = \mathbf{y}(t)$, and (c) follows from $\hat{\mathbf{x}}_1(0) = \hat{\mathbf{x}}_2(0) = \frac{\mathbf{x}(0)}{2}$. The plant's state at time $t = 1$ is as follows:

$$\mathbf{x}(1) = \mathbf{A}\mathbf{x}(0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_1(0) + \mathbf{K}\hat{\mathbf{x}}_2(0) + \mathbf{r}_1(0) + \mathbf{r}_2(0)\right)$$

$$= \mathbf{A}_{cl}\mathbf{x}(0) + \mathbf{B}\mathbf{r}(0)\ . \tag{4.13}$$

Using (4.12) and (4.13), we have $\hat{\mathbf{x}}_1(1) + \hat{\mathbf{x}}_2(1) = \mathbf{A}_{cl}\mathbf{x}(0) + \mathbf{B}\mathbf{r}(0) = \mathbf{x}(1)$. *For the inductive step, we assume that the claim is true up to time* t_0 *and then prove it for time* $t_0 + 1$ *as shown below:*

$$\begin{aligned}
&\hat{\mathbf{x}}_1(t_0+1) + \hat{\mathbf{x}}_2(t_0+1) \\
&= \mathbf{A}\hat{\mathbf{x}}_1(t_0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_1(t_0) + \mathbf{r}_1(t_0)\right) + \mathbf{L}\left(\mathbf{y}_1(t_0) - \mathbf{C}\hat{\mathbf{x}}_1(t_0)\right) \\
&\quad + \mathbf{A}\hat{\mathbf{x}}_2(t_0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_2(t_0) + \mathbf{r}_2(t_0)\right) + \mathbf{L}\left(\mathbf{y}_2(t_0) - \mathbf{C}\hat{\mathbf{x}}_2(t_0)\right) \\
&\overset{(a)}{=} \mathbf{A}\mathbf{x}(t_0) + \mathbf{B}\left(\mathbf{K}\mathbf{x}(t_0) + \mathbf{r}(t_0)\right) + \mathbf{L}\left(\mathbf{y}(t_0) - \mathbf{C}\mathbf{x}(t_0)\right) \\
&= \mathbf{A}_{cl}\mathbf{x}(t_0) + \mathbf{B}\mathbf{r}(t_0)
\end{aligned} \tag{4.14}$$

where (a) follows from the induction hypothesis. Similarly,

$$\begin{aligned}
&\mathbf{x}(t_0+1) \\
&= \mathbf{A}\mathbf{x}(t_0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_1(t_0) + \mathbf{K}\hat{\mathbf{x}}_2(t_0) + \mathbf{r}_1(t_0) + \mathbf{r}_2(t_0)\right) \\
&= \mathbf{A}_{cl}\mathbf{x}(t_0) + \mathbf{B}\mathbf{r}(t_0).
\end{aligned} \tag{4.15}$$

This completes the proof of the claim.

Since $\mathbf{x}(t+1) = \mathbf{A}_{cl}\mathbf{x}(t) + \mathbf{B}\mathbf{r}(t)$, the plant's state sequence $\mathbf{x}_{1:l}$ given input sequence $\mathbf{r}_{0:l-1}$ is exactly as shown in (4.5).

4.3.3 Secrecy

In order to perform the secrecy analysis we start by listing the observations of a 1-passive adversary (we consider the case when the 1-passive adversary taps observer 1; the analysis for observer 2 follows by symmetry). The adversary knows the initial state estimate $\hat{\mathbf{x}}_1(0) = \frac{\mathbf{x}(0)}{2} = \mathbf{0}$ and observes, up to time l, the sequence of encoded reference inputs $\mathbf{r}_1(0), \mathbf{r}_1(1), \ldots \mathbf{r}_1(l-1)$ and the sequence of encoded sensor measurements $\mathbf{y}_1(1), \mathbf{y}_1(2), \ldots \mathbf{y}_1(l)$ fed to observer 1. Hence, the

information available to the adversary can be summarized as the vector $\mathbf{v}_l$:

$$\begin{aligned}
\mathbf{v}_l &= \begin{bmatrix} \mathbf{C}\mathbf{x}(1)+2\boldsymbol{\delta}(1) \\ \mathbf{C}\mathbf{x}(2)+2\boldsymbol{\delta}(2) \\ \vdots \\ \mathbf{C}\mathbf{x}(l)+2\boldsymbol{\delta}(l) \\ \mathbf{r}(0)+2\boldsymbol{\theta}(0) \\ \mathbf{r}(1)+2\boldsymbol{\theta}(1) \\ \vdots \\ \mathbf{r}(l-1)+2\boldsymbol{\theta}(l-1) \end{bmatrix} \\
&\overset{(a)}{=} \begin{bmatrix} (\mathbf{I}_l \otimes \mathbf{C})\,\mathbf{J}_l \\ \mathbf{I}_{ml} \end{bmatrix} \mathbf{r}_{0:l-1} + 2 \begin{bmatrix} \boldsymbol{\delta}_{1:l} \\ \boldsymbol{\theta}_{0:l-1} \end{bmatrix} \\
&= \mathbf{H}_l \mathbf{r}_{0:l-1} + \mathbf{z}_l
\end{aligned} \tag{4.16}$$

where (a) follows from correctness, proved in Section 4.3.2 ($\mathbf{J}_l$ is defined in (4.5)), and $\otimes$ denotes Kronecker product. Equation (4.16) shows that the adversary's observations are affine on the reference input $\mathbf{r}$. The adversary's objective is then to estimate $\mathbf{r}_{0:l-1}$. Note that $\mathbf{z}_l$ is unknown to the adversary although it knows the distribution from which the elements of $\mathbf{z}_l$ are drawn. In this context, there can be several choices for the estimation criterion (*e.g.,* biased or unbiased) [27, 28]. For concreteness, in this chapter we give guarantees on the accuracy of a minimum variance unbiased (MVU) estimate [27] made by the adversary; the guarantees can be easily extended for biased estimators using results in [28]. Given (4.16), the accuracy of the adversary's MVU estimate of $\mathbf{r}_{0:l-1}$ is fundamentally limited by the Cramer-Rao lower bound (CRLB) [27]. The CRLB for the affine model (4.16) can be easily evaluated (see [27] for details) as shown below:

$$\begin{aligned}
\mathbf{E}_{r,0:l-1} &\succeq \left(\mathbf{H}_l^T \boldsymbol{\Sigma}_z^{-1} \mathbf{H}_l\right)^{-1} \\
&= 4\sigma^2 \left(\mathbf{H}_l^T \mathbf{H}_l\right)^{-1}
\end{aligned} \tag{4.17}$$

where $\mathbf{E}_{r,0:l-1}$ is the error covariance matrix for the adversary's MVU estimate of $\mathbf{r}_{0:l-1}$, and $\boldsymbol{\Sigma}_z$ is the covariance matrix of $\mathbf{z}_l$ (in (4.16)). The above result also implies that the trace of $\mathbf{E}_{r,0:l-1}$ is not less than $4\sigma^2 \mathrm{tr}\left(\left(\mathbf{H}_l^T \mathbf{H}_l\right)^{-1}\right)$.

The plant's state sequence $\mathbf{x}_{1:l}$ is the linear function $\mathbf{x}_{1:l} = \mathbf{J}_l \mathbf{r}_{0:l-1}$ of the input sequence. Hence, the CRLB for $\mathbf{x}_{1:l}$ can be derived from the CRLB for $\mathbf{r}_{0:l-1}$ [27] as shown below:

$$\mathbf{E}_{x,1:l} \succeq 4\sigma^2 \mathbf{J}_l \left(\mathbf{H}_l^T \mathbf{H}_l\right)^{-1} \mathbf{J}_l^T. \tag{4.18}$$

Equations (4.17) and (4.18) show that by suitably adjusting σ we can impose any desired lower bound on the accuracy of the reference input and state estimates made by the adversary. Therefore, the secrecy constraint defined in (4.7) is satisfied. As a final remark we note that the Gaussian distribution is the best choice to generate the vector $\mathbf{z}_l$ since it is shown in [29] that it leads to the *worst* CRLB for an MVU estimator.

4.4 1-active adversary

As mentioned in Section 4.2.4, $d_{min} \geq 4$ is necessary for a 1-active adversary. In this section, we show that $d_{min} = 4$ is sufficient for a 1-active adversary by designing a 4-observer setup (in Section 4.4.1) and showing that the correctness and secrecy constraints are satisfied (in Sections 4.4.2 and 4.4.3 respectively).

4.4.1 4-observer setup

The operations of the encoders, observers (indexed by i) and decoder in the 4-observer setup are described below.

Encoders For $i \in \{1,2,3,4\}$, the following operation is done at the input encoder ENC_I which receives $\mathbf{r}(t)$ as input:

$$\mathbf{r}_i(t) = \mathbf{r}(t) + \lambda_i \boldsymbol{\theta}(t) \tag{4.19}$$

where $\boldsymbol{\theta}(t) \in \mathbb{R}^m$ is a random vector $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_m\right)$ generated by ENC_I and is distributed i.i.d. over time. The scaling factor $\lambda_i \in \mathbb{R} - \{0\}$ is the same for all time t for observer i. Similarly, the following operation is done at the output encoder ENC_O:

$$\mathbf{y}_i(t) = \mathbf{y}(t) + \lambda_i \boldsymbol{\delta}(t) \tag{4.20}$$

where $\boldsymbol{\delta}(t) \in \mathbb{R}^p$ is a random vector $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_p\right)$ generated by ENC_O and is distributed i.i.d. over time. The scaling factor λ_i is same as the one used by ENC_I for observer i. The adversary is assumed to have knowledge of the scaling factor λ_i for each observer. Also, $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are assumed to be distinct (needed for proving correctness in Section 4.4.2). The random vectors generated by ENC_I and ENC_O are assumed to be independent.

Observers The operations done at an observer which is not under the influence of an adversary are described below. For $i \in \{1,2,3,4\}$, observer i receives $\mathbf{r}_i(t)$ and $\mathbf{y}_i(t)$ at time t and uses update rule (4.6) for its state estimate $\hat{\mathbf{x}}_i(t)$. Since we assume that $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$ observer i sets its initial state estimate as $\hat{\mathbf{x}}_i(0) = \mathbf{x}(0)$. However, a 1-active adversary can attack any of the observers and arbitrarily change its operation.

Decoder For $i \in \{1,2,3,4\}$, the decoder DEC receives $\tilde{\mathbf{r}}_i(t)$ and $\tilde{\mathbf{k}}_i(t)$ at time t. Under normal operation (with no adversarial errors) $\tilde{\mathbf{r}}_i(t) = \mathbf{r}_i(t)$ and $\tilde{\mathbf{k}}_i(t) = \mathbf{K}\hat{\mathbf{x}}_i(t)$. When an adversary injects errors in the outputs of observer i, the decoder receives $\tilde{\mathbf{r}}_i(t) = \mathbf{r}_i(t) + \mathbf{e}_{i,r}(t)$ and $\tilde{\mathbf{k}}_i(t) = \mathbf{K}\hat{\mathbf{x}}_i(t) + \mathbf{e}_{i,k}(t)$, where $\mathbf{e}_{i,r}(t)$ and $\mathbf{e}_{i,k}(t)$ are errors (of arbitrary magnitude) introduced by the adversary. In this 1-active adversary setting, the decoder does not know a priori which observer is under the adversary's influence. Having received $\tilde{\mathbf{r}}_i(t)$ and $\tilde{\mathbf{k}}_i(t)$, the decoder computes the following for all pairs (i,i') such that $i,i' \in \{1,2,3,4\}$ and $i < i'$:

$$\mathbf{s}_{ii',r}(t) = \frac{\lambda_{i'}}{\lambda_{i'} - \lambda_i}\tilde{\mathbf{r}}_i(t) - \frac{\lambda_i}{\lambda_{i'} - \lambda_i}\tilde{\mathbf{r}}_{i'}(t) \tag{4.21}$$

$$\mathbf{s}_{ii',k}(t) = \frac{\lambda_{i'}}{\lambda_{i'} - \lambda_i}\tilde{\mathbf{k}}_i(t) - \frac{\lambda_i}{\lambda_{i'} - \lambda_i}\tilde{\mathbf{k}}_{i'}(t). \tag{4.22}$$

There are $\binom{4}{2} = 6$ possible $\mathbf{s}_{ii',r}(t)$ and the majority value (most frequently occurring) among these is denoted by $\mathbf{s}_r^*(t)$. Similarly, the majority value for $\mathbf{s}_{ii',k}(t)$ is denoted by $\mathbf{s}_k^*(t)$. We show in Section 4.4.2 that the majority value for both $\mathbf{s}_{ii',r}(t)$ and $\mathbf{s}_{ii',k}(t)$ is always unique (*i.e.,* a tie never occurs). The decoder adds $\mathbf{s}_r^*(t)$ and $\mathbf{s}_k^*(t)$ to obtain $\mathbf{u}(t)$ (fed to the plant) as shown below:

$$\mathbf{u}(t) = \mathbf{s}_r^*(t) + \mathbf{s}_k^*(t). \tag{4.23}$$

4.4.2 Correctness

We first prove the following claim which we use in the proof of correctness.

Claim 2 *Assuming $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are distinct and non-zero, and the operations of ENC_I, ENC_O, DEC and observers are as described in Section 4.4.1, the following are true (even in the presence of a 1-active adversary):*

(a) For time $t \geq 0$, $\mathbf{s}_r^*(t) = \mathbf{r}(t)$.

(b) If observer i is not under the adversary's influence and $\mathbf{K}\hat{\mathbf{x}}_i(t) = \mathbf{K}\mathbf{x}(t) + \lambda_i \mathbf{K}\boldsymbol{\Delta}(t)$ holds at time t, then $\mathbf{s}_k^*(t) = \mathbf{K}\mathbf{x}(t)$.

where $\boldsymbol{\Delta}(t) \in \mathbf{R}^n$ in (b) is arbitrary.

Proof 2 *We first describe the proof of (a) as follows. When the adversary does not inject errors in $\tilde{\mathbf{r}}_i(t)$ (*i.e., *$\tilde{\mathbf{r}}_i(t) = \mathbf{r}_i(t) = \mathbf{r}(t) + \lambda_i \boldsymbol{\theta}(t)$), it is easy to verify that all the 6 possible $\mathbf{s}_{ii',r}(t)$ are equal to $\mathbf{r}(t)$; hence $\mathbf{s}_r^*(t) = \mathbf{r}(t)$. When there is a 1-active adversary, the majority value $\mathbf{s}_r^*(t)$ is still unique and equal to $\mathbf{r}(t)$. To check this, consider the case when a non-zero error $\mathbf{e}_{1,r}(t)$ is introduced by the adversary in $\tilde{\mathbf{r}}_1(t)$ (*i.e., *$\tilde{\mathbf{r}}_1(t) = \mathbf{r}_1(t) + \mathbf{e}_{1,r}(t)$). Due to this error $\mathbf{e}_{1,r}(t)$:*

$$\mathbf{s}_{12,r}(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1}\mathbf{e}_{1,r}(t) + \mathbf{r}(t)$$
$$\mathbf{s}_{13,r}(t) = \frac{\lambda_3}{\lambda_3 - \lambda_1}\mathbf{e}_{1,r}(t) + \mathbf{r}(t)$$
$$\mathbf{s}_{14,r}(t) = \frac{\lambda_4}{\lambda_4 - \lambda_1}\mathbf{e}_{1,r}(t) + \mathbf{r}(t)$$
$$\mathbf{s}_{23,r}(t) = \mathbf{s}_{34,r}(t) = \mathbf{s}_{24,r}(t) = \mathbf{r}(t).$$

Due to having distinct $\lambda_1, \lambda_2, \lambda_3$ and λ_4, $\mathbf{s}_{12,r}(t) \neq \mathbf{s}_{13,r}(t) \neq \mathbf{s}_{14,r}(t)$ while $\mathbf{s}_{23,r}(t), \mathbf{s}_{34,r}(t)$ and $\mathbf{s}_{24,r}(t)$ lead to the majority value $\mathbf{r}(t)$. Similarly, it can be easily verified for the case when the adversary attacks observer $i \in \{2,3,4\}$ that the majority value $\mathbf{s}_r^*(t)$ is unique and equal to $\mathbf{r}(t)$. The proof of (b) is similar to the proof of (a), and we skip it for brevity.

The following claim is sufficient to show correctness.

Claim 3 *Assuming $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are distinct and non-zero, and the operations of ENC_I, ENC_O, DEC and observers are as described in Section 4.4.1, the following are true for time $t \geq 0$:*

(a) $\mathbf{u}(t) = \mathbf{K}\mathbf{x}(t) + \mathbf{r}(t)$.

(b) If observer i is not under the adversary's influence, $\hat{\mathbf{x}}_i(t) = \mathbf{x}(t) + \lambda_i \boldsymbol{\Delta}(t)$. *In addition,* $\boldsymbol{\Delta}(t) \in \mathbf{R}^n$ *satisfies the following:* $\boldsymbol{\Delta}(0) = \mathbf{0}$ *and* $\boldsymbol{\Delta}(t+1) = (\mathbf{A} + \mathbf{B}\mathbf{K} - \mathbf{L}\mathbf{C})\boldsymbol{\Delta}(t) + \mathbf{B}\boldsymbol{\theta}(t) + \mathbf{L}\boldsymbol{\delta}(t)$.

Proof 3 *The proof is by induction. For the base case $t = 0$, if observer i is not under the adversary's influence, $\hat{\mathbf{x}}_i(0) = \mathbf{x}(0)$ and $\boldsymbol{\Delta}(t) = \mathbf{0}$; hence $\hat{\mathbf{x}}_i(0) = \mathbf{x}(0) + \lambda_i \boldsymbol{\Delta}(0)$. Since $\mathbf{K}\hat{\mathbf{x}}_i(0) = \mathbf{K}\mathbf{x}(0) + \lambda_i \mathbf{K}\boldsymbol{\Delta}(0)$ holds in this case, we have $\mathbf{s}_k^*(0) = \mathbf{K}\mathbf{x}(0)$ and $\mathbf{s}_r^*(0) = \mathbf{r}(0)$ (by Claim 2). Hence, $\mathbf{u}(0) = \mathbf{s}_k^*(0) + \mathbf{s}_r^*(0) = \mathbf{K}\mathbf{x}(0) + \mathbf{r}(0)$. This completes proof of claim for the the base case.*

For the inductive step, we assume that the claim is true up to time t_0 and then show that it holds for time $t_0 + 1$. Using the inductive hypothesis $\mathbf{u}(t_0) = \mathbf{K}\mathbf{x}(t_0) + \mathbf{r}(t_0)$,

$$\mathbf{x}(t_0+1) = \mathbf{A}\mathbf{x}(t_0) + \mathbf{u}(t_0) = \mathbf{A}_{cl}\mathbf{x}(t_0) + \mathbf{B}\mathbf{r}(t_0). \tag{4.24}$$

The state estimate update at time t_0 for observer i (which is not under the adversary's influence) is as shown below:

$$\begin{aligned}
&\hat{\mathbf{x}}_i(t_0+1) \\
&\overset{(a)}{=} \mathbf{A}\hat{\mathbf{x}}_i(t_0) + \mathbf{B}\left(\mathbf{K}\hat{\mathbf{x}}_i(t_0) + \mathbf{r}_i(t_0)\right) + \mathbf{L}\left(\mathbf{y}_i(t_0) - \mathbf{C}\hat{\mathbf{x}}_i(t_0)\right) \\
&\overset{(b)}{=} \mathbf{A}\left(\mathbf{x}(t_0) + \lambda_i\boldsymbol{\Delta}(t_0)\right) + \mathbf{B}\left(\mathbf{K}\mathbf{x}(t_0) + \lambda_i\mathbf{K}\boldsymbol{\Delta}(t_0) + \mathbf{r}(t_0) + \lambda_i\boldsymbol{\theta}(t_0)\right) \\
&\quad + \mathbf{L}\left(\mathbf{C}\mathbf{x}(t_0) + \lambda_i\boldsymbol{\delta}(t_0) - \mathbf{C}\mathbf{x}(t_0) - \lambda_i\mathbf{C}\boldsymbol{\Delta}(t_0)\right) \\
&= \mathbf{A}_{cl}\mathbf{x}(t_0) + \mathbf{B}\mathbf{r}(t_0) + \lambda_i\boldsymbol{\Delta}(t_0+1) \\
&\overset{(c)}{=} \mathbf{x}(t_0+1) + \lambda_i\boldsymbol{\Delta}(t_0+1)
\end{aligned} \tag{4.25}$$

where (a) follows from the estimate update rule (4.6) at observer i, (b) follows from the inductive hypothesis $\hat{\mathbf{x}}_i(t_0) = \mathbf{x}(t_0) + \lambda_i\boldsymbol{\Delta}(t_0)$, and (c) follows from (4.24). Also, (4.25) implies $\mathbf{K}\hat{\mathbf{x}}_i(t_0+1) = \mathbf{K}\mathbf{x}(t_0+1) + \lambda_i\mathbf{K}\boldsymbol{\Delta}(t_0+1)$. Hence, we have $\mathbf{s}_k^*(t_0+1) = \mathbf{K}\mathbf{x}(t_0+1)$ and $\mathbf{s}_r^*(t_0+1) = \mathbf{r}(t_0+1)$ (by Claim 2). Using the above results, we have the following:

$$\mathbf{u}(t_0+1) = \mathbf{s}_k^*(t_0+1) + \mathbf{s}_r^*(t_0+1) = \mathbf{K}\mathbf{x}(t_0+1) + \mathbf{r}(t_0+1).$$

This completes the proof of the claim for time t_0+1.

