# On Distributed Estimation for Resource Constrained Wireless Sensor Networks 

Alireza Sani<br>University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd
University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

## STARS Citation

Sani, Alireza, "On Distributed Estimation for Resource Constrained Wireless Sensor Networks" (2017). Electronic Theses and Dissertations, 2004-2019. 5684.
https://stars.library.ucf.edu/etd/5684


# ON DISTRIBUTED ESTIMATION FOR RESOURCE CONSTRAINED WIRELESS SENSOR NETWORKS 

by

ALIREZA SANI
B.S. in Electrical Engineering, University of Tabriz, Tabriz, Iran, 2009
M.S. in Electrical Engineering University of Tehran, Tehran, Iran, 2012

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2017

Major Professor: Azadeh Vosoughi
© 2017 Alireza Sani


#### Abstract

We study Distributed Estimation (DES) problem, where several agents observe a noisy version of an underlying unknown physical phenomena (which is not directly observable), and transmit a compressed version of their observations to a Fusion Center (FC), where collective data is fused to reconstruct the unknown. One of the most important applications of Wireless Sensor Networks (WSNs) is performing DES in a field to estimate an unknown signal source. In a WSN battery powered geographically distributed tiny sensors are tasked with collecting data from the field. Each sensor locally processes its noisy observation (local processing can include compression, dimension reduction, quantization, etc) and transmits the processed observation over communication channels to the FC, where the received data is used to form a global estimate of the unknown source such that the Mean Square Error (MSE) of the DES is minimized.

The accuracy of DES depends on many factors such as intensity of observation noises in sensors, quantization errors in sensors, available power and bandwidth of the network, quality of communication channels between sensors and the FC, and the choice of fusion rule in the FC. Taking into account all of these contributing factors and implementing a DES system which minimizes the MSE and satisfies all constraints is a challenging task. In order to probe into different aspects of this challenging task we identify and formulate the following three problems and address them accordingly: - Consider an inhomogeneous WSN where the sensors' observations is modeled linear with additive Gaussian noise. The communication channels between sensors and FC are orthogonal powerand bandwidth-constrained erroneous wireless fading channels. The unknown to be estimated is a Gaussian vector. Sensors employ uniform multi-bit quantizers and BPSK modulation. Given this setup, we ask: what is the best fusion rule in the FC? what is the best transmit power and quantization rate (measured in bits per sensor) allocation schemes that minimize the MSE? In order


to answer these questions we derive some upper bounds on global MSE and through minimizing those bounds, we propose various resource allocation schemes for the problem, through which we investigate the effect of contributing factors on the MSE.

- Consider an inhomogeneous WSN with an FC which is tasked with estimating an scalar Gaussian unknown. The sensors are equipped with uniform multi-bit quantizers and the communication channels are modeled as Binary Symmetric Channels (BSC). In contrast to former problem the sensors experience independent multiplicative noises (in addition to additive noise). The natural question in this scenario is: how does multiplicative noise affect the DES system performance? how does it affect the resource allocation for sensors, with respect to the case where there is no multiplicative noise? We propose a linear fusion rule in the FC and derive the associated MSE in closed-form. We propose several rate allocation schemes with different levels of complexity which minimize the MSE. Implementing the proposed schemes lets us study the effect of multiplicative noise on DES system performance and its dynamics. We also derive Bayesian Cramér-Rao Lower Bound (BCRLB) and compare the MSE performance of our porposed methods against the bound. As a dual problem we also answer the question: what is the minimum required bandwidth of the network to satisfy a predetermined target MSE?
- Assuming the framework of Bayesian DES of a Gaussian unknown with additive and multiplicative Gaussian noises involved, we answer the following question: Can multiplicative noise improve the DES performance in any case/scenario? the answer is yes, and we call the phenomena as 'enhancement mode' of multiplicative noise. Through deriving different lower bounds, such as BCRLB, Weiss-Weinstein Bound (WWB), Hybrid CRLB (HCRLB), Nayak Bound (NB), Yatarcos Bound (YB) on MSE, we identify and characterize the scenarios that the enhancement happens. We investigate two situations where variance of multiplicative noise is known and unknown. We also compare the performance of well-known estimators with the derived bounds, to ensure practicability of the mentioned enhancement modes.
to my dad Behrouz,
that If I become only half the thinker, half the person that he is, it will surely be one of my greatest accomplishments, and to my mom Shayan,
who is ridiculously kind to me and always stood beside me in gloomy days, and to my sister Afsane,
who wants nothing but the best for me.


## ACKNOWLEDGMENTS

First, I want to express my sincere gratitude to my advisor Prof. Azadeh Vosoughi of the UCF College of Engineering for the continuous support of my PhD study and research. I owe her more than I can pay off. She is an adorable blend of intelligence, kindness, patience, laughter, and professional attitude. She was always there to answer my questions, I knew sometimes I asked silly questions and every time she answered those in the most professional way. She was more than an adviser to me, a life compass in stormy days, a friend in difficult times and family in gloomy seasons. As an international student living here 9000 miles away from home, having an adviser like Azadeh was the best gift of world to me.

I also appreciate the continued help and useful comments of my other committee members: Dr. George Atia, Dr. Nazanin Rahnavard, Dr. Lei Wei of the Electrical Engineering program and Dr. Mainak Chatterjee of the Computer Science program.

I also want to dedicate this dissertation to Dr. Alireza Seyedi who has left the earth before us, but his memory will remain with me for ever. It was my misfortune that I knew him briefly, however he stays an inspiration to me as long as I live.

Another thanks goes to my close friends. I always appreciate their stress-relieving company in my PhD life. I also want to thank my lab mates for the brains-storms we did and for sharing their valuable research related idea.

Last but not leas I want thank National Science Foundation for supporting this research under grants CCF-1336123, CCF-1341966, and CCF-1319770.

## TABLE OF CONTENTS

LIST OF FIGURES ..... xiii
LIST OF ACRONYMS ..... xvii
CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW ..... 1
1.1 Centralized versus Distributed Estimation ..... 1
1.2 Literature Survey and Related Works ..... 3
1.2.1 Bandwidth Constrained Distributed Estimation ..... 4
1.2.2 Power Constrained Distributed Estimation ..... 6
1.2.3 Distributed Estimation in Environments with Multiplicative Noise ..... 9
1.3 Dissertation Organization and Contributions ..... 11
CHAPTER 2: DISTRIBUTED VECTOR ESTIMATION FOR POWER- AND BANDWIDTH- CONSTRAINED WIRELESS SENSOR NETWORKS ..... 16
2.1 System Model and Problem Statement ..... 17
2.2 Characterization of MSE ..... 19
2.2.1 Characterization of First Bound $\mathcal{D}_{a}$ ..... 21
2.2.2 Characterization of Second Bound $\mathcal{D}_{b}$ ..... 23
2.3 "Coupled" Scheme for Resource Allocation ..... 25
2.3.1 Coupled Scheme for Minimizing $\mathcal{D}_{a}$ ..... 26
2.3.2 Migration from Continuous to Integer Solutions for Rates ..... 33
2.3.3 Coupled Scheme for Minimizing $\mathcal{D}_{b}$ ..... 34
2.4 "Decoupled" Scheme For Resource Allocation ..... 35
2.4.1 Decoupled Scheme for Minimizing $\mathcal{D}_{a}$ ..... 36
2.4.2 Does Depleting $B_{t o t}$ always reduce $\mathcal{D}_{a}$ ? ..... 37
2.4.3 Decoupled Scheme for Minimizing $\mathcal{D}_{b}$ ..... 40
2.5 Complexity Comparison of Algorithms ..... 41
2.6 Discussion on Observation Model ..... 42
2.7 Numerical and Simulation Results ..... 43
2.8 Conclusions ..... 58
CHAPTER 3: CLOSED FORM APPROXIMATION FOR CRAMÉR RAO LOWER BOUND FOR DISTRIBUTED ESTIMATION WITH QUANTIZED OBSERVATIONS ..... 60
3.1 System Model and Problem Statement ..... 60
3.2 Bayesian Cramér Rao Lower Bound ..... 61
3.2.1 BCRLB closed-form approximation ..... 64
3.3 Simulation Results ..... 66
3.4 Conclusions ..... 68
CHAPTER 4: ON DISTRIBUTED LINEAR ESTIMATION IN MULTIPLICATIVE NOISE ENVIRONMENT AND OBSERVATION MODEL UNCERTAINTIES ..... 69
4.1 System Model and Problem Statement ..... 70
4.2 Characterizing MSE for LMMSE estimator ..... 71
4.3 Solving Constrained Problem (P1) ..... 75
4.3.1 LRLP Method ..... 76
4.3.2 Greedy Method ..... 78
4.3.3 Integer Relaxation Method ..... 80
4.3.4 IRA Method ..... 84
4.4 Solving Constrained Problem (P2) ..... 86
4.4.1 Greedy Method ..... 86
4.4.2 Integer Relaxation Method ..... 87
4.4.3 IRA Method ..... 88
4.5 BCRLB ..... 89
4.6 Extension to Erroneous Channels ..... 91
4.6.1 LMMSE Estimator and its corresponding MSE ..... 91
4.6.2 BCRLB and Fisher Information Expressions ..... 92
4.7 Numerical and Simulation Results ..... 93
4.8 Conclusions ..... 107
CHAPTER 5: NOISE ENHANCEMENT IN BAYESIAN DISTRIBUTED ESTIMATION ..... 108
5.1 System Model and Problem Statement ..... 109
5.2 Lower Bounds on MSE ..... 110
5.2.1 Lower bound on the MSE for known $\sigma_{h}^{2}$ ..... 110
5.2.1.1 WWB ..... 110
5.2.1.2 BCRLB ..... 112
5.2.2 Lower bounds on the MSE for unknown $\sigma_{h}^{2}$ ..... 116
5.2.2.1 HCRLB ..... 116
5.2.2.2 YB ..... 118
5.2.2.3 NB ..... 119
5.2.2.4 RUB ..... 119
5.3 Estimation Schemes ..... 121
5.3.1 Estimation Schemes for known $\sigma_{h}^{2}$ ..... 121
5.3.1.1 MMSE estimator ..... 121
5.3.1.2 MAP estimator ..... 122
5.3.2 Estimation Schemes for unknown $\sigma_{h}^{2}$ ..... 123
5.4 Extension of Derivations of WWB and BCRLB to the Case of Quantized Obser- vations ..... 124
5.4.1 WWB derivation for quantized observations ..... 124
5.4.2 BCRLB derivation for quantized observations ..... 126
5.5 Simulation Results ..... 128
5.6 Conclusions ..... 136
CHAPTER 6: CONCLUSION AND FUTURE WORK DIRECTIONS ..... 138
6.1 Conclusions ..... 138
6.2 Future Work Directions ..... 141
6.2.1 Fusion Rules based on channel outputs ..... 141
6.2.2 Effect of Correlated Multiplicative Noise on DES Design and Performance ..... 141
APPENDIX A: APPENDIX FOR CHAPTER 2 ..... 142
A. 1 Finding Upper Bound on $\mathbb{E}\left\{\left(\hat{m}_{k}-m_{k}\right)^{2}\right\}$ ..... 143
A. 2 Properties of $\mathcal{D}_{2}^{u p b}$ and $\mathcal{D}_{2}^{u u p b}$ ..... 143
A. 3 Properties of $\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$ ..... 144
A. 4 More Properties on $\mathcal{D}_{1}$ and $\mathcal{D}_{2}^{u p b}$ ..... 145
A. 5 Rate Discretization Algorithm ..... 147
A. 6 "a-decoupled" algorithm ..... 148
A. 7 Ellipsoid Method is Guaranteed to Converge ..... 148
APPENDIX B: APPENDIX FOR CHAPTER 4 ..... 151
B. 1 On one-bit MLE ..... 152
B. 2 Fisher information for non Gaussian Multiplicative Noise ..... 155
APPENDIX C: APPENDIX FOR CHAPTER 5 ..... 157
C. $1 \quad \mu(t)$ and $\mu_{q}(t)$ are even functions of $t$ ..... 158
C. 2 Proof for lemma 8 ..... 159
C. 3 Derivation of MLE for $\sigma_{h}^{2}$ in (5.25) ..... 161
APPENDIX D: BIOGRAPHICAL SKETCH ..... 163
APPENDIX E: PUBLICATIONS ..... 165
LIST OF REFERENCES ..... 167

## LIST OF FIGURES

2.1 system model ..... 18
2.2 Modified ellipsoid method for constrained optimization problem ..... 31
$2.3 \quad \mathcal{D}_{1}, \mathcal{D}_{2}^{u p b}, \mathcal{D}_{a}$ versus $B_{t o t}$ (measured in bits) ..... 38
2.4 Behavior of " $a$-decoupled" algorithm ..... 39
2.5 " $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$ ..... 43
2.6 "a-coupled" algorithm $\left\{L_{k}\right\}_{k=1}^{3}$ vs. $P_{t o t}$ ..... 44
2.7 " $a$-decoupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$ ..... 45
$2.8 \quad$ " $a$-decoupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $P_{\text {tot }}$ ..... 46
2.9
" $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $B_{\text {tot }}$ ..... 47
2.10 " $a$-coupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$ ..... 47
2.11 " $a$-decoupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $B_{\text {tot }}$ ..... 48
2.12
" $a$-decoupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$ ..... 48
2.13 $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $P_{\text {tot }}$ for all algorithms ..... 49
2.14 Simulated MSE vs. $P_{\text {tot }}$ for all algorithms ..... 50
2.15 $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $B_{t o t}$ for all algorithms ..... 51
Simulated MSE vs. $B_{\text {tot }}$ for all algorithms ..... 52
2.17 " $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}, B_{\text {tot }}=30$ bits, $K=3$ ..... 52
2.18$\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $P_{t o t}$ for all algorithms for $K=20$54
2.20
Simulated MSE vs. $P_{\text {tot }}$ for all algorithms for $K=20$ ..... 55
2.212.22
3.1 Comparison of Fisher information of clairvoyant and the one associated withlimited dynamic range of sensors67
3.2 CRLB and proposed approximation for different values of quantization rate ..... 67
4.1 LRLP method ..... 77
4.2 Greedy method ..... 79
4.3 Analytical and simulated MSE of greedy vs. $B_{t o t}$ ..... 94
4.4 MSE of different rate allocation methods vs. $B_{t o t}$ for $K=5$ ..... 94
4.5 MSE performance for different allocation schemes for $K=50$ ..... 95
4.6 MSE and CRLB vs. $B_{\text {tot }}$ for $K=5$ ..... 96
4.7 MSE and CRLB for proposed schemes for $K=50$ ..... 97
4.8 Number of active sensors vs. $B_{t o t}$ ..... 98
4.9 Average quantization rate of active sensors vs. $B_{t o t}$ ..... 98
4.10
required bandwidth versus a target MSE for different allocation schemes for
$K=5$ ..... 99
4.11 Required bandwidth of different rate allocation methods vs. target MSE for $K=50$ ..... 100
4.12 Number of active sensors vs. target MSE ..... 100
4.13 Quantization rate of active sensors vs. target MSE ..... 101
4.14 MSE and CRLB of greedy vs. $B_{t o t}$ for $K=50$ ..... 102
4.15 R-D and BCRLB for $K=5,50$ and $\sigma_{\theta}^{2}=1$ ..... 103
4.16 CRLB correspoding to one bit and two bit quantizers ..... 104
4.17 a ..... 105
4.18 Performance comparison of one bit and multi bit estimation vs. $B_{t o t}, K=50$, $\theta_{\max }=2 \sigma_{\theta}$ ..... 106
5.1 WWB and BCRLB vs known $\sigma_{h}^{2}$ ..... 128
5.2 a ..... 129
5.3 Comparison of quantized bounds and clairvoyant benchmark, $\sigma_{n}^{2}=1$ ..... 130
5.4 Performance of MMSE and MAP estimator ..... 131
5.5
comparison of HCRLB and NB and RUB vs unknown $\sigma_{h}^{2}$, for $K=10$ and different $\sigma_{n}^{2}$ values ..... 132
5.6 HCRLB, NB and RUB vs unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different $K$ 's ..... 133
5.7 comparison of WWB and RUB vs known and unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network size $K$ ..... 133
5.8 Performance of MMSE-ML and MAP-ML estimator, for network size $K=$20134
5.9 Performance of MMSE-ML and MAP-ML estimator, for $\sigma_{n}^{2}=1$ and differ- ent network size $K=10,50,100$ ..... 135
5.10 comparison of MSE of various estimators for case of known and unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network size $K=10,50,100$ ..... 136
B. $1 \quad P_{s}$ with $\sigma_{n}^{2}=1, \theta_{\max }=5$ ..... 153
B. $2 \quad \hat{\theta}_{M L}$ and $\hat{\theta}_{M A P}$ vs. realizations of $\theta$ ..... 154

## LIST OF ACRONYMS

AF Amplify and Forward. 7, 8

AWGN Additive White Gaussian Noise. 6-8

BCRLB Bayesian Cramér-Rao Lower Bound. 12-14, 60, 61, 66, 68, 69, 89, 95, 96, 101, 102, $107,108,110,112,114,115,118,126,128-132,136,139,140$

BLUE Best Linear Unbiased Estimator. 6-8

BNC Binary Natural Coding. 91-93

BPSK Binary Phase Shift Keying. 2, 16, 18, 22, 58, 138

BSC Binary Symmetric Channel. 8, 13, 69, 70, 91, 92, 107, 140

CDF Cumulative Distribution Function. 63, 64, 91, 113

CE Centralized Estimation. 1-3, 9, 16, 38, 58, 74, 139

CNR Channel to Noise Ratio. 22, 27

CRLB Cramér-Rao Lower Bound. 4-6, 9-11, 13, 103, 104

CSI Channel State Information. 141

DES Distributed Estimation. 1-15, 58, 60, 68-70, 102, 107-110, 128, 136, 138-140

DFS Depth First Search. 77

FC Fusion Center. 1-14, 16-20, 58, 60, 61, 64, 65, 69-71, 74, 80, 82, 91, 103, 107-109, 124, 136, 138-141

HCRLB Hybrid Cramér-Rao Lower Bound. 14, 108, 110, 116-118, 132, 134, 136, 140

IRA Individual Rate Allocation. 76, 86, 93-95, 97, 99-101, 107, 140

KKT Karush-Kuhn-Tucker. 26, 27, 34, 36, 81, 87, 149

LMMSE Linear Minimum Mean Square Error. 7, 8, 13, 20, 21, 69, 71, 73, 74, 91, 92, 96, 103, 104, 107, 139-141

LRLP Longest Root to Leaf Path. 76

M-QAM M-ary Quadrature Amplitude Modulation. 7

MAC Multiple Access Channel. 7, 8

MAP Maximum A Posterior. 14, 103, 104, 107, 108, 122, 123, 128, 131, 134-136, 140, 153

MAP-ML Maximum A Posterior-Maximum Likelihood. 14, 108, 123, 128, 134, 135, 137, 141

MINLP Mixed Integer Nonlinear Programming. 25

ML Maximum Likelihood. 9, 122, 123, 161

MLE Maximum Likelihood Estimator. 5, 9-11, 103, 104, 107, 122, 140, 152, 153, 161, 162

MMSE Minimum Mean Square Error. 5, 10, 12, 14, 96, 108, 118, 119, 121, 123, 128, 131, 134-136, 140, 159

MMSE-ML Minimum Mean Square Error-Maximum Likelihood. 14, 108, 123, 128, 134, 135, 137, 141

MSE Mean Square Error. 3, 4, 6-10, 12-16, 19, 20, 35, 43, 50, 51, 53, 55, 57, 58, 61, 66, 69, 71, $73,74,76,86-89,91,92,94-97,99-101,103,104,107,108,110,112,114,116,119,128$, 131, 134-141, 149

NB Nayak Bound. 14, 108, 110, 116, 119, 120, 132, 134, 136, 140

NLIP Non Linear Integer Programming. 75

OA Order Aware. 93, 95, 99, 100

PAM Pulse Amplitude Modulated. 141

PDF Probability Density Function. 10, 63, 90, 91, 104, 127, 153, 155, 156

R-D Rate Distortion. 102, 104

RU Risk Unbiased. 120, 123

RUB Risk Unbiased Bound. 14, 108, 110, 116, 119, 120, 132, 134, 136, 140

SNR Signal to Noise Ratio. 6, 83

WSN Wireless Sensor Network. 1, 2, 4, 7, 8, 11-13, 16, 17, 59, 60, 69, 70, 107, 108, 136, 138-140

WSNs Wireless Sensor Networks. 1, 2, 4

WW Weiss-Weinstein. 13

WWB Weiss-Weinstein Bound. 13, 14, 108, 110, 114, 115, 126, 128-131, 136, 140

YB Yatarcos Bound. 14, 116, 118, 119

## CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Distributed Estimation (DES) serves as one of the major and main applications in Wireless Sensor Network (WSN)s, where large scale WSNs consisting of battery operated devices with limited sensing, computation and communication capabilities are deployed over a sensing region to monitor physical or environmental phenomenas. DES in WSNs can benefit many practices such as surveillance [1], target tracking, fire detection, border protection and target localization [2,3]. The vision is that once the sensors are deployed in the sensing field, each sensor makes a noisy measurement which depends on the unknown parameter, processes locally its measurement and transmits the relevant information to a common Fusion Center (FC). Having the collective information from sensors, the FC is tasked with estimating the unknown parameter via fusing the received data. The main challenges for designing a DES schemes in Wireless Sensor Networks (WSNS) can be categorized as following: $i$ ) designing local encoders in sensors while taking into account the practical limitations of sensors, ii) proposing optimal resource (power and rate) allocation methods for transmission of locally processed data to the FC, iii) designing optimal fusion rules in the FC to estimate the unknown via fusing received data from sensors. Our proposed novel schemes addressing the mentioned challenges, enable an accurate and robust estimation subject to practical limitations such as network bandwidth and power constraints.

### 1.1 Centralized versus Distributed Estimation

In Centralized Estimation (CE), it is assumed that the observations data are transmitted to the FC with full precision. The communication channels between the sensors and the FC are assumed to be error-free (without power or bandwidth constraint on channels). Having all full precision observations, the FC is tasked with making a global estimation of the unknown via fusing the col-
lective data. Thus the quality and reliability of the global estimation depends on the number of available observations, noise levels in sensors and fusion rule. This is more like the classic estimation philosophy where it is assumed that all observations are available with full precision without any further processing in a central unit, where a single estimator fuses the noisy observations to produce an estimate of the unknown [4]. In practical WSNS there is limitations on transmit power and bandwidth. Hence assuming to have the error-free full-precision observations in the FC is not realistic. To address power and bandwidth limitations, it is desirable that the sensors locally process (compress) their observations and then send their locally processed (compressed) observations to the FC. This is the core idea of DES. In contrast to CE, in DES framework, each sensors' observations are locally compressed to few information bits, then the bits are transmitted via digital transmission schemes over error-free or erroneous channels to the FC. Next, the FC is tasked with reliably estimating the unknown via fusing the received bits. Designing DES system involves different challenges at sensor level, channel level and FC level. One needs to design: i) a local processing unit (for instance a quantizer) at the sensors, considering the limitations of the sensors and individual observation qualities, $i i$ ) a power allocation scheme taking into account the channel qualities and available power, iii) a global fusion rule in the FC to fuse the received data and produce a reliable estimate of unknown.

In this thesis we study the DES in a WSN where the available bandwidth and/or power are limited. Regarding the communication channels between the sensors and the FC, we consider both errorfree and erroneous channel models. As we mentioned earlier the sensors in WSNs have limited sensing, computation and communication capabilities which are all integrated into a tiny electronic board. Sensors are powered with batteries which usually are not replaceable. This motivates us to study the problem where each sensor node in the network, quantizes its noisy observation to few bits and transmits the bits via digital modulation schemes like Binary Phase Shift Keying (BPSK) to the FC. Generally the received quantization levels (assuming a hard decoding at the

FC, the quantization levels can be reconstructed by fusing the bits received from each sensor) are not necessarily equal to the transmitted ones due to wireless channel errors. This implies that the accuracy of estimation in the FC depends on the quality of transmission channels as well as the quality of sensors' observations, which should be taken into account in designing the DES schemes. Thus in designing an optimal DES scheme, all system parameters associated with the sensors' observation model and also parameters corresponding to the communication channels needs to be considered.