Since $\mathbf{u}(t) = \mathbf{K}\mathbf{x}(t) + \mathbf{r}(t)$ leads to the plant's state sequence shown in (4.5), the correctness constraint is satisfied.

4.4.3 Secrecy

The observations of a 1-active adversary in the 4-observer setup are similar to that of a 1-passive adversary in Section 4.3.3, *i.e.,* observations are in the form of an affine model in the parameter $\mathbf{r}_{0:l-1}$, similar to (4.16)). For an adversary attacking observer i, the CRLB leads to the following bound:

$$\mathbf{E}_{r,0:l-1} \succeq \lambda_i^2 \sigma^2 \left(\mathbf{H}_l^T \mathbf{H}_l\right)^{-1} \tag{4.26}$$

where $\mathbf{E}_{r,0:l-1}$ is the error covariance matrix for adversary's MVU estimate of $\mathbf{r}_{0:l-1}$, and $\mathbf{H}_l$ is as defined in (4.16).

4.5 ρ-active adversary

In this section, we generalize the results in Section 4.4 from a 1-active adversary to a ρ-active adversary. This generalization is based on a class of error correcting codes called Reed-Solomon codes [30, 31, 19]; we briefly describe the idea behind this generalization in Section 4.5.1. We then describe the proposed $3\rho+1$-observer setup (in Section 4.5.2) and prove that it satisfies the correctness and secrecy constraints (in Sections 4.5.3 and 4.5.4 respectively) against a ρ-active adversary. As a result, $d_{min} = 3\rho+1$ for a ρ-active adversary.

4.5.1 Reed-Solomon codes

Consider the following polynomial in $\lambda \in \mathbb{R}$:

$$f(\lambda) = \sum_{j=0}^{\rho} c_j \lambda^j \tag{4.27}$$

with coefficients $c_j \in \mathbb{R}$ and degree at most ρ. For $i \in \{1,2,\ldots,w\}$ let d_i be the evaluation of f at distinct and non-zero points λ_i, *i.e.,* $d_i = f(\lambda_i)$. Clearly, when $w = \rho + 1$, the evaluations $d_1, d_2, \ldots d_w$ are sufficient to reconstruct the polynomial. Finding the value of $c_0 = f(0)$ from $d_1, d_2, \ldots d_w$ involves Lagrange interpolation (a linear combination of $d_1, d_2, \ldots d_w$). Now, consider the problem of finding c_0 from evaluations $d_1, d_2, \ldots d_w$ when any q of the evaluations are erroneous (arbitrarily different from the true evaluation). It can be shown that c_0 can still be recovered in the presence of such erroneous evaluations if the following constraint holds [30]:

$$q < \frac{w-\rho}{2} \, . \tag{4.28}$$

The process of finding c_0 in the above problem is equivalent to finding the polynomial which *fits* the maximum number of evaluations [30]. But the above problem is also the same as decoding a Reed-Solomon code where c_0 is a message symbol and $d_1, d_2, \ldots d_w$ are codeword symbols [30, 31]. In such a setting, the problem translates to decoding the message from codeword symbols despite errors in some of the codeword symbols. This connection primarily offers an intuitive insight into the design and error correcting nature of Reed-Solomon codes.

The above connection provides an alternative interpretation of the decoder's operation in the 4-observer setup in Section 4.4.1. In the absence of adversarial corruptions, the decoder receives $\mathbf{r}_i(t) = \mathbf{r}(t) + \lambda_i \boldsymbol{\theta}(t)$ which is essentially a system of polynomials (of degree at most 1) evaluated at $\lambda = \lambda_i$. Hence, the task of finding $\mathbf{r}(t)$ using evaluations $\tilde{\mathbf{r}}_1(t), \tilde{\mathbf{r}}_2(t), \ldots \tilde{\mathbf{r}}_4(t)$ (with at most one erroneous evaluation) is equivalent to a decoding a Reed-Solomon code. For decoding a Reed-Solomon code, the approach of finding the best fitting polynomial still works but there exist faster methods (*e.g.,* Berlekamp-Welch algorithm [32]) whose time complexity is polynomial in number of evaluations. Similarly, due to the invariant $\hat{\mathbf{x}}_i(t) = \mathbf{x}(t) + \lambda_i \boldsymbol{\Delta}(t)$ (Claim 3) in the absence of adversarial corruptions, the same interpretation (based on polynomial evaluations) can be made for recovering $\mathbf{K}\mathbf{x}(t)$ at the decoder. Hence, the invariants ($\mathbf{r}_i(t) = \mathbf{r}(t) + \lambda_i \boldsymbol{\theta}(t)$ and $\hat{\mathbf{x}}_i(t) = \mathbf{x}(t) + \lambda_i \boldsymbol{\Delta}(t)$) for the 4-observer setup essentially realize a Reed-Solomon code over reals and enable the decoder to recover $\mathbf{r}(t)$ and $\mathbf{K}\mathbf{x}(t)$ despite the presence of a 1-active adversary.

From the perspective of secrecy, following the Reed-Solomon code interpretation, an adver-

sary also gets to observe some of the codewords (*i.e.,* $\mathbf{r}_i(t)$) and tries to estimate the message $\mathbf{r}(t)$. As proved in Section 4.4.3, the presence of intentionally generated random vectors $\boldsymbol{\theta}(t)$ limits the adversary's accuracy in recovering $\mathbf{r}(t)$. We generalize the ideas discussed above for ensuring correctness and secrecy in a $3\rho+1$-observer setup against a ρ-active adversary (by using polynomials of degree at most ρ).

4.5.2 $3\rho+1$-observer setup

The operations of the encoders, observers (indexed by i) and decoder are described below.

Encoders For $i \in \{1,2,\ldots 3\rho+1\}$, the following operation is done at ENC_I which receives $\mathbf{r}(t)$ as input:

$$\mathbf{r}_i(t) = \mathbf{r}(t) + \sum_{j=1}^{\rho} \lambda_i^j \boldsymbol{\theta}_j(t) \tag{4.29}$$

where $\boldsymbol{\theta}_j(t)$ is a random vector $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_m\right)$ generated by ENC_I and is distributed i.i.d. over time. Also, for $j \neq j'$, $\boldsymbol{\theta}_j(t)$ and $\boldsymbol{\theta}_{j'}(t)$ are independent. The scaling factor $\lambda_i \in \mathbb{R} - \{0\}$ is the same for all time t for observer i (and is assumed to be distinct across the observers *i.e.,* $\lambda_i \neq \lambda_{i'}$ where $i \neq i'$). Clearly, $\mathbf{r}_i(t)$ corresponds to the evaluation of $\mathbf{r}(t) + \sum_{j=1}^{\rho} \lambda^j \boldsymbol{\theta}_j(t)$ at $\lambda = \lambda_i$. Similarly, the following operation is done by the output encoder ENC_O:

$$\mathbf{y}_i(t) = \mathbf{y}(t) + \sum_{j=1}^{\rho} \lambda_i^j \boldsymbol{\delta}_j(t) \tag{4.30}$$

where $\boldsymbol{\delta}_j(t)$ is a random vector $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_p\right)$ generated by ENC_O and is distributed i.i.d. over time. Also, for $j \neq j'$, $\boldsymbol{\delta}_j(t)$ and $\boldsymbol{\delta}_{j'}(t)$ are independent. The adversary is assumed to have knowledge of the scaling factor λ_i for each observer. The random vectors generated by ENC_I and ENC_O are assumed to be independent.

Observers The operations done at an observer which is not under the influence of an adversary are described below. For $i \in \{1,2,\ldots 3\rho+1\}$, observer i receives $\mathbf{r}_i(t)$ and $\mathbf{y}_i(t)$ at time t and uses update rule (4.6) for its state estimate $\hat{\mathbf{x}}_i(t)$. Observer i has knowledge of $\mathbf{x}(0)$ and sets its initial

state estimate as $\hat{\mathbf{x}}_i(0) = \mathbf{x}(0)$. A ρ-active adversary can attack any ρ observers and arbitrarily change their operations.

Decoder For $i \in \{1,2,\ldots 3\rho+1\}$, the decoder *DEC* receives $\tilde{\mathbf{r}}_i(t)$ and $\tilde{\mathbf{k}}_i(t)$ at time t. Under normal operation (with no adversarial errors) $\tilde{\mathbf{r}}_i(t) = \mathbf{r}_i(t)$ and $\tilde{\mathbf{k}}_i(t) = \mathbf{K}\hat{\mathbf{x}}_i(t)$. When an adversary injects errors in the outputs of observer i, the decoder receives $\tilde{\mathbf{r}}_i(t) = \mathbf{r}_i(t) + \mathbf{e}_{i,r}(t)$ and $\tilde{\mathbf{k}}_i(t) = \mathbf{K}\hat{\mathbf{x}}_i(t) + \mathbf{e}_{i,k}(t)$, where $\mathbf{e}_{i,r}(t)$ and $\mathbf{e}_{i,k}(t)$ are errors of arbitrary magnitude introduced by the adversary. As a consequence of correctness proved in Section 4.5.3, $\tilde{\mathbf{r}}_i(t)$ and $\tilde{\mathbf{k}}_i(t)$ correspond to evaluations of a system of polynomials (of degree at most ρ) at λ_i; these evaluations are erroneous if observer i is attacked by an adversary. The decoder does not know a priori which ρ (out of $3\rho+1$) observers are under the adversary's influence. In this setting, the task of finding $\mathbf{r}(t)$ and $\mathbf{Kx}(t)$ is equivalent to decoding a Reed-Solomon code. For the sake of clarity, we describe below a decoding method which finds the best fitting polynomial to decode the underlying Reed-Solomon code[2]. Having received $\tilde{\mathbf{r}}_i(t)$ and $\tilde{\mathbf{k}}_i(t)$ from all the $3\rho+1$ observers, the decoder computes the following for all $\rho+1$-tuples $\left(i_1, i_2, \ldots i_{\rho+1}\right)$ such that $i_1, i_2, \ldots i_{\rho+1} \in \{1,2,\ldots 3\rho+1\}$ and $i_1 < i_2 < \ldots < i_{\rho+1}$:

$$\begin{aligned}&\mathbf{s}_{i_1 i_2 \ldots i_{\rho+1},r}(t)\\ &= POLY\left(0, \tilde{\mathbf{r}}_{i_1}(t), \tilde{\mathbf{r}}_{i_2}(t), \ldots, \tilde{\mathbf{r}}_{i_{\rho+1}}(t), \lambda_{i_1}, \lambda_{i_2} \ldots \lambda_{i_{\rho+1}}\right) \qquad (4.31)\\ &\mathbf{s}_{i_1 i_2 \ldots i_{\rho+1},k}(t)\\ &= POLY\left(0, \tilde{\mathbf{k}}_{i_1}(t), \tilde{\mathbf{k}}_{i_2}(t), \ldots, \tilde{\mathbf{k}}_{i_{\rho+1}}(t), \lambda_{i_1}, \lambda_{i_2} \ldots \lambda_{i_{\rho+1}}\right) \qquad (4.32)\end{aligned}$$

where $POLY\left(b_0, \mathbf{d}_1, \mathbf{d}_2, \ldots \mathbf{d}_{\rho+1}, \lambda_{i_1}, \lambda_{i_2} \ldots \lambda_{i_{\rho+1}}\right)$ considers the function $\mathbf{f}(\lambda) = \sum_{j=0}^{\rho} \mathbf{c}_j \lambda^j$ (where λ is a scalar parameter and $\mathbf{c}_j$ are vector coefficients) and finds the value of $\mathbf{f}(b_0)$ by assuming that $\mathbf{f}(\lambda_{i_1}) = \mathbf{d}_1, \mathbf{f}(\lambda_{i_2}) = \mathbf{d}_2, \ldots \mathbf{f}\left(\lambda_{i_{\rho+1}}\right) = \mathbf{d}_{\rho+1}$. This process simply involves taking a linear combination of $\mathbf{d}_1, \mathbf{d}_2, \ldots \mathbf{d}_{\rho+1}$ (*i.e.,* Lagrange interpolation).

There are $\binom{3\rho+1}{\rho+1}$ possible $\mathbf{s}_{i_1 i_2 \ldots i_{\rho+1},r}(t)$ and the majority[3] value (most frequently occurring)

[2]This is similar to the decoder's operation in the 4-observer setup in Section 4.4.1 but can be replaced by faster Reed-Solomon decoding methods like Berlekamp-Welch [32].

[3]It can be shown that the majority value is always unique in this case; a tie never occurs.

among these is denoted by $\mathbf{s}_r^*(t)$. Similarly, the majority value for $\mathbf{s}_{i_1 i_2 \ldots i_{\rho+1},k}(t)$ is denoted by $\mathbf{s}_k^*(t)$. The decoder adds $\mathbf{s}_r^*(t)$ and $\mathbf{s}_k^*(t)$ to obtain $\mathbf{u}(t)$ (fed to the plant) as shown below:

$$\mathbf{u}(t) = \mathbf{s}_r^*(t) + \mathbf{s}_k^*(t). \tag{4.33}$$

4.5.3 Correctness

For correctness of the $3\rho+1$-observer setup, we need to show that despite the presence of ρ active adversaries, the plant's state sequence $\mathbf{x}_{1:l}$ is exactly as shown in (4.5) for any external reference input sequence $\mathbf{r}_{0:l-1}$. The proof follows along the same lines as that for a 1-active adversary in Section 4.4.2, and utilizes the connection to Reed-Solomon codes described in Section 4.5.1. We skip the details here for brevity.

4.5.4 Secrecy

For the secrecy analysis, consider the case when the adversary attacks observers $a_1, a_2, \ldots a_\rho \in \{1, 2, \ldots 3\rho+1\}$. By listing all the inputs to the observers under the adversary's influence, the observations of the adversary form an affine model (similar to (4.16)) with parameter $\mathbf{r}_{0:l-1}$. The CRLB for the MVU estimator for $\mathbf{r}_{0:l-1}$ in this case is as shown below:

$$\mathbf{E}_{r,0:l-1} \succeq \frac{\sigma^2 \left(\mathbf{H}_l^T \mathbf{H}_l\right)^{-1}}{\boldsymbol{\eta} \left(\Lambda \Lambda^T\right)^{-1} \boldsymbol{\eta}^T} \tag{4.34}$$

where $\mathbf{E}_{r,0:l-1}$ is the error covariance matrix for the MVU estimate of $\mathbf{r}_{0:l-1}$, $\boldsymbol{\eta} = [1\ 1\ \ldots 1]$, $\mathbf{H}_l$ is as defined in (4.16) and matrix Λ is as shown below:

$$\Lambda = \begin{bmatrix} \lambda_{a_1} & \lambda_{a_1}^2 & \ldots & \lambda_{a_1}^\rho \\ \lambda_{a_2} & \lambda_{a_2}^2 & \ldots & \lambda_{a_2}^\rho \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{a_\rho} & \lambda_{a_\rho}^2 & \ldots & \lambda_{a_\rho}^\rho \end{bmatrix}. \tag{4.35}$$

Remark 3 *It can be shown that without secrecy constraints, $d_{min} = 2\rho+1$ against a ρ-active adversary; if $2\rho+1 < d < 3\rho+1$, secrecy against $d-(2\rho+1)$ compromised observers can still*

be guaranteed. This can be realized by using a polynomial of degree $d-(2\rho+1)$ *(instead of a polynomial of degree* ρ *as done in the case* $d=3\rho+1$*).*

CHAPTER 5

Secure State Estimation against Sensor Attacks in the Presence of Noise

5.1 Introduction

In this chapter[1], we focus on securely estimating the state of a linear dynamical system from a set of noisy and maliciously corrupted sensor measurements. We restrict the sensor attacks to be sparse in nature, *i.e.*, an adversary can arbitrarily corrupt an unknown subset of sensors in the system but is restricted by an upper bound on the number of attacked sensors.

Several recent works have studied the problem of secure state estimation against sensor attacks in linear dynamical systems. For setups with no noise in sensor measurements, the results reported in [33, 34, 35, 36] show that, given a strong notion of observability, (sparse) sensor attacks can always be detected and isolated, and we can exactly estimate the state of the system. However, with noisy sensors, it is not trivial to distinguish between the noise and the attacks injected by an adversary. Prior work on state estimation with sensor attacks in the presence of noise can be broadly divided into two categories depending on the noise model: 1) bounded non-stochastic noise, and 2) Gaussian noise. Results reported in [37, 38, 39] deal with bounded non-stochastic noise. Though they provide sufficient conditions for distinguishing the sparse attack vector from bounded noise, they do not guarantee the optimality of their estimation algorithm. The problem we focus on in this chapter falls in the second category, *i.e.*, sensor attacks in the presence of Gaussian noise. Prior work in this category includes [40, 41, 42, 43]. In [40], the focus is on detecting a class of sensor attacks called *replay* attacks where the attacker replaces legitimate sensor outputs with

[1] Joint work with Yasser Shoukry, Nikhil Karamchandani, Paulo Tabuada and Suhas Diggavi.

outputs from previous time instants. In [41], the performance degradation of a scalar Kalman filter (*i.e.*, scalar state and a single sensor) is studied when the (single) sensor is under attack. They do not study attack sparsity across multiple sensors, and in addition, they focus on an adversary whose objective is to degrade the estimation performance without being detected (leading to a restricted class of sensor attacks). In [42] and [43], robustification approaches for state estimation against sparse sensor attacks are studied. However, they lack optimality guarantees against arbitrary sensor attacks.

In this chapter, we study a general linear dynamical system with process and sensor noises having a Gaussian distribution, and give (optimal) guarantees on the achievable state estimation error against arbitrary sensor attacks. The following toy example is illustrative of the nature of the problem addressed in this chapter and some of the ideas behind our solution.

Example 4 *Consider a linear dynamical system with a scalar state $x(t)$ such that $x(t+1) = x(t) + w(t)$, and three sensors (indexed by $d \in \{1,2,3\}$) with outputs $y_d(t) = x(t) + v_d(t)$; where $w(t)$ and $v_d(t)$ are the process noise and sensor noise at sensor d respectively. The process and sensor noises follow a zero mean Gaussian distribution with i.i.d. instantiations over time. The sensor noise is also independent across sensors. Now, consider an adversary which can attack any one of the sensors in the system and arbitrarily change its output. In the absence of sensor noise, it is trivial to detect such an attack since the two good sensors (not attacked by the adversary) will have the same output. Hence, a majority based rule on the outputs leads to the exact state. However, in the presence of sensor noise, a difference in outputs across sensors can also be attributed to the noise, and thus cannot be considered an attack indicator. As a consequence of results in this chapter, in this example we can identify a subset of two sensors which can be reliably used for state estimation despite an adversary who can attack any one of the three noisy sensors. In particular, our approach for this example would be to search for a subset of two sensors which satisfy the following check: over a large enough time window, the outputs from the two sensors are consistent with the Kalman state estimate based on outputs from the same subset of sensors. Furthermore, we can show that such an approach leads to the optimal state estimation error for the given adversarial setup.*

In this chapter, we generalize the Kalman filter based approach in the above example to a general linear dynamical system with sensor and process noise. The Kalman estimate based check mentioned in the above example forms the basis of a detector for an *effective* attack; a notion that we introduce in this chapter. For state estimation, we search for a sensor subset which passes such an effective attack detector, and then use outputs from such a sensor subset for state estimation. We also derive impossibility results (lower bounds) on the state estimation error in our adversarial setup, and show that our proposed state estimation algorithm is optimal in the sense that it achieves these lower bounds. To further reduce the sensor subset search time for the state estimator, we propose Satisfiability Modulo Theory (SMT) based techniques to harness the combinatorial nature of the search problem, and demonstrate the improvements in search time through numerical experiments.

As a result of independent interest, we give a coding theoretic interpretation (alternate proof) for the necessary and sufficient conditions for secure state estimation in the absence of noise [34, 35, 38] (known as the sparse observability condition). In particular, we relate the sparse observability condition required for attack detection and secure state estimation in dynamical systems to the Hamming distance requirements for error detection and correction [19] in classical coding theory.

The remainder of this chapter is organized as follows. Section 5.2 deals with the setup and problem formulation. In Section 5.3, we describe our effective attack detector followed by Section 5.4 on our main results for effective attack detection and secure state estimation. Section 5.5 deals with SMT based techniques and Section 5.6 with the experimental results. Finally, Section 5.7 describes the coding theoretic view for attack detection and secure state estimation.

5.2 Setup

In this section, we discuss the adversarial setup along with assumptions on the underlying dynamical system, and provide a mathematical formulation of the state estimation problem considered in this chapter.