### 1.2 Literature Survey and Related Works

The research on DES problem for deterministic and random unknown parameters has a long history and dates back to 1988 when Zhang and Berger [5] introduced the notion of estimation via compressed information for an unknown deterministic scalar and formulated a novel problem by extending CE theory to the case in which the observations are transmitted to the FC over bandwidth-constrained (otherwise error-free) orthogonal channels. Toby Berger and his colleague's work was followed by [6, 7]. Next, multi-terminal source coding was introduced in [8] and studied further in [9-11] for DES of an unknown random scalar. Another inherently related problem to DES, widely known as CEO problem was introduced in [12,13]. In quadratic Gaussian CEO problem [13] a team of agents observes a Gaussian source, corrupted by independent additive Gaussian noises. Agents communicate their coded messages over error-free channels to the CEO (FC), that is tasked with estimating the Gaussian source with minimal Mean Square Error (MSE). The authors conjectured a rate-distortion region, where the decay rate of the MSE is inversely proportional to the total source coding rate of agents when the number of agents goes to infinity. Later, [14] found the rate-distortion region explicitly for an arbitrary number of agents. These works consider the problem from the information theoretic perspective, and build upon funda-
mental assumptions that the sensors sample their observations continuously and use infinite-length complex source codes (jointly typical sequences) to compress their samples into bits, and the FC utilizes the best decoder (estimator), in terms of minimizing the MSE. These assumptions disregard several key characteristics of WSN, including hardware complexity, computational capability, power, and delay constraints.

On the other hand the distributed signal processing approaches towards DES are based on the assumptions that the delay and computational complexity are low and each sensor makes only few observations. Hence the proposed schemes in information theoretic works (including the rate allocation schemes in [15-17]) cannot be directly applied to WSNS. However these information theoretic works provide useful benchmarks for the latter approaches and insights for algorithmic designs.

In particular, [18] designed optimal quantizers that maximize Bayesian Fisher information for estimating a random parameter. Assuming identical one-bit quantizers, $[19,20]$ found the minimum achievable Cramér-Rao Lower Bound (CRLB) and the optimal quantizers for estimating a deterministic parameter.

### 1.2.1 Bandwidth Constrained Distributed Estimation

The study of bandwidth-constrained DES has a long history in information theory [5-13,21] and signal processing [18-20,22-46] literature ${ }^{1}$. The common assumption is to translate the bandwidth constraint to limiting the number of bits per observation that a sensor can transmit to the FC, rendering quantization and quantizer design a critical issue of DES. Focusing on studying the effect of quantization and bandwidth constraint on DES problem, several works [18-20,22-44] assumed

[^0]that the quantized versions of observations are transmitted to the FC on error-free channels. Early works of Reibman [22,23] discussed iterative algorithms to find local quantizers at each sensor to optimize the Bayesian Fisher information for estimating an unknown random scalar. Considering a network-wide bandwidth constraint, [67] investigated the trade-off between fine quantization of observations of only a few sensors and coarse quantization of observations of as many sensors as possible, as well as its effect on Fisher information. In [18] assuming that the joint distribution of unknown and observations is available, the authors designed optimal quantizers that maximize Bayesian Fisher information for estimating a random parameter. Assuming identical one-bit quantizers, $[19,20]$ found the minimum achievable CRLB and the optimal quantizers for estimating a deterministic parameter.

For some special cases of linear fusions [25,26] established the optimal Minimum Mean Square Error (MMSE) local quantizers. When estimating a deterministic unknown, the estimation accuracy usually depends on the unknown itself, and makes it infeasible to design quantizers to be optimal for all values of unknown. To circumvent this [27-29] considered a MinMax problem which minimizes the maximum difference between Fisher information of quantized and unquantified observations. The authors in $[30,31]$ formulated a MaxMin problem to maximize the minimum asymptotic relative efficiency between Maximum Likelihood Estimator (MLE) based on quantized and unquantized observations for a class of score-function quantizers. One-bit quantizers has attracted the attention of many researchers for DES of a scalar deterministic parameter for several years. For example in [35,36] the authors employed adaptive one-bit quantizers and in [32-34] they used the fixed one-bit quantizers. We will discuss in Chapter 2 that in some scenarios one-bit quantizers have a critical drawback that can deteriorate the estimation performance drastically. In [32] it is shown that the value of optimal quantization threshold, assuming identical one-bit quantizers in sensors, depends on the unknown parameter. To circumvent this, the authors in $[33,34]$ suggested to divide the sensors in groups and assign fixed but distinct thresholds for these groups, hoping that
some thresholds would be close the value of unknown parameter. In $[35,36]$ the authors proposed adaptive quantizers such that each sensor adjusts its quantization threshold according to the earlier transmission from other sensors. Universal DES schemes were first proposed by Luo in [37], in which the distribution of additive observation noise is assumed to be unknown. He continued his work in $[38,39]$ expanding the idea to inhomogeneous environments where sensors have different Signal to Noise Ratio (SNR). Later, Giannakis and his colleague investigated the fixed one bit quantizers for estimating a deterministic parameter and nonlinear observation models [40].

In some literature the researchers interpreted the bandwidth constraint as limiting the number of real valued messages that a sensor can send per observation to the FC. For example in [41-44] the authors discussed DES of an unknown deterministic [44] or random [41-43] vector. The sensors make a vector observation and reduce its dimension by using a linear transformation and transmit that to the FC, where a linear estimator is used to reconstruct the unknown vector.

Several researchers relaxed the assumption of communication channels being error free $[68,69]$. For estimating a vector of deterministic parameters, [68] investigated an Expectation-Maximization [4] algorithm and compared its MSE performance with CRLB, when sensors employ fixed and identical multi-bit quantizers and the communication channels are modeled as Additive White Gaussian Noise (AWGN). A related problem was studied by [69], where the FC employs a spatial Best Linear Unbiased Estimator (BLUE) for field reconstruction and the MSE distortion is compared with a posterior CRLB.

### 1.2.2 Power Constrained Distributed Estimation

As mentioned in earlier subsection, error-free channel assumption is widely adopted in many signal processing literature, however a near error-free and reliable communication can be achieved at the cost of resource consuming channel coding and consequently increasing transmit power
consumption. Such an increasing demand for power, exceeds the energy constraints of a typical WSN. Fading channels and adverse channel qualities will worsen the challenge of power scarcity of the network. Motivated by this many researchers studied power constrained and energy-efficient DES problem, some examples are [70-78] for analog transmission and [79-89] digital transmission, over an AWGN Multiple Access Channel (MAC) [90-92]. Particularly for DES of a scalar unknown $[70,73]$ proposed an optimal power allocation scheme on an orthogonal MAC, which minimizes the MSE subject to total network power constraint. The authors in [72] explored DES of an unknown vector with non-linear observation model and correlated additive noises and analog transmission over an orthogonal MAC. Assuming a non-orthogonal coherent MAC and Amplify and Forward (AF) transmission [71,76] suggested optimal linear transformations at the sensors subject to network power constraint in order to minimize the MSE distortion. In [78] a power allocation scheme was proposed for minimizing the distortion outage probability of the WSN, subject to total and individual power constraints. The FC employs Linear Minimum Mean Square Error (LMMSE) estimator for estimating a Gaussian random parameter. The communication channels are modeled as coherent MAC with fading.

The authors in [77] studied MSE minimization of DES of a random scalar subject to network power constraint. The BLUE estimation rule is employed at the FC and communication channels are modeled as coherent MAC. The authors also investigated the effect of correlated additive noise on DES performance. From an information theoretic point of view in [93] Gastpar and his colleagues probe into the trade-offs between the MSE distortion and power consumption for the quadratic CEO problem on an AWGN MAC. Few studies [80,82,83] also exist on energy efficient DES with digital transmission (sensors transmit quantized data to the FC).

The authors in [80] extended the studies in $[33,34]$ to noisy channels. An optimal power allocation scheme for M-ary Quadrature Amplitude Modulation (M-QAM) transmission was proposed in [83] in oder to minimize the $L^{2}$-norm of the transmission power vector of sensors subject to a target

MSE distortion constraint, where the FC employs BLUE to reconstruct a scalar deterministic unknown parameter. A converse problem was considered in [82], which minimized the MSE subject to a network transmit power constraint.

For a homogeneous WSN, [79] investigated a bit and power allocation scheme that minimizes the MSE, subject to a total transmit power constraint, when communication channels are modeled as Binary Symmetric Channel (BSC). Note that [79] did not include a total bit constraint in its problem formulation. The authors in [84] proposed a resource allocation scheme for sensors in order to minimize an upper bound on MSE of the LMMSE estimation of a scalar Gaussian parameter, where sensors transmit via digital modulations on fading channels. The authors in [85] derived a Fihser information matrix for DES of a Gaussian vector in a WSN with digital transmission and total power constraint.

It is noteworthy to mention that the authors in [93] concluded that for estimating a scalar Gaussian random unknown in a WSN with Independent and Identically Distributed (i.i.d) Gaussian additive observation noises and communication channels modeled as AWGN non-orthogonal coherent MAC, the optimal transmission scheme is analog AF (in contrast to Quantize and Forward or relaying techniques [94-98]). The optimality results however do not hold true for other channel models. As a matter of fact in [99] the authors from an information theoretic point of view contended that digital transmission is optimal in a WSN with orthogonal MAC between the sensors and the FC, where the unknown parameter to be estimated is modeled as a Gaussian scalar. This result and also the advantages of digital communication motivated us to employ the digital scheme in our study in the succeeding Chapters.

Consensus algorithms for an ad-hoc WSN with no FC have been discussed in [100-103] . Note that in this thesis we focus on WSNs with a FC, hence consensus algorithms are out of scope of this thesis.

Several literature [11,104-106] studied DES problem in hierarchal network structures, where sensors are grouped in clusters. Each cluster has one cluster-head which acts as a local FC and collects the information from sensors and forms a local estimate of the unknown parameter. All clusterheads transmit their local estimates to the global FC, where the global FC is tasked with estimating the unknown parameter by fusing all collective data received from the cluster-heads. In this thesis we focus on star topology with a single FC with connectivity to all sensors.

### 1.2.3 Distributed Estimation in Environments with Multiplicative Noise

The bulk of literature on DES considered an observation (sensing) model with only additive noise [18,20,33,67-69,72,77,78,80, 81, 83, 107-113], while some assumed that the statistical knowledge of the additive observation noise is incomplete (referred to as noise model uncertainty) [33, 109].

One of the early works that considered multiplicative observation noise in a sensor array, is [114] that proposed an approximate MLE for localizing a source, where the received signal is corrupted by multiplicative observation noise. Later, CE with both multiplicative and additive observation noises was studied in [115-118]. In particular [115] assumed that there is a bounded perturbation in sensing/observation matrix, and proposed a linear MinMax estimator for estimation of a deterministic vector, that minimizes worst case MSE over all perturbations. In [116] assuming that sensing matrix is random with known first and second order statistics, the authors proposed a linear MinMAx estimator that minimizes the worst case MSE over all possible unknown deterministic vectors to be estimated. In $[117,118]$ the authors studied Maximum Likelihood (ML) CE of a deterministic unknown vector with linear observation model where the mixing matrix is modeled as a random Gaussian matrix with known second order statistics. In addition the CRLB was derived in [118] and the effect of uncertainty in mixing matrix on the estimation performance was investigated. The authors reported scenarios that for some specific values of the deterministic unknown
parameter, randomness in model matrix may improve the MSE performance, however in most cases the multiplicative observation noise exacerbates the estimation performance. Another related work is [119] which derived and analyzed the CRLB for estimation of a deterministic sparse vector where both sensing matrix and measurement vector are corrupted with Gaussian noise.

Despite its great importance, few researchers have studied DES with both multiplicative and additive observation noises. DES with Gaussian multiplicative and additive observation noises, one-bit quantizers at the sensors and MLE at the FC has been investigated very recently in [120, 121], respectively, for vector and scalar unknown deterministic parameters. In [120] the authors reported that the multiplicative noise exacerbates the performance of the MLE in most cases. However, provided that the variance of additive observation noise is small compared to the energy of the unknown parameter, some values of multiplicative noise variance may improve the MSE performance for some special values of the deterministic unknown parameter. Similar studies have been included in [121], where the authors also reported that employing binary quantizers with nonidentical nonzero thresholds in the sensors, improves the performance of MLE in comparison to the case of zero thresholds for all sensors. These works rely on MLEs which require perfect knowledge of Probability Density Function (PDF) of the model uncertainties which may not always be justifiable.

As another quite related topic, some papers have reported similar enhancement effects for additive observation noise, such that increasing the variance of additive noise can enhance the estimation performance in some scenarios. For instance [122] reported a case with a special non-linear observation model, where increasing the additive noise intensity enhances the estimation accuracy of the MMSE estimator. In [123] the authors derived the optimal distribution of the additive observation noise that minimizes the CRLB for estimation of a scalar unknown parameter based on quantized observations.

We also note that most DES literature is focused on one-bit quantization [18, 20, 33, 72, 80, 107$109,121,124]$, assuming that the dynamic range of the unknown parameter to be estimated is equal to or less than that of the additive observation noise [109]. Interestingly, [33] argued that there can be a significant gap between the CRLB performance based on one-bit quantization and the clairvoyant benchmark (when unquantized observations are available at the FC), when the dynamic range of the unknown is large with respect to the additive observation noise variance. Also, [124] reported that in presence of the multiplicative observation noise, low power additive observation noise can negatively impact the performance of MLE based on one-bit quantization. The works in $[33,124]$ motivate us to consider DES based on multi-bit quantization in Chapters 4 and 5. Different from [33,124], we assume that either the distribution of the multiplicative noise is unknown or it has a known distribution with unknown variance. We consider both error-free and erroneous communication channel models, however we focus on the effect of the multiplicative observation noise and multi-bit quantization errors on the estimation performance of DES.

### 1.3 Dissertation Organization and Contributions

Thanks to the collective efforts of many researchers significant progress has been made toward understanding of DES of a scalar unknown parameter with a linear observation model and additive observation noise. The works on DES of an unknown vector are mainly concentrated on reducing the dimensionality of observation vectors over error-free channels. The literature falls short of studying DES of an unknown vector with correlated observation noises. These knowledge gaps motivated us to study the DES of an unknown Gaussian vector with correlated additive noises and digital communications on fading channels in Chapter 2. Assuming an inhomogeneous WSN and power and bandwidth constrained wireless channels, we proposed various power and rate allocation schemes with different levels of complexity. In contrast to the aforementioned works that
considered either transmit power or bandwidth constraints, we consider DES subject to both total transmit power and bandwidth constraints. From practical perspectives, having a total transmit power constraint enhances energy efficiency in battery-powered WSNs. On the other hand placing a cap on the total bandwidth can further improve energy efficiency because data communication is a major contributor to the network energy consumption. We investigate the effects of observation noise, quantization errors, fading and communication channel noises on estimation accuracy and explore the trade-offs between transmit power, rate and estimation performance.

In Chapter 3 we derive an accurate closed-form approximation for Bayesian Cramér-Rao Lower Bound (BCRLB) on MSE of a Bayesian DES in a heterogeneous WSN. It is axiomatic that greatest lower bound ${ }^{2}$ under the MSE criterion is the MSE of MMSE estimator, however its implementation is often not practical and calculating its MSE usually requires multiple integrations which may be computationally infeasible [125]. Thus providing a closed-form expression for BCRLB, or an alternative accurate approximation is always intriguing [126,127]. In Chapter 3 we contribute to this topic by deriving a compact expression for the BCRLB. The unknown parameter to be estimated is modeled as a Gaussian random variable with known mean and variance and the observation model is linear with Gaussian additive noise. Each sensor separately quantizes its noisy observation with a uniform multi-bit quantizer and transmits the bits to the FC, where the received data is fused to estimate the unknown parameter. We take into account practical limitations of sensors and assume that the sensing dynamic range of the sensors are limited. In sequel we provide an easy-to-manipulate closed-form approximation for the BCRLB, which provides us a better understanding of the behavior of the bound. The simulation examples verify the accuracy of our proposed approximation. We also investigate the effect of sensing range of sensors, quantization rates and additive noise variance on the BCRLB with simulation examples.

[^1]In Chapter 4 we consider DES of a Gaussian source, corrupted by independent multiplicative and additive observation noises, in a heterogeneous bandwidth-constrained network. Similar to [33, 81], we choose the total number of quantization bits as the measure of network bandwidth. Different from $[121,124]$, we assume that the distribution of the multiplicative observation noise is unknown (observation model uncertainties) and only its mean and variance are known. To overcome the limitations caused by the observation model uncertainties, the FC employs LMMSE estimator to fuse the quantization bits received from the sensors over the orthogonal channels. We consider both error-free and erroneousness communication channels, using BSC model. We focus on the effects of observation model uncertainties and quantization errors on the accuracy of estimating the Gaussian source. We derive a closed-form expression for the MSE of the LMMSE estimator and consider two system-level constrained optimization problems with respect to the sensors' quantization rates: $i$ ) we minimize the MSE given a network-wide bandwidth constraint, and $i i$ ) we minimize the required network-wide bandwidth given a target MSE. To address these two problems, we propose several rate allocation schemes. In addition, we compare the MSE performance of the proposed schemes against the BCRLB. Note that focus of Chapter 4 is on investigating the effect of multiplicative noise, observation model uncertainties and available bandwidth, on the DES performance, which makes the Chapter different from energy efficient and power constrained DES problems addressed in [68-71,77,78, 80, 82-84, 110-112, 128].

In Chapter 5 we explore lower bounds on the MSE of the Bayesian DES of a Gaussian unknown parameter in a WSN, where both multiplicative and additive Gaussian noises are considered in the observation model. To assess the Bayesian estimation performance bounds, and analyze their behaviors with respect to (w.r.t) multiplicative noise variance, two well-known Bayesian bounds from Weiss-Weinstein (WW) family are derived for the problem in hand. First one is Weiss-Weinstein Bound (WWB) [125] which is known to be the tightest bound of the family. The second one is the Bayesian version of classical well-known CRLB (BCRLB) [129]. Analyzing the derived
bounds we discover some scenarios that the multiplicative observation noise enhances the estimation accuracy. We call the phenomena enhancement mode of the multiplicative noise. We contend that in the enhancement mode as the variance of multiplicative noise increases the lower bounds on the MSE of DES decreases which is an unintuitive result. In addition we derived the WWB and BCRLB for the case that sensor measurements are separately quantized with uniform quantizers at the senors and transmitted on error-free channels to the FC. Analyzing the latter bounds we observe the enhancement mode of the multiplicative noise even for the case of quantized observations. However for this case the quantization rate and additive noise variance need to be large enough to generate enhancement mode. Afterwards we compare the MSE performance of two well-known Bayesian estimators (i.e., MMSE and Maximum A Posterior (MAP) estimators) with the derived bounds in different scenarios. The comparisons verify the existence of enhancement modes for these estimators as-well, provided that the network size is large enough. In sequel we consider the case where the variance of multiplicative noise is unknown. Modeling this unknown as a deterministic nuisance parameter we establish lower bounds on MSE of the Bayesian DES. In particular, we characterize and analyze Hybrid Cramér-Rao Lower Bound (HCRLB) [130, 131], Yatarcos Bound (YB) [132], Nayak Bound (NB) [133, 134] and recently proposed Risk Unbiased Bound (RUB) [135] for different scenarios. According to the bounds the enhancement mode can still occur in this case, however it needs larger network sizes, compared with those of known multiplicative noise variance. Next we compare the MSE performance of Minimum Mean Square Error-Maximum Likelihood (MMSE-ML) and Maximum A PosteriorMaximum Likelihood (MAP-ML) with the bounds in various scenarios. The comparisons reveal that although that the bounds may suggest existence of enhancement mode, according to MSE performance of estimators there is no enhancement modes for the case of unknown multiplicative noise variance.

The results in Chapter 5 for Bayesian DES are interestingly different from those reported in [120]
[121] [118] for non-Bayesian DES. The reported improvements in [120] [121] [118] depend on the value of the unknown parameter, which makes exploiting the enhancement modes elusive. Depending the enhancements to be upon on values of unknown makes it elusive to exploit the enhancement modes. On the other hand our reported enhancement mode does not depend on the values of the unknown parameter and therefore is more convenient to assess and characterize. The reported improvements in [120] [121] [118] are also expected to be maximum for a special value of the multiplicative noise variance. In contrast our reported enhancement mode in Chapter 5 is essentially different from that of these papers, such that increasing the multiplicative noise variance improves the estimation accuracy unboundedly till the MSE reaches zero. The authors in [120] and [121] also argued that enhancement mode occurs when the variance of additive observation noise is small in comparison to the energy of the unknown parameter. In contrast our results suggest that for the Bayesian DES, the enhancement mode is more likely to happen for larger additive noise variances. Furthermore for the case of quantized observations, according to our results there is no enhancement mode for binary quantizers in Bayesian DES, in contrast [120] and [121] reported some scenarios where the presence of the multiplicative noise may improve the MSE or estimation performance when sensors employ binary quantizers.

## CHAPTER 2: DISTRIBUTED VECTOR ESTIMATION FOR POWERAND BANDWIDTH-CONSTRAINED WIRELESS SENSOR NETWORKS

We consider the MSE of a Gaussian vector with a linear observation model in an inhomogeneous WSN, in which a FC reconstructs the unknown vector using a linear estimator. Sensors employ uniform multi-bit quantizers and BPSK modulation, and they communicate with the FC over orthogonal power- and bandwidth-constrained wireless channels. We study transmit power and quantization rate (measured in bits per sensor) allocation schemes that minimize the MSE. In particular, we derive two closed-form upper bounds on the MSE in terms of the optimization parameters and propose "coupled" and "decoupled" resource allocation schemes that minimize these bounds. The proposed schemes enables us to find the best resource allocation (i.e., power and bit) in extreme cases where (i) we have scarce total transmit power and ample total bandwidth and (ii) we have plentiful total transmit power and scarce total bandwidth. We show that the bounds are good approximations of the simulated MSE and that the performance of the proposed schemes approaches the clairvoyant CE when the total transmit power or bandwidth is very large. We investigate how the power and rate allocations and overall estimation accuracy are dependent on the sensors' observation qualities and channel gains and on the total transmit power and bandwidth constraints. Our simulations corroborate our analytical results and demonstrate the superior performance of the proposed algorithms.

The remainder of this chapter is organized as follows. Section 2.1 introduces our system model and establishes our optimization problem. In Section 2.2, we derive two closed-form upper bounds, $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$, on the MSE corresponding to the linear estimator at the FC in terms of the optimization parameters (i.e., transmit power and quantization rate per sensor). In Section 2.3, we propose "coupled" resource allocation schemes that minimize these bounds using the iterative ellipsoid method.

This method conducts a multi-dimensional search to find the quantization rate vector. In Section 2.4, we propose "decoupled" resource allocation schemes, which rely on a one-dimensional search to find the quantization rates. Section 2.7 discusses our numerical results. Section 2.8 concludes this chapter.