5.2.1 Notation

The symbols $\mathbb{N}, \mathbb{R}$ and $\mathbb{B}$ denote the sets of natural, real, and Boolean numbers respectively. The symbol $\wedge$ denotes the logical AND operator. The support of a vector $\mathbf{x} \in \mathbb{R}^n$, denoted by $\text{supp}(\mathbf{x})$, is the set of indices of the non-zero elements of $\mathbf{x}$. If $\mathbf{s}$ is a set, $|\mathbf{s}|$ is the cardinality of $\mathbf{s}$. For the matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, unless stated otherwise, we denote by $\mathbf{M}_i \in \mathbb{R}^{1 \times n}$ the ith row of the matrix. For the set $\mathbf{s} \subseteq \{1, \ldots, m\}$, we denote by $\mathbf{M}_\mathbf{s} \in \mathbb{R}^{|\mathbf{s}| \times n}$ the matrix obtained from $\mathbf{M}$ by removing all the rows except those indexed by $\mathbf{s}$. We use $tr(\mathbf{M})$ to denote the trace of the matrix $\mathbf{M}$. If the matrix $\mathbf{M}$ is symmetric, we use $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ to denote the minimum and maximum eigenvalue of $\mathbf{M}$ respectively. We denote by $\mathbb{S}^n_+$ the set of all $n \times n$ positive semi-definite matrices. For a random variable $\mathbf{x} \in \mathbb{R}^n$, we denote its mean by $\mathbb{E}(\mathbf{x}) \in \mathbb{R}$ and its covariance by $Var(\mathbf{x}) \in \mathbb{S}^n_+$. For a discrete time random process $\{\mathbf{x}(t)\}_{t \in \mathbb{N}}$, the sample average of $\mathbf{x}$ using N samples starting at time t_1 is defined as follows:

$$\mathbb{E}_{N,t_1}(\mathbf{x}) = \frac{1}{N} \sum_{t=t_1}^{t_1+N-1} \mathbf{x}(t). \tag{5.1}$$

We denote by $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ and $\mathbf{1}_m \in \mathbb{R}^{m \times 1}$ the identity matrix of dimension m and the vector of all ones respectively. The notation $\mathbf{x}(t) \sim \mathcal{N}(\mu, \Omega)$ is used to denote an i.i.d. Gaussian random process with mean μ and covariance matrix Ω. Finally, we use the symbol $\preccurlyeq$ for element-wise comparison between matrices. That is, for two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size, $\mathbf{A} \preccurlyeq \mathbf{B}$ is true if and only if each element $a_{i,j}$ is smaller than or equal to $b_{i,j}$.

5.2.2 System model

We consider a linear dynamical system $\mathbf{\Sigma}_a$ with sensor attacks as shown below:

$$\mathbf{\Sigma}_a \begin{cases} \mathbf{x}(t+1) & = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{w}(t), \\ \mathbf{y}(t) & = \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t) + \mathbf{a}(t), \end{cases} \tag{5.2}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the state of the plant at time $t \in \mathbb{N}$, $\mathbf{u}(t) \in \mathbb{R}^m$ denotes the input at time t, $\mathbf{w}(t) \sim \mathcal{N}\left(\mathbf{0}, \sigma_w^2 \mathbf{I}_n\right)$ denotes the process noise at time t, $\mathbf{y}(t) \in \mathbb{R}^p$ denotes the output of the plant at time t and $\mathbf{v}(t) \sim \mathcal{N}\left(\mathbf{0}, \sigma_v^2 \mathbf{I}_p\right)$ denotes the sensor noise at time t. Both $\mathbf{v}(t)$ and $\mathbf{w}(t)$ have

i.i.d. instantiations over time, and $\mathbf{v}(t)$ is independent of $\mathbf{w}(t)$. In addition, we denote the output and (sensor) noise at sensor $i \in \{1,2,\ldots,p\}$ at time t as $y_i(t) \in \mathbb{R}$ and $v_i(t) \in \mathbb{R}$ respectively. We assume that the input $\mathbf{u}(t)$ is known at all time. Hence, its contribution to the output $\mathbf{y}(t)$ is also known, and therefore, $\mathbf{u}(t)$ can be ignored. That is, for the rest of the chapter, and without loss of generality, we consider the case of $\mathbf{u}(t) = 0$ for all time $t \in \mathbb{N}$.

The sensor attack vector $\mathbf{a}(t) \in \mathbb{R}^p$ in (5.2) is introduced by a k-adversary defined as follows.

Assumption 1 *A k-adversary can corrupt any k out of the p sensors in the system.*

Specifically, let $\boldsymbol{\kappa} \subseteq \{1,2,\ldots,p\}$ denote the set of attacked sensors (with $|\boldsymbol{\kappa}| = k$). The k-adversary can observe the actual outputs in the k attacked sensors and change them arbitrarily. For an attack free sensor $j \notin \boldsymbol{\kappa}$, $\mathbf{a}_j(t) = 0$ for all time $t \in \mathbb{N}$.

Assumption 2 *The adversary's choice of* $\boldsymbol{\kappa}$ *is unknown but is assumed to be constant over time (static adversary).*

Assumption 3 *The adversary is assumed to have unbounded computational power, and knows the system parameters (*e.g., $\mathbf{A}$ *and* $\mathbf{C}$*) and noise statistics (*e.g., σ_w^2 *and* σ_v^2*).*

However, the adversary is limited to have only causal knowledge of the process and sensor noise as stated by the following two assumptions.

Assumption 4 *The adversary's knowledge at time t is statistically independent of* $\mathbf{w}(t')$ *for* $t' > t$*, i.e.,* $\mathbf{a}(t)$ *is statistically independent of* $\{\mathbf{w}(t')\}_{t'>t}$.

Assumption 5 *For an attack-free sensor* $i \in \{1,2,\ldots,p\} \setminus \boldsymbol{\kappa}$*, the adversary's knowledge at time t (and hence* $\mathbf{a}(t)$*) is statistically independent of* $\{v_i(t')\}_{t'>t}$.

Intuitively, Assumptions 4 and 5 limit the adversary to have only causal knowledge of the process noise and the sensor noise in *good* sensors (not attacked by the adversary). Note that, apart from Assumptions 4 and 5, we do not impose any restrictions on the statistical properties, boundedness and the time evolution of the corruptions introduced by the k-adversary.

In the following subsections, we first introduce the (effective) attack detection problem, followed by the (optimal) secure state estimation problem. As we show later in the chapter (in Section 5.4), our solution for the effective attack detection problem is used as a crucial component for solving the secure state estimation problem.

5.2.3 Effective Attack Detection Problem

In this section, we introduce our notion of effective (sensor) attacks and formulate the problem of detecting them. Recall that in the absence of sensor attacks, using a Kalman filter for estimating the state in (5.2) leads to the (optimal) minimum mean square error (MMSE) covariance asymptotically [44]. In this context, our notion of effective attacks is based on the following intuition: if we naively use a Kalman filter for state estimation in the presence of an adversary, an attack is effective when it causes a higher empirical error variance compared to the attack-free case. Before we formally state our definition of effective attacks, we first setup some notation for Kalman filters as described below.

We denote by $\hat{\mathbf{x}}_{\mathbf{s}}(t)$ the state estimate of a Kalman filter at time t using outputs till time $t-1$ from the sensor subset $\mathbf{s} \subseteq \{1,2,\ldots,p\}$. Since we use outputs till time $t-1$, we essentially use the *prediction* version of a Kalman filter as opposed to *filtering* where outputs till time t are used to compute $\hat{\mathbf{x}}_{\mathbf{s}}(t)$. In this chapter, we state our results using the prediction version of the Kalman filter; the extension for the filtering version is straightforward (for details about the filtering version of our results, see Appendix A.0.14). In addition to $\hat{\mathbf{x}}_{\mathbf{s}}(t)$, we denote by $\hat{\mathbf{x}}^{\star}_{\mathbf{s}}(t)$ the Kalman filter state estimate at time t using sensor subset $\mathbf{s}$ when all the sensors in $\mathbf{s}$ are attack-free. We eliminate the subscript $\mathbf{s}$ from the previous notation whenever the Kalman filter uses all sensor measurements, *i.e.*, when $\mathbf{s} = \{1,\ldots,p\}$. In this chapter, for the sake of simplicity, we assume that all the Kalman filters we consider (in our proposed algorithms and their analysis) are in steady state [44] when they use uncorrupted sensor outputs. Hence, in the absence of attacks, the error covariance matrix $\mathbf{P}^{\star}(t) \in \mathbb{S}_n^+$ defined as:

$$\mathbf{P}^{\star}(t) = \mathbf{P}^{\star} = \mathbb{E}\Big(\left(\mathbf{x}(t) - \hat{\mathbf{x}}^{\star}(t)\right) \left(\mathbf{x}(t) - \hat{\mathbf{x}}^{\star}(t)\right)^T \Big),$$

does not depend on time. In a similar spirit, we define the error covariance matrix $\mathbf{P}_{\mathbf{s}}^{\star} \in \mathbb{S}_n^+$ corresponding to sensor subset $\mathbf{s} \subseteq \{1,2,\ldots,p\}$ as:

$$\mathbf{P}_{\mathbf{s}}^{\star} = \mathbb{E}(\mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}^{\star}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}^{\star}(t))^T.$$

Note that the error covariance matrix depends on the set of sensors involved in estimating the state. Also, the steady state error has zero mean, *i.e.*, $\mathbb{E}\left(\mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}^{\star}(t)\right) = 0$. Using the above notation, we define an $(\varepsilon, \mathbf{s})$-effective attack as follows.

Definition 1 ($(\varepsilon, \mathbf{s})$-Effective Attack) *Consider the linear dynamical system under attack $\mathbf{\Sigma_a}$ as defined in* (5.2)*, and a k-adversary satisfying Assumptions 1-5. For the set of sensors* $\mathbf{s}$*, an $\varepsilon > 0$, and a large enough $N \in \mathbb{N}$, an attack signal is called $(\varepsilon, \mathbf{s})$-effective at time t_1 if the following bound holds:*

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e_s}\mathbf{e_s}^T\right)\right) > tr(\mathbf{P}_{\mathbf{s}}^{\star}) + \varepsilon,$$

where $\mathbf{e_s}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}(t)$, and $\mathbb{E}_{N,t_1}(\cdot)$ denotes the sample average as defined (5.1).

In other words, an attack is called $(\varepsilon, \mathbf{s})$-effective if it can lead to a higher estimation error compared to the optimal estimation error in the absence of sensor attacks, using the same set of sensors $\mathbf{s}$. An attack is called $(\varepsilon, \mathbf{s})$-ineffective if it is not $(\varepsilon, \mathbf{s})$-effective. Essentially, we use $\mathbb{E}_{N,t_1}\left(\mathbf{e_s}\mathbf{e_s}^T\right)$ as a *proxy* for the state estimation error covariance matrix in the presence of attacks; a sample average is used instead of an expectation because the resultant error in the presence of attacks may not be ergodic. Also, since $\hat{\mathbf{x}}_{\mathbf{s}}(t)$ is computed using all measurements from time 0 till time $t-1$, Definition 1 implicitly takes into consideration the effect of attack signal $\mathbf{a}(t)$ for the time window starting from 0 till time $t_1 + N - 1$.

Using the above notion of an $(\varepsilon, \mathbf{s})$-effective attack, we define the ε-effective attack detection problem as follows.

Problem 1 ***[ε-Effective Attack Detection Problem]*** *Consider the linear dynamical system under attack $\mathbf{\Sigma_a}$ as defined in* (5.2)*, and a k-adversary satisfying Assumptions 1-5. Let $\mathbf{s}_{all}$ be the set of all*

sensors, i.e., $\mathbf{s}_{all} = \{1, \ldots, p\}$. *Given an* $\varepsilon > 0$*, construct an attack indicator* $\hat{d}_{attack} \in \{0,1\}$ *such that:*

$$\hat{d}_{attack}(t_1) = \begin{cases} 1 & \textbf{\textit{if}} \text{ the attack is } (\varepsilon, \mathbf{s}_{all})\text{-effective at time } t_1 \\ 0 & \textbf{\textit{otherwise}}. \end{cases}$$

5.2.4 Optimal Secure State Estimation Problem

We now focus on the problem of estimating the state from the adversarially corrupted sensors. We start by showing a negative result stating that a certain estimation error bound may be impossible to achieve in the presence of a k-adversary. To do so, we define the sensor set that contains $p-k$ sensors and corresponds to the worst case Kalman estimate as:

$$\mathbf{s}_{\text{worst},p-k} = \arg \max_{\substack{\mathbf{s} \subseteq \{1,2,\ldots,p\}, \\ |\mathbf{s}| = p-k}} tr(\mathbf{P}^{\star}_{\mathbf{s}}). \tag{5.3}$$

The impossibility result can now be stated as follows.

Theorem 8 (Impossibility) *Consider the linear dynamical system under attack* $\mathbf{\Sigma_a}$ *as defined in* (5.2)*, and an oracle MMSE estimator that has knowledge of* $\boldsymbol{\kappa}$*, i.e., the set of sensors attacked by a k-adversary. Then, there exists a choice of sensors* $\boldsymbol{\kappa}$ *and an attack sequence* $\mathbf{a}(t)$ *such that the trace of the error covariance of the oracle estimator is bounded from below as follows:*

$$tr\left(\mathbb{E}\left(\mathbf{e}(t)\mathbf{e}^T(t)\right)\right) \geq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right), \tag{5.4}$$

where $\mathbf{e}(t)$ *above is the oracle estimator's error.*

Proof 4 *Consider the attack scenario where the outputs from all attacked sensors are equal to zero,* i.e.*, the corruption* $\mathbf{a}_j(t) = -\mathbf{C}_j\mathbf{x}(t) - v_j(t)$, $\forall j \in \boldsymbol{\kappa}$*. In such a scenario, the information collected from the attacked sensors cannot enhance the estimation performance, and the oracle estimator can simply use the remaining (attack free) sensors to achieve the best possible error performance. Hence, the result follows by picking* $\boldsymbol{\kappa}$ *such that* $\boldsymbol{\kappa} = \{1, \ldots, p\} \setminus \mathbf{s}_{worst,p-k}$.

In the context of Theorem 8, we define a state estimate to be optimal if it is guaranteed to achieve the lower bound shown in (5.4). This can be formalized as follows.

Problem 2 ***[Optimal Secure State Estimation Problem]*** *Consider the linear dynamical system under attack* $\mathbf{\Sigma_a}$ *as defined in* (5.2)*, and a k-adversary satisfying Assumptions 1-5. For a time window* $G = \{t_1, t_1+1, \ldots, t_1+N-1\}$*, construct the state estimates* $\{\hat{\mathbf{x}}(t)\}_{t\in G}$ *such that:*

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}\mathbf{e}^T\right)\right) \leq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right),$$

where $\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t)$ *is the state estimation error.*

Similarly to Definition 1, we use the sample average $\mathbb{E}_{N,t_1}\left(\mathbf{e}\mathbf{e}^T\right)$ in Problem 2 (and not expectation) since the resultant error in the presence of attacks may not be ergodic.

5.3 Sparse observability and $(\varepsilon, \mathbf{s})$-effective attack detection

In this section, we first describe the notion of k-sparse observability [35]. This notion plays an important role in determining when Problems 1 and 2 are solvable. After describing sparse observability, we describe an algorithm for $(\varepsilon, \mathbf{s})$-effective attack detection which leverages sparse observability for its performance guarantees.

5.3.1 k-Sparse Observability

Definition 2 ***(k-Sparse Observable System)*** *The linear dynamical system under attack* $\mathbf{\Sigma_a}$ *as defined in* (5.2)*, is said to be k-sparse observable if for every set* $\mathbf{s} \subseteq \{1, \ldots, p\}$ *with* $|\mathbf{s}| = p-k$*, the pair* $(A, C_{\mathbf{s}})$ *is observable.*

In other words, a system is k-sparse observable if it remains observable after eliminating any choice of k sensors. In the absence of sensor and process noise, the conditions under which exact (*i.e.*, zero error) state estimation can be done despite sensor attacks have been studied in [34, 35, 38] where it is shown that $2k$-sparse observability is necessary and sufficient for exact state estimation

against a k-adversary. In Section 5.7, we provide a coding theoretic interpretation for this condition in the context of attack detection and secure state estimation in any noiseless dynamical system.

5.3.2 $(\varepsilon, \mathbf{s})$-Effective Attack Detector

In this section, we describe an algorithm based on the sparse observability condition for detecting an $(\varepsilon, \mathbf{s})$-effective attack. We first introduce some additional notation, followed by the description of the algorithm and its performance guarantee.

5.3.2.1 Additional notation

Let the sensors be indexed by $i \in \{1, 2, \ldots, p\}$. We define the following observability matrices:

$$\mathscr{O}_i = \begin{bmatrix} \mathbf{C}_i \\ \mathbf{C}_i\mathbf{A} \\ \vdots \\ \mathbf{C}_i\mathbf{A}^{\mu_i - 1} \end{bmatrix}, \quad \mathscr{O} = \begin{bmatrix} \mathscr{O}_1 \\ \mathscr{O}_2 \\ \vdots \\ \mathscr{O}_p \end{bmatrix}, \tag{5.5}$$

where $\mathscr{O}_i$ is the observability matrix for sensor i (with observability index μ_i as shown in (5.5)) and $\mathscr{O}$ is the observability matrix for the entire system (*i.e.*, p sensors) formed by stacking the observability matrices for the sensors. Similarly, for any sensor subset $\mathbf{s} \subseteq \{1, 2, \ldots, p\}$, we denote the observability matrix for $\mathbf{s}$ by $\mathscr{O}_{\mathbf{s}}$ (formed by stacking the observability matrices of sensors in $\mathbf{s}$). Without loss of generality, we will consider the observability index $\mu_i = n$ for each sensor. For any sensor subset $\mathbf{s}$ with $|\mathbf{s}| > k$, we define $\lambda_{\min,\mathbf{s}\setminus k}$ as follows:

$$\lambda_{\min,\mathbf{s}\setminus k} = \min_{\mathbf{s}_1 \subset \mathbf{s},\ |\mathbf{s}_1| = |\mathbf{s}| - k} \lambda_{\min}\left(\mathscr{O}_{\mathbf{s}_1}^T \mathscr{O}_{\mathbf{s}_1}\right), \tag{5.6}$$

where $\lambda_{\min}\left(\mathscr{O}_{\mathbf{s}_1}^T \mathscr{O}_{\mathbf{s}_1}\right)$ denotes the minimum eigenvalue of $\mathscr{O}_{\mathbf{s}_1}^T \mathscr{O}_{\mathbf{s}_1}$. We define matrices $\mathbf{J}_i$, $\mathbf{J}$ and $\mathbf{M}$ as shown below:

$$\mathbf{J}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}_i & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}_i\mathbf{A} & \mathbf{C}_i & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_i\mathbf{A}^{\mu_i-2} & \mathbf{C}_i\mathbf{A}^{\mu_i-3} & \dots & \mathbf{C}_i \end{bmatrix}, \ \mathbf{J} = \begin{bmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \vdots \\ \mathbf{J}_p \end{bmatrix},$$

$$\mathbf{M} = \sigma_w^2 \mathbf{J}\mathbf{J}^T + \sigma_v^2 \mathbf{I}_{np}. \tag{5.7}$$

In a similar spirit, $\mathbf{J}_\mathbf{s}$ is defined for a sensor subset $\mathbf{s}$ by stacking $\mathbf{J}_i$ for $i \in \mathbf{s}$, and $\mathbf{M}_\mathbf{s} = \sigma_w^2 \mathbf{J}_\mathbf{s}\mathbf{J}_\mathbf{s}^T + \sigma_v^2 \mathbf{I}_{n|s|}$. We use the following notation for sensor outputs and noises corresponding to a time window of size $\mu_i = n$ (observability index):

$$\mathbf{y}_i(t) = \begin{bmatrix} y_i(t) \\ y_i(t+1) \\ \vdots \\ y_i(t+\mu_i-1) \end{bmatrix}, \ \mathbf{v}_i(t) = \begin{bmatrix} v_i(t) \\ v_i(t+1) \\ \vdots \\ v_i(t+\mu_i-1) \end{bmatrix},$$

$$\bar{\mathbf{y}}(t) = \begin{bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \\ \vdots \\ \mathbf{y}_p(t) \end{bmatrix}, \ \bar{\mathbf{v}}(t) = \begin{bmatrix} \mathbf{v}_1(t) \\ \mathbf{v}_2(t) \\ \vdots \\ \mathbf{v}_p(t) \end{bmatrix}, \ \bar{\mathbf{w}}(t) = \begin{bmatrix} \mathbf{w}(t) \\ \mathbf{w}(t+1) \\ \vdots \\ \mathbf{w}(t+n-2) \end{bmatrix}, \tag{5.8}$$

where $y_i(t)$ and $v_i(t)$ denote the output and sensor noise at sensor i at time t respectively.

5.3.2.2 Attack Detection Algorithm

We consider the attack detection problem for a time window $G = \{t_1, t_1+1, \dots, t_1+N-1\}$, and assume without loss of generality that the window size N is divisible by n. For a sensor subset $\mathbf{s}$ with $|\mathbf{s}| > k$, we start by computing the state estimate $\hat{\mathbf{x}}_\mathbf{s}(t_1)$ obtained through a Kalman filter that uses measurements collected from time 0 up to time $t_1 - 1$ from all sensors indexed by the subset $\mathbf{s}$. Using this estimate, we can calculate the *block* residue $\mathbf{r}_\mathbf{s}(t_1)$ which is the discrepancy between

Algorithm 1 ATTACK-DETECT$(\mathbf{s}, t_1)$

1: Run a Kalman filter that uses all measurements from sensors indexed by $\mathbf{s}$ until time $t_1 - 1$ and compute the estimate $\hat{\mathbf{x}}_{\mathbf{s}}(t_1) \in \mathbb{R}^n$.

2: Recursively repeat the previous step $N-1$ times to calculate all estimates $\hat{\mathbf{x}}_{\mathbf{s}}(t) \in \mathbb{R}^n$, $\forall t \in G = \{t_1, t_1+1, \ldots, t_1+N-1\}$.

3: For time $t \in G$, calculate the *block* residue:

$$\mathbf{r}_{\mathbf{s}}(t) = \bar{\mathbf{y}}_{\mathbf{s}}(t) - \mathscr{O}_{\mathbf{s}} \hat{\mathbf{x}}_{\mathbf{s}}(t) \quad \forall t \in G.$$

4: **if** block residue test defined below holds,

$$\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}}\mathbf{r}_{\mathbf{s}}^T\right) - \left(\mathscr{O}_{\mathbf{s}}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}}^T + \mathbf{M}_{\mathbf{s}}\right) \preccurlyeq \eta\ \mathbf{1}_{n|\mathbf{s}|}\mathbf{1}_{n|\mathbf{s}|}^T,$$

where $0 < \eta \leq \left(\frac{\lambda_{\min,\mathbf{s}\setminus k}}{3n(|\mathbf{s}|-k)}\right)\varepsilon$, **then**

5: assert $\hat{d}_{\text{attack},\mathbf{s}}(t_1) := 0$

6: **else**

7: assert $\hat{d}_{\text{attack},\mathbf{s}}(t_1) := 1$

8: **end if**

9: **return** $(\hat{d}_{\text{attack},\mathbf{s}}(t_1), \{\hat{\mathbf{x}}_{\mathbf{s}}(t)\}_{t\in G})$

the estimated output $\hat{\bar{\mathbf{y}}}_{\mathbf{s}}(t_1) = \mathscr{O}_{\mathbf{s}}\hat{\mathbf{x}}_{\mathbf{s}}(t_1)$ and the actual output $\bar{\mathbf{y}}_{\mathbf{s}}(t_1)$, *i.e.*,

$$\mathbf{r}_{\mathbf{s}}(t_1) = \bar{\mathbf{y}}_{\mathbf{s}}(t_1) - \hat{\bar{\mathbf{y}}}_{\mathbf{s}}(t_1) = \bar{\mathbf{y}}_{\mathbf{s}}(t_1) - \mathscr{O}_{\mathbf{s}}\hat{\mathbf{x}}_{\mathbf{s}}(t_1). \tag{5.9}$$

By repeating the previous procedure $N-1$ times, we can obtain the sequence of residues $\{\mathbf{r}_{\mathbf{s}}(t)\}_{t\in G}$. The next step is to calculate the sample average of $\mathbf{r}_{\mathbf{s}}(t)\mathbf{r}_{\mathbf{s}}^T(t)$, and compare the sample average with the expected value of $\mathbf{r}_{\mathbf{s}}(t)\mathbf{r}_{\mathbf{s}}^T(t)$ in the case when sensor subset $\mathbf{s}$ is attack-free. This can be done using the following (block) residue test:

$$\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}}\mathbf{r}_{\mathbf{s}}^T\right) - \left(\mathscr{O}_{\mathbf{s}}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}}^T + \mathbf{M}_{\mathbf{s}}\right) \preccurlyeq \eta\ \mathbf{1}_{n|\mathbf{s}|}\mathbf{1}_{n|\mathbf{s}|}^T, \tag{5.10}$$

for some $\eta > 0$. Simply put, the residue test checks whether the sample average of $\mathbf{r}_{\mathbf{s}}(t)\mathbf{r}_{\mathbf{s}}^T(t)$ over time window G is *close* to its attack-free expected value $\mathscr{O}_{\mathbf{s}}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}}^T + \mathbf{M}_{\mathbf{s}}$. This is similar in spirit to

a Chi-squared test [46] (with stronger guarantees as shown in Section 5.3.3), and the time window essentially *averages out* the effect of noise. Note the attack-free estimation error covariance matrix $\mathbf{P}_{\mathbf{s}}^{\star}$ used in (5.10) can be computed offline [44] without the need for any data collected from attack-free sensors. If the element-wise comparison in the residue test (5.10) is valid, we set the attack detection flag $\hat{d}_{\text{attack},\mathbf{s}}(t_1)$ to zero indicating that no attack was detected in sensor subset $\mathbf{s}$. This procedure is summarized in Algorithm 1.

5.3.3 Performance Guarantees

In this subsection, we describe our first main result which is concerned with the correctness of Algorithm 1.

Lemma 1 *Let the linear dynamical system as defined in* (5.2) *be* $2k$*-sparse observable. Consider a* k*-adversary satisfying Assumptions* $1-5$ *and a sensor subset* $\mathbf{s} \subseteq \{1,2,\ldots,p\}$ *with* $|\mathbf{s}| \geq p-k$*. For any* $\varepsilon > 0$ *and* $\delta > 0$*, there exists a large enough time window length* N *such that when Algorithm 1 terminates with* $\hat{d}_{attack,\mathbf{s}}(t_1) = 0$*, the following probability bound holds:*

$$\mathbb{P}\Big(tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right) - \mathbf{P}_{\mathbf{s}}^{\star}\right) \leq \varepsilon\Big) \geq 1-\delta, \tag{5.11}$$

where $\mathbf{e}_{\mathbf{s}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}(t)$*. In other words, for large enough* N*, the bound* $tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right) - \mathbf{P}_{\mathbf{s}}^{\star}\right) \leq \varepsilon$ *holds with high probability*[2] *(w.h.p.). Moreover, in the context of* $(\varepsilon,\mathbf{s})$*-effective attacks, the following also holds:*

$$\mathbb{P}\left(\hat{d}_{attack,\mathbf{s}}(t_1) = d_{attack,\mathbf{s}}(t_1)\right) \geq 1-\delta, \tag{5.12}$$

where $\hat{d}_{attack,\mathbf{s}}(t_1)$ *is the output of Algorithm 1 while* $d_{attack,\mathbf{s}}(t_1)$ *is the output of an oracle detector that knows the exact set of attacked sensors. Hence, Algorithm 1 can detect any* $(\varepsilon,\mathbf{s})$*-effective attack w.h.p. for large enough* N*.*

Proof 5 (Proof of Lemma 1) *We focus only on showing that* (5.11) *holds whenever Algorithm 1 terminates with* $\hat{d}_{attack,\mathbf{s}}(t_1) = 0$*; the rest of the lemma easily follows from the proof of* (5.11) *and*

[2] *By stating that the bound holds with high probability for large enough* N*, we mean that for any* $\delta > 0$ *and* $\varepsilon > 0$*,* $\exists N_{\delta,\varepsilon} \in \mathbb{N}$ *such that for* $N > N_{\delta,\varepsilon}$*,* $\mathbb{P}\Big(tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right) - \mathbf{P}_{\mathbf{s}}^{\star}\right) \leq \varepsilon\Big) \geq 1-\delta$*.*

Definition 1. Since we assume that the set $\mathbf{s}$ *has cardinality* $|\mathbf{s}| \geq p-k$*, we can conclude that there exists a subset* $\mathbf{s}_g \subset \mathbf{s}$ *with cardinality* $|\mathbf{s}_g| \geq p-2k$ *sensors such that all its sensors are attack-free (subscript* g *in* $\mathbf{s}_g$ *stands for good sensors in* $\mathbf{s}$*). Hence, by decomposing the set* $\mathbf{s}$ *into an attack-free set* $\mathbf{s}_g$ *and a potentially attacked set* $\mathbf{s} \setminus \mathbf{s}_g$*, we can conclude that, after a permutation similarity transformation for* (5.10)*, the following holds for the attack-free subset* $\mathbf{s}_g$*:*

$$\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right) - \mathscr{O}_{\mathbf{s}_g}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T - \mathbf{M}_{\mathbf{s}_g} \preccurlyeq \eta\ \mathbf{1}_{n(|\mathbf{s}|-k)}\mathbf{1}_{n(|\mathbf{s}|-k)}^T.$$

Therefore,

$$\begin{aligned} tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right) - \mathscr{O}_{\mathbf{s}_g}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T - \mathbf{M}_{\mathbf{s}_g}\right) \\ \leq n(|\mathbf{s}|-k)\eta = \varepsilon_1. \end{aligned} \tag{5.13}$$

Similarly, after a suitable permutation Π*, we can decompose the block residue* $\mathbf{r}_{\mathbf{s}}(t)$ *defined in equation* (5.9) *as follows:*

$$\begin{aligned} \Pi(\mathbf{r}_{\mathbf{s}}(t)) &= \begin{bmatrix} \mathbf{r}_{\mathbf{s}_g}(t) \\ \mathbf{r}_{\mathbf{s}\setminus\mathbf{s}_g}(t) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}}_{\mathbf{s}_g}(t) - \mathscr{O}_{\mathbf{s}_g}\hat{\mathbf{x}}_{\mathbf{s}}(t) \\ \bar{\mathbf{y}}_{\mathbf{s}\setminus\mathbf{s}_g}(t) - \mathscr{O}_{\mathbf{s}\setminus\mathbf{s}_g}\hat{\mathbf{x}}_{\mathbf{s}}(t) \end{bmatrix} \\ &= \begin{bmatrix} \mathscr{O}_{\mathbf{s}_g}\mathbf{x}(t) + \mathbf{J}_{\mathbf{s}_g}\bar{\mathbf{w}}(t) + \bar{\mathbf{v}}_{\mathbf{s}_g}(t) - \mathscr{O}_{\mathbf{s}_g}\hat{\mathbf{x}}_{\mathbf{s}}(t) \\ \bar{\mathbf{y}}_{\mathbf{s}\setminus\mathbf{s}_g}(t) - \mathscr{O}_{\mathbf{s}\setminus\mathbf{s}_g}\hat{\mathbf{x}}_{\mathbf{s}}(t) \end{bmatrix} \\ &= \begin{bmatrix} \mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t) + \mathbf{z}_{\mathbf{s}_g}(t) \\ \bar{\mathbf{y}}_{\mathbf{s}\setminus\mathbf{s}_g}(t) - \mathscr{O}_{\mathbf{s}\setminus\mathbf{s}_g}\hat{\mathbf{x}}_{\mathbf{s}}(t) \end{bmatrix}, \end{aligned} \tag{5.14}$$

where $\mathbf{z}_{\mathbf{s}_g}(t) = \mathbf{J}_{\mathbf{s}_g}\bar{\mathbf{w}}(t) + \bar{\mathbf{v}}_{\mathbf{s}_g}(t)$*. Using* (5.14)*, we can rewrite* $tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right)\right)$ *as:*

$$\begin{aligned} &tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right)\right) \\ &\quad = tr\left(\mathscr{O}_{\mathbf{s}_g}\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\mathscr{O}_{\mathbf{s}_g}^T\right) + tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}\mathbf{z}_{\mathbf{s}_g}^T\right)\right) \\ &\qquad + 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}\right). \end{aligned} \tag{5.15}$$

By combining (5.13) *and* (5.15)*:*

$$tr\left(\mathscr{O}_{\mathbf{s}_g}\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\mathscr{O}_{\mathbf{s}_g}^T - \mathscr{O}_{\mathbf{s}_g}\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T\right)$$

$$
\begin{aligned}
&\leq tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}\mathbf{z}_{\mathbf{s}_g}^T\right)\right) + \varepsilon_1 \\
&\quad - 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T \mathscr{O}_{\mathbf{s}_g}^T \mathbf{z}_{\mathbf{s}_g}\right) \\
&\overset{(a)}{\leq} 2\varepsilon_1 - 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T \mathscr{O}_{\mathbf{s}_g}^T \mathbf{z}_{\mathbf{s}_g}\right) \qquad (5.16) \\
&\overset{(b)}{\leq} 3\varepsilon_1, \qquad (5.17)
\end{aligned}
$$

where (a) *follows w.h.p. due to the law of large numbers (LLN) for large enough N (details in Appendix A.0.12.1), and* (b) *follows w.h.p. by showing that the cross term* $2\mathbb{E}_{N,t_1}\left(\mathbf{e}^T \mathscr{O}_{\mathbf{s}_g}^T \mathbf{z}_{\mathbf{s}_g}\right)$ *has zero mean and vanishingly small variance for large enough N. The cross term analysis is described in detail in Appendix A.0.12.2. Now recall that for any two matrices,* $\mathbf{A}$ *and* $\mathbf{B}$ *of appropriate dimensions,* $tr(\mathbf{AB}) = tr(\mathbf{BA})$*. Using this fact along with* (5.17)*, the following holds:*

$$
tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T - \mathbf{P}_{\mathbf{s}}^{\star}\right)\mathscr{O}_{\mathbf{s}_g}^T \mathscr{O}_{\mathbf{s}_g}\right) \leq 3\varepsilon_1, \qquad (5.18)
$$

and hence, we get the following bound which completes the proof:

$$
tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right) - \mathbf{P}_{\mathbf{s}}^{\star}\right) \overset{(c)}{\leq} \frac{3\varepsilon_1}{\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T \mathscr{O}_{\mathbf{s}_g}\right)} \overset{(d)}{\leq} \frac{3\varepsilon_1}{\lambda_{\min,\mathbf{s}\setminus k}} \leq \varepsilon \qquad (5.19)
$$

where (c) *follows from Lemma 3 in Appendix A.0.13 and* (d) *follows from the definition of* $\lambda_{\min,\mathbf{s}\setminus k}$*. Note that, it follows from* $|\mathbf{s}_g| \geq p - 2k$ *and* $2k$*-sparse observability, that both* $\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T \mathscr{O}_{\mathbf{s}_g}\right)$ *and* $\lambda_{\min,\mathbf{s}\setminus k}$ *are bounded away from zero. This completes the proof of* (5.11)*. Based on the proof of* (5.11) *and Definition 1, it is now straightforward to show* (5.12)*. Intuitively, the probability of mismatch between* $\hat{d}_{attack,\mathbf{s}}(t_1)$ *and* $d_{attack,\mathbf{s}}(t_1)$ *in* (5.12) *stems from the chances of a false detection; this happens when the noise realizations deviate from LLN and the detector's threshold check fails despite the absence of an adversary. As seen in the proof of* (5.11)*, w.h.p. the noise realizations obey LLN, and hence w.h.p.* $\hat{d}_{attack,\mathbf{s}}(t_1)$ *and* $d_{attack,\mathbf{s}}(t_1)$ *are equal.*

5.4 Effective Attack Detection and Secure State Estimation

Based on the performance guarantees for the ATTACK-DETECT algorithm described in Section 5.3, in this section we describe our main results for Problems 1 and 2.

5.4.1 Attack detection

We start by showing a solution to Problem 1 (ε-effective attack detection), which follows directly from Lemma 1.

Theorem 9 *Let the linear dynamical system defined in* (5.2) *be k-sparse observable system. Consider a k-adversary satisfying Assumptions* 1-5*, and the detector* $\hat{d}_{attack}(t_1) = \text{ATTACK-DETECT}(\mathbf{s}_{all}, t_1)$ *where the set* $\mathbf{s}_{all} = \{1, \ldots, p\}$*. Then, for large enough time window length N, w.h.p.* $\hat{d}_{attack}(t_1)$ *is equal to the attack indicator which solves Problem 1.*

Proof 6 *The proof is similar to the proof of Lemma 1. In the proof of Lemma 1, we basically required the set of good sensors* $\mathbf{s}_g$ *to form an observable system. Similarly, while checking for effective attacks on a sensor set of size p, we require the set of good sensors (of size* $\geq p-k$*) to form an observable system in order to repeat the steps in the proof for Lemma 1; this requirement is guaranteed by the k-sparse observability condition. On a related note, in Section 5.7, we give a coding theoretic interpretation for the k-sparse observability requirement for attack detection.*

5.4.2 Secure State Estimation

Algorithm 2 describes our proposed solution for Problem 2 (secure state estimation). As described in Algorithm 2, we exhaustively enumerate $\binom{p}{p-k}$ sensor subsets of size $p-k$, and then apply ATTACK-DETECT on each sensor subset until we find one subset $\mathbf{s}^*$ for which ATTACK-DETECT returns $\hat{d}_{\text{attack},\mathbf{s}^*}(t_1) = 0$ indicating that the subset is (ε-effective) attack-free. The following theorem states the performance guarantees associated with Algorithm 2.

Theorem 10 *Let the linear dynamical system defined in* (5.2) *be* $2k$*-sparse observable system. Consider a k-adversary satisfying Assumptions 1-5. Consider the state estimate* $\hat{\mathbf{x}}_{\mathbf{s}^*}(t)$ *computed by Algorithm 2. Then, for any* $\varepsilon > 0$ *and* $\delta > 0$*, there exists a large enough N such that:*

$$\mathbb{P}\left(tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}^*}\mathbf{e}_{\mathbf{s}^*}^T\right)\right) \leq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right) + \varepsilon\right) \geq 1-\delta, \tag{5.20}$$

Algorithm 2 EXHAUSTIVE SEARCH

1: Enumerate all sets $\mathbf{s} \in \mathbf{S}$ such that:

$$\mathbf{S} = \{\mathbf{s} | \mathbf{s} \subset \{1, 2, \ldots, p\},\ |\mathbf{s}| = p - k\}.$$

2: Exhaustively search for $\mathbf{s}^* \in \mathbf{S}$ for which $d_{\text{attack},\mathbf{s}^*}(t_1) = 0$ and use $\hat{\mathbf{x}}_{\mathbf{s}^*}(t)$ for $t \in G$ as the state estimate.

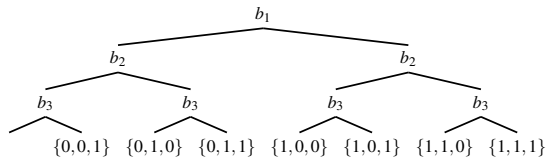
where $\mathbf{e}_{\mathbf{s}^*}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}^*}(t)$ *is the estimation error using* $\hat{\mathbf{x}}_{\mathbf{s}^*}(t)$ *as the state estimate. In other words, w.h.p. Algorithm 2 achieves the bound* $\limsup_{N\to\infty} \frac{1}{N} \sum_{t\in G} \mathbf{e}_{\mathbf{s}^*}^T(t) \mathbf{e}_{\mathbf{s}^*}(t) \leq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right)$.

Proof 7 *The result follows from Lemma 1 which ensures that, in the absence of the* $(\varepsilon, \mathbf{s})$*-effective attack property, the calculated state estimate still guarantees the bound* (5.11)*. This in turn implies that, in the worst case,* $\limsup_{N\to\infty} \frac{1}{N} \sum_{t\in G} \mathbf{e}_{\mathbf{s}^*}^T(t) \mathbf{e}_{\mathbf{s}^*}(t) = tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right)$ *is achievable. However, since the k-adversary may not always attack the worst case set of sensors* $\mathbf{s}_{worst,p-k}$*, we can replace the equality sign above with an inequality, leading to* $\limsup_{N\to\infty} \frac{1}{N} \sum_{t\in G} \mathbf{e}_{\mathbf{s}^*}^T(t) \mathbf{e}_{\mathbf{s}^*}(t) \leq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right)$.

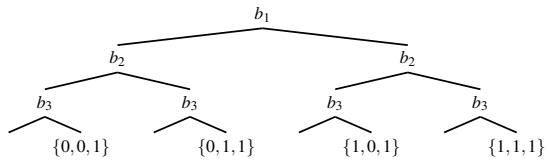
5.5 Reducing Search Time Using Satisfiability Modulo Theory Solving

Algorithm 2 exhaustively explores all combinations of $p-k$ sensors until a set $\mathbf{s}^*$ satisfying $d_{\text{attack},\mathbf{s}^*}(t_1) = 0$ is found. In this section, we explore the idea of using sophisticated search techniques in order to harness the underlying combinatorial aspect of the secure state estimation problem. In particular, we extend previous work by the authors and co-workers on using Satisfiability Modulo Theory (SMT)-like solvers [37], developed for the noiseless case, in order to improve the search time while preserving optimality of the solution.

The driving concept behind SMT solvers can be summarized as follows. First, the search space of all sensor subsets with cardinality $p-k$, is encoded using Boolean variables (the number of Boolean variables increases linearly with the number of sensors), and a Boolean search engine (*e.g.*, SAT solver) is used in order to traverse the search space. Whenever the SAT solver suggests one possible solution in the search space, a higher level solver (typically referred to as the Theory-solver) is used to check the correctness of that particular solution. Finally, in order to



(a) A tree showing all the combinations of three Boolean indicator variables b_1, b_2, b_3 when a conflicting certificate of the form $\phi_{\text{cert}} := b_1 + b_2 + b_3 \geq 1$ is generated. The missing combination $\{0,0,0\}$ is the only one that is eliminated as a result of this certificate.



(b) A tree showing all the combinations of three Boolean indicator variables b_1, b_2, b_3 when a conflicting certificate of the form $\phi_{\text{cert}} := b_3 \geq 1$ is generated. The missing four combinations $\{0,0,0\}, \{0,1,0\}, \{1,0,0\}, \{1,1,0\}$ are eliminated as a result of this certificate.

Figure 5.1: Pictorial example illustrating the effect of generating smaller conflicting certificates.

prevent the SAT solver from enumerating all possible solutions in the search space, the Theory-solver generates counter examples (certificates), explaining why a particular solution is not valid. Each certificate is used by the SAT solver in order to prune the search space and hence enhance the performance of the overall algorithm. This methodology of "counter-example guided search" effectively breaks the secure state estimation problem into two simpler tasks over the Boolean and Reals domain. Further details about this technique are described below.

5.5.1 Overall Architecture

We start by introducing a Boolean indicator variable $b = (b_1, \ldots, b_p) \in \mathbb{B}^p$ where the assignment $b_i = 1$ hypothesizes that the ith sensor is under attack while the assignment $b_i = 0$ hypothesizes that the ith sensor is attack-free. Using this indicator variable, b, we start by asking the (pseudo-)Boolean SAT solver to assign values to b in order to satisfy the following formula:

$$\phi(0) ::= \sum_{i=1}^{p} b_i \leq k, \tag{5.21}$$

which ensures that at most k sensors are going to be hypothesized as being under attack (the addition in (5.21) is over Reals).