### 2.1 System Model and Problem Statement

We consider a WSN with $K$ spatially distributed inhomogeneous sensors. Each sensor makes a noisy observation, which depends on an unobservable vector of random parameters $\boldsymbol{\theta}$, locally processes its observation, and transmits a summary of its observation to a FC over erroneous wireless channels. The FC is tasked with estimating $\boldsymbol{\theta}$ via fusing the collective received data from sensors (see Fig. 2.1). We assume that $\boldsymbol{\theta}=\left[\theta_{1}, \ldots, \theta_{q}\right]^{T} \in \mathbb{R}^{q}$ is zero-mean Gaussian with covariance matrix $\mathcal{C}_{\theta}=\mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{\theta}^{\boldsymbol{T}}\right\}$. Let random scalar $x_{k}$ denote the noisy observation of sensor $k$. We assume the following linear observation model:

$$
\begin{equation*}
x_{k}=\mathbf{a}_{k}^{T} \boldsymbol{\theta}+n_{k}, \quad k=1, \ldots, K, \tag{2.1}
\end{equation*}
$$

where $\mathbf{a}_{k}=\left[a_{k_{1}}, \ldots, a_{k_{q}}\right]^{T} \in \mathbb{R}^{q}$ is a known observation gain vector. We assume that the observation noise vector $\boldsymbol{n}=\left[n_{1}, \ldots, n_{K}\right]^{T}$ is zero-mean Gaussian with covariance $\mathcal{C}_{\boldsymbol{n}}$ and that $\boldsymbol{n}$ and $\boldsymbol{\theta}$ are uncorrelated. Define observation vector $\boldsymbol{x}=\left[x_{1}, \ldots, x_{K}\right]^{T}$ and matrix $\mathbf{A}=\left[\mathbf{a}_{1}, \ldots, \mathbf{a}_{K}\right]$. Suppose that $\mathcal{C}_{\boldsymbol{x}}=\mathbb{E}\left\{\boldsymbol{x} \boldsymbol{x}^{T}\right\}$ and $\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}=\mathbb{E}\left\{\boldsymbol{x} \boldsymbol{\theta}^{T}\right\}$ represent the covariance matrix of $\boldsymbol{x}$ and cross-covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{\theta}$, respectively. It is easy to verify that

$$
\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}=\mathbf{A}^{T} \mathcal{C}_{\boldsymbol{\theta}}, \quad \mathcal{C}_{\boldsymbol{x}}=\mathbf{A}^{T} \mathcal{C}_{\boldsymbol{\theta}} \mathbf{A}+\mathcal{C}_{\boldsymbol{n}} .
$$



Figure 2.1: system model

Sensor $k$ employs a uniform quantizer with $M_{k}$ quantization levels and quantization step size $\Delta_{k}$ to map $x_{k}$ into a quantization level $m_{k} \in\left\{m_{k, 1}, \ldots, m_{k, M_{k}}\right\}$, where $m_{k, i}=\frac{\left(2 i-1-M_{k}\right) \Delta_{k}}{2}$ for $i=1, \ldots, M_{k}$. Considering our observation model, we assume that $x_{k}$ almost surely lies in the interval $\left[-\tau_{k}, \tau_{k}\right]$ for some reasonably large value of $\tau_{k}$, i.e., the probability $p\left(\left|x_{k}\right| \geq \tau_{k}\right) \approx 0$. Consequently, we let $\Delta_{k}=\frac{2 \tau_{k}}{M_{k}-1}$. These imply that the quantizer maps $x_{k}$ as the following: if $x_{k} \in\left[m_{k, i}-\frac{\Delta_{k}}{2}, m_{k, i}+\frac{\Delta_{k}}{2}\right]$, then $m_{k}=m_{k, i}$; if $x_{k} \geq \tau_{k}$, then $m_{k}=\tau_{k}$; and if $x_{k} \leq-\tau_{k}$, then $m_{k}=-\tau_{k}$. Following quantization, sensor $k$ maps the index $i$ of $m_{k, i}$ into a bit sequence of length $L_{k}=\log _{2} M_{k}$ and finally modulates these $L_{k}$ bits into $L_{k}$ BPSK modulated symbols [79]. Sensors send their symbols to the FC over orthogonal wireless channels, where transmission is subject to both transmit power and bandwidth constraints. The $L_{k}$ symbols sent by sensor $k$ experience flat fading with a fading coefficient $h_{k}$ and are corrupted by a zero-mean complex Gaussian receiver noise $w_{k}$ with variance $\sigma_{w_{k}}^{2}$. We assume that $w_{k} \mathrm{~s}$ are mutually uncorrelated and that $h_{k}$ does not change during the transmission of $L_{k}$ symbols. Let $P_{k}$ denote the transmit power corresponding to $L_{k}$ symbols from sensor $k$, which we assume is distributed equally among $L_{k}$ symbols. Suppose that there are constraints on the total transmit power and bandwidth of this network, i.e., $\sum_{k=1}^{K} P_{k} \leq P_{t o t}$ and $\sum_{k=1}^{K} L_{k} \leq B_{t o t}$.

To describe the estimation operation at the FC , let $\hat{m}_{k}$ denote the recovered quantization level
corresponding to sensor $k$, where in general $\hat{m}_{k} \neq m_{k}$ due to communication channel errors. The FC first processes the received signals from the sensors individually to recover the transmitted quantization levels. In the absence of knowledge of the joint distribution of $\hat{m}_{k} \mathrm{~s}$ and $\boldsymbol{\theta}$, the FC resorts to a linear estimator $[77,82]$ to form the estimate $\hat{\boldsymbol{\theta}}$ using $\hat{m}_{1}, \ldots, \hat{m}_{K}$. The linear estimator has a low computational complexity [136], and only requires knowledge of $\mathbf{A}$ and second-order statistics $\mathcal{C}_{\boldsymbol{\theta}}$ and $\mathcal{C}_{\boldsymbol{n}}$ to estimate $\boldsymbol{\theta}$. Let $\mathcal{D}_{0}=\mathbb{E}\left\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^{T}\right\}$ denote the error correlation matrix corresponding to this linear estimator, whose $i$-th diagonal entry, $\left[\mathcal{D}_{0}\right]_{i, i}$, is the MSE corresponding to the $i$-th entry of vector $\boldsymbol{\theta}$. We choose $\mathcal{D}_{0}=\operatorname{tr}\left(\mathcal{D}_{0}\right)$ as our MSE distortion metric [4]. Our goal is to find the optimal resource allocation scheme, i.e., quantization rate $L_{k}$ and transmit power $P_{k}$ $\forall k$, that minimizes $\mathcal{D}_{0}$. In other words, we are interested in solving the following optimization problem:

$$
\begin{align*}
& \underset{L_{k}, P_{k}, \forall k}{\operatorname{minimize}} \mathcal{D}_{0}\left(\left\{L_{k}, P_{k}\right\}_{k=1}^{K}\right)  \tag{2.2}\\
& \text { s.t. } \sum_{k=1}^{K} L_{k} \leq B_{\text {tot }}, \sum_{k=1}^{K} P_{k} \leq P_{\text {tot }}, L_{k} \in \mathbb{Z}_{+}, P_{k} \in \mathbb{R}_{+}, \forall k .
\end{align*}
$$

### 2.2 Characterization of MSE

We wish to characterize $\mathcal{D}_{0}$ in terms of the optimization parameters $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$. For this purpose, we take a two-step approach [83]: in the first step, we assume that the quantization levels transmitted by the sensors are received error free at the FC. Based on the error-free transmission assumption, we characterize the MSE due to observation noises and quantization errors. In the second step, we take the contribution of wireless communication channel errors on the MSE into account. This approach provides us with an upper bound on $\mathcal{D}_{0}$, which can be expressed in terms of $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$. Define vector $\boldsymbol{m}=\left[m_{1}, \ldots, m_{K}\right]^{T}$, which includes the transmitted quantization
levels for all sensors, and vector $\hat{\boldsymbol{m}}=\left[\hat{m}_{1}, \ldots, \hat{m}_{K}\right]^{T}$, which consists of the recovered quantization levels at the FC. Let

$$
\begin{equation*}
\check{\boldsymbol{\theta}}=\boldsymbol{G} \boldsymbol{m}, \quad \text { where } \boldsymbol{G}=\mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{m}^{T}\right\}\left(\mathbb{E}\left\{\boldsymbol{m} \boldsymbol{m}^{T}\right\}\right)^{-1} \tag{2.3}
\end{equation*}
$$

Note that $\check{\boldsymbol{\theta}}$ is the LMMSE estimator if the FC had received the transmitted quantization levels without errors [4]. Having $\hat{\boldsymbol{m}}$, the FC employs the same linear operator $\boldsymbol{G}$ to obtain the linear estimate $\hat{\boldsymbol{\theta}}=\boldsymbol{G} \hat{\boldsymbol{m}}$. To characterize the MSE, we define covariance matrices $\mathcal{D}_{1}=\mathbb{E}\{(\check{\boldsymbol{\theta}}-\boldsymbol{\theta})(\check{\boldsymbol{\theta}}-$ $\left.\boldsymbol{\theta})^{T}\right\}$ and $\mathcal{D}_{2}=\mathbb{E}\left\{(\hat{\boldsymbol{\theta}}-\check{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\check{\boldsymbol{\theta}})^{T}\right\}$. One can verify that

$$
\mathcal{D}_{0}=\mathcal{D}_{1}+\mathcal{D}_{2}+2 \mathbb{E}\left\{(\check{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\check{\boldsymbol{\theta}})^{T}\right\}
$$

or equivalently,

$$
\mathcal{D}_{0}=\mathcal{D}_{1}+\mathcal{D}_{2}+2 \mathbb{E}\left\{(\check{\boldsymbol{\theta}}-\boldsymbol{\theta})^{T}(\hat{\boldsymbol{\theta}}-\check{\boldsymbol{\theta}})\right\}
$$

where $\mathcal{D}_{1}=\operatorname{tr}\left(\mathcal{D}_{1}\right)$ and $\mathcal{D}_{2}=\operatorname{tr}\left(\mathcal{D}_{2}\right)$. Applying the Cauchy-Schwarz inequality and using the fact that $(x+y)^{2} \leq 2\left(x^{2}+y^{2}\right)$ for $x, y \geq 0$, we establish an upper bound on $\mathcal{D}_{0}$ as follows:

$$
\begin{equation*}
\mathcal{D}_{0} \leq\left(\sqrt{\mathcal{D}_{1}}+\sqrt{\mathcal{D}_{2}}\right)^{2} \leq 2\left(\mathcal{D}_{1}+\mathcal{D}_{2}\right)=2 \mathcal{D} \tag{2.4}
\end{equation*}
$$

Note that the upper bound on $\mathcal{D}_{0}$ in (2.4) consists of two terms: the first term $2 \mathcal{D}_{1}$ represents the MSE due to observation noises and quantization errors, whereas the second term $2 \mathcal{D}_{2}$ is the MSE due to communication channel errors. In other words, the contributions of observation noises and quantization errors in the upper bound are decoupled from those of communication channel errors.

Relying on (2.4), in the remainder of this section, we derive two upper bounds on $\mathcal{D}$, denoted as $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ in Sections 2.2.1 and 2.2.2, respectively, in terms of $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$. Although $\mathcal{D}_{a}$ is a tighter bound than $\mathcal{D}_{b}$, its minimization demands a higher computational complexity. Leveraging on the
expressions of $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ derived in this section, in Sections 2.3 and 2.4, we propose two distinct schemes, which we refer to as "coupled" and "decoupled" schemes, to address the optimization problem formulated in (2.2) when $\mathcal{D}_{0}$ is replaced with $\mathcal{D}_{a}$ or $\mathcal{D}_{b}$.

### 2.2.1 Characterization of First Bound $\mathcal{D}_{a}$

Recall that $\mathcal{D}=\mathcal{D}_{1}+\mathcal{D}_{2}$ in (2.4). In the following, we first derive $\mathcal{D}_{1}$. Deriving an exact expression for $\mathcal{D}_{2}$ remains elusive. Hence, we derive an upper bound on $\mathcal{D}_{2}$, represented as $\mathcal{D}_{2}^{u p b}$. Let $\mathcal{D}_{a}=$ $\mathcal{D}_{1}+\mathcal{D}_{2}^{u p b}$. Based on (2.4), we have $\mathcal{D}_{0} \leq 2 \mathcal{D} \leq 2 \mathcal{D}_{a}$.

- Derivation of $\mathcal{D}_{1}$ in $\mathcal{D}_{a}$ : Because $\check{\boldsymbol{\theta}}$ is the LMMSE estimator of $\boldsymbol{\theta}$ given $\boldsymbol{m}, \mathcal{D}_{1}$ is the corresponding error covariance matrix. Consequently, $\mathcal{D}_{1}=\operatorname{tr}\left(\mathcal{D}_{1}\right)$ is [4]

$$
\begin{equation*}
\mathcal{D}_{1}=\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}-\mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{m}^{T}\right\}\left(\mathbb{E}\left\{\boldsymbol{m} \boldsymbol{m}^{T}\right\}\right)^{-1} \mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{m}^{T}\right\}^{T}\right) \tag{2.5}
\end{equation*}
$$

where $\mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{m}^{T}\right\}$ and $\mathbb{E}\left\{\boldsymbol{m} \boldsymbol{m}^{T}\right\}$ are cross-covariance and covariance matrices, respectively. To find these matrices, we need to delve into the statistics of quantization errors. For sensor $k$, let the difference between observation $x_{k}$ and its quantization level $m_{k}$, i.e., $\epsilon_{k}=x_{k}-m_{k}$, be the corresponding quantization noise. In general, $\epsilon_{k} \mathrm{~S}$ are mutually correlated and are also correlated with $x_{k} \mathrm{~s}$. However, in [137], it is shown that when correlated Gaussian random variables are quantized with uniform quantizers of step sizes $\Delta_{k} \mathrm{~s}$, quantization noises can be approximated as mutually independent random variables that are uniformly distributed in the interval $\left[-\frac{\Delta_{k}}{2}, \frac{\Delta_{k}}{2}\right]$ and are also independent of quantizer inputs. In this chapter, because the $\boldsymbol{\theta}$ and $n_{k} \mathrm{~s}$ in (2.1) are assumed to be Gaussian, $x_{k} \mathrm{~s}$ are correlated Gaussian that are quantized with uniform quantizers of quantization step sizes $\Delta_{k} \mathrm{~s}$. Hence, $\epsilon_{k} \mathrm{~s}$ are approximated as mutually independent zero-mean uniform random variables with variance $\sigma_{\epsilon_{k}}^{2}=\frac{\Delta_{k}^{2}}{12}$ that are also independent of $x_{k} \mathrm{~S}$ (and thus independent of $\boldsymbol{\theta}$ and
$\left.n_{k} \mathbf{s}\right)$. These imply that $\mathbb{E}\left\{\boldsymbol{\theta} m_{k}\right\}=\mathbb{E}\left\{\boldsymbol{\theta}\left(x_{k}-\epsilon_{k}\right)\right\}=\mathcal{C}_{\boldsymbol{\theta}} \mathbf{a}_{k}$. Therefore,

$$
\begin{equation*}
\mathbb{E}\left\{\boldsymbol{\theta} \boldsymbol{m}^{T}\right\}=\mathcal{C}_{\boldsymbol{\theta}} \mathbf{A}=\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} . \tag{2.6}
\end{equation*}
$$

Additionally, it is straightforward to verify that

$$
\begin{equation*}
\mathbb{E}\left\{\boldsymbol{m} \boldsymbol{m}^{T}\right\}=\mathbf{A}^{T} \mathcal{C}_{\boldsymbol{\theta}} \mathbf{A}+\mathcal{C}_{\boldsymbol{n}}+\boldsymbol{Q}=\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}, \tag{2.7}
\end{equation*}
$$

where $\boldsymbol{Q}=\operatorname{diag}\left(\sigma_{\epsilon_{1}}^{2}, \ldots, \sigma_{\epsilon_{K}}^{2}\right)$. Substituting (2.6) and (2.7) into (2.3) and (2.5) yields

$$
\begin{align*}
\boldsymbol{G} & =\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}  \tag{2.8}\\
\mathcal{D}_{1} & =\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}\right)-\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right) \tag{2.9}
\end{align*}
$$

- Derivation of $\mathcal{D}_{2}^{u p b}$ in $\mathcal{D}_{a}$ : By substituting $\check{\boldsymbol{\theta}}=\boldsymbol{G} \boldsymbol{m}$ and $\hat{\boldsymbol{\theta}}=\boldsymbol{G} \hat{\boldsymbol{m}}$ into $\mathcal{D}_{2}$, we obtain $\mathcal{D}_{2}=$ $\boldsymbol{G \mathcal { M }} \boldsymbol{G}^{\boldsymbol{T}}$, where we define matrix $\boldsymbol{\mathcal { M }}=\mathbb{E}\left\{(\hat{\boldsymbol{m}}-\boldsymbol{m})(\hat{\boldsymbol{m}}-\boldsymbol{m})^{T}\right\}$. Because communication channel noises are mutually uncorrelated, $\mathbb{E}\left\{\left(\hat{m}_{i}-m_{i}\right)\left(\hat{m}_{j}-m_{j}\right)\right\}=0$ for $i \neq j$. Hence, $\boldsymbol{\mathcal { M }}$ is a diagonal matrix, whose $k$-th entry, $[\mathcal{M}]_{k, k}=\mathbb{E}\left\{\left(\hat{m}_{k}-m_{k}\right)^{2}\right\}$, depends on the employed modulation scheme, channel gain $\left|h_{k}\right|$, channel noise variance $\sigma_{\omega_{k}}^{2}$, value of $\tau_{k}$, transmit power $P_{k}$, and quantization rate $L_{k}$. For our system model depicted in Section 5.1, in which sensors utilize BPSK modulation, we obtain (see Appendix A.1)

$$
\begin{equation*}
[\mathcal{M}]_{k, k} \leq\left(\frac{4 \tau_{k}^{2} L_{k}}{3}\right) \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right)=u_{k} \tag{2.10}
\end{equation*}
$$

where $\gamma_{k}=\frac{\left|h_{k}\right|^{2}}{2 \sigma_{\omega_{k}}}$ is the Channel to Noise Ratio (CNR) for sensor $k$. Because $\mathcal{M}$ and $\mathcal{D}_{\mathbf{2}}$ are semipositive definite matrices, i.e., $\mathcal{M} \succeq 0, \mathcal{D}_{\mathbf{2}} \succeq 0$, the bound in (2.10) can provide us with an upper bound on $\mathcal{D}_{2}=\operatorname{tr}\left(\mathcal{D}_{\mathbf{2}}\right)$. Let $\mathcal{M}^{\prime}=\operatorname{diag}\left(u_{1}, \ldots, u_{K}\right)$. Because $\mathcal{M}^{\prime}-\boldsymbol{\mathcal { M }} \succeq 0$ and $\boldsymbol{G} \mathcal{M}^{\prime} \boldsymbol{G}^{\boldsymbol{T}} \succeq 0$, we
find that [138]

$$
\begin{equation*}
\mathcal{D}_{2}=\operatorname{tr}\left(\boldsymbol{G \mathcal { M }} \boldsymbol{G}^{\boldsymbol{T}}\right) \leq \operatorname{tr}\left(\boldsymbol{G} \mathcal{M}^{\prime} \boldsymbol{G}^{\boldsymbol{T}}\right)=\mathcal{D}_{2}{ }^{u p b} \tag{2.11}
\end{equation*}
$$

A remark follows regarding $\mathcal{D}_{1}$ and $\mathcal{D}_{2}^{u p b}$.

- Remark 1: Note that $\mathcal{D}_{1}$ in (2.9) only depends on quantization rates $L_{k} \mathrm{~s}$ through the variances of quantization noises $\epsilon_{k} \mathrm{~s}$ in $\boldsymbol{Q}$. However, $\mathcal{D}_{2}^{u p b}$ in (2.11) depends on transmit powers $P_{k} \mathrm{~s}$ through $\boldsymbol{\mathcal { M }}^{\prime}$, as well as $L_{k}$ s through $\mathcal{M}^{\prime}$ and $\boldsymbol{Q}$ in $\boldsymbol{G}$. Hence, we derive the upper bound $\mathcal{D}_{a}=\mathcal{D}_{1}+\mathcal{D}_{2}^{u p b}$ in terms of the optimization parameters $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$.


### 2.2.2 Characterization of Second Bound $\mathcal{D}_{b}$

Recall from Section 2.2.1 that $\mathcal{D}_{a}=\mathcal{D}_{1}+\mathcal{D}_{2}^{u p b}$, where $\mathcal{D}_{0} \leq 2 \mathcal{D} \leq 2 \mathcal{D}_{a}$. Note that both $\mathcal{D}_{1}$ and $\mathcal{D}_{2}^{u p b}$ involve the inversion of matrix $\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}$, incurring a high computational complexity that increases with $K$. For large $K$, such a matrix inversion required to find the optimal resource allocation (see Section 2.3.1) is burdensome. To reduce the computational complexity, in this section, we derive upper bounds on $\mathcal{D}_{1}$ and $\mathcal{D}_{2}^{\text {upb }}$, which are represented as $\mathcal{D}_{1}^{\text {upb }}$ and $\mathcal{D}_{2}^{\text {uupb }}$, respectively, that do not involve such a matrix inversion. Let $\mathcal{D}_{b}=\mathcal{D}_{1}^{u p b}+\mathcal{D}_{2}^{u u p b}$. Based on (2.4), we have $\mathcal{D}_{0} \leq 2 \mathcal{D} \leq 2 \mathcal{D}_{a} \leq$ $2 \mathcal{D}_{b}$.

- Derivation of $\mathcal{D}_{1}^{u p b}$ in $\mathcal{D}_{b}$ : To find an upper bound on $\mathcal{D}_{1}$ in (2.9), we use the following inequality [72]:

$$
\begin{equation*}
\operatorname{tr}\left(\mathbf{E}^{T} \mathbf{F}^{-1} \mathbf{E}\right) \geq \frac{\left(\operatorname{tr}\left(\mathbf{E}^{T} \mathbf{E}\right)\right)^{2}}{\operatorname{tr}\left(\mathbf{E}^{T} \mathbf{F} \mathbf{E}\right)} \tag{2.12}
\end{equation*}
$$

where $\mathbf{E}$ is arbitrary and $\mathbf{F} \succeq 0$. Recall that $\boldsymbol{Q}$ is a diagonal matrix with non-negative entries, i.e., $\boldsymbol{Q} \succeq 0$. Additionally, $\mathcal{C}_{\boldsymbol{x}} \succeq 0$ because it is a covariance matrix. These imply that $\boldsymbol{Q}+\mathcal{C}_{\boldsymbol{x}} \succeq 0$ [138]. Applying (2.12) to (2.9), we obtain

$$
\begin{equation*}
\mathcal{D}_{1} \leq \operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}\right)-\frac{\left(\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)\right)^{2}}{\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right) \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)}=\mathcal{D}_{1}^{u p b} \tag{2.13}
\end{equation*}
$$

- Derivation of $\mathcal{D}_{2}^{u u p b}$ in $\mathcal{D}_{b}$ : To find an upper bound on $\mathcal{D}_{2}^{u p b}$ in (2.11), we take the following steps:

$$
\begin{align*}
& \mathcal{D}_{2}^{u p b} \stackrel{(a)}{=} \sum_{k=1}^{K} \lambda_{k}\left(\boldsymbol{G}^{T} \boldsymbol{G} \boldsymbol{\mathcal { M }}^{\prime}\right) \stackrel{(b)}{\leq} \sum_{k=1}^{K} \lambda_{k}\left(\boldsymbol{G}^{T} \boldsymbol{G}\right) \lambda_{k}\left(\boldsymbol{\mathcal { M }}^{\prime}\right) \\
& \stackrel{(c)}{\leq} \lambda_{\max }\left(\boldsymbol{G}^{T} \boldsymbol{G}\right) \sum_{k=1}^{K} \lambda_{k}\left(\boldsymbol{\mathcal { M }}^{\prime}\right)=\lambda_{\max }\left(\boldsymbol{G}^{T} \boldsymbol{G}\right) \sum_{k=1}^{K} u_{k}, \tag{2.14}
\end{align*}
$$

where $(a)$ in (2.14) is obtained using the facts that $\operatorname{tr}(\mathbf{E F})=\operatorname{tr}(\mathbf{F E})$ for arbitrary $\mathbf{E}, \mathbf{F}$ with matching sizes and $\operatorname{tr}(\mathbf{E})=\sum_{k} \lambda_{k}(\mathbf{E})=\sum_{k}[\mathbf{E}]_{k, k}$ for a square matrix $\mathbf{E}$ with eigenvalues $\lambda_{k} \mathrm{~s}$, $(b)$ is found using Theorem 9 in [139], and $(c)$ is true because $\lambda_{k}\left(\boldsymbol{G}^{T} \boldsymbol{G}\right) \leq \lambda_{\max }\left(\boldsymbol{G}^{T} \boldsymbol{G}\right)$ for $\forall k$. Next, we derive an upper bound on $\lambda_{\max }\left(\boldsymbol{G}^{T} \boldsymbol{G}\right)$ in (2.14). Using $\boldsymbol{G}$ in (2.8), we find that $\boldsymbol{G}^{T} \boldsymbol{G}=\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$. Note that $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ and $\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}$ are symmetric matrices. Therefore,

$$
\begin{align*}
\lambda_{\max }\left(\boldsymbol{G}^{T} \boldsymbol{G}\right) & \stackrel{(d)}{=}\left\|\boldsymbol{G}^{T} \boldsymbol{G}\right\|_{2} \leq\left[\left\|\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}\right\|_{2}\right]^{2}\left\|\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\right\|_{2} \\
& =\left[\lambda_{\max }\left(\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}\right)\right]^{2} \lambda_{\max }\left(\mathcal{C}_{\boldsymbol{x}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\right) \\
& \stackrel{(e)}{=}\left[\frac{1}{\lambda_{\min }\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)}\right]^{2} \lambda_{\max }\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\right)  \tag{2.15}\\
& \stackrel{(f)}{\leq}\left[\frac{1}{\lambda_{\min }\left(\mathcal{C}_{\boldsymbol{x}}\right)+\min _{k}\left(\sigma_{\epsilon_{k}}^{2}\right)}\right]^{2} \lambda_{\max }\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\right)=\tilde{\lambda}
\end{align*}
$$

(d) holds due to the norm equality [140], (e) is true because $\lambda_{\max }\left(\mathbf{E}^{-1}\right)=\frac{1}{\lambda_{\min }(\mathbf{E})}$ for an invertible $\mathbf{E}$, and $(f)$ is obtained because for $\mathbf{E}, \mathbf{F} \succeq 0$, Weyl's inequality [140] states that $\lambda_{\text {min }}(\mathbf{E}+\mathbf{F}) \geq$ $\lambda_{\min }(\mathbf{E})+\lambda_{\min }(\mathbf{F})$. Additionally, $\lambda_{\min }(\mathbf{Q})=\min _{k}\left(\sigma_{\epsilon_{k}}^{2}\right)$. Combining (2.14) and (2.15), we obtain

$$
\mathcal{D}_{2}^{u p b} \leq \tilde{\lambda} \sum_{k=1}^{K} u_{k}=\mathcal{D}_{2}^{u u p b}
$$

A remark follows regarding $\mathcal{D}_{1}^{u p b}$ and $\mathcal{D}_{2}^{u u p b}$.

- Remark 2: Note that $\mathcal{D}_{1}^{u p b}$ in (2.13) only depends on $L_{k}$ s through the variances of quantization noises $\epsilon_{k} \mathrm{~s}$ in $\boldsymbol{Q}$. However, $\mathcal{D}_{2}^{\text {uupb }}$ depends on $P_{k} \mathrm{~s}$ through $u_{k} \mathrm{~s}$, as well as $L_{k} \mathrm{~s}$ through $u_{k} \mathrm{~s}$ and $\sigma_{\epsilon_{k}}^{2} \mathrm{~s}$ in $\tilde{\lambda}$. Hence, we derive the upper bound $\mathcal{D}_{b}=\mathcal{D}_{1}^{u p b}+\mathcal{D}_{2}^{u u p b}$ in terms of $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$. Proposition 1 summarizes the upper bounds $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ on $\mathcal{D}_{0}$.
- Proposition 1: $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ are $^{1}$

$$
\begin{aligned}
& \mathcal{D}_{0} \leq 2 \mathcal{D}_{a} \leq 2 \mathcal{D}_{b} \\
& \mathcal{D}_{a}=\mathcal{D}_{1}+\mathcal{D}_{2}^{u p b}=\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}\right)-\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)+\operatorname{tr}\left(\boldsymbol{G} \mathcal{M}^{\prime} \boldsymbol{G}^{\boldsymbol{T}}\right) \\
& \mathcal{D}_{b}=\mathcal{D}_{1}^{u p b}+\mathcal{D}_{2}^{u u p b}=\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}\right)-\frac{\left(\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)\right)^{2}}{\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right) \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)}+\tilde{\lambda} \sum_{k=1}^{K} u_{k}
\end{aligned}
$$

## 2.3 "Coupled" Scheme for Resource Allocation

Thus far, we have established that $\mathcal{D}_{0} \leq 2 \mathcal{D} \leq 2 \mathcal{D}_{a} \leq 2 \mathcal{D}_{b}$, where $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ are derived in Sections 2.2.1 and 2.2.2 in terms of $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$. In this section, we address the optimization problem formulated in (2.2) when $\mathcal{D}_{0}$ is replaced with $\mathcal{D}_{a}$ (in Section 2.3.1) or $\mathcal{D}_{b}$ (in Section 2.3.3). Note that (2.2) is a Mixed Integer Nonlinear Programming (MINLP) problem with exorbitant computational complexity [141]. To simplify the problem, we temporarily relax the integer constraint on $L_{k} \mathrm{~s}$ and allow them to be positive numbers, i.e., we consider

$$
\begin{align*}
& \underset{L_{k}, P_{k}, \forall k}{\operatorname{minimize}} \mathcal{D}_{0}\left(\left\{L_{k}, P_{k}\right\}_{k=1}^{K}\right)  \tag{2.16}\\
& \text { s.t. } \sum_{k=1}^{K} L_{k} \leq B_{\text {tot }}, \sum_{k=1}^{K} P_{k} \leq P_{\text {tot }}, L_{k}, P_{k} \in \mathbb{R}_{+}, \forall k
\end{align*}
$$

1 When correlated channels are considered, the non-diagonal entries of matrix $\boldsymbol{\mathcal { M }}^{\prime}$ are $\left[\boldsymbol{\mathcal { M }}^{\prime}\right]_{i, j}=$ $8 \tau^{2}\left[Q\left(\vartheta_{i}, \vartheta_{j} ; \rho_{i, j}^{c}\right)-Q\left(\vartheta_{i}, \vartheta_{j} ;-\rho_{i, j}^{c}\right)\right], i \neq j$, where $\vartheta_{i}=\sqrt{\frac{2 \gamma_{i} P_{i}}{L_{i}}}, \rho_{i, j}^{c}$ is the correlation coefficient between $w_{i}$ and $w_{j}$, and $Q(., . ;$.$) is the two-dimensional Q$-function. However, this does not lead to closed-form solutions for the following resource allocation schemes.

We propose the "coupled" scheme to solve the relaxed problem in (2.16), where the objective function is replaced with $\mathcal{D}_{a}$ or $\mathcal{D}_{b}$. In Section 2.3.2, we discuss a novel approach to migrate from relaxed continuous $L_{k} \mathrm{~s}$ to integer $L_{k} \mathrm{~s}$ solutions.

### 2.3.1 Coupled Scheme for Minimizing $\mathcal{D}_{a}$

The quintessence of this "coupled" scheme is as follows. We replace $\mathcal{D}_{0}$ with $\mathcal{D}_{a}$ in (2.16) and decompose the relaxed problem into two sub-problems (SP1) and (SP2), as follows:

$$
\begin{array}{ll}
\text { (SP1) } \quad \operatorname{given}\left\{L_{k}\right\}_{k=1}^{K}, \operatorname{minimize}_{P_{k}, \forall k} \mathcal{D}_{a}\left(\left\{P_{k}\right\}_{k=1}^{K}\right) \\
& \text { s.t. } \sum_{k=1}^{K} P_{k} \leq P_{t o t}, P_{k} \in \mathbb{R}_{+}, \forall k, \\
\text { (SP2) } \operatorname{given}\left\{P_{k}\right\}_{k=1}^{K}, \operatorname{minimize}_{L_{k}, \forall k} \mathcal{D}_{a}\left(\left\{L_{k}\right\}_{k=1}^{K}\right)  \tag{2.18}\\
& \text { s.t. } \sum_{k=1}^{K} L_{k} \leq B_{t o t}, L_{k} \in \mathbb{R}_{+}, \forall k .
\end{array}
$$

We iterate between solving these two sub-problems until we reach the solution [142].

- Solving (SP1) Given in (2.17): Considering Remark 1, we note that only $\mathcal{D}_{2}^{u p b}$ in $\mathcal{D}_{a}$ depends on $P_{k}$ s. Hence, we replace the objective function in (2.17) with $\mathcal{D}_{2}^{u p b}$. Because $\mathcal{D}_{2}^{u p b}$ is a jointly convex function of $P_{k} \mathrm{~s}$ (see Appendix A.2), we use the Lagrange multiplier method and solve the corresponding Karush-Kuhn-Tucker (KKT) conditions to find the solution. Substituting $u_{k}$ of (2.10) into $\mathcal{M}^{\prime}$ and noting that $\boldsymbol{G}$ does not depend on $P_{k} \mathrm{~s}$, we rewrite (SP1) as follows:

$$
\begin{align*}
& \text { given }\left\{L_{k}\right\}_{k=1}^{K}, \operatorname{minimize}_{P_{k}, \forall k} \sum_{k=1}^{K} \alpha_{k} L_{k} \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right)  \tag{2.19}\\
& \text { s.t. } \sum_{k=1}^{K} P_{k} \leq P_{t o t}, P_{k} \in \mathbb{R}_{+}, \forall k,
\end{align*}
$$

where $\alpha_{k}=\left(4 \tau_{k}^{2} / 3\right)\left\|\boldsymbol{g}_{k}\right\|^{2}$ and $\left\|\boldsymbol{g}_{k}\right\|^{2}$ is the squared Euclidean norm of the $k$-th column of $\boldsymbol{G}$. Let
$\mathcal{L}\left(\left\{L_{k}, P_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)$ be the Lagrangian for (2.19), where $\mu_{k}$ and $\lambda$ are the Lagrange multipliers. The corresponding KKT conditions are as follows:

$$
\begin{aligned}
& \frac{\partial \mathcal{L}}{\partial P_{k}}=-\alpha_{k} \gamma_{k} \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right)-\mu_{k}+\lambda=0, \forall k, \\
& P_{k} \mu_{k}=0, \quad \mu_{k} \geq 0, \quad P_{k} \geq 0, \quad \forall k, \\
& \lambda\left(\sum_{k=1}^{K} P_{k}-P_{t o t}\right)=0, \quad \lambda \geq 0, \quad \sum_{k=1}^{K} P_{k} \leq P_{t o t} .
\end{aligned}
$$

Because $\mathcal{D}_{2}^{u p b}$ is a decreasing function of $P_{k} \mathrm{~s}$ and $P_{\text {tot }}$ (see Appendix A.2), solving (2.19) for $P_{k} \mathrm{~s}$, we find that

$$
\begin{equation*}
P_{k}=\left[\frac{L_{k}}{\gamma_{k}} \ln \left(\frac{\gamma_{k} \alpha_{k}}{\lambda^{*}}\right)\right]^{+}, \quad \forall k, \tag{2.20}
\end{equation*}
$$

where $[x]^{+}=\max (0, x), \sum_{k=1}^{K} P_{k}=P_{\text {tot }}$, and $\ln \lambda^{*}$ is

$$
\begin{equation*}
\ln \lambda^{*}=\left(\sum_{k \notin \mathcal{I}} \frac{L_{k}}{\gamma_{k}}\right)^{-1}\left[-P_{t o t}+\sum_{k \notin \mathcal{I}} \frac{L_{k}}{\gamma_{k}} \ln \left(\gamma_{k} \alpha_{k}\right)\right] . \tag{2.21}
\end{equation*}
$$

Set $\mathcal{I}=\left\{k: P_{k}=0, k=0, \ldots, K\right\}$ in (2.21) as the set of inactive sensors: sensors whose $L_{k}=0$ or $\gamma_{k} \alpha_{k}<\lambda^{*}$, where $\gamma_{k}$ is the CNR of sensor $k$ and $\alpha_{k}$ depends on the parameters of the observation model. Eq. (2.20) indicates that $P_{k}$ depends on both sensor observation and communication channel qualities through $\gamma_{k}$ and $\alpha_{k}$. Additionally, sensor $k$ with a larger $L_{k}$ is allocated a larger $P_{k}$. For the asymptotic regime of large $P_{\text {tot }}$, we substitute (2.21) into (2.20) and obtain the following:

$$
\begin{equation*}
P_{k}=\frac{L_{k} P_{\text {tot }}}{\gamma_{k} \sum_{k \notin \mathcal{I}} \frac{L_{k}}{\gamma_{k}}}, \quad \forall k \tag{2.22}
\end{equation*}
$$

Eq. (2.22) implies that in this asymptotic regime, $P_{k}$ is proportional to $\frac{L_{k}}{\gamma_{k}}$. When $L_{k} \mathrm{~s}$ are equal, a sensor with a smaller CNR is allotted a larger $P_{k}$ (inverse of water filling). When $\gamma_{k} \mathrm{~s}$ are equal,
a sensor with a larger $L_{k}$ is assigned a larger $P_{k}$. Regarding the solution in (2.20), two remarks follow.

- Remark 3: For sensor $k$, we examine how $P_{k}$ varies as $\gamma_{k}$ changes for a given $L_{k}$. We obtain the following:

$$
\begin{equation*}
\frac{\partial P_{k}}{\partial \gamma_{k}}=\frac{L_{k}}{\gamma_{k}^{2}}\left(1-\ln \left(\frac{\gamma_{k} \alpha_{k}}{\lambda^{*}}\right)\right), \quad \forall k \tag{2.23}
\end{equation*}
$$

Examining (2.23) shows that when $\gamma_{k} \alpha_{k}<e \lambda^{*}$, as $\gamma_{k}$ increases, $P_{k}$ increases (water filling). In contrast, when $\gamma_{k} \alpha_{k}>e \lambda^{*}$, as $\gamma_{k}$ increases, $P_{k}$ decreases (inverse of water filling).

- Remark 4: For sensors $i, j$, we examine how $P_{i}, P_{j}$ are related to $\gamma_{i}, \gamma_{j}$. Suppose that $L_{i}=L_{j}=L$ and $\alpha_{i}=\alpha_{j}=\alpha$. When $\frac{e \lambda^{*}}{\alpha}<\gamma_{j}<\gamma_{i}$, then $P_{i}<P_{j}$ (inverse of water filling). In contrast, when $\gamma_{j}<\gamma_{i}<\frac{e \lambda^{*}}{\alpha}$, then $P_{i}>P_{j}$ (water filling).
- Solving (SP2) Given in (2.18): Finding a closed-form solution for this problem remains elusive due to the non-linearity of the cost function and the fact that the inequality constraint on $L_{k} \mathrm{~s}$ is not necessarily active. Let $\mathcal{F}=\left\{L_{k}: \sum_{k=1}^{K} L_{k} \leq B_{t o t}, L_{k} \in \mathbb{R}_{+}, \forall k\right\}$ be the feasible set of (SP2). To solve (SP2), we use a modified version of the ellipsoid method [143, 144]. This cutting-plane optimization method is the generalized form of the one-dimensional bisection method for higher dimensions, and it is theoretically efficient with guaranteed convergence [143]. The description of the method follows. Suppose that the solution of (SP2) is contained in an initial ellipsoid $\epsilon^{0}$ with center $\mathbf{L}^{\prime(0)}$ and shaped by matrix $\mathbf{S}^{(0)} \succeq 0$. The definition of $\epsilon^{0}$ is as follows:

$$
\epsilon^{0}=\left\{\mathbf{z}:\left(\mathbf{z}-\mathbf{L}^{\prime(0)}\right)^{T} \mathbf{S}^{(0)^{-1}}\left(\mathbf{z}-\mathbf{L}^{\prime(0)}\right) \leq 1\right\} .
$$

For $\epsilon^{0}$, we choose a sphere that contains $\mathcal{F}$, with center $\mathbf{L}^{\prime(0)}=\frac{B_{\text {tot }}}{2}[1, \ldots, 1]$, radius $\frac{B_{\text {tot }}}{2} \sqrt{K}$, and thus $\mathbf{S}^{(0)}=\left(\frac{B_{\text {tot }} \sqrt{K}}{2}\right) \mathbf{I}_{K}$. Essentially, this method uses gradient evaluation at iteration $i$ to discard half of $\epsilon^{i}$ and to form $\epsilon^{i+1}$ with center $\mathbf{L}^{(i+1)}$, which is the minimum volume ellipsoid covering the remaining half of $\epsilon^{i}$. Note that $\epsilon^{i+1}$ can be larger than $\epsilon^{i}$ in diameter; however, it is proven that
the volume of $\epsilon^{i+1}$ is smaller than that of $\epsilon^{i}$ and center $\mathbf{L}^{\prime(i)}$ eventually converges to the solution of (SP2). To elaborate this method, suppose that at iteration $i$, we have ellipsoid $\epsilon^{i}$ with center $\mathbf{L}^{\prime(i)}$ and shaped by matrix $\mathbf{S}^{(i)}$. Ellipsoid $\epsilon^{i+1}$ at iteration $i+1$ is obtained by evaluating the gradient $\nabla^{(i)}$, defined below. When $\mathbf{L}^{\prime(i)} \in \mathcal{F}$, then $\nabla^{(i)}$ is the gradient of the objective function (socalled objective cut) evaluated at $\mathbf{L}^{\prime(i)}$. When $\mathbf{L}^{\prime(i)} \notin \mathcal{F}$, then $\nabla^{(i)}$ is the gradient of the inequality constraint that is being violated (so-called feasibility cut) evaluated at $\mathbf{L}^{\prime(i)}$. The update steps are as follows:

$$
\begin{aligned}
\epsilon^{i+1} & =\epsilon^{i} \cap\left\{\mathbf{z}: \nabla^{(i) T}\left(\mathbf{z}-\mathbf{L}^{\prime(i)}\right) \leq 0\right\}, \\
\mathbf{L}^{\prime(i+1)} & =\mathbf{L}^{\prime(i)}-\frac{1}{K+1} \mathbf{S}^{(i)} \tilde{\nabla}^{(i)}, \quad \tilde{\nabla}^{(i)}=\frac{\nabla^{(i)}}{\sqrt{\nabla^{(i) T} \mathbf{S}^{(i)} \nabla^{(i)}}}, \\
\mathbf{S}^{(i+1)} & =\frac{K^{2}}{K^{2}-1}\left(\mathbf{S}^{(i)}-\frac{2}{K+1} \mathbf{S}^{(i)} \tilde{\nabla}^{(i)} \tilde{\nabla}^{(i) T} \mathbf{S}^{(i)}\right),
\end{aligned}
$$

in which

$$
\nabla^{(i)}=\left\{\begin{array}{lll}
\nabla_{o c}^{(i)}, & \text { if } & \sum_{k=1}^{K} L_{k}^{\prime(i)} \leq B_{t o t}, L_{k}^{\prime(i)} \in \mathbb{R}_{+}, \forall k \\
\nabla_{s f c}^{(i)}, & \text { if } & \sum_{k=1}^{K} L_{k}^{\prime(i)}>B_{t o t}, L_{k}^{\prime(i)} \in \mathbb{R}_{+}, \forall k \\
\nabla_{n f c}^{(i)} & \text { if } & L_{j}^{\prime(i)} \leq 0, \text { for some } j \in\{1, \ldots, K\}
\end{array}\right.
$$

where $\nabla_{o c}^{(i)}, \nabla_{s f c}^{(i)}$ and $\nabla_{n f c}^{(i)}$ are the objective cut, rate-sum constraint feasibility cut and nonnegative rate feasibility cut, respectively, evaluated at $\mathbf{L}^{\prime(i)}$ :

$$
\begin{gather*}
\nabla_{o c}^{(i)}=\left[\left.\frac{\partial \mathcal{D}_{a}}{\partial L_{1}}\right|_{L_{1}=L_{1}^{\prime(i)}}, \ldots,\left.\frac{\partial \mathcal{D}_{a}}{\partial L_{K}}\right|_{L_{K}=L_{K}^{\prime(i)}}\right]^{T},  \tag{2.24}\\
\nabla_{s f c}^{(i)}=\left[\left.\frac{\partial\left(\sum_{j=1}^{K} L_{j}\right)}{\partial L_{1}}\right|_{L_{1}=L_{1}^{\prime(i)}}, \ldots,\left.\frac{\partial\left(\sum_{j=1}^{K} L_{j}\right)}{\partial L_{K}}\right|_{L_{K}=L_{K}^{\prime(i)}}\right]^{T},  \tag{2.25}\\
\nabla_{n f c}^{(i)}=\left[\left.\frac{-\partial L_{j}}{\partial L_{1}}\right|_{L_{1}=L_{1}^{\prime(i)}}, \ldots,\left.\frac{-\partial L_{j}}{\partial L_{K}}\right|_{L_{K}=L_{K}^{\prime(i)}}\right]^{T} . \tag{2.26}
\end{gather*}
$$

After some mathematical manipulations and using the fact that $\frac{\partial \mathbf{E}^{-1}}{\partial x}=-\mathbf{E}^{-1} \frac{\partial \mathbf{E}}{\partial x} \mathbf{E}^{-1}$, we find that $\frac{\partial \mathcal{D}_{a}}{\partial L_{k}} \forall k$ in (2.24) is equal to the following:

$$
\operatorname{tr}\left(\boldsymbol{G}\left[\frac{\partial \boldsymbol{Q}}{\partial L_{k}}-\frac{\partial \boldsymbol{Q}}{\partial L_{k}}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \boldsymbol{\mathcal { M }}^{\prime}+\frac{\partial \boldsymbol{\mathcal { M }}^{\prime}}{\partial L_{k}}-\boldsymbol{\mathcal { M }}^{\prime}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \frac{\partial \boldsymbol{Q}}{\partial L_{k}}\right] \boldsymbol{G}^{T}\right)
$$

in which $\frac{\partial \boldsymbol{Q}}{\partial L_{k}}$ and $\frac{\partial \mathcal{M}^{\prime}}{\partial L_{k}}$ are all-zero matrices, except for one non-zero element in each matrix $\left[\frac{\partial \boldsymbol{Q}}{\partial L_{k}}\right]_{k, k}=\frac{\partial \sigma_{\epsilon_{k}}^{2}}{\partial L_{k}}=-\frac{2 \ln 2 \tau_{k}^{2} 2^{L_{k}}}{3\left(2^{L_{k}}-1\right)^{3}}$ and $\left[\frac{\partial \boldsymbol{\mathcal { M }}^{\prime}}{\partial L_{k}}\right]_{k, k}=\frac{\partial u_{k}}{\partial L_{k}}=\frac{4 \tau_{k}^{2}}{3} \exp \left(\frac{-\gamma_{k} P_{k}}{L_{k}}\right)\left[1+\frac{\gamma_{k} P_{k}}{L_{k}}\right]$. Furthermore, $\nabla_{s f c}^{(i)}$ in (2.25) is a vector of all ones, and $\nabla_{n f c}^{(i)}$ in (2.26) is a vector of all zeros and -1 for its $j$-th entry.

As the stopping criterion, we check whether $\sqrt{\nabla^{(i) T} \mathbf{S}^{(i)} \nabla^{(i)}}<\varepsilon$, where $\varepsilon$ is a predetermined error threshold, or whether the number of iterations exceeds a predetermined maximum $I_{\max }$. Fig. 2.2 illustrates the above ellipsoid method for $K=2$ sensors, where the feasible set $\mathcal{F}$ is the triangle with the hatch pattern.

In a nutshell, we have analytically solved (SP1) and provided an iterative ellipsoid method with guaranteed convergence to address (SP2). With these results, we address the problem in (2.16) when $\mathcal{D}_{0}$ is replaced with $\mathcal{D}_{a}$. Let vectors $\mathbf{L}^{c}=\left[L_{1}^{c}, \ldots, L_{K}^{c}\right], \mathbf{P}^{c}=\left[P_{1}^{c}, \ldots, P_{K}^{c}\right]$ denote the solutions to (2.16). We take an iterative approach that switches between solving (SP1) and (SP2) until we converge to vectors $\mathbf{L}^{c}$ and $\mathbf{P}^{c}$ [142]. As the stopping criterion, we check whether the decrease in $\mathcal{D}_{a}$ in two consecutive iterations is less than a predetermined error threshold $\eta$ or whether the number of switches between solving (SP1) and (SP2) exceeds a predetermined maximum $J_{\text {max }}$. The " $a$-coupled" algorithm summarizes the steps described above, guaranteeing that $\mathcal{D}_{a}$ decreases in each iteration $j$. A remark follows regarding the " $a$-coupled" algorithm.