In the next step, this hypothesized assignment is then checked by the theory solver. This is done by running the ATTACK-DETECT algorithm (Algorithm 1) using only the set of hypothe-

sized attack-free sensors $\mathbf{s}(b) = \{1, \ldots, p\} - \text{supp}(b)$. If the ATTACK-DETECT algorithm returns $\hat{d}_{\text{attack},\mathbf{s}(b)} = 0$ then our solver approves this hypothesis and the algorithm terminates. Otherwise, an UNSAT certificate (also known as a counter-example) is generated explaining why this assignment of b is not valid (*i.e.*, a conflict). A trivial UNSAT certificate that can always be generated takes the following form (in iteration j):

$$\phi_{\text{cert}}(j) ::= \sum_{i \in \mathbf{s}(b)} b_i \geq 1, \tag{5.22}$$

which ensures that the current assignment of the variable b is excluded. Once this UNSAT certificate is generated, the (pseudo-)Boolean SAT solver is then invoked again in the next iteration with the following constraints:

$$\phi(j+1) ::= \phi(j) \wedge \phi_{\text{cert}}(j),$$

until one assignment of the variable b passes the attack detection test. This procedure is summarized in Algorithm 3.

5.5.2 Conflicting Certificates

The generated UNSAT certificates heavily affect the overall execution time. Smaller UNSAT certificates prune the search space faster. For simplicity, consider the example shown in Figure 5.1 where the vector b has only three elements. On one hand, an UNSAT certificate that has the form $\phi_{\text{cert}} = b_1 + b_2 + b_3 \geq 1$ leads to pruning only one sample in the search space. On the other hand, a smaller UNSAT certificate that has the form $\phi_{\text{cert}} = b_1 \geq 1$ eliminates four samples in the search space which is indeed a higher reduction, and hence leads to better execution time.

To generate a compact (*i.e.*, smaller) Boolean constraint that explains a conflict, we aim to find a small set of sensors that cannot all be attack-free. To do so, we start by removing one sensor from the set $\mathbf{s}(b)$ and run the ATTACK-DETECT algorithm on the reduced set $\mathbf{s}'(b)$ to obtain $\hat{d}_{\text{attack},\mathbf{s}'(b)}$. If $\hat{d}_{\text{attack},\mathbf{s}'(b)}$ still equals one (which indicates that set $\mathbf{s}'(b)$ still contains a conflicting set of sensors), we generate the more compact certificate:

$$\phi_{\text{cert}}(j) ::= \sum_{i \in \mathbf{s}'(b)} b_i \geq 1. \tag{5.23}$$

We continue removing sensors one by one until we cannot find any more conflicting sensor sets. Indeed, the order in which the sensors are removed is going to affect the overall execution time. In Algorithm 4 we implement a heuristic (for choosing this order) which is inspired by the strategy we adopted in the noiseless case [37].

Note that the reduced sets $\mathbf{s}'(b)$ are used only to generate the UNSAT certificates. Hence, it is direct to show that Algorithm 3 still preserves the optimality of the state estimate as stated by the following result.

Theorem 11 *Let the linear dynamical system defined in* (5.2) *be* $2k$*-sparse observable system. Consider a k-adversary satisfying Assumptions 1-5. Consider the state estimate* $\hat{\mathbf{x}}_{\mathbf{s}^*}(t)$ *computed by Algorithm 3. Then, for any* $\varepsilon > 0$ *and* $\delta > 0$*, there exists a large enough N such that:*

$$\mathbb{P}\left(tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}^*}\mathbf{e}_{\mathbf{s}^*}^T\right)\right) \leq tr\left(\mathbf{P}^{\star}_{\mathbf{s}_{worst,p-k}}\right) + \varepsilon\right) \geq 1-\delta. \tag{5.24}$$

Note that although, for the sake of brevity, we did not analyze analytically the worst case execution time (in terms on number of iterations) of Algorithm 3, we show numerical results in Section 5.6 that support the claim that the proposed SMT-like solver works much better in practice compared to the exhaustive search procedure (Algorithm 2).

Algorithm 3 SMT-BASED SEARCH

status := UNSAT;

$\phi_B := \sum_{i \in \{1,\ldots,p\}} b_i \leq k$;

while status == UNSAT **do**

  $b :=$ SAT-SOLVE(ϕ_B);

  $\mathbf{s}(b) := \{1,2,\ldots,p\} - \text{supp}(b)$;

  $(\hat{d}_{\text{attack},\mathbf{s}(b)}, \{\hat{\mathbf{x}}_{\mathbf{s}(b)}(t)\}_{t \in G})$
  $:=$ ATTACK-DETECT($\mathbf{s}(b), t_1$);

  if $\hat{d}_{\text{attack},\mathbf{s}(b)} == 1$ **then**

    ϕ_{cert}
    $:=$ GENERATE-CERTIFICATE($\mathbf{s}(b), \{\hat{\mathbf{x}}_{\mathbf{s}(b)}(t)\}_{t \in G}$);

    $\phi_B := \phi_B \wedge \phi_{\text{cert}}$;

  end if

end while

$\mathbf{s}^* = \mathbf{s}(b)$;

return $\{\hat{\mathbf{x}}_{\mathbf{s}^*}(t)\}_{t \in G}$;

Algorithm 4 GENERATE-CERTIFICATE($\mathbf{s}, \{\hat{\mathbf{x}}_{\mathbf{s}}(t)\}_{t\in G}$)

Compute the residues for $i \in \mathbf{s}$

$\mathbf{r}_i(t) := \mathbf{y}_i(t) - \mathscr{O}_i\hat{\mathbf{x}}_{\mathbf{s}}(t), \forall t \in G = \{t_1, \ldots, t_1 + N - 1\}$

$\mu_i(t_1) := \left| tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_i\mathbf{r}_i^T\right) - \mathscr{O}_i\mathbf{P}_{\mathbf{s}}^{\star}\mathscr{O}_i^T - \mathbf{M}_i\right) - \eta n\right|;$

Normalize the residues

$\mu_i(t_1) := \mu_i(t_1)/\lambda_{\max}\left(\mathscr{O}_i^T\mathscr{O}_i\right),$

$\mu(t_1) := \{\mu_i(t_1)\}_{i\in\mathbf{s}}$;

Sort the residues in ascending order

$\mu\_sorted(t_1) :=$ sortAscendingly$(\mu(t_1))$;

Choose sensor indices of $p - 2k + 1$ **smallest residues**

$\mu\_min\_r :=$ Index $(\mu\_sorted[1 : p - 2k + 1])$;

Search linearly for the UNSAT certificate

status = `UNSAT`; counter $= 1$; $\phi_{\text{conf-cert}} = 1$; $\mathbf{s}' = \mathbf{s}$

while status == `UNSAT` **do**

  $\mathbf{s}' := \mathbf{s}' \setminus \mu\_min\_r$[counter];

  $(\hat{d}_{\text{attack},\mathbf{s}'}, \{\hat{\mathbf{x}}_{\mathbf{s}'}(t)\}_{t\in G}) :=$ ATTACK-DETECT$(\mathbf{s}', t_1)$;

  if $\hat{d}_{\text{attack},\mathbf{s}'} == 1$ **then**

    $\phi_{\text{conf-cert}} := \phi_{\text{conf-cert}} \wedge \sum_{i\in\mathbf{s}'} b_i \geq 1$;

    counter := counter + 1;

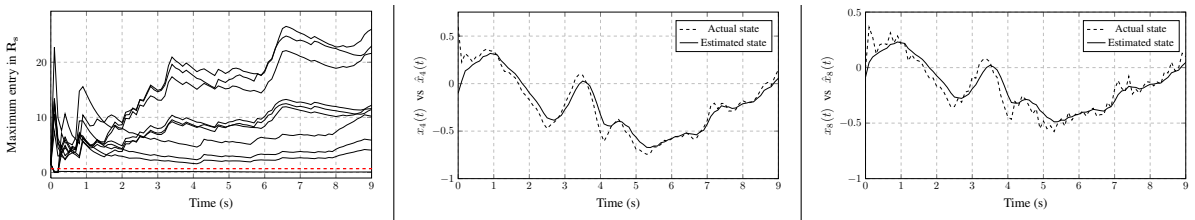
  else

    status := `SAT`;

  end if

end while

return $\phi_{\text{conf-cert}}$



(a) (b)

Figure 5.2: Figure showing results of Experiment 1: (a) the maximum entry in the residue test matrix $\mathbf{R_s} = \mathbb{E}_{N,t_1}\left(\mathbf{r_s}\mathbf{r_s}^T\right) - \left(\mathscr{O}_\mathbf{s}\mathbf{P}_\mathbf{s}^\star\mathscr{O}_\mathbf{s}^T + \mathbf{M_s}\right)$ for the 10 Kalman filters versus the threshold $\eta = 0.7$ (indicated by the dashed red line). As shown in the figure, there is only one subset of sensors which satisfies the threshold η, and this corresponds to the attack-free set of sensors, and (b) the estimated state trajectory (of state x_4 and x_8, *i.e.*, dimension 4 and 8 of $\mathbf{x}$) from the subset of sensors which satisfy the threshold versus the actual state trajectory.

5.6 Numerical Experiments

In this section, we report numerical results for Algorithms 2 and 3 as described by the experiments below.

5.6.1 Experiment 1: Residue test performance in Algorithm 2

In this experiment, we numerically check the performance of the residue test involved in Algorithm 2 while checking for effective attacks across sensor subsets. We generate a stable system randomly with $n = 20$ (state dimension) and $p = 5$ sensors. We select $k = 2$ sensors at random, and apply a random attack signal to the two sensors. We apply Algorithm 2 by running all the $\binom{5}{3} = 10$ Kalman filters (one for each distinct sensor subset of size 3) and do the residue test corresponding to each sensor subset. Figure 5.2(a) shows the maximum entry in the residue test matrix $\mathbf{R_s} = \mathbb{E}_{N,t_1}\left(\mathbf{r_s}\mathbf{r_s}^T\right) - \left(\mathscr{O}_\mathbf{s}\mathbf{P}_\mathbf{s}^\star\mathscr{O}_\mathbf{s}^T + \mathbf{M_s}\right)$ for the 10 different Kalman filters. It is apparent from Figure 5.2(a) that only one Kalman filter produces a state estimate that passes the residue test defined in Algorithm 1. This indeed corresponds to the set of attack-free sensors in the experiment.

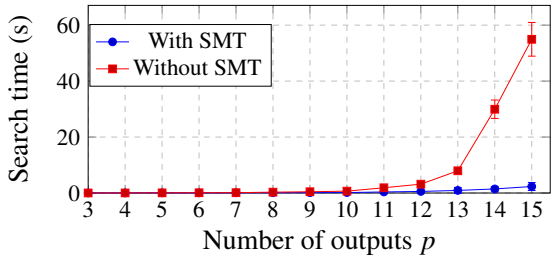


Figure 5.3: Comparison of sensor subset search times for exhaustive search and SMT based search.

5.6.2 Experiment 2: Performance of SMT-based Search

In this experiment, we compare the sensor subset search time for the SMT-based approach (Algorithm 3) with that for the exhaustive search approach (Algorithm 2). For this experiment, we fix $n = 50$ (state dimension) and vary the number of sensors from $p = 3$ to $p = 15$. For each system, we pick one third of the sensors to be under attack, *i.e.*, $k = \lfloor p/3 \rfloor$. The attack signal is chosen as a linear function of the measurement noise. For each system, we run the bank of $\binom{p}{p-k}$ Kalman filters to generate the state estimates corresponding to all sensor subsets of size $p-k$. We then use both exhaustive search as well as the SMT-based search to find the sensor subset that satisfies the residue test in Algorithm 1. Figure 5.3 shows the average time needed to perform the search across 50 runs of the same experiment. Figure 5.3 suggests that the SMT-based search has an exponential improvement over exhaustive search as the number of sensors increases. In particular, for $p = 15$, the SMT-based search out-performs exhaustive search by an order of magnitude.

5.7 Sparse observability: Coding theoretic view

In this section, we revisit the sparse observability condition against a k-adversary and give a coding theoretic interpretation for the same. We first describe our interpretation for a linear system, and then discuss how it can be generalized for non-linear systems.

Consider the linear dynamical system in (5.2) without the process and sensor noise (*i.e.*, $\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t)$, $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{a}(t)$). If the system's initial state is $\mathbf{x}(0) \in \mathbb{R}^n$ and the system is θ-sparse observable, then clearly in the absence of sensor attacks, by observing the outputs from any $p - \theta$

sensors for n time instants ($t = 0, 1, \ldots, n-1$) we can exactly recover $\mathbf{x}(0)$ and hence, *exactly* estimate the state of the plant. A coding theoretic view of this can be given as follows. Consider the outputs from sensor $d \in \{1, 2, \ldots, p\}$ for n time instants as a symbol $\mathcal{Y}_d \in \mathbb{R}^n$. Thus, in the (symbol) observation vector $\mathcal{Y} = \begin{bmatrix} \mathcal{Y}_1 & \mathcal{Y}_2 \ldots \mathcal{Y}_p \end{bmatrix}$, due to θ-sparse observability, any $p - \theta$ symbols are sufficient (in the absence of attacks) to recover the initial state $\mathbf{x}(0)$. Now, let us consider the case of a k-adversary which can arbitrarily corrupt any k sensors. In the coding theoretic view, this corresponds to arbitrarily corrupting any k (out of p) symbols in the observation vector. Intuitively, based on the relationship between error correcting codes and the Hamming distance between codewords in classical coding theory [19], one can expect the recovery of the initial state despite such corruptions to depend on the (symbol) Hamming distance between the observation vectors corresponding to two distinct initial states (say $\mathbf{x}^{(1)}(0)$ and $\mathbf{x}^{(2)}(0)$ with $\mathbf{x}^{(1)}(0) \neq \mathbf{x}^{(2)}(0)$). In this context, the following lemma relates θ-sparse observability to the minimum Hamming distance between observation vectors in the absence of attacks.

Lemma 2 *For a θ-sparse observable system, the minimum (symbol) Hamming distance between observation vectors corresponding to distinct initial states is $\theta + 1$.*

Proof 8 *Consider a system with p sensors, and observation vectors $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ corresponding to distinct initial states $\mathbf{x}^{(1)}(0)$ and $\mathbf{x}^{(2)}(0)$. Due to θ-sparse observability, at most $p - \theta - 1$ symbols in $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ can be identical; if any $p - \theta$ of the symbols are identical, this would imply $\mathbf{x}^{(1)}(0) = \mathbf{x}^{(2)}(0)$. Hence, the (symbol) Hamming distance between the observation vectors $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ (corresponding to $\mathbf{x}^{(1)}(0)$ and $\mathbf{x}^{(2)}(0)$) is at least $p - (p - \theta - 1) = \theta + 1$ symbols. Also, there exists a pair of initial states $\left(\mathbf{x}^{(1)}(0), \mathbf{x}^{(2)}(0)\right)$, such that the corresponding observation vectors $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ are identical in exactly $p - \theta - 1$ symbols[3] and differ in the rest $\theta + 1$ symbols. Hence, the minimum (symbol) Hamming distance between the observation vectors is $\theta + 1$.*

[3] *If there is no such pair of initial states, the initial state can be recovered by observing any $p - \theta - 1$ sensors. By definition, in a θ-sparse observable system, θ is the largest positive integer, such that the initial state can be recovered by observing any $p - \theta$ sensors.*

For a θ-sparse observable system, since the minimum Hamming distance between the observation vectors corresponding to distinct initial states is $\theta+1$, we can:

(1) correct up to $k < \frac{\theta+1}{2}$ sensor corruptions,

(2) detect up to $k \leq \theta$ sensor corruptions.

Note that (1) above is equivalent to $2k \leq \theta$ (sparse observability condition for secure state estimation [35]). It should be noted that a k-adversary can attack *any* set of k (out of p) sensors, and the condition $k < \frac{\theta+1}{2}$ is both necessary and sufficient for exact state estimation despite such attacks. When $k \geq \frac{\theta+1}{2}$, it is straightforward to show a scenario where the observation vector (after attacks) can be explained by multiple initial states, and hence exact state estimation is not possible. The following example illustrates such an attack scenario.

Example 5 *Consider a θ-sparse observable system with $\theta = 3$, number of sensors $p = 5$, and a k-adversary with $k = 2$. Clearly, the condition $k < \frac{\theta+1}{2}$ is not satisfied in this example. Let $\mathbf{x}^{(1)}(0)$ and $\mathbf{x}^{(2)}(0)$ be distinct initial states, such that the corresponding observation vectors $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ have (minimum) Hamming distance $\theta+1 = 4$ symbols. Figure 5.4 depicts the observation vectors $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$, and for the sake of this example, we assume that the observation vectors have the same first symbol (i.e., $\mathcal{Y}_1^{(1)} = \mathcal{Y}_1^{(2)} = \mathcal{Y}_1$) and differ in the rest 4 symbols (hence, a Hamming distance of 4). Now, as shown in Figure 5.4, suppose the observation vector after attacks was $\mathcal{Y} = \left[\mathcal{Y}_1 \;\; \mathcal{Y}_2^{(1)} \;\; \mathcal{Y}_3^{(1)} \;\; \mathcal{Y}_4^{(2)} \;\; \mathcal{Y}_5^{(2)}\right]$. Clearly, there are two possible explanations for this (attacked) observation vector: (a) the initial state was $\mathbf{x}^{(1)}(0)$ and sensors 4 and 5 were attacked, or (b) the initial state was $\mathbf{x}^{(2)}(0)$ and sensors 2 and 3 were attacked. Since there are two possibilities, we cannot estimate the initial state exactly given the attacked observation vector. This example can be easily generalized to show the necessity of the condition $k < \frac{\theta+1}{2}$.*

For (noiseless) non-linear systems, by analogously defining θ-sparse observability, the same coding theoretic interpretation holds. This leads to the necessary and sufficient conditions for attack detection and secure state estimation in any noiseless dynamical system with sensor attacks.

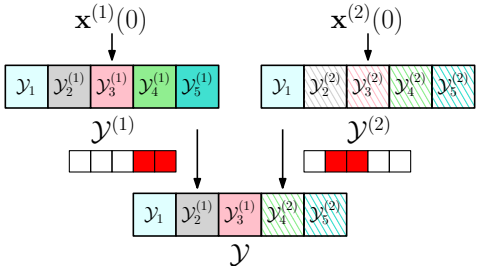


Figure 5.4: Example with $\theta = 3$, $p = 5$ and $k = 2$. For distinct initial states $\mathbf{x}^{(1)}(0)$ and $\mathbf{x}^{(2)}(0)$, the corresponding observation vectors are $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$. Given (attacked) observation vector $\mathcal{Y} = \left[\mathcal{Y}_1 \;\; \mathcal{Y}_2^{(1)} \;\; \mathcal{Y}_3^{(1)} \;\; \mathcal{Y}_4^{(2)} \;\; \mathcal{Y}_5^{(2)}\right]$, there are two possibilities for the initial state: (a) $\mathbf{x}^{(1)}(0)$ with attacks on sensors 4 and 5, or (b) $\mathbf{x}^{(2)}(0)$ with attacks on sensors 2 and 3.

CHAPTER 6

Conclusions and Future Work

In this chapter, we discuss the conclusions and directions for future research in the context of results in this dissertation. We start with a discussion on our results related to harnessing bursty interference, followed by a discussion on our results for secure state estimation.

6.1 Harnessing bursty interference

Burstiness model In Chapter 2, we consider an arbitrary joint distribution governing the interference states across subcarriers but constrain the marginal probabilities to be the same for each subcarrier. This constraint enables us to define a good fractional partition for the Madiman-Tetali subset inequality leading to tight outer bounds. From a practical perspective, our results with this constraint imply that even when the subcarriers experience a similar level of burstiness, some subcarriers can help others to achieve higher rates. In this context, considering the case of different marginal probabilities (of burstiness) for each subcarrier will be a natural extension of our results. Extending the results to non-i.i.d. bursts is another aspect which needs investigation. As an additional remark, the multicarrier system model in this dissertation can be interpreted as a single carrier setup with transmission blocks of M symbols, and feedback after every block. With such an interpretation, the interference can span multiple symbols in a block during the instantiation of a *burst*; this essentially provides a way to model interference bursts of longer duration in the single carrier setup.

Symmetric interference setup In Chapter 2, we focus on a symmetric interference channel for each subcarrier. This allows us to build on the capacity results from the single carrier setup [1]. For

the case of asymmetric interference, even in the single carrier setup, the capacity characterization is not complete [1]. Hence, for extending our multicarrier results for asymmetric interference, a better understanding of asymmetric (bursty) interference in the single carrier setup is needed.

Output feedback In Chapter 2, we assumed output feedback at the transmitter. From a practical view point, enabling channel state information (CSI) feedback is significantly cheaper than enabling output feedback. Hence, it will be of practical interest to extend the multicarrier schemes developed in this dissertation for the case when only CSI feedback is available at the transmitter.

Constant gap result For the setup with Gaussian noise (GN setup) in Chapter 2, we only give a tight generalized degrees of freedom characterization. Hence, the question of whether we can give a constant gap result for the GN setup remains to be explored. In this direction, using lattice codes along with ideas developed for GDoF achievability in this dissertation, may be a good candidate for the achievability scheme.

Coding without structure A practical drawback of the achievability schemes in Chapters 2 and 3 is the use of structured coding for carefully aligning the interfered signal; finding alternative schemes which do not require such structured coding is another direction which remains to be explored.

6.2 Secure state estimation

Combining observer and sensor attacks In Chapter 4 we studied the problem of observer attacks, and in Chapter 5 we studied the problem of sensor attacks. A natural direction for future research is considering both sensor and observer attacks at the same time, and then leveraging dynamics to come up with secure state estimation algorithms which are resilient against such attacks. We believe that, in principle, the tools developed in this dissertation can be combined to handle such attacks.

Attacks in non-linear systems In this dissertation, we primarily focused on attacks in systems with linear dynamics. We developed fundamental results which demonstrate how one can use the knowledge of the underlying dynamics towards mitigating observer and sensor attacks. Since CPS deployed in practice usually have non-linear dynamics, the next step forward will be to extend our results for non-linear systems. In this context, our coding theoretic results on the amount of (sensor and observer) redundancy required against attacks also hold for non-linear systems, and will serve as a guiding principle in designing secure state estimation algorithms for non-linear systems.

Model-free (data driven) attack detection and secure state estimation in CPS In this dissertation, we assumed knowledge of the dynamics (*i.e.*, system model). In many large scale practical scenarios, we do not have a precise understanding of the system model and may just have access to the input and output data streams. It will be of theoretical as well as practical interest to extend our results for attack detection and secure state estimation in such data driven scenarios. It would also be interesting to mix ideas from our secure state estimation algorithms with tools from machine learning like hidden Markov model (HMM) and recurrent neural network (RNN) based anomaly detection algorithms.

Privacy preserving attack detection against sensor attacks in CPS In Chapter 5, we just focused on accurate state estimation against sensor attacks without considering privacy constraints associated with sensor data. In a case where each sensor has privacy constraints, it will be interesting to see if encryption schemes (for sensor data) can be used in conjunction with our attack detection algorithms (developed for secure state estimation).

Attack detection in CPS using compressed sensor data In many industrial scenarios, the data generated from sensors is not processed immediately for anomalies and is stored in a compressed format for future processing. It would be interesting to study lossy compression schemes in conjunction with our attack detection algorithms (given the knowledge of dynamics) towards the development of attack detection algorithms which leverage dynamics and work on compressed data.

Part III

Appendix

APPENDIX A

Appendix

A.0.1 Proof of outer bound (2.3)

The proof of outer bound (2.3) is a straightforward (multicarrier) extension of the corresponding proof in the single carrier setup [1]. The proof details for (2.3) are described below.

Using Fano's inequality for Rx_1, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
& NR^{(1)} - N\varepsilon \\
& \leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}\right) \\
& = I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) \\
& \leq H\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) \\
& \leq \sum_{t=1}^{N} H\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t]\right) \\
& = \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t] = \mathbf{s}\right) \\
& \leq \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \sum_{j=1}^{M} n_j \mathbb{I}_{j \notin \mathbf{s}} + \max(n_j, k_j) \mathbb{I}_{j \in \mathbf{s}} \\
& = N \sum_{j=1}^{M} n_j + p\left(\max(n_j, k_j) - n_j\right) \\
& = Np\Delta + N \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p, \qquad \text{(A.1)}
\end{aligned}
$$

where $\Delta = \sum_{j=1}^{M} \max(n_j, k_j) + (n_j - k_j)^+ - 2n_j$. The outer bound on $R^{(2)}$ follows by symmetry and this completes the proof of outer bound (2.3).

A.0.2 Proof of outer bound (2.5)

Like the proof of (2.3), the proof of outer bound (2.5) is also a simple (multicarrier) extension of the corresponding sum rate bound in the single carrier setup [1]. The proof details are described below.

Using Fano's inequality for Rx_1 and Rx_2, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
& NR^{(1)} + NR^{(2)} - 2N\varepsilon \\
& \leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}\right) + I\left(W^{(2)}; W^{(1)}, \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)}, \mathbf{S}_{1:N}\right) \\
& = I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + I\left(W^{(2)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) \\
& = H\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - H\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) + H\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) \\
& = H\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + H\left(\mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
& = H\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + H\left(\hat{\mathbf{X}}_{1:N}^{(2)} \middle| \mathbf{V}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
& \leq \sum_{t=1}^{N} H\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t]\right) + \sum_{t=1}^{N} H\left(\hat{\mathbf{X}}^{(2)}[t] \middle| \mathbf{V}_{\mathbf{S}[t]}^{(1)}[t], \mathbf{S}[t]\right) \\
& = \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t] = \mathbf{s}\right) + \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) H\left(\hat{\mathbf{X}}^{(2)}[t] \middle| \mathbf{V}_{\mathbf{S}[t]}^{(1)}[t], \mathbf{S}[t] = \mathbf{s}\right) \\
& \leq \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \sum_{j=1}^{M} n_j \mathbb{I}_{j \notin \mathbf{s}} + \max(n_j, k_j) \mathbb{I}_{j \in \mathbf{s}} \\
& \quad + \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}\left(\mathbf{S}[t] = \mathbf{s}\right) \sum_{j=1}^{M} n_j \mathbb{I}_{j \notin \mathbf{s}} + (n_j - k_j)^{+} \mathbb{I}_{j \in \mathbf{s}} \\
& = \sum_{t=1}^{N} \sum_{j=1}^{M} n_j (1-p) + \max(n_j, k_j) p + \sum_{t=1}^{N} \sum_{j=1}^{M} n_j (1-p) + (n_j - k_j)^{+} p \\
& = Np\Delta + 2N \sum_{j=1}^{M} n_j, \qquad \text{(A.2)}
\end{aligned}
$$

where $\Delta = \sum_{j=1}^{M} \max(n_j, k_j) + (n_j - k_j)^{+} - 2n_j$. This completes the proof of outer bound (2.5).

A.0.3 Proof of outer bound (2.7)

Using Fano's inequality for Rx_1, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
& NR^{(1)} - N\varepsilon \\
& \leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)}, W^{(2)}, \mathbf{S}_{1:N}\right) \\
& = I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) - h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(1)}, W^{(2)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t], \mathbf{Y}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, W^{(1)}, W^{(2)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Z}^{(1)}[t], \mathbf{Z}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, W^{(1)}, W^{(2)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Z}^{(1)}[t], \mathbf{Z}^{(2)}[t]\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| W^{(2)}, \mathbf{S}_{1:N}\right) - 2NM\log(\pi e) \\
& = \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t], \mathbf{Y}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{Y}_{1:t-1}^{(2)}, W^{(2)}, \mathbf{S}_{1:N}\right) - 2NM\log(\pi e) \\
& \leq \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t], \mathbf{Y}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{Y}_{1:t-1}^{(2)}, W^{(2)}, \mathbf{S}[t]\right) - 2NM\log(\pi e) \\
& = \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t] = \mathbf{s})\, h\left(\mathbf{Y}^{(1)}[t], \mathbf{Y}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{Y}_{1:t-1}^{(2)}, W^{(2)}, \mathbf{S}[t] = \mathbf{s}\right) - 2NM\log(\pi e) \\
& \leq \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t] = \mathbf{s}) \times \\
& \quad \sum_{j=1}^{M} \mathbb{I}_{j \notin \mathbf{s}} \left(\log\left(\pi e\left(1 + |g_{D,j}|^2\right)\right) + \log(\pi e)\right) + \mathbb{I}_{j \in \mathbf{s}} \log\left((\pi e)^2\left(1 + |g_{D,j}|^2 + |g_{I,j}|^2\right)\right) \\
& \quad - 2NM\log(\pi e) \\
& = N \sum_{j=1}^{M} (1-p)\log\left(1 + |g_{D,j}|^2\right) + p\log\left(1 + |g_{D,j}|^2 + |g_{I,j}|^2\right). \qquad \text{(A.3)}
\end{aligned}
$$

The bound on $R^{(2)}$ follows by symmetry and this completes the proof of outer bound (2.7). We also prove a looser bound on $R^{(i)}$ as shown below (the proof for this looser bound is used in the proof of outer bounds (2.8) and (2.9)).

Using Fano's inequality for Rx_1, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
& NR^{(1)} - N\varepsilon \\
& \leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}\right) \\
& = I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - h\left(\mathbf{Y}_{1:N}^{(1)} \middle| W^{(1)}, \mathbf{S}_{1:N}\right) \\
& \leq h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - h\left(\mathbf{Y}_{1:N}^{(1)} \middle| W^{(2)}, W^{(1)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, W^{(2)}, W^{(1)}, \mathbf{S}_{1:N}\right) \\
& \leq h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h(\mathbf{Y}^{(1)}[t] | \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, W^{(2)}, W^{(1)}, \mathbf{S}_{1:N}) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Z}^{(1)}[t] \middle| \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{Y}_{1:t-1}^{(1)}, W^{(2)}, W^{(1)}, \mathbf{S}_{1:N}\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - \sum_{t=1}^{N} h\left(\mathbf{Z}^{(1)}[t]\right) \\
& = h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - NM\log(\pi e) \\
& \leq \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t]\right) - NM\log(\pi e) \\
& = \sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t] = \mathbf{s})\, h\left(\mathbf{Y}^{(1)}[t] \middle| \mathbf{S}[t] = \mathbf{s}\right) - NM\log(\pi e) \\
& \leq \left(\sum_{t=1}^{N} \sum_{\mathbf{s}} \mathbb{P}(\mathbf{S}[t] = \mathbf{s}) \sum_{j=1}^{M} \log(\pi e(1 + |g_{D,j}|^2)) \mathbb{I}_{j \notin \mathbf{s}} + \log\left(\pi e\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right)\right) \mathbb{I}_{j \in \mathbf{s}} \right) \\
& \quad - NM\log(\pi e) \\
& = \left(\sum_{t=1}^{N} \sum_{j=1}^{M} (1-p)\log\left(\pi e\left(1 + |g_{D,j}|^2\right)\right) + p\log\left(\pi e\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right)\right) \right) - NM\log(\pi e) \\
& = N \sum_{j=1}^{M} (1-p)\log\left(1 + |g_{D,j}|^2\right) + p\log\left(1 + \left(|g_{D,j}| + |g_{I,j}|\right)^2\right). \qquad \text{(A.4)}
\end{aligned}
$$

As mentioned above, (A.4) is a looser bound compared to (2.7), but the above proof is used in proving outer bounds (2.8) and (2.9).

A.0.4 Proof of outer bound (2.9)

Using Fano's inequality for Rx_1 and Rx_2, for any $\varepsilon > 0$, there exists a large enough N such that:

$$
\begin{aligned}
&NR^{(1)} + NR^{(2)} - 2N\varepsilon \\
&\leq I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}\right) + I\left(W^{(2)}; W^{(1)}, \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)}, \mathbf{S}_{1:N}\right) \\
&= I\left(W^{(1)}; \mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + I\left(W^{(2)}; \mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) - h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) \\
&\quad + h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}\right) - h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}, W^{(2)}\right) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
&\quad - h\left(\mathbf{Y}_{1:N}^{(1)}, \mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{S}_{1:N}, W^{(1)}, W^{(2)}\right) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
&\quad - \sum_{t=1}^{N} h\left(\mathbf{Y}^{(1)}[t], \mathbf{Y}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{S}_{1:N}, W^{(1)}, W^{(2)}\right) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) \\
&\quad - \sum_{t=1}^{N} h\left(\mathbf{Z}^{(1)}[t], \mathbf{Z}^{(2)}[t] \middle| \mathbf{Y}_{1:t-1}^{(1)}, \mathbf{Y}_{1:t-1}^{(2)}, \mathbf{S}_{1:N}, W^{(1)}, W^{(2)}\right) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\mathbf{Y}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) - 2NM\log(\pi e) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\hat{\mathbf{X}}_{1:N}^{(2)} \biguplus \mathbf{Z}_{1:N}^{(2)} \middle| \mathbf{Y}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) - 2NM\log(\pi e) \\
&= h\left(\mathbf{Y}_{1:N}^{(1)} \middle| \mathbf{S}_{1:N}\right) + h\left(\hat{\mathbf{X}}_{1:N}^{(2)} \biguplus \mathbf{Z}_{1:N}^{(2)} \middle| \mathbf{V}_{1:N}^{(1)} \biguplus \mathbf{Z}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) - 2NM\log(\pi e) \\
&\overset{(a)}{\leq} N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right) + p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right) \\
&\quad + h\left(\hat{\mathbf{X}}_{1:N}^{(2)} \biguplus \mathbf{Z}_{1:N}^{(2)} \middle| \mathbf{V}_{1:N}^{(1)} \biguplus \mathbf{Z}_{1:N}^{(1)}, \mathbf{S}_{1:N}, W^{(1)}\right) - NM\log(\pi e) \\
&\leq N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right) + p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right) \\
&\quad + \sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right) h\left(\hat{\mathbf{X}}^{(2)}[t] \biguplus \mathbf{Z}^{(2)}[t] \middle| \mathbf{V}_{\mathbf{s}}^{(1)}[t] \biguplus \mathbf{Z}^{(1)}[t], W^{(1)}, \mathbf{S}[t]=\mathbf{s}\right) - NM\log(\pi e)
\end{aligned}
$$

$$
\begin{aligned}
&\leq N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)\\
&\quad+\sum_{t=1}^{N}\sum_{\mathbf{s}}\mathbb{P}\left(\mathbf{S}[t]=\mathbf{s}\right)\sum_{j=1}^{M}\log(\pi e(1+|g_{D,j}|^2))\mathbb{I}_{j\notin\mathbf{s}}+\log\left(\pi e\left(1+\frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right)\right)\mathbb{I}_{j\in\mathbf{s}}\\
&\quad-NM\log(\pi e)\\
&=N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)\\
&\quad+N\sum_{j=1}^{M}(1-p)\log\left(1+|g_{D,j}|^2\right)+p\log\left(1+\frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right)\\
&=N\left(2\sum_{j=1}^{M}\log\left(1+|g_{D,j}|^2\right)+p\Delta_G\right) \qquad \text{(A.5)}
\end{aligned}
$$

where (a) follows from the proof of (A.4) (see Appendix A.0.3) and
$\Delta_G=\sum_{j=1}^{M}\log\left(1+\left(|g_{D,j}|+|g_{I,j}|\right)^2\right)+\log\left(1+\frac{|g_{D,j}|^2}{1+|g_{I,j}|^2}\right)-2\log\left(1+|g_{D,j}|^2\right)$.

A.0.5 Verification of multicarrier separability

The verification for the non-bursty case (*i.e.*, when $p\in\{0,1\}$) is straightforward, and we focus on verifying the separability for the case when $0<p<1$ (*i.e.*, when the interfering link is bursty). We first verify the separability for the case when all $\alpha_j\leq 2$, followed by the verification for the case when all $\alpha_j\geq 2$.

All $\alpha_j\leq 2$ In this case, we can re-write Δ as shown below,

$$
\begin{aligned}
\Delta&=\sum_{j=1}^{M}\max(n_j,k_j)+(n_j-k_j)^+-2n_j\\
&=\sum_{j:\alpha_j>2}(k_j-2n_j)-\sum_{j:\alpha_j\leq 1}k_j-\sum_{j:1<\alpha_j\leq 2}(2n_j-k_j)\\
&=-\sum_{j:\alpha_j\leq 1}k_j-\sum_{j:1<\alpha_j\leq 2}(2n_j-k_j). \qquad \text{(A.6)}
\end{aligned}
$$

The sum of symmetric rates across subcarriers (using the optimal single carrier scheme [1] for each subcarrier) can be written as shown below:

$$\left(\sum_{j:\alpha_j\leq 1} n_j - \frac{p}{1+p}k_j\right) + \left(\sum_{j:1<\alpha_j\leq 2} n_j - \frac{p}{1+p}(2n_j - k_j)\right) \overset{(a)}{=} \frac{p}{1+p}\Delta + \sum_{j=1}^{M} n_j = R_C, \qquad \text{(A.7)}$$

where (a) follows from (A.6). Hence, the symmetric capacity R_C (since $\Delta \leq 0$ in this case) can be achieved by simply using the optimal single carrier scheme [1] for each subcarrier.

All $\alpha_j \geq 2$ In this case,

$$\begin{aligned}\Delta &= \sum_{j=1}^{M} \max(n_j, k_j) + (n_j - k_j)^+ - 2n_j \\ &= \sum_{j:\alpha_j\geq 2} (k_j - 2n_j). \end{aligned} \qquad \text{(A.8)}$$

The sum of symmetric rates across subcarriers (using the optimal single carrier scheme [1] for each subcarrier) can be written as shown below:

$$\sum_{j=1}^{M} n_j + \frac{p}{2}(k_j - 2n_j) \overset{(a)}{=} \frac{p}{2}\Delta + \sum_{j=1}^{M} n_j = R_{NC}, \qquad \text{(A.9)}$$

where (a) follows from (A.8). Hence, the symmetric capacity R_{NC} (since $\Delta \geq 0$ in this case) can be achieved by simply using the optimal single carrier scheme [1] for each subcarrier.

A.0.6 Achievability of corner points D_1 and D_2

As shown in Figure 2.3, these corner points appear when $\Delta > 0$. We will describe the achievability of D_1 and achievability of D_2 follows by symmetry. The achievability of D_1 is similar to achieving $R_{NC} = \frac{p}{2}\Delta + \sum_{j=1}^{M} n_j$ (described in Section 2.4.3); with a slight modification for subcarriers with $\alpha_j > 2$. The additive term $\frac{p}{2}\Delta$ appears in R_{NC} because of bursty relaying in the leftover helper levels (Δ in number). For D_1, to achieve $R^{(1)} = p\Delta + \sum_{j=1}^{M} n_j$, we use an asymmetric version of bursty relaying as follows: In every block Tx_1 sends $N_B\Delta$ linear combinations of $pN_B\Delta$ fresh symbols in the leftover helper levels. Rx_2 receives $pN_B\Delta$ such linear combinations in every block; it recovers the constituent symbols and forwards them to Tx_2. In the next block, Tx_2 creates $N_B\Delta$

linear combinations of the constituent symbols sent by Rx_1 and sends them on its leftover helper levels. Rx_1 receives $pN_B\Delta$ of these linear combinations and thus recovers the constituent symbols. So compared to R_{NC}, Rx_1 now gains an additional rate $\frac{p}{2}\Delta$ but Rx_2 loses[1] rate $\frac{p}{2}\Delta$. This completes the achievability of D_1.

A.0.7 Achievability of corner points Q_1 and Q_2

Both Q_1 and Q_2 are achieved using a separation based scheme (*i.e.,* no coding across subcarriers). We first describe the achievability of Q_1; achievability of Q_2 follows by symmetry. In

$$Q_1 = \left(R^{(1)}, R^{(2)}\right) = \left(p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p,\ \sum_{j=1}^{M} (n_j - k_j)^+\right),$$

we can rewrite rate $R^{(1)}$ as follows:

$$\begin{aligned} & p\Delta + \sum_{j=1}^{M} n_j(1+p) - (n_j - k_j)^+ p \\ &= \sum_{j:\alpha_j \le 1} n_j + \sum_{j:\alpha_j > 1} n_j + (k_j - n_j)p. \end{aligned}$$

Also, from the single carrier schemes in [1], the following rate tuples $\left(R^{(1)}, R^{(2)}\right)$ are achievable for a single carrier setup:

- $\left(n_j, n_j - k_j\right)$ for $\alpha_j \le 1$.
- $\left(n_j + (k_j - n_j)p, 0\right)$ for $\alpha_j > 1$.

Clearly, achieving the above rate tuple for each subcarrier and summing rates across subcarriers leads to corner point Q_1. The achievability of Q_2 follows by symmetry.

A.0.8 Proof of Corollary 3

For $\{L \ge \frac{M}{2},\ 0 \le \alpha \le \frac{1}{2}\}$, inequality (3.7) is not active in presence of inequalities (3.5) and (3.6). This can be proved as follows.

[1]The loss stems from Tx_2 not using its leftover helper levels for its own messages; it just uses them to relay messages for Rx_1.

In this regime, inequalities (3.5), (3.6) and (3.7) can be rewritten (shown below) as (A.10), (A.11) and (A.12) respectively.

$$MR_L + (M-L)R_0 \leq M(M-L\alpha)n \tag{A.10}$$

$$R_L + R_0 \leq Mn \tag{A.11}$$

$$2R_L + R_0 \leq M(2-\alpha)n \tag{A.12}$$

Figure A.1 shows the situation in this regime[2]; it is clear that (A.12) (dashed red line in Figure A.1) is not active in presence of (A.10) and (A.11) (solid green lines in Figure A.1). Since inequalities (A.10) and (A.11) are inner bounds as well as outer bounds in this regime, we have a tight characterization.

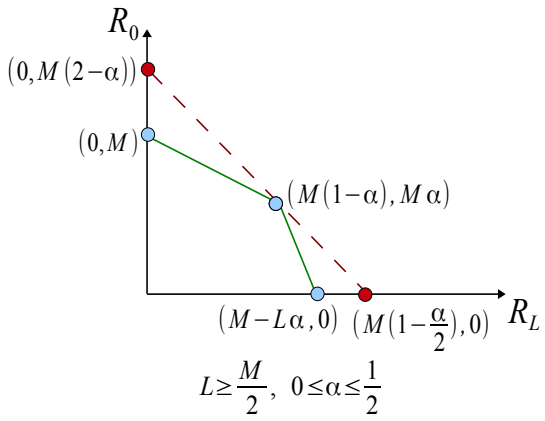


Figure A.1: Rate inequalities (normalized with respect to n) for the regime $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ in the $(R_0, R_L, 0)$-setup. Inequality (A.12) (dashed red line) is not active in presence of (A.10) and (A.11) (solid green lines).

A.0.9 Proof of Corollary 4

For $\{L \leq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$, inequality (3.10) is not active in presence of inequalities (3.8) and (3.9). This can be proved as follows.

[2]For $L \geq \frac{M}{2}$, $M - L\alpha = M(1-\frac{\alpha}{2}) + (\frac{M}{2} - L)\alpha \leq M(1-\frac{\alpha}{2})$

In this regime, inequalities (3.8), (3.9) and (3.10) can be rewritten (shown below) as (A.13), (A.14) and (A.15) respectively.

$$R_L + R_M \leq (M - L\alpha)n \tag{A.13}$$

$$R_M \leq M(1-\alpha)n \tag{A.14}$$

$$MR_L + 2(M-L)R_M \leq M(M-L)(2-\alpha)n \tag{A.15}$$

Figure A.2 shows the situation in this regime; it is clear that (A.15) (dashed red line in Figure A.2) is not active in presence of (A.13) and (A.14) (solid green lines in Figure A.2). Since inequalities (A.13) and (A.14) are inner bounds as well as outer bounds in this regime, we have a tight characterization.