(a) $\boldsymbol{L}^{\prime(i)} \in \mathcal{F} \Rightarrow \nabla^{(i)}=\nabla_{o c}^{(i)}$, update $\epsilon^{i}$ with center $\boldsymbol{L}^{\prime(i)}$ to $\epsilon^{i+1}$ with center $\boldsymbol{L}^{\prime(i+1)}$

(b) $\boldsymbol{L}^{\prime(i)} \notin \mathcal{F}$ because $L_{1}^{\prime(i)}+L_{2}^{\prime(i)}>$ $B_{t o t} \Rightarrow \nabla^{(i)}=\nabla_{s f c}^{(i)}$, update $\epsilon^{i}$ with center $\boldsymbol{L}^{\prime(i)}$ to $\epsilon^{i+1}$ with center $\boldsymbol{L}^{\prime(i+1)}$

(c) $\boldsymbol{L}^{\prime(i)} \notin \mathcal{F}$ because $L_{1}^{\prime(i)}<0 \Rightarrow \nabla^{(i)}=$ $\nabla_{n f c}^{(i)}$, update $\epsilon^{i}$ with center $\boldsymbol{L}^{\prime(i)}$ to $\epsilon^{i+1}$ with center $\boldsymbol{L}^{\prime(i+1)}$

Figure 2.2: Modified ellipsoid method for constrained optimization problem

Data: System parameters defined in Section 5.1
Result: Continuous solutions for optimization parameters $\mathbf{L}^{c}=\left[L_{1}^{c}, \ldots, L_{K}^{c}\right], \mathbf{P}^{c}=\left[P_{1}^{c}, \ldots, P_{K}^{c}\right]$ initialization;
$j=0, \mathbf{L}^{c(0)}=\frac{B_{\text {tot }}}{2}[1, \ldots, 1], \mathbf{P}^{c(0)}=[0, \ldots, 0]$
while 1 do
$\boldsymbol{Q}^{(j)}=\operatorname{diag}\left(\frac{\tau_{1}^{2}}{3\left(2_{1}^{L_{1}^{(c j}}-1\right)^{2}}, \ldots, \frac{\tau_{K}^{2}}{3\left(2^{L_{K}^{C(j)}}-1\right)^{2}}\right)$
$\boldsymbol{G}^{(j)}=\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}^{(j)}\right)^{-1}$ as defined in (2.8)
$\alpha_{k}^{(j)}=\left(4 \tau_{k}^{2} / 3\right)\left\|\boldsymbol{g}_{k}^{(j)}\right\|^{2}, \quad \forall k$
$\lambda^{(j)}=\exp \left(\left(\sum_{k \notin \mathcal{I}} \frac{L_{k}^{c(j)}}{\gamma_{k}}\right)^{-1}\left[-P_{\text {tot }}+\sum_{k \notin \mathcal{I}} \frac{L_{k}^{c(j)}}{\gamma_{k}} \ln \left(\gamma_{k} \alpha_{k}^{(j)}\right)\right]\right)$ as defined in (2.21) $P_{k}^{c(j)}=\left[\frac{L_{k}^{c(j)}}{\gamma_{k}} \ln \left(\frac{\gamma_{k} \alpha_{k}^{(j)}}{\lambda^{(j)}}\right)\right]^{+}, \forall k$, as defined in (2.20)

## Modified ellipsoid method

modified ellipsoid method initialization;
$i=0, \mathbf{L}^{\prime(0)}=\frac{B_{t o t}}{2}[1, \ldots, 1], \quad \mathbf{S}^{(0)}=\left(\frac{B_{t o t} \sqrt{K}}{2}\right) \mathbf{I}$
while 1 do
$\mathbf{L}^{\prime(i+1)}=\mathbf{L}^{\prime(i)}-\frac{1}{K+1} \mathbf{S}^{(i)} \tilde{\nabla}^{(i)}$
$\mathbf{S}^{(i+1)}=\frac{K^{2}}{K^{2}-1}\left(\mathbf{S}^{(i)}-\frac{2}{K+1} \mathbf{S}^{(i)} \tilde{\nabla}^{(i)} \tilde{\nabla}^{(j) T} \mathbf{S}^{(i)}\right)$
$i=i+1$
if $\sqrt{\nabla^{(i+1) T} \mathbf{S}^{(i+1)} \nabla^{(i+1)}}<\varepsilon \vee i>I_{\max }$ then
break
end
end

$$
\mathbf{L}^{c(j+1)}=\mathbf{L}^{\prime(i+1)}
$$

$j=j+1$
if $\mathcal{D}_{a}\left(\mathbf{P}^{c(j)}, \mathbf{L}^{c(j)}\right)-\mathcal{D}_{a}\left(\mathbf{P}^{c(j-1)}, \mathbf{L}^{c(j-1)}\right)<\eta \vee j>J_{\max }$ then
break
end
end
Algorithm: " $a$-coupled" algorithm for minimizing $\mathcal{D}_{a}$

- Remark 5: Implementing the ellipsoid method requires a $K$-dimensional search. Moreover, in the inner loop where $\boldsymbol{L}^{\prime(i)} \in \mathcal{F}$, we have $\nabla^{(i)}=\nabla_{o c}^{(i)}$, which requires inversion of $\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}$. In the outer loop, inversion of $\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}$ is required to update $\alpha_{k}^{(j)}$.


### 2.3.2 Migration from Continuous to Integer Solutions for Rates

We describe an approach for migrating from continuous solution $\mathbf{L}^{c}$ to integer solution $\mathbf{L}^{d}$. Let $\mathcal{X}$ be the index set of sensors with discretized quantization rates. Initially, $\mathcal{X}$ is an empty set. We consider two scenarios: (i) $\sum_{j \in \mathcal{X}} L_{j}^{d}+\sum_{j \notin \mathcal{X}} L_{j}^{c}<B_{t o t}$ and (ii) $\sum_{j \in \mathcal{X}} L_{j}^{d}+\sum_{j \notin \mathcal{X}} L_{j}^{c}=B_{t o t}$. When case (i) occurs, it means that minimizing $\mathcal{D}_{a}$ has not been negatively impacted by a bandwidth shortage. Hence, we discretize the rate of sensor $j$ with the smallest $L_{j}^{c}$ because this sensor is more likely to be the weakest player in the network in the sense that it has the least contribution to $\mathcal{D}_{a}$. We round $L_{j}^{c}$ "up" or "down", depending on which one yields a smaller $\mathcal{D}_{a}$, and consider the discretized $L_{j}^{d}$ as a fixed and final value. When case (ii) occurs, it is very likely that minimizing $\mathcal{D}_{a}$ has been negatively impacted by a bandwidth shortage and that some sensors were imposed smaller rates compared with an unlimited $B_{\text {tot }}$ scenario. Note that in case (ii), rounding up the rate of any sensor would enforce decreasing the rates of some other sensors. Hence, we should discretize in a way such that the positive impact of rounding up a rate on $\mathcal{D}_{a}$ would dominate the negative effect of decreasing the rates of some other sensors on $\mathcal{D}_{a}$. Therefore, we discretize the rate of sensor $j$ with the largest $L_{j}^{c}$ because this sensor is more likely to be the strongest player in the sense that it has the greatest contribution to $\mathcal{D}_{a}$. We round $L_{j}^{c}$ "up" and consider the discretized $L_{j}^{d}$ as a fixed and final value. After each discretization, we need to update $\mathcal{X}$ and the available bandwidth to $B_{t o t}-\sum_{j \in \mathcal{X}} L_{j}^{d}$ and apply the " $a$-coupled" algorithm to reallocate $P_{\text {tot }}$ among all sensors and $B_{t o t}-\sum_{j \in \mathcal{X}} L_{j}^{d}$ among those sensors with continuous valued rates. We continue this procedure until $\mathcal{X}$ includes all sensors. To see the algorithmic view of discretization process see the Appendix A.5.

### 2.3.3 Coupled Scheme for Minimizing $\mathcal{D}_{b}$

Similar to Section 2.3.1, in this section, we consider two sub-problems, which we refer to as (SP3) and (SP4). They are similar to (2.17) and (2.18), with the difference that $\mathcal{D}_{a}$ is replaced by $\mathcal{D}_{b}$. Solutions to (SP3) and (SP4) follow.

- Solving (SP3): Considering Remark 2, we note that only $\mathcal{D}_{2}^{u u p b}$ in $\mathcal{D}_{b}$ depends on $P_{k}$ s. Hence, we replace the objective function in (SP3) with $\mathcal{D}_{2}^{\text {uupb }}$. Because $\mathcal{D}_{2}^{\text {uupb }}$ is a jointly convex function of $P_{k} \mathrm{~S}$ (see Appendix A.2), we use the Lagrange multiplier method and solve the corresponding KKT conditions to find the solution. Additionally, $\mathcal{D}_{2}^{u u p b}$ is a decreasing function of $P_{k} \mathrm{~s}$ and $P_{\text {tot }}$ (see Appendix A.2). Therefore, solving (SP3) for $P_{k} \mathrm{~s}$, we find that

$$
\begin{align*}
P_{k} & =\left[\frac{L_{k}}{\gamma_{k}} \ln \left(\frac{\gamma_{k} \tau_{k}^{2}}{\mu^{*}}\right)\right]^{+}, \forall k, \text { where } \sum_{k=1}^{K} P_{k}=P_{\text {tot }},  \tag{2.27}\\
\ln \mu^{*} & =\left(\sum_{k \notin \mathcal{J}} \frac{L_{k}}{\gamma_{k}}\right)^{-1}\left[-P_{t o t}+\sum_{k \notin \mathcal{J}} \frac{L_{k}}{\gamma_{k}} \ln \left(\gamma_{k} \tau_{k}^{2}\right)\right] . \tag{2.28}
\end{align*}
$$

Set $\mathcal{J}=\left\{k: P_{k}=0, k=0, \ldots, K\right\}$ in (2.28) as the set of inactive sensors: sensors whose $L_{k}=0$ or $\gamma_{k} \tau_{k}^{2}<\mu^{*}$. Note that for the asymptotic regime of large $P_{\text {tot }}$, we have the same power allocation policy as in (2.22).

- Solving (SP4): We apply the modified ellipsoid method that we used for (SP2) to solve (SP4). The feasible set $\mathcal{F}$ and the update steps are similar, with the following difference: the gradient of the objective function $\nabla_{o c}^{(i)}$ changes, and rather than $\frac{\partial \mathcal{D}_{a}}{\partial L_{k}}, \frac{\partial \mathcal{D}_{b}}{\partial L_{k}}$ needs to be derived. However, $\frac{\partial \mathcal{D}_{b}}{\partial L_{k}}=\frac{\partial \mathcal{D}_{1}^{u p b}}{\partial L_{k}}+\frac{\partial \mathcal{D}_{2}^{u u p b}}{\partial L_{k}}$, where

$$
\frac{\partial \mathcal{D}_{1}^{u p b}}{\partial L_{k}}=\frac{-\left(\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)\right)^{2} \operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} \frac{\partial \mathbf{Q}}{\partial L_{k}} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)}{\left(\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right) \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)\right)^{2}}
$$

$$
\frac{\partial \mathcal{D}_{2}^{u u p b}}{\partial L_{k}}= \begin{cases}\tilde{\lambda} \frac{\partial u_{k}}{\partial L_{k}}, & \text { if } k \notin \mathcal{A} \\ \tilde{\lambda}\left[\frac{\partial u_{k}}{\partial L_{k}}-\frac{2 \frac{\partial \sigma_{\epsilon_{k}}^{2}}{\lambda_{m i n}} \sum_{j=1}^{K} u_{j}}{\left.\operatorname{Cox}_{x}\right)+\sigma_{\epsilon_{k}}^{2}}\right], & \text { if } k \in \mathcal{A}\end{cases}
$$

in which set $\mathcal{A}=\left\{k: k=\underset{i}{\operatorname{argmin}}\left(\sigma_{\epsilon_{i}}^{2}\right)\right\}, \tilde{\lambda}$ is defined in (2.15), and $\frac{\partial \mathbf{Q}}{\partial L_{k}}, \frac{\partial u_{k}}{\partial L_{k}}$ are given in Section 2.3.1. Having the solutions for (SP3) and (SP4), we can address the problem in (2.16) when $\mathcal{D}_{0}$ is replaced with $\mathcal{D}_{b}$ utilizing an iterative algorithm similar to the " $a$-coupled" algorithm outlined in Section 2.3.1, which we call the " $b$-coupled" algorithm, and a discretization approach similar to Section 2.3.2. In contrast to the " $a$-coupled" algorithm, the " $b$-coupled" algorithm does not require matrix inversion, although implementing the modified ellipsoid method still requires a $K$ dimensional search.

## 2.4 "Decoupled" Scheme For Resource Allocation

In Section 2.3, we proposed two iterative coupled schemes that minimize $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$. In both schemes, we resorted to the iterative modified ellipsoid method to find $L_{k} \mathrm{~s}$ because finding a closed-form solution for $L_{k} \mathrm{~s}$ remained elusive. We recall the discussion at the beginning of Section 2.2, which indicates that $\mathcal{D}_{1}$ (and its bound $\mathcal{D}_{1}^{\text {upb }}$ ) represent the MSE due to observation noises and quantization errors, whereas $\mathcal{D}_{2}$ (and its bounds $\mathcal{D}_{2}^{\text {upb }}, \mathcal{D}_{2}^{\text {uupb }}$ ) are the MSE due to communication channel errors. Leveraging on this decoupling of the contributions of observation noises and quantization errors from those of communication channel errors, we propose a "decoupled" scheme to minimize these decoupled contributions separately and to find the optimization parameters $\left\{L_{k}, P_{k}\right\}_{k=1}^{K}$ in closed-form expressions, thereby eliminating the computational burden of the modified ellipsoid method for conducting a $K$-dimensional search and finding the $\mathbf{L}^{c}$ vector. Similar to Section 2.3, we start with the "decoupled" scheme to solve the relaxed problem by allowing $L_{k} \mathrm{~s}$ to be positive numbers.

### 2.4.1 Decoupled Scheme for Minimizing $\mathcal{D}_{a}$

The essence of the "decoupled" scheme is to solve the following two sub-problems in a sequential order:

$$
\begin{array}{ll} 
& \text { (SP5) } \\
& \operatorname{minimize}_{L_{k}, \forall k} \mathcal{D}_{1}\left(\left\{L_{k}\right\}_{k=1}^{K}\right) \\
& \text { s.t. } \sum_{k=1}^{K} L_{k} \leq B_{t o t}, L_{k} \in \mathbb{R}_{+}, \forall k,  \tag{2.30}\\
\text { (SP6) } \quad \text { given }\left\{L_{k}\right\}_{k=1}^{K}, \operatorname{minimize}_{P_{k}, \forall k} \mathcal{D}_{2}^{u p b}\left(\left\{P_{k}\right\}_{k=1}^{K}\right) \\
& \text { s.t. } \sum_{k=1}^{K} P_{k} \leq P_{t o t}, P_{k} \in \mathbb{R}_{+}, \forall k .
\end{array}
$$

In contrast to Section 2.3, there is no iteration between (SP5) and (SP6). After solving (SP5) and (SP6), we take a similar approach to Section 2.3.2 to migrate from continuous to integer solutions for $L_{k} \mathrm{~s}$.

- Solving (SP5): We minimize $\mathcal{D}_{1}^{u p b}$ in (SP5) rather than $\mathcal{D}_{1}$ because it yields a closed-form solution for $L_{k} \mathrm{~s}$ given in (2.33). In Section 2.4.2, we discuss minimizing $\mathcal{D}_{a}=\mathcal{D}_{1}+\mathcal{D}_{2}^{u p b}$ when we substitute (2.33) into $\mathcal{D}_{1}$ and (2.20), (2.21), (2.33) into $\mathcal{D}_{2}^{u p b}$. Considering (2.13), we realize that minimizing $\mathcal{D}_{1}^{u p b}$ is equivalent to minimizing $\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T} \boldsymbol{Q} \mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}\right)=\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$, where $\delta_{k}$ which is the squared Euclidean norm of the $k$-th row of $\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}$. Because $\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$ is a jointly convex function of $L_{k} \mathrm{~s}$ (see Appendix A.3), we use the Lagrange multiplier method and solve the corresponding KKT conditions to find the solution. Additionally, $\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$ is a decreasing function of $L_{k} \mathrm{~s}$ and $B_{t o t}$ (see Appendix A.3). Therefore, solving (SP5) in (2.29) for $L_{k} \mathrm{~S}$ and using the approximation $2^{L_{k}}-1 \approx 2^{L_{k}}$, we obtain

$$
\begin{equation*}
L_{k}=\left[\log _{2}\left(\sqrt{\frac{\tau_{k}^{2} \delta_{k} 2 \ln 2}{3 \nu^{*}}}\right)\right]^{+}, \forall k \text { where } \sum_{k=1}^{K} L_{k}=B_{t o t}, \tag{2.31}
\end{equation*}
$$

$$
\begin{equation*}
\nu^{*}=\frac{2 \ln 2}{3}\left[4^{-B_{t o t}} \prod_{k=1}^{K} \delta_{k} \tau_{k}^{2}\right]^{\frac{1}{K}} . \tag{2.32}
\end{equation*}
$$

By substituting (2.32) into (2.31), one can verify that

$$
\begin{equation*}
L_{k}=\left[\frac{B_{t o t}}{K}+\log _{2}\left(\sqrt{\frac{\delta_{k} \tau_{k}^{2}}{\left(\prod_{i=1}^{K} \delta_{i} \tau_{i}^{2}\right)^{\frac{1}{K}}}}\right)\right]^{+}, \forall k \tag{2.33}
\end{equation*}
$$

Examining (2.33), we note that the first term inside the bracket is common among all sensors, whereas the second term differs among sensors and depends on $\delta_{k}, \tau_{k}$. For $\tau_{k}=\tau, \forall k$, a sensor $k$ with a larger $\delta_{k}$ (i.e., better observation quality) is allocated a larger $L_{k}$. Note that because $\mathcal{D}_{1}^{u p b}$ does not capture communication channel errors, $L_{k}$ in (2.33) is independent of the sensor communication channel quality. This is different from the solution of (SP2) in (2.18), where $L_{k} \mathrm{~s}$ depend on both sensor observation and communication channel qualities.

- Solving (SP6): Similar to (SP1) in (2.17), the objective function in (2.30) is a jointly convex function of $P_{k} \mathrm{~s}$ and also decreases as $P_{k} \mathrm{~s}$ and $P_{\text {tot }}$ increase (see Appendix A.2). Indeed, solving (SP6) yields the same solution as that of (SP1), provided in (2.20) and (2.21).


### 2.4.2 Does Depleting $B_{\text {tot }}$ always reduce $\mathcal{D}_{a}$ ?

To answer this question, we consider the solution in (2.33) for the asymptotic regime of large $B_{t o t}$. For large $B_{t o t}$, we have $L_{k} \approx \frac{B_{t o t}}{K}, \forall k$, i.e., we should equally distribute $B_{t o t}$ among sensors. In situations where $B_{\text {tot }}$ is large and the communication channel quality of sensor $k$ is poor (i.e., small $\gamma_{k}$ due to small channel gain $\left|h_{k}\right|$ or small $P_{k}$ due to low $P_{\text {tot }}$ ), the solution in (2.33) can lead to a large value for $\mathcal{D}_{a}$ because sending a large number of bits $L_{k}$ over poor quality channels increases communication channel errors and thus $\mathcal{D}_{2}^{u p b}$.


Figure 2.3: $\mathcal{D}_{1}, \mathcal{D}_{2}^{u p b}, \mathcal{D}_{a}$ versus $B_{\text {tot }}$ (measured in bits)

This observation suggests that we should perhaps first find $B^{o p t}$ (for $B^{o p t}<B_{t o t}$ ), where $B^{o p t}$ depends on channel gains and $P_{t o t}$, and then distribute $B^{o p t}$ among sensors to control the growth of $\mathcal{D}_{2}^{u p b}$ in $\mathcal{D}_{a}$. In fact, when we substitute (2.33) into $\mathcal{D}_{1}$ and (2.20), (2.21), (2.33) into $\mathcal{D}_{2}^{u p b}$, we find that $\mathcal{D}_{1}$ is a decreasing function of $B_{t o t}$, whereas $\mathcal{D}_{2}^{u p b}$ is an increasing function of $B_{t o t}$. As $B_{t o t} \rightarrow 0$, $\mathcal{D}_{1}$ approaches its maximum value $\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{\theta}}\right)$, i.e., trace of covariance of unknowns, whereas $\mathcal{D}_{2}^{u p b}$ goes to zero. However, as $B_{\text {tot }} \rightarrow \infty, \mathcal{D}_{1}$ approaches its minimum value $d_{0}=\operatorname{tr}\left(\mathcal{C}_{\theta}\right)-\operatorname{tr}\left(\mathcal{C}_{x \theta}^{T} \mathcal{C}_{\boldsymbol{x}}^{-1} \mathcal{C}_{x \theta}\right)$, i.e., clairvoyant CE where unquantized sensor observations are available at the FC , whereas $\mathcal{D}_{2}^{\text {upb }}$ increases unboundedly (see Appendix A.4). This trade-off suggests that there should be a value $B^{\text {opt }}$ that minimizes $\mathcal{D}_{a}$. Fig. 2.3 illustrates $\mathcal{D}_{1}, \mathcal{D}_{2}^{u p b}, \mathcal{D}_{a}$ versus $B_{\text {tot }}$ for two values $P_{\text {tot }}^{1}, P_{\text {tot }}^{2}$, where $P_{t o t}^{2} \geq P_{t o t}^{1}$. Fig. 2.3 shows that $B_{2}^{\text {opt }} \geq B_{1}^{\text {opt }}$. This observation can be explained as follows. Note that $\mathcal{D}_{1}$ is independent of $P_{\text {tot }}$, whereas $\mathcal{D}_{2}^{u p b}$ decreases as $P_{\text {tot }}$ increases (Appendix A. 2 shows that $\frac{\partial \mathcal{D}_{p}^{u p b}}{P_{k}} \leq 0 \forall k$, and thus, $\frac{\partial \mathcal{D}_{2}^{u p b}}{P_{t o t}} \leq 0$ ). Hence, as $P_{\text {tot }}$ increases, we can transmit a larger number of bits, i.e., larger $B^{o p t}$, without incurring an increase in communication channel errors.

Motivated by these facts, we propose the " $a$-decoupled" algorithm. This algorithm starts by initi-


Figure 2.4: Behavior of " $a$-decoupled" algorithm
ating $B^{o p t}=1$ and increasing the value of $B^{o p t}$ by one bit at each iteration, where the maximum number of iterations is $B_{t o t}$. At iteration $i$, we find $L_{k}^{(i)}$ s using (2.33) and $P_{k}^{(i)} \operatorname{s}$ using (2.20), (2.21), substitute these $L_{k}^{(i)}, P_{k}^{(i)}$ s into $\mathcal{D}_{a}$, and check whether the decrease in $\mathcal{D}_{a}$ in two consecutive iterations is less than zero, i.e., $\mathcal{D}_{a}\left(\left\{L_{k}^{(i)}, P_{k}^{(i)}\right\}_{k=1}^{K}\right)-\mathcal{D}_{a}\left(\left\{L_{k}^{(i-1)}, P_{k}^{(i-1)}\right\}_{k=1}^{K}\right)<0$. If this inequality holds for $i<B_{\text {tot }}$, we let $B^{o p t}=i$ and find the solution for $L_{k} \mathrm{~s}$ using (2.33) and $P_{k} \mathrm{~s}$ using (2.20), (2.21), when $B_{t o t}$ is substituted with $B^{o p t}$. Otherwise, we let $B^{o p t}=B_{t o t}$ and find the solution for $L_{k} \mathrm{~s}$ using (2.33) and $P_{k} \mathrm{~s}$ using (2.20), (2.21) (see appendix A. 6 for an algorithmic view). Finally, we take the approach in Section 2.3.2 to migrate from continuous to integer solutions for rates and find the corresponding powers.