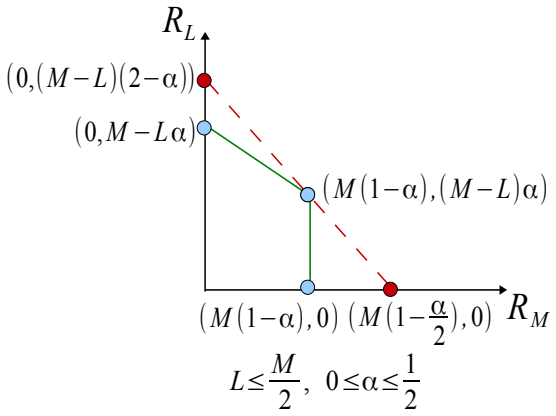


Figure A.2: Rate inequalities (normalized with respect to n) for the regime $\{L \leq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ in the $(0, R_L, R_M)$-setup. Inequality (A.15) (dashed red line) is not active in presence of (A.13) and (A.14) (solid green lines).

A.0.10 Proof of Corollary 5

In the regime $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{2}{3}\}$, inequality (3.13) is not active in presence of inequalities (3.11) and (3.12). This can be shown as follows.

For this regime, inequalities (3.11), (3.12) and (3.13) can be rewritten (shown below) as (A.16),

(A.17) and (A.18) respectively.

$$\begin{aligned} R_L + R_M &\leq ((M-L)(2-\alpha) \\ &\quad +(2L-M)\max(1-\alpha,\alpha))n \qquad \text{(A.16)} \\ R_M &\leq M\max(1-\alpha,\alpha)n \qquad \text{(A.17)} \\ R_L + R_M &\leq \frac{M}{2}(2-\alpha)n \qquad \text{(A.18)} \end{aligned}$$

To show (A.18) is not active in presence of (A.16) and (A.17), it is sufficient to prove (A.16) *dominates*[3] (A.18) in this regime. We prove this in two steps as shown below (analysis for $0 \leq \alpha \leq \frac{1}{2}$ followed by analysis for $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$).

For $0 \leq \alpha \leq \frac{1}{2}$, (A.16) can be simplified to

$$R_L + R_M \leq (M - L\alpha)n$$

Since $L \geq \frac{M}{2}$; $(M-L\alpha) \leq \frac{M}{2}(2-\alpha)$. Thus, (A.16) dominates (A.18) for $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$.

For $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$, (A.16) can be simplified to

$$\begin{aligned} R_L + R_M &\leq (M(2-2\alpha) - L(2-3\alpha))n \\ &= (\frac{M}{2}(2-\alpha) + (\frac{M}{2} - L)(2-3\alpha))n \end{aligned}$$

Since $\alpha \leq \frac{2}{3}$ and $\frac{M}{2} \leq L$, $\frac{M}{2}(2-\alpha) + (\frac{M}{2} - L)(2-3\alpha) \leq \frac{M}{2}(2-\alpha)$. Thus, (A.16) dominates (A.18) for $\{L \geq \frac{M}{2},\ \frac{1}{2} \leq \alpha \leq \frac{2}{3}\}$.

As shown above, (A.16) dominates (A.18) for both $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{1}{2}\}$ and $\{L \geq \frac{M}{2},\ \frac{1}{2} \leq \alpha \leq \frac{2}{3}\}$. Since inequalities (A.16) and (A.17) are inner bounds as well as outer bounds, we have a tight characterization in the regime $\{L \geq \frac{M}{2},\ 0 \leq \alpha \leq \frac{2}{3}\}$.

A.0.11 Effect of error in initial state estimate

In the single observer setup described in Section 4.2.1, we assumed $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$. We show in this appendix that all the secrecy guarantees remain unchanged even if this assumption is not satisfied.

[3] gives a smaller bound for $R_L + R_M$

In general, the initial state estimate can be written as:

$$\hat{\mathbf{x}}(0) = \mathbf{x}(0) - \mathbf{e}(0) \tag{A.19}$$

where $\mathbf{e}(0) \neq \mathbf{0}$ is the error in the state estimate at time $t = 0$. Since we are using Luenberger observers, the dynamics of the estimate error $\mathbf{e}(t)$ is given by:

$$\mathbf{e}(t+1) = \mathbf{x}(t+1) - \hat{\mathbf{x}}(t+1) = (\mathbf{A} - \mathbf{LC})\,\mathbf{e}(t) \tag{A.20}$$

while the state evolution is governed by:

$$\mathbf{x}(t) = \mathbf{A}_{cl}^{t}\mathbf{x}(0) + \sum_{j=0}^{t-1} \mathbf{A}_{cl}^{t-1-j}\mathbf{B}\mathbf{r}(j) - \sum_{j=0}^{t-1} \mathbf{A}_{cl}^{t-1-j}\mathbf{B}\mathbf{K}\mathbf{e}(j)\ . \tag{A.21}$$

Let us now describe the implications of $\mathbf{e}(0) \neq \mathbf{0}$ for a 1-passive adversary in the 2-observer setup described in Section 4.3.1. It can be easily shown that the dynamics of the plant in the 2-observer setup is same as with a single observer (as shown in (A.21)); in this case $\mathbf{e}(0) = \mathbf{x}(0) - \hat{\mathbf{x}}_1(0) - \hat{\mathbf{x}}_2(0)$ where $\hat{\mathbf{x}}_1(0)$ and $\hat{\mathbf{x}}_2(0)$ are arbitrary initial state estimates at observers 1 and 2. The information an adversary tapping observer 1 receives from the encoded outputs $\mathbf{y}_1(1), \mathbf{y}_1(2), \ldots \mathbf{y}_1(l)$ can be written as shown below:

$$\begin{aligned}
&\begin{bmatrix} \mathbf{C}\mathbf{x}(1) \\ \mathbf{C}\mathbf{x}(2) \\ \vdots \\ \mathbf{C}\mathbf{x}(l) \end{bmatrix} + 2 \begin{bmatrix} \boldsymbol{\delta}(1) \\ \boldsymbol{\delta}(2) \\ \vdots \\ \boldsymbol{\delta}(l) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{C}\mathbf{A}_{cl} \\ \mathbf{C}\mathbf{A}_{cl}^{2} \\ \vdots \\ \mathbf{C}\mathbf{A}_{cl}^{l} \end{bmatrix} \mathbf{x}(0) + \begin{bmatrix} \mathbf{CB} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{C}\mathbf{A}_{cl}\mathbf{B} & \mathbf{CB} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{C}\mathbf{A}_{cl}^{l-1}\mathbf{B} & \mathbf{C}\mathbf{A}_{cl}^{l-2}\mathbf{B} & \ldots & \mathbf{CB} \end{bmatrix} \mathbf{r}_{0:l-1} \\
&\quad + \begin{bmatrix} \mathbf{CBK} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{C}\mathbf{A}_{cl}\mathbf{BK} & \mathbf{CBK} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{C}\mathbf{A}_{cl}^{l-1}\mathbf{BK} & \mathbf{C}\mathbf{A}_{cl}^{l-2}\mathbf{BK} & \ldots & \mathbf{CBK} \end{bmatrix} \mathbf{e}_{0:l-1} + 2\boldsymbol{\delta}_{1:l}\ .
\end{aligned}$$

If we assume that the adversary knows $\mathbf{e}(0)$ (and hence knows $\mathbf{x}(0)$ and $\mathbf{e}(1), \mathbf{e}(2), \ldots \mathbf{e}(l-1)$), the resulting set of equations seen by the adversary is same as in Section 4.3.3 where the assumption $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$ was used. Hence, any secrecy guarantee under the assumption $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$ also holds when $\mathbf{x}(0) \neq \hat{\mathbf{x}}(0)$. This observation can be easily generalized for all the results in this paper, hence justifying the assumption $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$.

A.0.12 Proof details for Theorem 9

A.0.12.1 Proof of (5.16) using LLN

$$
\begin{aligned}
& tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}\mathbf{z}_{\mathbf{s}_g}^T\right)\right) \\
&= \frac{1}{n}\sum_{l=0}^{n-1} tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - \frac{1}{N}\sum_{t\in G} tr\left(\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\right) \\
&\overset{(a)}{=} \sum_{l=0}^{n-1}\frac{1}{n}\left(tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - \frac{1}{N_B}\sum_{t\in G_l} tr\left(\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\right)\right) \\
&\leq \sum_{l=0}^{n-1}\frac{1}{n}\left|tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - \frac{1}{N_B}\sum_{t\in G_l} tr\left(\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\right)\right| \\
&\overset{(b)}{\leq} \varepsilon_1,
\end{aligned}
$$

where (a) follows from partitioning time window G (of size N) into n groups $G_0, G_1, \ldots G_{n-1}$ (each of size N_B) such that $G_l = \{t | ((t-t_1) \ mod \ n) = l\}$, and (b) follows w.h.p. from LLN (for different time indices in G_l, $tr\left(\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\right)$ corresponds to i.i.d. realizations of the same random variable).

A.0.12.2 Cross term analysis and proof of (5.17)

The cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)$ can be written down as a sum of n terms as shown below:

$$
\begin{aligned}
2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right) &\overset{(a)}{=} \frac{2}{n}\sum_{l=0}^{n-1}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t)\right) \\
&= \frac{2}{n}\sum_{l=0}^{n-1}\zeta_l, \qquad \text{(A.22)}
\end{aligned}
$$

where (a) follows from partitioning time window G (of size N) into n groups $G_0, G_1, \ldots G_{n-1}$ (each of size N_B) such that $G_l = \{t | ((t-t_1) \ mod \ n) = l\}$. Now, we will show that each ζ_l has zero mean and vanishingly small variance for large enough N. The mean analysis can be done as shown below:

$$\begin{aligned}\mathbb{E}(\zeta_l) &= \mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t)\right) \\ &\overset{(a)}{=} \frac{1}{N_B}\sum_{t\in G_l}\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}^T(t)\right)\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t)\right) = 0, \end{aligned} \tag{A.23}$$

where (a) follows from the independence of $\mathbf{e}_{\mathbf{s}}(t)$ from $\mathbf{z}_{\mathbf{s}_g}^T(t)$ (due to assumptions 4 and 5). This implies that the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)$ has zero mean. As a consequence of (A.23) and (5.16),

$$\begin{aligned}2\varepsilon_1 &\geq \mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathscr{O}_{\mathbf{s}_g}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T - \mathbf{P}_{\mathbf{s}}^\star\right)\mathscr{O}_{\mathbf{s}_g}^T\right)\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T - \mathbf{P}_{\mathbf{s}}^\star\right)\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)\right)\right) \\ &\overset{(a)}{\geq} \lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T - \mathbf{P}_{\mathbf{s}}^\star\right)\right)\right), \end{aligned} \tag{A.24}$$

where (a) follows from Lemma 3 (discussed in Appendix A.0.13). Using (A.24),

$$\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\right)\right) \leq \frac{2\varepsilon_1}{\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)} + tr\left(\mathbf{P}_{\mathbf{s}}^\star\right). \tag{A.25}$$

We will use the intermediate result (A.25) in the variance analysis of $\zeta_l = \frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t) = \frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)$ described below.

For any $\varepsilon_2 > 0$, there exists a large enough N_B such that:

$$\begin{aligned}&Var\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right) \\ &= \mathbb{E}\left(\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right) \\ &\quad - \left(\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\right)^2\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} \mathbb{E}\left(\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad+\mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l,\,t<t'}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)\\
&\overset{(b)}{=}\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad+\frac{2}{N_B^2}\sum_{t,t'\in G_l,\,t<t'}\mathbb{E}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T\right)\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}(t')\right)\\
&\overset{(c)}{=}\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^T\mathbf{e}_{\mathbf{s}}(t)\right)\\
&\overset{(d)}{=}\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}tr\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^T\mathbf{e}_{\mathbf{s}}(t)\right)\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}tr\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^T\mathbf{e}_{\mathbf{s}}(t)\mathbf{e}_{\mathbf{s}}^T(t)\right)\right)\\
&\overset{(e)}{=}\frac{1}{N_B^2}\sum_{t\in G_l}tr\left(\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\right)\mathbb{E}\left(\mathbf{e}_{\mathbf{s}}(t)\mathbf{e}_{\mathbf{s}}^T(t)\right)\right)\\
&\overset{(f)}{\leq}\frac{1}{N_B^2}\sum_{t\in G_l}\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)tr\left(\mathbb{E}\left(\mathbf{e}_{\mathbf{s}}(t)\mathbf{e}_{\mathbf{s}}^T(t)\right)\right)\\
&=\frac{\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)}{N_B}\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}tr\left(\mathbf{e}_{\mathbf{s}}(t)\mathbf{e}_{\mathbf{s}}^T(t)\right)\right)\\
&=\frac{\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)}{N_B}\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}^T(t)\mathbf{e}_{\mathbf{s}}(t)\right)\\
&=\frac{n\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)}{N_B}\mathbb{E}\left(\frac{1}{N}\sum_{t\in G_l}\mathbf{e}_{\mathbf{s}}{}^T(t)\mathbf{e}_{\mathbf{s}}(t)\right)\\
&\leq\frac{n\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)\mathbb{E}\left(\frac{1}{N}\sum_{t\in G}\mathbf{e}_{\mathbf{s}}^T(t)\mathbf{e}_{\mathbf{s}}(t)\right)}{N_B}
\end{aligned}
$$

$$\overset{(g)}{\leq} \varepsilon_2, \tag{A.26}$$

where (a) follows from (A.23), (b) follows from the independence of $\mathbf{z}_{\mathbf{s}_g}(t')$ from $\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T$ for $t' > t$, (c) follows from $\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}(t')\right) = \mathbf{0}$, (d) follows from $\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)$ being a scalar, (e) follows from the independence of $\mathbf{z}_{\mathbf{s}_g}(t)$ from $\mathbf{e}_{\mathbf{s}}(t)$, (f) follows from Lemma 3 (discussed in Appendix A.0.13) with,

$$\begin{aligned}&\lambda_{max}\left(\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^T\right)\right)\\ &= \lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right).\end{aligned} \tag{A.27}$$

Finally, (g) follows from (A.25) for large enough N_B. This completes the variance analysis of ζ_l, and clearly ζ_l has vanishingly small variance as $N_B \to \infty$. As a consequence, the variance of the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right) = \frac{2}{n}\sum_{l=0}^{n-1}\zeta_l$ is also vanishingly small for $N_B \to \infty$ (follows from the Cauchy-Schwarz inequality). Since the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)$ has zero mean and vanishingly small variance, by the Chebyshev inequality, $\left|2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)\right| \leq \varepsilon_1$ holds w.h.p., and this completes the proof of (5.17).

A.0.13 Bounds on the trace of product of symmetric matrices

A useful lemma from [47] can be described as follows.

Lemma 3 *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are two symmetric matrices in* $\mathbb{R}^{n\times n}$*, and* $\mathbf{B}$ *is positive semi-definite:*

$$\lambda_{min}(\mathbf{A})tr(\mathbf{B}) \leq tr(\mathbf{AB}) \leq \lambda_{max}(\mathbf{A})tr(\mathbf{B}). \tag{A.28}$$

A.0.14 Results for the filtering version of the Kalman filter

As stated in Section 5.2, we use the prediction version of the Kalman filter for deriving our main results in this paper. Proving similar results for the filtering version can be done using the same techniques used for the prediction version. In the remainder of this Section, we will first describe the filtering setup with some additional notation, and then describe our effective attack detector for the filtering setup.

A.0.14.1 Filtering setup and additional notation

The state estimate update rule for the filtering version of the Kalman filter (in steady state) is as shown below [44]:

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^{(P)}(t) + \mathbf{L}\left(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}^{(P)}(t)\right), \tag{A.29}$$

$$\hat{\mathbf{x}}^{(P)}(t+1) = \mathbf{A}\hat{\mathbf{x}}(t), \tag{A.30}$$

where $\mathbf{L}$ is the steady state Kalman filter gain, and $\hat{\mathbf{x}}(t)$ is the (filtered) state estimate at time t (which also depends on the output at time t). We denote by $\mathbf{L_s}$ the steady state Kalman filter gain when only outputs from sensor subset $\mathbf{s} \subseteq \{1,2,\ldots,p\}$ are used, and use $\mathbf{F}_\mathbf{s}^\star$ for the corresponding filtering error covariance matrix. In addition, we use the following notation for the senor noise in subset $\mathbf{s} = \{i_1, i_2, \ldots, i_{|\mathbf{s}|}\}$ at time t:

$$\tilde{\mathbf{v}}_\mathbf{s}(t) = \begin{bmatrix} v_{i_1}(t) \\ v_{i_2}(t) \\ \vdots \\ v_{i_{|\mathbf{s}|}}(t) \end{bmatrix}, \tag{A.31}$$

and define $\Delta_\mathbf{s}$ as shown below:

$$\Delta_\mathbf{s} = \mathbb{E}\left(\mathbf{z}_\mathbf{s}(t)\tilde{\mathbf{v}}_\mathbf{s}^T(t)\mathbf{L}_\mathbf{s}^T\mathscr{O}_\mathbf{s}^T\right), \tag{A.32}$$

where $\mathbf{z}_\mathbf{s}(t) = \mathbf{J}_\mathbf{s}\bar{\mathbf{w}}(t) + \bar{\mathbf{v}}_\mathbf{s}(t)$ (as defined for the prediction setup in Section 5.3). Note that $\Delta_\mathbf{s}$ can be easily expressed (after evaluating the expectation in (A.32)) in terms of σ_v^2, $\mathbf{L}_\mathbf{s}^T$, and $\mathscr{O}_\mathbf{s}^T$; we define $\Delta_\mathbf{s}$ just for conveniently describing our detector and proving its performance guarantees. In the prediction setup, we limited the the adversary through Assumptions 1-5; however, in the filtering setup, to show similar results, we require stronger versions of Assumptions 4 and 5 as described below:

Assumption 6 *The adversary's knowledge at time t is statistically independent of $\mathbf{w}(t')$ for $t' \geq t$, i.e., $\mathbf{a}(t)$ is statistically independent of $\{\mathbf{w}(t')\}_{t' \geq t}$.*

Assumption 7 *For an attack-free sensor $i \in \{1,2,\ldots,p\} \setminus \boldsymbol{\kappa}$, the adversary's knowledge at time t (and hence $\mathbf{a}(t)$) is statistically independent of $\{v_i(t')\}_{t' \geq t}$.*

Using these assumptions, we define the effective attack detection problem for the filtering setup as follows.

Definition 3 ($(\varepsilon,\mathbf{s})$-Effective Attack (filtering)) *Consider the linear dynamical system under attack $\boldsymbol{\Sigma_a}$ as defined in* (5.2)*, and a k-adversary satisfying Assumptions 1-3 and Assumptions 6-7. For the set of sensors $\mathbf{s}$, an $\varepsilon > 0$, and a large enough $N \in \mathbb{N}$, an attack signal is called $(\varepsilon,\mathbf{s})$-effective at time t_1 if the following bound holds:*

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e_s}\mathbf{e_s}^T\right)\right) > tr(\mathbf{F_s^\star}) + \varepsilon,$$

where $\mathbf{e_s}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_\mathbf{s}(t)$ (with $\hat{\mathbf{x}}_\mathbf{s}(t)$ being the filtered state estimate).

Note that, compared to Definition 1 for the prediction setup, we have just replaced $\mathbf{P_s^\star}$ by $\mathbf{F_s^\star}$ as shown above. In the following subsections, we describe the effective attack detector for the filtering setup and prove its performance guarantees.

A.0.14.2 ε-effective attack detector for filtering setup

The effective attack detector for the filtering setup is described in Algorithm 5. Compared to Algorithm 1 for the prediction setup, Algorithm 5 mainly differs in the residue test; $\mathbf{F_s^\star}$ is used in place of $\mathbf{P_s^\star}$, and the extra terms $-\Delta_\mathbf{s} - \Delta_\mathbf{s}^T$ account for the dependence of the estimation error at time t on the sensor noise at time t (details in the following subsection on performance guarantees). Note that the expected value of $\mathbf{r_s}(t)\mathbf{r_s}^T(t)$ in the absence of attacks is exactly equal to $\mathscr{O}_\mathbf{s}\mathbf{F_s^\star}\mathscr{O}_\mathbf{s}^T + \mathbf{M_s} - \Delta_\mathbf{s} - \Delta_\mathbf{s}^T$; as in the prediction version, the residue test basically checks if the sample average of $\mathbf{r_s}(t)\mathbf{r_s}^T(t)$ in the presence of attacks is close to its expected value in the absence of attacks.

A.0.14.3 Performance guarantees for Algorithm 5

The following lemma states the performance guarantees for Algorithm 5 in the context of detecting ε-effective attacks in the filtering setup.

Algorithm 5 FILTERING ATTACK-DETECT$(\mathbf{s}, t_1)$

1: Run a Kalman filter that uses all measurements from sensors indexed by $\mathbf{s}$ until time t_1 and compute the estimate $\hat{\mathbf{x}}_{\mathbf{s}}(t_1) \in \mathbb{R}^n$.

2: Recursively repeat the previous step $N-1$ times to calculate all estimates $\hat{\mathbf{x}}_{\mathbf{s}}(t) \in \mathbb{R}^n$, $\forall t \in G = \{t_1, t_1+1, \ldots, t_1+N-1\}$.