Fig. 2.4 illustrates the behavior of the " $a$-decoupled" algorithm and in particular how $B^{\text {opt }}, \mathcal{D}_{1}, \mathcal{D}_{2}^{u p b}$, $\mathcal{D}_{a}$ vary as $P_{t o t}$ increases. Note that as $P_{\text {tot }}$ increases, $B^{\text {opt }}$ remains constant for certain (long) intervals and increases for some other (short) intervals of $P_{\text {tot }}$ values. The behavior of $B^{\text {opt }}$ versus $P_{t o t}$ also dictates the behavior of $\mathcal{D}_{1}, \mathcal{D}_{2}^{u p b}, \mathcal{D}_{a}$. For the $P_{\text {tot }}$ intervals where $B^{o p t}$ is fixed, $\mathcal{D}_{1}$ is also fixed because it is independent of $P_{\text {tot }}$, whereas $\mathcal{D}_{2}^{u p b}$ and thus $\mathcal{D}_{a}$ decrease as $P_{\text {tot }}$ increases. For
the $P_{\text {tot }}$ intervals where $B^{\text {opt }}$ increases, $\mathcal{D}_{1}$ decreases and $\mathcal{D}_{2}^{u p b}$ increases (Appendix A. 4 shows that $\frac{\partial \mathcal{D}_{1}}{\partial B_{\text {tot }}} \leq 0$ and $\frac{\partial \mathcal{D}_{2}^{u p b}}{\partial B_{\text {tot }}} \geq 0$ ). Particularly, the multiple peaks for $\mathcal{D}_{2}^{u p b}$ occur at the exact points where we have a one bit increase in $B^{\text {opt }}$. Consider the first peak in $\mathcal{D}_{2}^{u p b}$. As $P_{t o t}$ increases from 7 dB to 8 dB , one would correctly expect to observe that $\mathcal{D}_{2}^{\text {upb }}$ decreases. However, because $B^{\text {opt }}$ also increases from 2 bits to 3 bits and $\frac{\partial \mathcal{D}_{2}^{u p b}}{\partial B_{t o t}} \geq 0$, it appears that $\mathcal{D}_{2}^{u p b}$ increases as $P_{\text {tot }}$ increases from 7 dB to 8 dB . A similar justification holds for the other peaks in $\mathcal{D}_{2}^{u p b}$. Overall, as $P_{\text {tot }}$ increases, $\mathcal{D}_{a}$ decreases because the decrease in $\mathcal{D}_{1}$ when $B^{\text {opt }}$ increases dominates the increase in $\mathcal{D}_{2}^{u p b}$. A remark follows regarding the modified " $a$-decoupled" algorithm.

- Remark 6: The modified " $a$-decoupled" algorithm requires only a one-dimensional search to find $B^{o p t}$ and thus $L_{k} \mathrm{~s}$ (2.33), as opposed to the $K$-dimensional search required by the modified ellipsoid method in the " $a$-coupled" algorithm. Additionally, finding $B^{o p t}$ at most requires $B_{\text {tot }}$ number of iterations, and no switching between solving (SP5) and (SP6) is needed.


### 2.4.3 Decoupled Scheme for Minimizing $\mathcal{D}_{b}$

Note that the " $a$-decoupled" algorithm still requires inversion of matrix $\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}$ to calculate $\alpha_{k}$ and find $P_{k}$ using (2.20), (2.21). To eliminate this matrix inversion, we propose minimizing $\mathcal{D}_{2}^{u u p b}$ rather than $\mathcal{D}_{2}^{u p b}$. Because $\mathcal{D}_{2}^{u u p b}$ is a jointly convex function of $P_{k} \mathrm{~s}$ (see Appendix A.2), substituting $L_{k} \mathrm{~s}$ of (2.33) into $\mathcal{D}_{2}^{u u p b}$ and minimizing it with respect to $P_{k} \mathrm{~s}$, we reach the solution provided in (2.27), (2.28). Let the " $b$-decoupled" algorithm be the one that minimizes $\mathcal{D}_{1}^{\text {upb }}$ and $\mathcal{D}_{2}^{\text {uupb }}$ separately in $\mathcal{D}_{b}$. This algorithm is very similar to the " $a$-decoupled" algorithm described in Section 2.4.2, with the difference that when finding $B^{o p t}$, at iteration $i$, we find $L_{k}^{(i)}$ s using (2.33) and $P_{k}^{(i)}$ s using (2.27), (2.28), substitute these $L_{k}^{(i)}, P_{k}^{(i)}$ s into $\mathcal{D}_{b}$, and check whether the decrease in $\mathcal{D}_{b}$ in two consecutive iterations is less than zero. After finding $B^{o p t}$, we find the solution for $L_{k} \mathrm{~s}$ using (2.33) and $P_{k} \mathrm{~s}$ using (2.27), (2.28). The behavior of the " $b$-decoupled" algorithm and in particular
how $B^{\text {opt }}, \mathcal{D}_{1}^{u p b}, \mathcal{D}_{2}^{\text {uupb }}, \mathcal{D}_{b}$ vary as $P_{\text {tot }}$ increases is similar to the " $a$-decoupled" algorithm depicted in Fig. 2.4 and is omitted due to the redundancy.

### 2.5 Complexity Comparison of Algorithms

We discuss the computational complexity of the " $a$-coupled", " $b$-coupled", " $a$-decoupled", and "b-decoupled" algorithms.

- " $a$-coupled" algorithm: in this algorithm, we switch between solving (SP1) and (SP2) until it reaches the stopping criteria. Suppose that the algorithm converges after $I_{a}$ iterations. In each iteration, we use (2.20) to solve (SP1) and the modified ellipsoid method to solve (SP2). When employing the matrix inversion algorithms in [145], the complexity of finding $P_{k} \mathrm{~s}$ via (2.20) is $\mathcal{O}\left(K^{2.37}\right)$ because we need to calculate $\alpha_{k} \mathrm{~s}$ and $\boldsymbol{G}$, which involve calculating $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$. The complexity order of the ellipsoid method is $\mathcal{O}\left(K^{2}\right)$ [143]. Thus, the overall complexity order of the " $a$-coupled" algorithm is $\mathcal{O}\left(I_{a}\left(K^{2.37}+K^{2}\right)\right)$. In our simulations, we find that $I_{a} \approx 6$ for $K=20$.
- " $b$-coupled" algorithm: the computational complexity of this algorithm is very similar to that of the " $a$-coupled" algorithm because the algorithm iterates between solving (SP3) and (SP4) until it reaches the solution. Suppose that the algorithm converges after $I_{b}$ iterations. No matrix inversion is involved for solving (SP3), and the complexity of finding $P_{k} \mathrm{~s}$ via (2.27) is $\mathcal{O}(K)$ because the complexity order of finding $\ln \mu^{*}$ in (2.28) is $\mathcal{O}(K)$. Hence, the overall complexity order of the "b-coupled" algorithm is $\mathcal{O}\left(I_{b}\left(K^{2}+K\right)\right)$. For $K=20$, we find $I_{b} \approx 6$.
- " $a$-decoupled" algorithm: the structures of the "decoupled" algorithms are different from those of the "coupled" algorithms. These algorithms include a one-dimensional search over $B_{t o t}$ to find $B^{\text {opt }}$ (where $0 \leq B^{\text {opt }} \leq B_{t o t}$ ). Starting with $B^{\text {opt }}=1$ and increasing the value of $B^{\text {opt }}$ by one bit,
for each candidate of $B^{o p t}$, we need (i) to find $P_{k} \mathrm{~s}$ via (2.20), which has a complexity order of $\mathcal{O}\left(K^{2.37}\right)$, and (ii) to calculate $\mathcal{D}_{a}$, which involves finding $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ that is found in (i). Thus, the overall complexity order of the " $a$-decoupled" algorithm is $\mathcal{O}\left(B^{o p t} K^{2.37}\right)$, and in the worst case, it is $\mathcal{O}\left(B_{t o t} K^{2.37}\right)$.
- $b$-decoupled" algorithm: the computational complexity of this algorithm is very similar to that of the " $a$-decoupled" algorithm, with a difference. In this algorithm, for each candidate of $B^{o p t}$, we calculate $\mathcal{D}_{b}$ rather than $\mathcal{D}_{a}$. We note that finding $\mathcal{D}_{b}$ does not require a matrix inversion and that the complexity order of calculating $\mathcal{D}_{b}$ is $\mathcal{O}(K)$. Hence, the overall complexity order of the " $b$-decoupled" algorithm is $\mathcal{O}\left(B^{\text {opt }} K\right)$.


### 2.6 Discussion on Observation Model

Adopting the linear observation model has several precedents in the distributed estimation literature $[70,72,73,76,77,83]$. We note that, regardless of the specific observation model that the system designer may adopt, our proposed algorithms only need $\mathcal{C}_{\boldsymbol{x}}$ and $\mathcal{C}_{\boldsymbol{\theta} \boldsymbol{x}}$ to find the resource allocations. As an example for nonlinear observation, consider $x_{k}=s_{k}+n_{k}, k=1, \ldots, K$, where $s_{k} \sim \mathcal{N}\left(0, \sigma_{k}^{2}\right)$ and $\boldsymbol{\theta}=\left[\theta_{1}, \ldots, \theta_{q}\right]^{T}$ are jointly Gaussian. The covariance of $s_{k}$ and $\theta_{i}$ is $\operatorname{Cov}\left(s_{k}, \theta_{i}\right)=\sigma_{k} \sigma_{\theta_{i}} \rho_{k, \theta_{i}}^{s}$ and $\operatorname{Cov}\left(s_{k}, s_{l}\right)=\sigma_{k} \sigma_{l} \rho_{k, l}^{s}$. One can employ the well-known "spatial correlation model" in [146] [147] to define the correlation coefficients as $\rho_{k, \theta_{i}}^{s}=\exp \left(-\left(\frac{d_{k, \theta_{i}}}{\nu_{1}}\right)^{\nu_{2}}\right)$ and $\rho_{k, l}^{s}=\exp \left(-\left(\frac{d_{k, l}}{\nu_{1}}\right)^{\nu_{2}}\right)$, where $d_{k, \theta_{i}}$ is the distance between sensor $k$ and source $\theta_{i}, d_{k, l}$ is the distance between sensors $k$ and $l, \nu_{1}>0$ is a parameter that controls how the spatial correlation changes with the distance, and $0<\nu_{2}<2$ is a roughness factor.

### 2.7 Numerical and Simulation Results

In this section, through simulations, we corroborate our analytical results. Additionally, we investigate the effects of observation noise correlation and size of the network on the upper bounds and simulated MSE and examine the tightness of the bounds. Without loss of generality and for the simplicity of the resource allocation presentation, we first consider a small network of $K=3$, $\mathcal{C}_{\boldsymbol{\theta}}=[1(\sqrt{2} / 2) ;(\sqrt{2} / 2) 2], \mathbf{a}_{1}=[11]^{T}, \mathbf{a}_{2}=\left[\begin{array}{ll}0.6 & 0.6\end{array}\right]^{T}, \mathbf{a}_{3}=\left[\begin{array}{ll}0.4 & 0.4\end{array}\right]^{T}, \sigma_{w_{k}}^{2}=1, h_{k}=1$, uncorrelated observation noises with the covariance matrix $\mathcal{C}_{\boldsymbol{n}}=\operatorname{diag}(1,1,1)$ for all algorithms. "CS" and "DS" in the legends of the figures indicate the continuous solutions and discrete solutions of the algorithms, respectively.

(a) $B_{\text {tot }}=30$ (bits)

(b) $B_{t o t}=3$ (bits)

Figure 2.5: " $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$


Figure 2.6: "a-coupled" algorithm $\left\{L_{k}\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$

Figs. 2.5 and 2.6 illustrate $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ and $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $P_{\text {tot }}$, respectively, for the " $a$ coupled" algorithm and $B_{t o t}=30,3$ bits. Figs. 2.5.a and 2.6.a for $B_{t o t}=30$ bits (abundant bandwidth) show that as $P_{\text {tot }}$ increases, both the power and rate allocation approach a uniform allocation. This is in agreement with (2.22). When $P_{\text {tot }}$ is small, only sensor 1 is active. As $P_{\text {tot }}$ increases, sensors 2 and 3 become active in sequential order. Figs. 2.5.b and 2.6.b for $B_{t o t}=3$ bits (scarce bandwidth) show that only sensor 1 is active and $P_{1}=P_{t o t}$. Overall, these observations imply that the power and rate allocation depends on both $P_{t o t}$ and $B_{t o t}$, e.g., when we have plentiful $P_{t o t}$ and scarce $B_{t o t}$, only the sensor with the largest observation gain is active. Moreover, the uniform power and rate allocation is near optimal when we have ample $P_{t o t}$ and $B_{\text {tot }}$.

Figs. 2.7 and 2.8 depict $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ and $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $P_{\text {tot }}$, respectively, for the " $a$-decoupled" algorithm and $B_{\text {tot }}=30,3$ bits. Similar to the " $a$-coupled" algorithm, we observe that when both $P_{t o t}$ and $B_{t o t}$ are abundant, the uniform power and rate allocation is near optimal. This result is in agreement with (2.22) and (2.33).


Figure 2.7: " $a$-decoupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$

However, when $B_{t o t}$ is scarce, the power and rate allocation is far from being uniform. In fact, when $P_{\text {tot }}$ is ample, (2.22) indicates that $P_{k}$ is proportional to $L_{k}$. Additionally, when $B_{t o t}$ is scarce, (2.33) states that $L_{k} \mathrm{~s}$ and consequently $P_{k} \mathrm{~s}$ are not uniformly distributed. These results indicate that the power and rate allocation is affected by the sensors' observation qualities and channel gains, as well as both $P_{\text {tot }}$ and $B_{\text {tot }}$. There are two slight differences between the " $a$ coupled" and " $a$-decoupled" algorithms: (i) for $B_{\text {tot }}=30$ bits (ample bandwidth), sensors 2 and 3 become active at smaller $P_{\text {tot }}$ values in the "a-decoupled" algorithm, and (ii) for $B_{t o t}=3$ bits (scarce bandwidth), the " $a$-decoupled" algorithm ultimately activates all sensors as $P_{\text {tot }}$ increases, whereas the " $a$-coupled" algorithm only activates sensor 1 . Note that for scarce bandwidth, even when $P_{\text {tot }}$ is very large, the power and rate allocation in " $a$-decoupled" algorithm is non-uniform.

(a) $B_{t o t}=30$ (bits)

(b) $B_{t o t}=3$ (bits)

Figure 2.8: " $a$-decoupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $P_{\text {tot }}$

Figs. 2.9 and 2.10 depict $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ and $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$, respectively, for the " $a$-coupled" algorithm and $P_{\text {tot }}=16,30 \mathrm{~dB}$. The observations in these figures are in full agreement with the former ones. In particular, for $P_{\text {tot }}=30 \mathrm{~dB}$ (large power), when $B_{t o t}$ is small, only sensor 1 is active. As $B_{\text {tot }}$ increases, sensors 2 and 3 also become active in a way such that the power and rate allocation approaches uniformity for large $B_{t o t}$. However, for $P_{t o t}=16 \mathrm{~dB}$, only sensor 1 is active and $P_{1}=P_{t o t}$.

Figs. 2.11 and 2.12 illustrate $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ and $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$, respectively, for the " $a$-decoupled" algorithm and $P_{\text {tot }}=16,30 \mathrm{~dB}$. While the behaviors of the " $a$-coupled" and " $a$ decoupled" algorithms have similarities, they have the following differences: (i) for $P_{\text {tot }}=30 \mathrm{~dB}$, sensors 2 and 3 become active at a smaller $B_{\text {tot }}$ value in the " $a$-decoupled" algorithm, and (ii) for $P_{\text {tot }}=16 \mathrm{~dB}$, the " $a$-decoupled" algorithm ultimately activates all sensors as $B_{\text {tot }}$ increases, whereas the " $a$-coupled" algorithm only activates sensor 1 .


Figure 2.9: " $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $B_{\text {tot }}$


Figure 2.10: " $a$-coupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$

Note that for $P_{\text {tot }}=16 \mathrm{~dB}$, even when $B_{\text {tot }}$ is very large, the power and rate allocation in the " $a$ decoupled" algorithm is non-uniform. This is because $B^{o p t}<B_{t o t}$ in this case, and according to (2.33), (2.20), (2.21), the power and rate allocation would be non-uniform.


Figure 2.11: " $a$-decoupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $B_{\text {tot }}$


Figure 2.12: " $a$-decoupled" algorithm $\left\{L_{k}\right\}_{k=1}^{K}$ vs. $B_{\text {tot }}$

(a) $B_{t o t}=30$ (bit)

(b) $B_{t o t}=3$ (bit)

Figure 2.13: $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $P_{\text {tot }}$ for all algorithms

In Fig. 2.13, we plot $\mathcal{D}_{a}$ when implementing the " $a$-coupled" and " $a$-decoupled" algorithms and $\mathcal{D}_{b}$ when implementing the " $b$-coupled" and " $b$-decoupled" algorithms vs. $P_{\text {tot }}$ for $B_{t o t}=30,3$ bits. We observe that the " $a$-coupled" and " $b$-decoupled" algorithms perform the best and the worst, respectively. Furthermore, the algorithms outperform uniform resource allocation (except for the " $b$-decoupled" algorithm when $B_{t o t}=3$ bits and $13<P_{\text {tot }}<18 \mathrm{~dB}$ ). For small $P_{\text {tot }}$, the performance of each algorithm does not change as we decrease $B_{t o t}=30$ bits to $B_{t o t}=3$ bits. This observation can be explained as follows. For small $P_{t o t}$, the communication channels cannot support reliable transmission of a large number of bits. Hence, the algorithms allocate few bits to sensors, and increasing $B_{\text {tot }}$ does not improve the performance. Another observation is that for $B_{t o t}=30$ bits (plentiful bandwidth) and large $P_{\text {tot }}$, the performance of all algorithms reaches the clairvoyant benchmark $d_{0}$, whereas for $B_{t o t}=3$ bits (scarce bandwidth) and large $P_{\text {tot }}$, there is a persistent gap with $d_{0}$ for each algorithm due to quantization errors, and the " $a$-coupled" and " $b$-decoupled" algorithms perform the best and the worst, respectively.


Figure 2.14: Simulated MSE vs. $P_{\text {tot }}$ for all algorithms

Fig. 2.14 depicts the Monte Carlo simulated MSE when the " $a$-coupled", " $b$-coupled", " $a$-decoupled" and " $b$-decoupled" algorithms are implemented for power and rate allocation vs. $P_{\text {tot }}$ for $B_{t o t}=$ 30,3 bits. Similar observations to those of Fig. 2.13 are made, with a few differences: (i) the " $b$ coupled" algorithm outperforms the " $a$-decoupled" algorithm in very low $P_{\text {tot }}$, and (ii) the simulated MSE obtained by the " $b$-coupled", " $a$-decoupled", and " $b$-decoupled" algorithms approaches the same value for $B_{t o t}=3$ bits and large $P_{t o t}$. The reason is perhaps that the discretized quantization rates are the same for these algorithms, and according to (2.22), for large $P_{t o t}$, the $P_{k}$ s of these algorithms become identical.

Fig. 2.15 depicts $\mathcal{D}_{a}$ when implementing the " $a$-coupled" and " $a$-decoupled" algorithms and $\mathcal{D}_{b}$ when implementing the " $b$-coupled" and " $b$-decoupled" algorithms vs. $B_{\text {tot }}$ for $P_{\text {tot }}=30,16 \mathrm{~dB}$. The bowl-shaped curves for uniform resource allocation are caused by the same phenomena that we explained in Fig. 2.3. Avoiding this effect is the reason why we find $B^{o p t} \leq B_{t o t}$ in the " $a$ decoupled" and " $b$-decoupled" algorithms.


Figure 2.15: $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $B_{t o t}$ for all algorithms

Fig. 2.16 depicts the simulated MSE when the " $a$-coupled", " $b$-coupled", " $a$-decoupled" and " $b$ decoupled" algorithms are implemented for power and rate allocation vs. $B_{t o t}$ for $P_{t o t}=30,16 \mathrm{~dB}$. We make similar observations and conclusions to those that we made for Fig. 2.14. For $P_{\text {tot }}=30$ dB and large $B_{t o t}$, the performance of all algorithms reaches the clairvoyant benchmark $d_{0}$. However, for $P_{\text {tot }}=16 \mathrm{~dB}$ and large $B_{t o t}$, there is a persistent gap with $d_{0}$ for each algorithm due to communication channel errors. When comparing Figs. 2.15 and 2.16, we note that the behavior of the bounds $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $B_{t o t}$ is very similar to that of the simulated MSE.

Effect of observation noise correlation: To investigate the effect of observation noise correlation on the resource allocation and the MSE upper bounds, we provide a numerical example with $\mathcal{C}_{n}=[1.5 .3 ; .510 .2 ; .30 .21]$. Fig. 2.17 presents $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}$ for the " $a$-coupled" algorithm and $B_{t o t}=30$ bits for both uncorrelated and correlated observation noises.


Figure 2.16: Simulated MSE vs. $B_{\text {tot }}$ for all algorithms


Figure 2.17: " $a$-coupled" algorithm $\left\{10 \log _{10}\left(P_{k}\right)\right\}_{k=1}^{3}$ vs. $P_{\text {tot }}, B_{\text {tot }}=30$ bits, $K=3$

In Fig. 2.18, we plot $\mathcal{D}_{a}$ when implementing the " $a$-coupled" and " $a$-decoupled" algorithms vs. $P_{t o t}$ for $B_{t o t}=30$ bits for both uncorrelated and correlated observation noises. For correlated noises, we observe that more resources are allocated to sensor 1 and that less resources are allotted to sensors 2 and 3 compared with those of uncorrelated noises. This result is intuitively expected because the observation noise of sensor 1 is highly correlated to those of sensors 2 and 3 , implying that observation $x_{1}$ is highly correlated to $x_{2}$ and $x_{3}$. Combined with the fact that sensor 1 has the largest observation gain, we conclude that because we have more information about $x_{1}$ (more resources are allocated to sensor 1), we need less information about $x_{2}, x_{3}$ (less resources are allocated to sensors 2 and 3 ) in this example. Consistent with the classical literature [4] (correlation leads to a higher MSE), we observe that the upper bounds have larger values for correlated observation noises. When $P_{\text {tot }}$ is small, because only the sensor with the largest observation gain is active, correlation between observation noises does not affect the values of the upper bounds.

Effect of network size and spatially correlated noises: We consider $K=20$ sensors randomly deployed in a $20 \times 20$ field. Assuming a Cartesian coordinate system with the origin at the center of the field, two unknown sources are deployed at coordinates $(-5,0)$ and $(5,0)$. We adopt the observation model in Section 2.6 with $\nu_{1}=10, \nu_{2}=1$ and $\sigma_{s_{k}}=1, \forall k$. We also assume correlated observation and communication channel noises based on the same spatial model, i.e., $\operatorname{Cov}\left(n_{i}, n_{j}\right)=\sigma_{n_{i}} \sigma_{n_{j}} \exp \left(-\left(\frac{d_{i, j}}{\beta_{1}}\right)^{\beta_{2}}\right)$ and $\operatorname{Cov}\left(w_{i}, w_{j}\right)=\sigma_{w_{i}} \sigma_{w_{j}} \exp \left(-\left(\frac{d_{i, j}}{\zeta_{1}}\right)^{\zeta_{2}}\right)$, with $\beta_{1}=1, \beta_{2}=1$, $\zeta_{1}=0.1$, and $\zeta_{2}=1$.