3: For time $t \in G$, calculate the *block* residue:

$$\mathbf{r}_{\mathbf{s}}(t) = \bar{\mathbf{y}}_{\mathbf{s}}(t) - \mathscr{O}_{\mathbf{s}} \hat{\mathbf{x}}_{\mathbf{s}}(t) \quad \forall t \in G.$$

4: **if** block residue test defined below holds,

$$\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}}\mathbf{r}_{\mathbf{s}}^T\right) - \left(\mathscr{O}_{\mathbf{s}}\mathbf{F}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}}^T + \mathbf{M}_{\mathbf{s}} - \Delta_{\mathbf{s}} - \Delta_{\mathbf{s}}^T\right) \preccurlyeq \eta\ \mathbf{1}_{n|\mathbf{s}|}\mathbf{1}_{n|\mathbf{s}|}^T, \tag{A.33}$$

where $0 < \eta \leq \left(\frac{\lambda_{\min,\mathbf{s}\setminus k}}{3n(|\mathbf{s}|-k)}\right)\varepsilon$, **then**

5: assert $\hat{d}_{\text{attack},\mathbf{s}}(t_1) := 0$

6: **else**

7: assert $\hat{d}_{\text{attack},\mathbf{s}}(t_1) := 1$

8: **end if**

9: **return** $(\hat{d}_{\text{attack},\mathbf{s}}(t_1), \{\hat{\mathbf{x}}_{\mathbf{s}}(t)\}_{t\in G})$

Lemma 4 *Let the linear dynamical system as defined in* (5.2) *be* $2k$*-sparse observable. Consider a* k*-adversary satisfying Assumptions* $1-3$ *and Assumptions* $6-7$*, and a sensor subset* $\mathbf{s} \subseteq \{1,2,\ldots,p\}$ *with* $|\mathbf{s}| \geq p-k$*. For any* $\varepsilon > 0$ *and* $\delta > 0$*, there exists a large enough time window length* N *such that when Algorithm 5 terminates with* $\hat{d}_{attack,\mathbf{s}}(t_1) = 0$*, the following probability bound holds:*

$$\mathbb{P}\Big(tr\left(\mathbb{E}_{t_1,N}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right) - \mathbf{F}_{\mathbf{s}}^{\star}\right) \leq \varepsilon\Big) \geq 1-\delta, \tag{A.34}$$

where $\mathbf{e}_{\mathbf{s}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_{\mathbf{s}}(t)$ *is the (filtering) state estimation error.*

Proof 9 *The proof is similar to that for the prediction version (Lemma 1). The main difference lies*

in the cross term analysis; it is more involved than in the prediction version due to the dependence of estimation error at time t on the sensor noise at time t. We describe the proof details below.

Since we assume that the set $\mathbf{s}$ *has cardinality* $|\mathbf{s}| \geq p-k$*, we can conclude that there exists a subset* $\mathbf{s}_g \subset \mathbf{s}$ *with cardinality* $|\mathbf{s}_g| \geq p-2k$ *sensors such that all its sensors are attack-free (subscript* g *in* $\mathbf{s}_g$ *stands for good sensors in* $\mathbf{s}$*). Hence, by decomposing the set* $\mathbf{s}$ *into an attack-free set* $\mathbf{s}_g$ *and a potentially attacked set* $\mathbf{s} \setminus \mathbf{s}_g$*, we can conclude that, after a permutation similarity transformation for* (A.33)*, the following holds for the attack-free subset* $\mathbf{s}_g$*:*

$$\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right) - \mathscr{O}_{\mathbf{s}_g}\mathbf{F}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T - \mathbf{M}_{\mathbf{s}_g} + \Delta_{\mathbf{s}_g} + \Delta_{\mathbf{s}_g}^T \preccurlyeq \eta\ \mathbf{1}_{n(|\mathbf{s}|-k)}\mathbf{1}_{n(|\mathbf{s}|-k)}^T, \tag{A.35}$$

where $\Delta_{\mathbf{s}_g} = \mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\right)$*. Therefore,*

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right) - \mathscr{O}_{\mathbf{s}_g}\mathbf{F}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T - \mathbf{M}_{\mathbf{s}_g} + \Delta_{\mathbf{s}_g} + \Delta_{\mathbf{s}_g}^T\right) \leq n(|\mathbf{s}|-k)\eta = \varepsilon_1. \tag{A.36}$$

As in the prediction version, we can rewrite $tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right)\right)$ *as:*

$$\begin{aligned} &tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{r}_{\mathbf{s}_g}\mathbf{r}_{\mathbf{s}_g}^T\right)\right) \\ &\quad = tr\left(\mathscr{O}_{\mathbf{s}_g}\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\mathscr{O}_{\mathbf{s}_g}^T\right) + tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}\mathbf{z}_{\mathbf{s}_g}^T\right)\right) \\ &\quad + 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}\right). \end{aligned} \tag{A.37}$$

By combining (A.36) *and* (A.37)*:*

$$\begin{aligned} &tr\left(\mathscr{O}_{\mathbf{s}_g}\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\mathscr{O}_{\mathbf{s}_g}^T - \mathscr{O}_{\mathbf{s}_g}\mathbf{F}_{\mathbf{s}}^{\star}\mathscr{O}_{\mathbf{s}_g}^T\right) \\ &\qquad \leq tr\left(\mathbf{M}_{\mathbf{s}_g}\right) - tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}\mathbf{z}_{\mathbf{s}_g}^T\right)\right) + \varepsilon_1 \\ &\qquad\quad - 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}\right) - 2tr\left(\Delta_{\mathbf{s}_g}\right) \\ &\qquad \overset{(a)}{\leq} 2\varepsilon_1 - 2\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}\right) - 2tr\left(\Delta_{\mathbf{s}_g}\right) \qquad (A.38) \\ &\qquad \overset{(b)}{\leq} 3\varepsilon_1, \qquad (A.39) \end{aligned}$$

where (a) *follows w.h.p. due to the law of large numbers (LLN) for large enough N (as shown in Appendix A.0.12.1), and* (b) *follows w.h.p. by showing that the cross term* $2\mathbb{E}_{N,t_1}\left(\mathbf{e}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}\right)$ *has mean equal to* $-2tr\left(\Delta_{\mathbf{s}_g}\right)$ *and vanishingly small variance for large enough N. The cross term analysis is described in detail in Appendix A.0.15. Using* (A.39)*, the following holds:*

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T-\mathbf{F}_{\mathbf{s}}^{\star}\right)\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)\leq 3\varepsilon_1, \tag{A.40}$$

and hence, we get the following bound which completes the proof:

$$tr\left(\mathbb{E}_{N,t_1}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)-\mathbf{F}_{\mathbf{s}}^{\star}\right)\overset{(c)}{\leq}\frac{3\varepsilon_1}{\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)}\overset{(d)}{\leq}\frac{3\varepsilon_1}{\lambda_{\min,\mathbf{s}\setminus k}}\leq\varepsilon \tag{A.41}$$

where (c) *follows from Lemma 3 in Appendix A.0.13 and* (d) *follows from the definition of* $\lambda_{\min,\mathbf{s}\setminus k}$*. Note that, it follows from* $|\mathbf{s}_g|\geq p-2k$ *and 2k-sparse observability, that both* $\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)$ *and* $\lambda_{\min,\mathbf{s}\setminus k}$ *are bounded away from zero.*

Using Lemma 4, deriving results for secure state estimation in the filtering setup is straightforward, and we skip the details for brevity.

A.0.15 Cross term analysis for filtering and proof of (A.39)

As in the prediction setup, the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)$ can be written down as a sum of n terms as shown below:

$$\begin{aligned}2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)&\overset{(a)}{=}\frac{2}{n}\sum_{l=0}^{n-1}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t)\right)\\&=\frac{2}{n}\sum_{l=0}^{n-1}\zeta_l,\end{aligned}$$

where (a) follows from partitioning time window G (of size N) into n groups $G_0,G_1,\ldots G_{n-1}$ (each of size N_B) such that $G_l=\{t|\,((t-t_1)\ mod\ n)=l\}$. Now, we will show that each ζ_l has mean equal to $-tr\left(\Delta_{\mathbf{s}_g}\right)$ and vanishingly small variance for large enough N. The mean analysis can be done as shown below:

$$\mathbb{E}\left(\zeta_l\right)=\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}(t)\right)$$

$$
\begin{aligned}
&\overset{(a)}{=} \mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t)-\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)\right)\\
&=\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\tilde{\mathbf{e}}_{\mathbf{s}}(t)\right)\\
&\quad-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)\\
&\overset{(b)}{=}\frac{1}{N_B}\sum_{t\in G_l}\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}^T(t)\right)\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}\tilde{\mathbf{e}}_{\mathbf{s}}(t)\right)\\
&\quad-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)\\
&=-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)\\
&=-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}tr\left(\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)\right)\\
&=-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}tr\left(\left(\mathscr{O}_{\mathbf{s}_g}\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\right)\\
&=-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}tr\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\right)\\
&=-\mathbb{E}\left(\frac{1}{N_B}\sum_{t\in G_l}tr\left(\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\right)\right)\\
&=-\frac{1}{N_B}\sum_{t\in G_l}tr\left(\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\right)\right)\\
&=-\frac{1}{N_B}\sum_{t\in G_l}tr\left(\Delta_{\mathbf{s}_g}\right)\\
&=-tr\left(\Delta_{\mathbf{s}_g}\right),
\end{aligned}
\tag{A.42}
$$

where (a) follows from expressing $\mathbf{e}_{\mathbf{s}}(t)$ as $\tilde{\mathbf{e}}_{\mathbf{s}}(t)-\mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)$ (*i.e.*, separating out the sensor noise at time t component in $\mathbf{e}_{\mathbf{s}}(t)$), and (b) follows from the independence of $\tilde{\mathbf{e}}_{\mathbf{s}}(t)$ from $\mathbf{z}_{\mathbf{s}_g}^T(t)$ (follows from assumptions 6 and 7). This implies that the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\mathbf{e}_{\mathbf{s}}\right)$ has mean equal to $-2tr\left(\Delta_{\mathbf{s}_g}\right)$. Also, using (A.42) and (A.38),

$$
2\varepsilon_1\geq\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathscr{O}_{\mathbf{s}_g}\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T-\mathbf{F}_{\mathbf{s}}^{\star}\right)\mathscr{O}_{\mathbf{s}_g}^T\right)\right)\right)
$$

$$= \mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\left(\mathbf{e_s}\mathbf{e_s}^T - \mathbf{F_s^\star}\right)\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)\right)\right)$$
$$\overset{(a)}{\geq} \lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathbf{e_s}\mathbf{e_s}^T - \mathbf{F_s^\star}\right)\right)\right), \tag{A.43}$$

where (a) follows from Lemma 3 (discussed in Appendix A.0.13). Using (A.43),

$$\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathbf{e_s}\mathbf{e_s}^T\right)\right)\right) \leq \frac{2\varepsilon_1}{\lambda_{min}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathscr{O}_{\mathbf{s}_g}\right)} + tr\left(\mathbf{F_s^\star}\right). \tag{A.44}$$

We will use the above intermediate result in the variance analysis done below.

The variance analysis for ζ_l can be done as shown below:

$$\mathbb{E}\left(\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\mathbf{e_s}(t)\right)^2\right)$$
$$= \mathbb{E}\left(\left(\frac{1}{N_B}\sum_{t\in G_l}\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)$$
$$= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)$$
$$+ \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{e_s}^T(t')\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)$$
$$= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)$$
$$+ \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)$$
$$- \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t')\mathbf{L_s}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)$$
$$\overset{(a)}{=} \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)$$
$$+ \frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\mathbb{E}\left(\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T\right)\mathbb{E}\left(\mathbf{z}_{\mathbf{s}_g}(t')\right)$$
$$- \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t')\mathbf{L_s}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)$$
$$= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e_s}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right) + \mathbf{0}$$

$$
\begin{aligned}
&- \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'} \mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t')\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\quad - \frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\Big(\mathbb{E}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\qquad\qquad \times \mathbb{E}\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t')\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right)\Big)\\
&\overset{(b)}{=} \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\quad + \frac{2}{N_B^2}\sum_{t,t'\in G_l, t<t'}\left(tr\left(\Delta_{\mathbf{s}_g}\right)\right)^2\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\quad + \frac{N_B(N_B-1)}{N_B^2}\left(tr\left(\Delta_{\mathbf{s}_g}\right)\right)^2, \qquad \text{(A.45)}
\end{aligned}
$$

where (a) follows from independence of $\mathbf{z}_{\mathbf{s}_g}(t')$ from $\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t')\mathscr{O}_{\mathbf{s}_g}^T$ for $t<t'$, and (b) follows from $\mathbb{E}\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t')\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t')\right) = tr\left(\Delta_{\mathbf{s}_g}\right)$ and $\mathbb{E}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right) = -tr\left(\Delta_{\mathbf{s}_g}\right)$. Now, we focus on analyzing the first term in (A.45) as shown below. For any $\varepsilon_2 > 0$, there exists a large enough N_B such that:

$$
\begin{aligned}
&\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\mathbf{e}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t) - \mathbf{L}_{\mathbf{s}}\tilde{\mathbf{v}}_{\mathbf{s}}(t)\right)^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t) - \tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\right)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t) - \tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\quad - \mathbb{E}\left(\frac{2}{N_B^2}\sum_{t\in G_l}\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&-\mathbb{E}\left(\frac{2}{N_B^2}\sum_{t\in G_l}\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&+\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&-2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&+\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&-2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&+\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\tilde{\mathbf{e}}_{\mathbf{s}}(t)\right)\\
&-2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&+\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&=\mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l}tr\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\tilde{\mathbf{e}}_{\mathbf{s}}(t)\right)\right)\\
&-2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&+\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}\left(\frac{1}{N_B^2}\sum_{t\in G_l} tr\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\right)\right)\\
&\quad -2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad +\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \frac{1}{N_B^2}\sum_{t\in G_l} tr\left(\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\mathbf{z}_{\mathbf{s}_g}^T(t)\mathscr{O}_{\mathbf{s}_g}\right)\right)\\
&\quad -2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad +\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&= \frac{1}{N_B^2}\sum_{t\in G_l} tr\left(\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)\right)\\
&\quad -2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad +\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\overset{(a)}{\leq} \frac{\lambda_{max}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{M}_{\mathbf{s}_g}\mathscr{O}_{\mathbf{s}_g}\right)}{N_B^2}\sum_{t\in G_l} tr\left(\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}(t)\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)\right)\\
&\quad -2\frac{\left(\sum_{t\in G_l}\frac{\mathbb{E}\left(\tilde{\mathbf{e}}_{\mathbf{s}}^T(t)\right)}{N_B}\right)}{N_B}\mathbb{E}\left(\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)\\
&\quad +\frac{1}{N_B}\mathbb{E}\left(\left(\tilde{\mathbf{v}}_{\mathbf{s}}^T(t)\mathbf{L}_{\mathbf{s}}^T\mathscr{O}_{\mathbf{s}_g}^T\mathbf{z}_{\mathbf{s}_g}(t)\right)^2\right)\\
&\overset{(b)}{\leq} \varepsilon_2,
\end{aligned}
\tag{A.46}
$$

where (a) follows from Lemma 3 (discussed in Appendix A.0.13), and (b) follows for large enough N_B from the boundedness of $\mathbb{E}\left(\mathbb{E}_{N,t_1}\left(tr\left(\mathbf{e}_{\mathbf{s}}\mathbf{e}_{\mathbf{s}}^T\right)\right)\right)$ as shown in (A.44).

Using (A.46) and (A.45), for any $\varepsilon_3 > 0$, there exists a large enough N_B such that:

$$Var(\zeta_l) = \mathbb{E}(\zeta_l^2) - (\mathbb{E}(\zeta_l))^2$$

$$\overset{(a)}{\leq} \varepsilon_3 + \left(tr\left(\Delta_{\mathbf{s}_g}\right)\right)^2 - \left(tr\left(\Delta_{\mathbf{s}_g}\right)\right)^2$$
$$= \varepsilon_3, \tag{A.47}$$

where (a) follows from (A.46) and (A.45). This completes the variance analysis of ζ_l, and clearly ζ_l has vanishingly small variance as $N_B \to \infty$. As a consequence, the variance of the cross term $2\mathbb{E}_{N,t_1}\left(\mathbf{z}_{\mathbf{s}_g}^T \mathscr{O}_{\mathbf{s}_g} \mathbf{e}_{\mathbf{s}}\right) = \frac{2}{n}\sum_{l=0}^{n-1} \zeta_l$ is also vanishingly small for $N_B \to \infty$ (follows from the Cauchy-Schwarz inequality). This completes the proof of (A.39).

REFERENCES

[1] I. H. Wang, C. Suh, S. Diggavi, and P. Viswanath, "Bursty interference channel with feedback," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 21–25.

[2] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, Dec 2008.

[3] C. Suh and D. N. C. Tse, "Feedback capacity of the gaussian interference channel to within 2 bits," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2667–2685, May 2011.

[4] M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2699–2713, June 2010.

[5] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[6] N. Khude, V. Prabhakaran, and P. Viswanath, "Opportunistic interference management," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 2076–2080.

[7] G. Bresler and D. Tse, "The two-user gaussian interference channel: a deterministic view," *European Transactions on Telecommunications*, vol. 19, pp. 333–354, 2008.

[8] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-Physical Systems Security*, Jul. 2009.

[9] Y. Mo, T.-H. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, Jan. 2012.

[10] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st ACM International Conference on High Confidence Networked Systems (HiCoNS)*, 2012, pp. 55–64.

[11] J. Villasenor, "Compromised by design? securing the defense electronics supply chain," *Brookings Institution Report*, Nov 2013. [Online]. Available: http://iis-db.stanford.edu/pubs/24484/Villasenor-Securing_the_Defense_Electronics_Supply_Chain.pdf

[12] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric gaussian k user interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3297–3303, July 2010.

[13] N. Khude, V. Prabhakaran, and P. Viswanath, "Harnessing bursty interference," in *Networking and Information Theory, 2009. ITW 2009. IEEE Information Theory Workshop on*, June 2009, pp. 13–16.

[14] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1059–1064, May 1997.

[15] J. Jiang, N. Marukala, and T. Liu, "Symmetrical multilevel diversity coding with an all-access encoder," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1662–1666.

[16] A. Carleial, "A case where interference does not reduce capacity (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 569–570, Sep 1975.

[17] A. E. Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[18] D. Dolev, C. Dwork, O. Waarts, and M. Yung, "Perfectly secure message transmission," *Journal of the ACM*, vol. 40, no. 1, pp. 17–47, Jan. 1993.

[19] R. Blahut, *Algebraic Codes for Data Transmission*. Cambridge University Press, 2003.

[20] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, Jan. 2012.

[21] S. S. Kia, J. Cortes, and S. Martinez, "Dynamic average consensus under limited control authority and privacy requirements," *preprint*, 2014. [Online]. Available: http://arxiv.org/pdf/1401.6463v1.pdf

[22] D. Chaum, C. Crépeau, and I. Damgard, "Multiparty unconditionally secure protocols," in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC)*, 1988, pp. 11–19.

[23] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, Jul. 2011.

[24] J. Le Ny and G. J. Pappas, "Differentially private Kalman filtering," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, 2012, pp. 1618–1625.

[25] W. A. Malik, N. C. Martins, and A. Swami, "LQ control under security constraints," in *Control of Cyber-Physical Systems*. Springer, 2013, pp. 101–120.

[26] P. Antsaklis and A. Michel, *Linear Systems*. Birkhäuser Boston, 2005.

[27] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall PTR, 1998.

[28] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1915–1930, Jul. 2004.

[29] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: The least favorable but the most useful [lecture notes]," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 183–186, Mar. 2013.

[30] J. Wolf, "An introduction to Reed-Solomon codes," Course notes. [Online]. Available: http://pfister.ee.duke.edu/courses/ecen604/rspoly.pdf

[31] R. J. McEliece and D. V. Sarwate, "On sharing secrets and Reed-Solomon codes," *Communications of the ACM*, vol. 24, no. 9, pp. 583–584, Sep. 1981.

[32] E. Berlekamp, "Bounded distance+1 soft-decision Reed-Solomon decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 704–720, May 1996.

[33] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *Control Systems, IEEE*, vol. 35, no. 1, pp. 110–127, Feb 2015.

[34] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, June 2014.

[35] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *arXiv pre-print*, Sep. 2013. [Online]. Available: http://arxiv.org/abs/1309.3511

[36] F. Pasqualetti, F. Dörfler, and F. Bullo, "A divide-and-conquer approach to distributed attack identification," in *IEEE Conference on Decision and Control*, Dec. 2015.

[37] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach," *arXiv pre-print*, Dec. 2014.

[38] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *American Control Conference (ACC)*, 2015.

[39] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, 2014.

[40] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Allerton Conference on Communication, Control, and Computing*, 2009.

[41] C.-Z. Bai and V. Gupta, "On kalman filtering in the presence of a compromised sensor: fundamental performance bounds," in *American Control Conference (ACC)*, 2014.

[42] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50–61, May 2010.

[43] S. Farahmand, G. B. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Trans. on Signal Processing*, vol. 59, no. 10, pp. 4529–4543, Oct. 2011.

[44] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[45] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *arXiv pre-print*, 2015. [Online]. Available: http://arxiv.org/abs/1510.02462

[46] A. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, Nov. 1976.

[47] S.-D. Wang, T.-S. Kuo, and C.-F. Hsu, "Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation," *IEEE Transactions on Automatic Control*, vol. 31, no. 7, pp. 654–656, Jul 1986.