Figure 2.18: $\mathcal{D}_{a}$ vs. $P_{\text {tot }}$ for " $a$-coupled" and " $a$-decoupled" algorithms, $B_{\text {tot }}=30$ bits, $K=3$


Figure 2.19: $\mathcal{D}_{a}$ and $\mathcal{D}_{b}$ vs. $P_{\text {tot }}$ for all algorithms for $K=20$

Fig. 2.19 depicts the results that are analogous to those of Fig. 2.13 for $B_{t o t}=120,10$ bits. We can infer similar conclusions as for Fig. 2.13, except for two differences: (i) for $B_{t o t}=30$ bits (large bandwidth) and large $P_{\text {tot }}$, all upper bounds converge to $d_{0}=\operatorname{tr}\left(\mathcal{C}_{\theta}\right)-\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \theta}^{T} \mathcal{C}_{\boldsymbol{x}}^{-1} \mathcal{C}_{\boldsymbol{x} \theta}\right)$ when $K=3$, whereas for $B_{t o t}=120$ bits (large bandwidth) and large $P_{t o t}, \mathcal{D}_{b}$ of the " $b$-coupled" and" $b$ decoupled" algorithms converge to $d_{0}^{\prime}=\operatorname{tr}\left(\mathcal{C}_{\theta}\right)-\frac{\left(\operatorname{tr}\left(\mathcal{C}_{x \theta}^{T} \mathcal{C}_{x \theta}\right)\right)^{2}}{\operatorname{tr}\left(\mathcal{C}_{x \theta}^{T} \mathcal{C}_{x} \mathcal{C}_{x \theta}\right)}, d_{0}^{\prime}>d_{0}$ when $K=20$. This is due to the larger difference between $\mathcal{D}_{1}$ and $\mathcal{D}_{1}^{u p b}$ in $K=20$ compared with that of $K=3$. As $B_{t o t} \rightarrow \infty$, we find that $\mathcal{D}_{1} \rightarrow d_{0}$ and $\mathcal{D}_{1}^{u p b} \rightarrow d_{0}^{\prime}$ and that the difference $d_{0}^{\prime}-d_{0}$ increases as $K$ increases. (ii) The gap between $\mathcal{D}_{b}$ of the " $b$-coupled" and " $b$-decoupled" algorithms and $\mathcal{D}_{a}$ of the " $a$-coupled" and " $a$-decoupled" algorithms are larger when $K=20$. This observation can be justified as follows. Compared with $\mathcal{D}_{a}, \mathcal{D}_{b}$ captures less statistical information about the observation model of the network. While $\tilde{\lambda}$ in $\mathcal{D}_{2}^{u u p b}=\tilde{\lambda} \sum_{k=1}^{K} u_{k}$ summarizes some of the statistical information embedded in $\boldsymbol{G}=\mathcal{C}_{\boldsymbol{x} \boldsymbol{\theta}}^{T}\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}, \mathcal{D}_{2}^{u p b}$ holds all the information embedded in $\boldsymbol{G}$.


Figure 2.20: Simulated MSE vs. $P_{t o t}$ for all algorithms for $K=20$

Furthermore, we minimize $\mathcal{D}_{1}^{u p b}$ in the " $b$-decoupled" algorithm (while we minimize $\mathcal{D}_{1}$ in the " $a$ decoupled" algorithm), which leads to a further loss of statistical information embedded in $\mathcal{C}_{x}$. As $K$ increases, this statistical information loss becomes more significant, and thus, the gap between $\mathcal{D}_{b}$ and $\mathcal{D}_{a}$ increases.

Fig. 2.20 plots results that are analogous to those of Fig. 2.14 for $K=20$. Similar comments can be made as those for Fig. 2.14. Note that in contrast to Fig. 2.19, the simulated MSE of all four algorithms in Fig. 2.20 converge to $d_{0}$ for $B_{\text {tot }}=30$ bits (large bandwidth) and large $P_{\text {tot }}$ (recall that $d_{0}$ is also the MSE benchmark corresponding to estimating $\boldsymbol{\theta}$ using unquantized observations $x_{k} \mathrm{~s}$ ). Figs. 2.14 and 2.20 show the simulated MSE for all the proposed algorithms for $K=3$ and $K=20$, respectively. As expected, the " $a$-coupled" and " $a$-decoupled" algorithms outperform the " $b$-coupled" and " $b$-decoupled" algorithms.


Figure 2.21: Simulated MSE and $\mathcal{D}_{a}$ vs. $P_{\text {tot }}$ for the " $a$-coupled" and " $a$-decoupled" algorithms for $K=3$


Figure 2.22: Simulated MSE and $\mathcal{D}_{a}$ vs. $P_{\text {tot }}$ for the " $a$-coupled" and " $a$-decoupled" algorithms for $K=20$

Furthermore, comparing Figs. 2.14 and 2.20 , we observe that as $K$ increases, the performance gap between these algorithms becomes larger.

Discussion on the tightness of the bounds: To highlight the tightness of the upper bounds and the implication over the true network performance when the proposed allocation schemes are utilized, we consider the same network of $K=20$ with the specified spatial correlated noises. Figs. 2.21
and 2.22 depict the simulated MSE and $\mathcal{D}_{a}$ when the " $a$-coupled" and " $a$-decoupled" algorithms are applied for resource allocation for $K=3$ and $K=20$, respectively. We observe that when active sensors can quantize with rates $L_{k}>1$ bit (i.e., when bandwidth is not scarce and $P_{\text {tot }}$ is large enough), $\mathcal{D}_{a}$ is very tight. However, when bandwidth is scarce or $P_{\text {tot }}$ is small, such that active sensors cannot quantize at rates larger than one, $\mathcal{D}_{a}$ is not very tight.

Regarding $\mathcal{D}_{b}$, Figs. 2.23 and 2.24 depict the simulated MSE and $\mathcal{D}_{b}$ when the " $b$-coupled" and " $b$-decoupled" algorithms are applied for resource allocation for $K=3$ and $K=20$, respectively. We observe that for $K=3$, when active sensors can quantize with rates $L_{k}>1$ bit (i.e., $B_{\text {tot }}=30$ bits and $P_{\text {tot }} \geq 20 \mathrm{~dB}$ ), the bound $\mathcal{D}_{b}$ is very tight.

However, for $K=20$, even when active sensors can quantize with rates $L_{k}>1$ bit, $\mathcal{D}_{b}$ is not as tight as $\mathcal{D}_{a}$ (compare Figs. 2.22 and 2.24). This observation is due to the fact that, compared with $\mathcal{D}_{a}, \mathcal{D}_{b}$ captures less statistical information about the observation model of the network. As $K$ increases, this statistical information loss becomes more significant, and thus, $\mathcal{D}_{b}$ becomes less tight, compared with $\mathcal{D}_{a}$ for $K=20$.


Figure 2.23: Simulated MSE and $\mathcal{D}_{b}$ vs. $P_{\text {tot }}$ for the " $b$-coupled" and " $b$-decoupled" algorithms for $K=3$


Figure 2.24: Simulated MSE and $\mathcal{D}_{b}$ vs. $P_{t o t}$ for the " $b$-coupled" and " $b$-decoupled" algorithms for $K=20$

### 2.8 Conclusions

We considered the DES of a Gaussian vector with a known covariance matrix and linear observation model, in which the FC is tasked with reconstruction of the unknowns using a linear estimator. Sensors employ uniform multi-bit quantizers and BPSK modulation, and they communicate with the FC over power- and bandwidth-constrained channels. We derived two closed-form upper bounds on the MSE in terms of the optimization parameters (i.e., transmit power and quantization rate per sensor). Each bound consists of two terms: the first term is the MSE due to observation noises and quantization errors, and the second term is the MSE due to communication channel errors. We proposed "coupled" and "decoupled" resource allocation schemes that minimize these bounds. The "coupled" schemes utilize the iterative modified ellipsoid method to conduct a $K$ dimensional search and find the quantization rate vector, whereas the "decoupled" ones rely on a one-dimensional search to find the quantization rates. Our simulations show that when $P_{\text {tot }}$ and $B_{t o t}$ are not too scarce, the bounds are good approximations of the actual MSE. Through simulations, we verified the effectiveness of the proposed schemes and confirmed that their performance approaches the clairvoyant CE for large $P_{t o t}$ and $B_{t o t}\left(P_{t o t} \approx 25 \mathrm{~dB}, B_{t o t} \approx 30\right.$ bits $)$. Our results
indicate that resource allocation is affected by the sensors' observation qualities, channel gains, and by $P_{t o t}$ and $B_{t o t}$, e.g., two WSNs with identical conditions and $P_{t o t}\left(B_{t o t}\right)$ and different $B_{t o t}$ $\left(P_{t o t}\right)$ require two different power (rate) allocations. Additionally, a greater number of bits and more transmit power are allotted to sensors with better observation qualities.

## CHAPTER 3: CLOSED FORM APPROXIMATION FOR CRAMÉR RAO LOWER BOUND FOR DISTRIBUTED ESTIMATION WITH QUANTIZED OBSERVATIONS

In this chapter we derive the corresponding BCRLB for DES of an unknown Gaussian random variable with known mean and variance, where observation model is linear with Gaussian additive noise. The sensors have limited dynamic sensing range and individual observations are separately quantized in sensors via uniform quantizers. In sequel we provide an accurate closed-form approximation for BCRLB expression which provides us with better understanding of its behavior. Afterwards through simulation examples we study the behavior of BCRLB with respect to dynamic sensing range of sensors, variance of additive noise and quantization rates. The simulations corroborate accuracy of proposed approximations.

### 3.1 System Model and Problem Statement

We consider a WSN of $K$ spatially distributed sensors and a FC, where the network is tasked with estimating a realization of a zero mean Gaussian source $\theta$, with known variance $\sigma_{\theta}^{2}$, i.e., $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$. Each sensor makes a noisy observation of $\theta$. In particular, we model the observation $x_{k}$ at sensor $k$ as:

$$
\begin{equation*}
x_{k}=\theta+n_{k}, \quad \text { for } k=1, \ldots, K, \tag{3.1}
\end{equation*}
$$

where $n_{k}$ 's are additive noises that are uncorrelated with each other and $\theta$. We assume $n_{k} \sim$ $\mathcal{N}\left(0, \sigma_{n_{k}}^{2}\right)$. Each sensor transmits its quantized observation over an error-free communication channel to the FC, where collective received data are fused to estimate $\theta$. Error-free communication channel model has been adopted before in $[33,37,81,121,148]$ in the context of DES.

Suppose sensor $k$ has a sensing dynamic range of $\left[-\tau_{k}, \tau_{k}\right]$ and employs a uniform quantizer with $M_{k}$ levels and boundaries $\left\{\zeta_{k, 1}, \ldots, \zeta_{k, M_{k}+1}\right\}$, where $\zeta_{k,\left(M_{k} / 2\right)+1}=0, \zeta_{k, i+1}-\zeta_{k, i}=\Delta_{k} \triangleq \frac{2 \tau_{k}}{M_{k}-1}$ for $i \in\left\{1, . . M_{k}\right\}$ (and $\Delta_{k}$ is the quantizer step size for sensor $k$ ). Sensor $k$ maps its observation $x_{k}$ into a quantization level $m_{k} \in\left\{m_{k, 1}, \ldots, m_{k, M_{k}}\right\}$ such that if $x_{k}$ lies in the interval $\left[\zeta_{k, i}, \zeta_{k, i+1}\right]$ then $x_{k}$ is mapped into $m_{k, i}$ where $m_{k, i}=\frac{\zeta_{k, i}+\zeta_{k, i+1}}{2}$. Next, sensor $k$ maps $m_{k}$ into a binary sequence of length $r_{k}=\log _{2} M_{k}$ (bits) and transmits this sequence to the FC . We refer to $r_{k}$ as quantization rate of sensor $k$.

Let $\boldsymbol{m}=\left[m_{1}, \ldots, m_{K}\right]^{T}$ denote the vector of quantized observations of all sensors. One can verify that the log-likelihood function of quantized observations satisfies the regularity condition, i.e., $\mathbb{E}\left\{\frac{\partial \ln _{p(\mathbf{m}, \theta)}}{\partial \theta}\right\}=0$. Let $F$ denote the Fisher information based on $\boldsymbol{m}$. Recall the MSE of any Bayesian estimator of $\theta$ based on $\boldsymbol{m}$ is at least as large as the BCRLB based on $\boldsymbol{m}$, i.e., the inverse of $F$, where $F$ is:

$$
\begin{equation*}
F=-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m}, \theta)}{\partial^{2} \theta}\right\}=-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta}\right\}-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\} . \tag{3.2}
\end{equation*}
$$

Our goal is to derive the BCRLB for any Bayesian estimator of random variable $\theta$, based on quantized observations. In sequel we employ some approximations which leads us to insightful closed-form expressions for the BCRLB that enable us to find a better understanding of the bound behavior w.r.t. the variations of noise variances and quantization rates.

### 3.2 Bayesian Cramér Rao Lower Bound

We start with finding the second term in (3.2). For $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$ it is easy to verify that the second term $\mathbb{E}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}=-\frac{1}{\sigma_{\theta}^{2}}$. Next we characterize the first term in (3.2). Since $n_{k}$ 's are all uncorrelated Gaussian and hence independent $m_{k}$ 's conditioned on $\theta$ are independent and
$\ln p(\mathbf{m} \mid \theta)=\sum_{k=1}^{K} \ln p\left(m_{k} \mid \theta\right)$. This allows us to write the first and second derivatives of the loglikelihood function as the following:

$$
\begin{equation*}
\frac{\partial \ln p(\mathbf{m} \mid \theta)}{\partial \theta}=\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta} \tag{3.3}
\end{equation*}
$$

$$
\begin{align*}
\frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta} & =\overbrace{\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial^{2} p\left(m_{k} \mid \theta\right)}{\partial^{2} \theta}}^{=F_{a}}- \\
& \underbrace{\sum_{k=1}^{K} \frac{1}{p^{2}\left(m_{k} \mid \theta\right)}\left(\frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta}\right)^{2}}_{=F_{b}} . \tag{3.4}
\end{align*}
$$

From (3.4) we find that the first term in (3.2) is equal to $-\mathbb{E}\left\{F_{a}\right\}+\mathbb{E}\left\{F_{b}\right\}$. Next we find $\mathbb{E}\left\{F_{a}\right\}$ and $\mathbb{E}\left\{F_{b}\right\}$. Let $S_{k, i}(\theta) \triangleq p\left(m_{k}=m_{k, i} \mid \theta\right)=\operatorname{Pr}\left\{\zeta_{k, i} \leq \theta+n_{k} \leq \zeta_{k, i+1} \mid \theta\right\}$ and $H_{k, i}(\theta) \triangleq \frac{\partial S_{k, i}(\theta)}{\partial \theta}$. We have $\mathbb{E}\left\{F_{a}\right\}=0$ since:

$$
\begin{aligned}
& \mathbb{E}\left\{F_{a}\right\}=\sum_{k=1}^{K} \mathbb{E}\left\{\frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial^{2} p\left(m_{k} \mid \theta\right)}{\partial^{2} \theta}\right\}= \\
& \sum_{k=1}^{K} \int p(\theta) \sum_{i=1}^{M_{k}} \frac{S_{k, i}(\theta)}{p\left(m_{k}=m_{k, i} \mid \theta\right)} \frac{\partial^{2} p\left(m_{k}=m_{k, i} \mid \theta\right)}{\partial^{2} \theta} d \theta= \\
& \sum_{k=1}^{K} \int p(\theta)(\frac{\overbrace{}^{2}(\overbrace{i=1}^{M_{k}} S_{k, i}(\theta))}{\partial^{2} \theta}) d \theta=0
\end{aligned}
$$

For $\mathbb{E}\left\{F_{b}\right\}$ we have:

$$
\begin{align*}
& \mathbb{E}\left\{F_{b}\right\}=\sum_{k=1}^{K} \mathbb{E}\left\{\frac{1}{p^{2}\left(m_{k} \mid \theta\right)}\left(\frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta}\right)^{2}\right\}=  \tag{3.5}\\
& \sum_{k=1}^{K} \int p(\theta) \sum_{i=1}^{M_{k}} \frac{\left(H_{k, i}(\theta)\right)^{2}}{S_{k, i}(\theta)} d \theta
\end{align*}
$$

To obtain $\mathbb{E}\left\{F_{b}\right\}$ in (3.5) we need to characterize $S_{k, i}(\theta)$ and $H_{k, i}(\theta)$. One can show that:

$$
\begin{gather*}
S_{k, i}(\theta)=\Phi\left(\frac{\zeta_{k, i+1}-\theta}{\sigma_{n_{k}}}\right)-\Phi\left(\frac{\zeta_{k, i}-\theta}{\sigma_{n_{k}}}\right)  \tag{3.6}\\
H_{k, i}(\theta)=\frac{1}{\sigma_{n_{k}}}\left[\phi\left(\frac{\zeta_{k, i}-\theta}{\sigma_{n_{k}}}\right)-\phi\left(\frac{\zeta_{k, i+1}-\theta}{\sigma_{n_{k}}}\right)\right] \tag{3.7}
\end{gather*}
$$

where $\Phi($.$) is the standard normal Cumulative Distribution Function (CDF) and \phi($.$) is the standard$ normal PDF. Combining all above we obtain $F$ in (3.2) as:

$$
\begin{equation*}
F=\sum_{k=1}^{K} \sum_{i=1}^{M_{k}} \mathbb{E}\left\{\frac{H_{k, i}^{2}(\theta)}{S_{k, i}(\theta)}\right\}+\frac{1}{\sigma_{\theta}^{2}} \tag{3.8}
\end{equation*}
$$

where the expressions for $S_{k, i}(\theta)$ and $H_{k, i}(\theta)$ are given in (3.6) and (3.7), and $S_{k, i}(\theta)>0, \forall \theta$. Note that the integral corresponding to the expectation in (3.8) does not render to a closed-form. In the following, we provide two approximations for this expectation, that lead into two closed-form expressions of $F$, corresponding to binary (coarse) quantizers and high rate (fine) quantizers at sensors. Our simulations verify that these two approximate closed-forms are very close to actual $F$ values.

$$
\begin{equation*}
F \approx \sum_{k=1}^{K} \sum_{i=1}^{M_{k}}\left(\frac{\left(\zeta_{k, i+1}-\zeta_{k, i}\right)\left(\sigma_{\theta}^{4}+\sigma_{n_{k}}^{2} \sigma_{\theta}^{2}+\sigma_{n_{k}}^{2} \zeta_{k, i}^{2}\right)}{\sigma_{n_{k}}^{2}\left(\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}\right)^{5 / 2}} \phi\left(\frac{\zeta_{k, i}}{\sqrt{\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}}}\right)\right)+\frac{1}{\sigma_{\theta}^{2}} \tag{3.9}
\end{equation*}
$$

### 3.2.1 BCRLB closed-form approximation

Suppose sensors employ binary quantizers with $r_{k}=1$, i.e., sensor $k$ transmits the sign of its observations to the FC, $m_{k}=\operatorname{sign}\left(x_{k}\right)$ where $m_{k}=1$ for $x_{k} \geq 0$ and $m_{k}=-1$ otherwise. Therefore, the quantizer boundaries are $\zeta_{k, 1}=-\infty, \zeta_{k, 2}=0, \zeta_{k, 3}=+\infty$. One can verify the following for binary quantizers:

$$
\begin{equation*}
\frac{H_{k, i}^{2}(\theta)}{S_{k, i}(\theta)}=\frac{\phi^{2}\left(\frac{\theta}{\sigma_{n_{k}}}\right)}{\sigma_{n_{k}}^{2} \Phi\left(\frac{\theta}{\sigma_{n_{k}}}\right) \Phi\left(\frac{-\theta}{\sigma_{n_{k}}}\right)} . \tag{3.10}
\end{equation*}
$$

Using the Chernoff bound for CDF as an approximation [149], we can approximate the denumerator in (3.10) as $\Phi\left(\frac{\theta}{\sigma_{n_{k}}}\right) \Phi\left(\frac{-\theta}{\sigma_{n_{k}}}\right) \approx 0.25 e^{-\frac{\theta^{2}}{2 \sigma_{n_{k}}}}$. Substituting this approximation in (3.10) and doing some integral math we obtain:

$$
\begin{equation*}
\mathbb{E}\left\{\frac{H_{k, i}^{2}(\theta)}{S_{k, i}(\theta)}\right\} \approx \frac{2}{\pi \sigma_{n_{k}} \sqrt{\sigma_{n_{k}}^{2}+\sigma_{n_{k}}^{2}}} \tag{3.11}
\end{equation*}
$$

For large quantization rates $r_{k}$ 's (small step sizes $\Delta_{k}$ 's given $\tau_{k}$ 's) we can use second order Taylor approximation for $S_{k, i}(\theta)$ and $H_{k, i}(\theta)$ :

$$
S_{k, i}(\theta) \approx\left(\frac{\zeta_{k, i+1}-\zeta_{k, i+1}}{\sigma_{n_{k}}}\right) \phi\left(\frac{\zeta_{k, i}-\theta}{\sigma_{n_{k}}}\right)
$$

$$
H_{k, i}(\theta) \approx\left(\frac{\left(\zeta_{k, i+1}-\zeta_{k, i}\right)\left(\zeta_{k, i}-\theta\right)}{\sigma_{n_{k}}^{3}}\right) \phi\left(\frac{\zeta_{k, i}-\theta}{\sigma_{n_{k}}}\right)
$$

substituting the above approximations in (3.8) and taking some tedious integral calculus steps we obtain the closed-form expression for the $F$ in (3.9).

Remark 1: Consider (3.9) when $r_{k} \rightarrow \infty$, in this case the internal sum can be interpreted as a Riemann Sum which must converge to following integral ${ }^{1}$ :

$$
\begin{array}{r}
\gamma=\int_{-\tau_{k}}^{\tau_{k}}\left[\frac{\sigma_{\theta}^{2}}{\sigma_{n_{k}}^{2}\left(\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}\right)^{3 / 2}}+\sigma_{n_{k}}^{2} \alpha^{2}\right] \phi\left(\frac{\alpha}{\sqrt{\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}}}\right) d \alpha= \\
\frac{2 \Phi\left(\frac{\tau_{k}}{\sqrt{\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}}}\right)-1}{\sigma_{n_{k}}^{2}}-\frac{2 \tau_{k} \phi\left(\frac{\tau_{k}}{\sqrt{\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}}}\right)}{\left(\sigma_{\theta}^{2}+\sigma_{n_{k}}^{2}\right)^{3 / 2}} . \tag{3.12}
\end{array}
$$

If we relax the simplifying assumption of $x_{k} \in\left[-\tau_{k}, \tau_{k}\right]$ and let $\tau_{k} \rightarrow+\infty$, it is easy to verify that $\gamma$ in (3.12) goes to $\frac{1}{\sigma_{n_{k}}^{2}}$ and renders the sum in (3.9) into the Fisher information corresponding to centralized estimation of $\theta$, where the FC has access to (unquantized) full precision observations $x_{k}$ 's, i.e., $F^{c v}=\sum_{k=1}^{K} \frac{1}{\sigma_{k}^{2}}+\frac{1}{\sigma_{\theta}^{2}}$ [129]. The expression in right side of equation (3.12) can also be interpreted as Fisher information associated to sensor $k$ where the sensor's sensing dynamic range is confined to $\left[-\tau_{k}, \tau_{k}\right]$ and the observation $x_{k}$ that lies within this range is available at the FC with full precision (no quantization).

[^2]
### 3.3 Simulation Results

In this section, with simulated examples, we demonstrate the behavior of $F$ and its inverse the BCRLB with respect to different system parameters and also illustrate the accuracy of the two proposed approximations of $F$. Without loss of generality we consider a homogeneous network where the variances of observation noises and quantization rates are equal, i.e., $\sigma_{n_{k}}^{2}=\sigma_{n}^{2}, r_{k}=r, \forall k$. We let $K=20$ and $\sigma_{\theta}^{2}=1$.

Fig. 3.1 depicts $F$, associated with expression in (3.12) shown as $F^{c v}(\tau)$ in the figure versus the ratio $\lambda=\tau / \sqrt{\sigma_{\theta}^{2}+\sigma_{n}^{2}}$, where sensors have a limited sensing dynamic range ${ }^{2}$ of $[-\tau, \tau]$ and no quantization is performed. We observe for $\tau>3.5 \sqrt{\sigma_{\theta}^{2}+\sigma_{n}^{2}}(\lambda>3.5), F^{c v}(\tau)$ converges to clairvoyant benchmark $F^{c v}$ for various values of $\sigma_{n}^{2}$. In other words, from practical point of view the information loss (in terms of increasing MSE) is negligible, if we choose a sensor with sensing dynamic range larger than $3.5 \sqrt{\sigma_{\theta}^{2}+\sigma_{n}^{2}}$. Based on this observation we set $\tau=3.5 \sqrt{\sigma_{\theta}^{2}+\sigma_{n}^{2}}$ to produce Fig. 3.2.

Fig. 3.2 compares the actual BCRLB and the proposed approximations associated with expressions in (3.11) and (3.9) (for $r=1$ and $r \geq 2$ respectively), versus $\sigma_{n}^{2}$. As can be seen the approximation for $r=1$ is satisfactory and for $r \geq 2$, as $r$ increases the approximation becomes more accurate, such that for $r=4$ it is very accurate.

[^3]

Figure 3.1: Comparison of Fisher information of clairvoyant and the one associated with limited dynamic range of sensors


Figure 3.2: CRLB and proposed approximation for different values of quantization rate

### 3.4 Conclusions

In this chapter we derived the BCRLB for DES of a Gaussian source where the individual observations at the sensors are separately quantized with uniform quantizers and sensors have limited sensing dynamic range. The observation model is assumed to be linear with additive Gaussian noise. We provided closed-form approximations for the BCRLB and studied the behavior of the BCRLB and the two corresponding approximations as quantization rates, variances of additive observation noises and sensing dynamic range of sensors vary. The simulation results corroborate the accuracy of the proposed approximations and verify that increasing the variance of the additive observation noise always degrades the estimation accuracy and increasing the quantization rates always improves the estimation accuracy. Our simulation results also illustrate that for a Bayesian Gaussian linear model, provided that the sensing dynamic range of sensors stays larger than 3.5 times the standard deviation of observations the information loss (in terms of increasing MSE) is negligible, in other words we can limit observations larger than 3.5 times the standard deviation of observations into a confined limited range, without noticeable degradation in the estimation accuracy.

# CHAPTER 4: ON DISTRIBUTED LINEAR ESTIMATION IN MULTIPLICATIVE NOISE ENVIRONMENT AND OBSERVATION MODEL UNCERTAINTIES 

In this chapter we consider DES of a Gaussian source in a heterogeneous bandwidth constrained WSN. Similar to $[33,81]$, we choose the total number of quantization bits as the measure of network bandwidth. The observations of the sensors is corrupted by independent multiplicative and additive observation noises, with incomplete statistical knowledge of the multiplicative noise. Actually we assume that the distribution of the multiplicative observation noise is unknown and only its mean and variance are known (observation model uncertainties). For multi-bit quantizers, we derive the closed-form MSE expression for the linear LMMSE estimator at the FC. For both error-free and erroneous (modeled as BSC) orthogonal communication channels, we consider two system-level constrained optimization problems with respect to the sensors' quantization rates: in (P1) we minimize the MSE given a network bandwidth constraint, and in (P2) we minimize the required network bandwidth given a target MSE. To address these two problems we propose several rate allocation methods named as longest root to leaf path, greedy and integer relaxation methods. We also derive the BCRLB and compare the MSE performance of our proposed methods against the BCRLB. Our results corroborate that, for low power multiplicative observation noises and adequate network bandwidth, the gaps between the MSE of our proposed methods and the BCRLB are negligible, while the performance of other methods like individual rate allocation and uniform is not satisfactory.

### 4.1 System Model and Problem Statement

We consider a WSN with $K$ spatially distributed heterogeneous sensors and a FC, where the FC is tasked with estimating a realization of a Gaussian source $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$, via fusing the collective received data from all sensors. Each sensor makes a noisy observation of $\theta$, where both multiplicative and additive observation noises are involved. Let $x_{k}$ denote the scalar noisy observation of $\theta$ at sensor $k$. We assume the following observation model:

$$
\begin{equation*}
x_{k}=h_{k} \theta+n_{k}, \quad \text { for } \quad k=1, \ldots, K, \tag{4.1}
\end{equation*}
$$

where $h_{k}$ and $n_{k}$ are multiplicative and additive observation noises, respectively. Also $h_{k}, n_{k}$, $\theta$ are uncorrelated. We assume $n_{k} \sim \mathcal{N}\left(0, \sigma_{n_{k}}^{2}\right), \mathbb{E}\left\{h_{k}\right\}=1 \forall k$, and ${ }^{1} \operatorname{var}\left(h_{k}\right)=\sigma_{h_{k}}^{2}$. Sensor $k$ employs a uniform quantizer with $M_{k}$ quantization levels and quantization step size $\Delta_{k}$. The quantizer maps $x_{k}$ into a quantization level $m_{k} \in\left\{m_{k, 1}, \ldots, m_{k, M_{k}}\right\}$, where $m_{k, i}=\frac{\left(2 i-1-M_{k}\right) \Delta_{k}}{2}$ for $i=1, \ldots, M_{k}$. We assume $x_{k}$ lies in the interval $\left[-\tau_{k}, \tau_{k}\right]$ almost surely, for some reasonably large value of $\tau_{k}$, and we let $\Delta_{k}=\frac{2 \tau_{k}}{M_{k}-1}$. These imply that the uniform quantization mapping rule can be described as the following: if $x_{k} \in\left[m_{k, i}-\frac{\Delta_{k}}{2}, m_{k, i}+\frac{\Delta_{k}}{2}\right]$, then $m_{k}=m_{k, i}$, if $x_{k} \geq \tau_{k}$, then $m_{k}=\tau_{k}$, and if $x_{k} \leq-\tau_{k}$, then $m_{k}=-\tau_{k}$. Following quantization, sensor $k$ maps the index $i$ of $m_{k, i}$ into a bit sequence of length $r_{k}=\log _{2} M_{k}$ and sends $r_{k}$ bits to the FC. Sensors transmit over orthogonal bandwidth-constrained error-free communication channels. Error-free communication channel model has been used before in several classical works on DES, examples are $[18,20,33,67,72,81,107,150]$. In Section 4.6 we extend our analytical results to the case where these channels are modeled as independent BSCs with different error probabilities. To capture the network bandwidth constraint we assume $\sum_{k=1}^{K} r_{k} \leq B_{t o t}$. In the absence of knowledge of

[^4]joint distribution of $m_{k}$ 's and $\theta$, we resort to the LMMSE estimator [4] to form the estimate $\hat{\theta}=\boldsymbol{G} \boldsymbol{m}$ at the FC, where $\boldsymbol{G}$ is the $1 \times K$ linear estimation operator and $\boldsymbol{m}=\left[m_{1}, \ldots, m_{K}\right]^{T}$ is the vector of transmitted quantization levels for all sensors. The LMMSE estimator has a low computational complexity and only requires the knowledge of moments $\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}$ and $\mathbb{E}\left\{\boldsymbol{m}^{T}\right\}$ to form $\hat{\theta}$. Let $\mathcal{D}=\mathbb{E}\left\{(\theta-\hat{\theta})^{2}\right\}$ denote the MSE corresponding to the LMMSE estimator, where $\mathcal{D}$ depends on $r_{k} \forall k$. We consider two system-level constrained optimization problems with respect to optimization variables $r_{k} \forall k$. In the first problem, we minimize $\mathcal{D}$ subject to the network bandwidth constraint. In the second problem, we minimize the total number of transmitted bits subject to the constraint on $\mathcal{D}$. In other words, we are interested to solve the following two constrained optimization problems:
\[

$$
\begin{array}{ll}
\text { (P1) } & \underset{r_{k}}{\operatorname{minimize}} \mathcal{D}\left(\left\{r_{k}\right\}_{k=1}^{K}\right) \\
& \text { s.t. } \sum_{k=1}^{K} r_{k} \leq B_{t o t}, r_{k} \in \mathbb{Z}_{+}, \forall k, \\
\text { (P2) } \quad & \underset{r_{k}}{\operatorname{minimize}} \sum_{k=1}^{K} r_{k}  \tag{4.3}\\
& \text { s.t. } \mathcal{D}\left(\left\{r_{k}\right\}_{k=1}^{K}\right) \leq \mathcal{D}_{0}, r_{k} \in \mathbb{Z}_{+}, \forall k,
\end{array}
$$
\]

where $\mathcal{D}_{0}$ is the pre-determined upper bound on $\mathcal{D}$.

### 4.2 Characterizing MSE for LMMSE estimator

We wish to characterize $\mathcal{D}$ in terms of the optimization variables $\left\{r_{k}\right\}_{k=1}^{K}$. From [4] we have:

$$
\begin{equation*}
\hat{\theta}=\boldsymbol{G} \boldsymbol{m} \text { where } \boldsymbol{G}=\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}\left(\mathbb{E}\left\{\boldsymbol{m}^{T}\right\}\right)^{-1} \tag{4.4}
\end{equation*}
$$

$$
\begin{equation*}
\left.\mathcal{D}=\sigma_{\theta}^{2}-\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}\left(\mathbb{E}\left\{\boldsymbol{m} \boldsymbol{m}^{T}\right\}\right)^{-1} \mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}^{T}\right) \tag{4.5}
\end{equation*}
$$

To find $\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}$ and $\mathbb{E}\left\{\boldsymbol{m}^{T} \boldsymbol{m}^{T}\right\}$ in (4.4), (4.5) we need to delve into statistics of the quantization errors.

- Characterizing $\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}$ and $\mathbb{E}\left\{\boldsymbol{m}^{T}\right\}$ : For sensor $k$, let the difference between observation $x_{k}$ and its quantized version $m_{k}$, i.e., $\epsilon_{k}=x_{k}-m_{k}$, be the corresponding quantization error. In general, $\epsilon_{k}$ 's are mutually correlated and also are correlated with $x_{k}$ 's. However, in [137] it is shown that, when highly correlated Gaussian random variables are coarsely quantized with uniform quantizers of step sizes $\Delta_{k}$ 's, quantization errors can be approximated as mutually independent random variables, that are uniformly distributed in the interval $\left[-\frac{\Delta_{k}}{2}, \frac{\Delta_{k}}{2}\right]$, and are also independent of quantizer inputs. Here, since $\theta$ and $n_{k}$ 's in (5.1) are Gaussian, conditioned on $h_{k}$ 's observations $x_{k}$ 's are correlated Gaussian that are quantized with uniform quantizers of quantization step sizes $\Delta_{k}$ 's. Thus $\epsilon_{k}$ 's are approximated as mutually independent zero mean uniform random variables with variance $\sigma_{\epsilon_{k}}^{2}=\frac{\Delta_{k}^{2}}{12}$, that are also independent of $x_{k}$ 's (and hence independent of $\theta$ and $n_{k}$ 's). Using the aforementioned assumptions and approximations for the quantization errors, $k$ th element of $\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}$ becomes:

$$
\begin{align*}
& \mathbb{E}\left\{\theta m_{k}\right\}=\mathbb{E}_{h_{k}}\left\{\mathbb{E}\left\{\theta m_{k} \mid h_{k}\right\}\right\}=\mathbb{E}_{h_{k}}\left\{\mathbb{E}\left\{\theta\left(x_{k}-\epsilon_{k}\right) \mid h_{k}\right\}\right\} \\
& =\mathbb{E}_{h_{k}}\left\{h_{k} \mathbb{E}\left\{\theta^{2}\right\}+\mathbb{E}\left\{n_{k}\right\} \mathbb{E}\{\theta\}-\mathbb{E}\left\{\epsilon_{k}\right\} \mathbb{E}\{\theta\}\right\}=\sigma_{\theta}^{2} \tag{4.6}
\end{align*}
$$

Hence we have $\mathbb{E}\left\{\theta \boldsymbol{m}^{T}\right\}=\sigma_{\theta}^{2} \mathbf{1}^{T}$, where $\mathbf{1}=[1, \ldots, 1]^{T}$. Similarly, for $(k, l)$ th element of $\mathbb{E}\left\{\boldsymbol{m m}^{T}\right\}$ we have:

$$
\begin{align*}
& \mathbb{E}\left\{m_{k} m_{l}\right\}=\mathbb{E}_{h_{k}, h_{l}}\left\{\mathbb{E}\left\{m_{k} m_{l} \mid h_{k}, h_{l}\right\}\right\}=  \tag{4.7}\\
& \mathbb{E}_{h_{k}, h_{l}}\left\{\mathbb{E}\left\{\left(h_{k} \theta+n_{k}-\epsilon_{k}\right)\left(h_{l} \theta+n_{l}-\epsilon_{l}\right) \mid h_{k}, h_{l}\right\}\right\} \stackrel{(a)}{=} \\
& \mathbb{E}_{h_{k}, h_{l}}\left\{h_{k} h_{l} \mathbb{E}\left\{\theta^{2}\right\}+\mathbb{E}\left\{n_{k} n_{l}\right\}+\mathbb{E}\left\{\epsilon_{k} \epsilon_{l}\right\}\right\}= \\
& \mathbb{E}_{h_{k}, h_{l}}\left\{h_{k} h_{l}\right\} \mathbb{E}\left\{\theta^{2}\right\}+\mathbb{E}\left\{n_{k} n_{l}\right\}+\mathbb{E}\left\{\epsilon_{k} \epsilon_{l}\right\},
\end{align*}
$$

where for (a) we have used the assumptions that (i) $n_{k}$ 's and $\theta$ are uncorrelated, (ii) $\epsilon_{k}$ 's and $\theta$ are uncorrelated (iii) $n_{k}$ 's and $\epsilon_{k}$ 's are uncorrelated. Having (4.7), and noting that $h_{k}$ 's are uncorrelated with unit means, we reach:

$$
\mathbb{E}\left\{m_{k} m_{l}\right\}= \begin{cases}\sigma_{\theta}^{2}+\sigma_{k}^{2}+\sigma_{\epsilon_{k}}^{2}, & \text { if } k=l \\ \sigma_{\theta}^{2}, & \text { if } k \neq l\end{cases}
$$

where $\sigma_{k}^{2}=\sigma_{\theta}^{2} \sigma_{h_{k}}^{2}+\sigma_{n_{k}}^{2}$. Consequently matrix $\mathbb{E}\left\{\boldsymbol{m}^{T}\right\}$ can be written as the following:

$$
\begin{equation*}
\mathbb{E}\left\{\boldsymbol{m}^{\boldsymbol{T}}\right\}=\sigma_{\theta}^{2} \mathbf{1 1}^{\boldsymbol{T}}+\operatorname{diag}\left(\frac{1}{\alpha_{1}}, \ldots, \frac{1}{\alpha_{K}}\right) \tag{4.8}
\end{equation*}
$$

where $\alpha_{k}^{-1}=\sigma_{k}^{2}+\sigma_{\epsilon_{k}}^{2}$. Applying matrix inversion Lemma [4] to (4.8) we find:

$$
\left[\left(\mathbb{E}\left\{\boldsymbol{m}^{T}\right\}\right)^{-1}\right]_{k, l}= \begin{cases}\alpha_{k}-\frac{\alpha_{k}^{2}}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \alpha_{k}}, & \text { if } k=l \\ -\frac{\alpha_{k} \alpha_{l}}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \alpha_{k}}, & \text { if } k \neq l\end{cases}
$$

Proposition 1 summarizes the expressions for $\hat{\theta}, \mathcal{D}$ in (4.4), (4.5).
Proposition 1. The LMMSE estimator $\hat{\theta}$ and its corresponding MSE $\mathcal{D}$, based on the quantized
observations $\left\{m_{k}\right\}_{k=1}^{K}$ are:

$$
\begin{gather*}
\hat{\theta}=\sum_{k=1}^{K} c_{k} m_{k} \text { where } c_{k}=\frac{\alpha_{k}}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \alpha_{k}} \\
\mathcal{D}=\frac{1}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \alpha_{k}} \tag{4.9}
\end{gather*}
$$

Examining (4.9), we note that $\alpha_{k}$ represents the contribution of sensor $k$ in reducing the overall MSE at the FC. Also, $\alpha_{k}$ can be viewed as an indicator for the quality of received message from sensor $k$ : the larger $\alpha_{k}$ is, the more reliable is the received message. It is easy to verify that $\alpha_{k}$ is increasing in $r_{k}$ and decreasing in $\sigma_{k}^{2}$.

Remark 1: When all observations $x_{k}$ 's are available at the FC with full precision (so-called CE), the LMMSE estimator would be $\breve{\theta}=\sum_{k=1}^{K} b_{k} x_{k}$, where $b_{k}=\frac{\sigma_{k}^{-2}}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \sigma_{k}^{-2}}$, with its corresponding MSE $\mathcal{D}^{c}=\frac{1}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \sigma_{k}^{-2}}$. This clairvoyant estimator can be used as our performance benchmark, since $\mathcal{D}>\mathcal{D}^{c}$.

Proposition 2. In a network with homogeneous sensors, i.e., $\sigma_{k}^{2}=\sigma^{2}, \forall k$, and all sensors quantize their observations with identical quantizers of step size $\Delta$, the MSE gap between two linear estimators $\hat{\theta}$ and $\breve{\theta}$ is:

$$
\begin{equation*}
\mathcal{D}-\mathcal{D}^{c}=\frac{K \Delta^{2}}{12\left(K+\sigma_{\theta}^{-2}\left(\sigma^{2}+\sigma_{\epsilon}^{2}\right)\right)\left(K+\sigma_{\theta}^{-2} \sigma^{2}\right)} \leq \frac{\Delta^{2}}{12 K} . \tag{4.10}
\end{equation*}
$$

Based on (4.10), if $\Delta \rightarrow 0$, then $\mathcal{D} \rightarrow \mathcal{D}^{c}$. Additionally, if $K \rightarrow \infty$, then $\mathcal{D} \rightarrow \mathcal{D}^{c}$ even for large $\Delta$. These conclusions still hold true in a network with heterogeneous sensors, where sensor $k$ quantizes with step size of $\Delta_{k}$. If $\left(\max \Delta_{k}\right) \rightarrow 0$, then $\alpha_{k} \rightarrow \sigma_{k}^{-2}$ and according to (4.9), $\mathcal{D} \rightarrow \mathcal{D}^{c}$. On the other hand, according to (4.9) and noting that $\alpha_{k}>0$ for active sensors, MSE always
decreases as the number of active sensors increases. Thus as $K \rightarrow \infty$, we have $\mathcal{D} \rightarrow \mathcal{D}^{c} \rightarrow 0$.

### 4.3 Solving Constrained Problem (P1)

Since the optimization variables $r_{k}$ 's are integer and $\mathcal{D}$ is a non-linear function of $r_{k}$ 's, ( $\left.\mathbf{P} 1\right)$ is a Non Linear Integer Programming (NLIP) problem and is NP-hard. Even if the inequality constrain holds with equality, i.e., $\sum_{k=1}^{K} r_{k}=B_{t o t}$, solving ( $\mathbf{P} 1$ ) requires a brute-force evaluation over $\binom{K+B_{t o t}-1}{K-1}$ choices. For $K=50$ and $B_{t o t}=60$ bits, the number of evaluations would be in the order of $10^{31}$. The following lemmas help us find strategies that reduce the computational complexity required for solving (P1).

Lemma 1. Minimizing $\mathcal{D}$ in (4.9), is equivalent to maximizing $\sum_{k=1}^{K} \alpha_{k}$ with the same constraints as in (4.2).

Proof. Since $\sigma_{\theta}^{-2}>0$, it is axiomatic.
Lemma 2. Suppose $\left\{r_{k}^{*}\right\}_{k=1}^{K}$ is the optimal solution to (P1). Then $\sum_{k=1}^{K} r_{k}^{*}=B_{\text {tot }}$.

Proof. Note $\alpha_{k}$ is a function of $r_{k}$ through $\sigma_{\epsilon_{k}}^{2}=\frac{\Delta_{k}^{2}}{12}=\frac{\tau_{k}^{2}}{3\left(2^{r} k-1\right)^{2}}$. It is easy to verify that $\frac{\partial \mathcal{D}}{\partial r_{k}} \leq 0$ and $\mathcal{D}$ is a decreasing function of $r_{k}$ 's. Thus the optimal solution satisfies the network bandwidth constraint, i.e., $\sum_{k=1}^{K} r_{k}^{*}=B_{t o t}$.

Lemma 3. Without loss of generality, suppose sensors are sorted ${ }^{2}$ such that $\sigma_{1}^{2} \leq \sigma_{2}^{2} \leq \ldots \leq \sigma_{K}^{2}$. Then the optimal solution satisfies $r_{i}^{*} \geq r_{j}^{*}$ for $i<j$.

Proof. Suppose $\left\{r_{k}\right\}_{k=1}^{K}$ is the optimal solution, such that $r_{i}<r_{j}$ for $i<j$. Also, suppose $\left\{r_{k}^{\prime}\right\}_{k=1}^{K}$ is a solution of (P1), which is not necessarily optimal, such that $r_{k}^{\prime}=r_{k}$ for $k \neq i, j$ and

[^5]$r_{i}^{\prime}=r_{j}, r_{j}^{\prime}=r_{i}$. Consider the following:
\[

$$
\begin{aligned}
& \sum_{k=1}^{K} \alpha_{k}\left(r_{k}^{\prime}\right)-\sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right)=\overbrace{\sum_{k \neq i, j}^{K} \alpha_{k}\left(r_{k}\right)-\sum_{\neq i, j}^{K} \alpha_{k}\left(r_{k}\right)}^{=\delta_{1}} \\
& +\overbrace{\frac{1}{\sigma_{i}^{2}+\frac{\tau^{2}}{3\left(2^{2}-1\right)^{2}}}-\frac{1}{\sigma_{i}^{2}+\frac{\tau^{2}}{3\left(2^{\left.r_{i}-1\right)^{2}}\right.}}}^{=\delta_{2}} \\
& +\overbrace{\frac{1}{\sigma_{j}^{2}+\frac{\tau^{2}}{3\left(2^{r_{j}}-1\right)^{2}}}-\frac{1}{\delta_{3}}}^{\sigma_{j}^{2}+\frac{\tau^{2}}{3\left(2^{\left.r_{i}-1\right)^{2}}\right.}}
\end{aligned}
$$>0 .
\]

One can verify $\delta_{1}=0, \delta_{2}>0, \delta_{3}>0$, thus $\sum_{k=1}^{K} \alpha_{k}\left(r_{k}^{\prime}\right)>\sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right)$. According to (4.9), the MSE associated with $\left\{r_{k}^{\prime}\right\}_{k=1}^{K}$ should be less than that of the optimal solution $\left\{r_{k}\right\}_{k=1}^{K}$, which is a contradiction. In this proof, we assumed $\tau_{k}=\tau, \forall k$, although the proof is still valid for unequal $\tau_{k}$ 's, provided that $\tau_{i} \leq \tau_{j}$, which is satisfied if we choose $\tau_{k}^{2} \propto \operatorname{var}\left(x_{k}\right)=\sigma_{\theta}^{2}+\sigma_{k}^{2}$.

In the next subsections, we propose four methods for solving (P1): $A$ ) Longest Root to Leaf Path (LRLP) method, which is optimal with less computational complexity than that of brute force, $B$ ) greedy method, $C$ ) integer relaxation method, $D$ ) Individual Rate Allocation (IRA) method. The suboptimal $B$ ), $C$ ), $D$ ) methods have moderate to low computational complexity.

### 4.3.1 LRLP Method

We view ( $\mathbf{P} 1$ ) as the problem of finding the longest "root to leaf" path in a weighted directed binary tree, where there is a constraint on the number of edges from "root to leaf" [151]. In fact our objective function $\sum_{k=1}^{K} \alpha_{k}$ can be viewed as the length of the path to be maximized, where the constraint on the number of edges is $\sum_{k=1}^{K} r_{k} \leq B_{t o t}$. Fig. 4.1 demonstrates the problem for $K=3$ and $B_{t o t}=5$ bits. The nodes are tagged with indices of sorted sensors and visiting
node $k$ is translated to "allocating one bit to sensor $k$ ". The edge ${ }^{3}$ weight $w_{k}(r)$ is the weight of the edge entering node $k$ and $r$ is the number of prior visits of node $k$, i.e., $w_{k}(r)=\alpha_{k}(r+1)-\alpha_{k}(r)=$ $\alpha_{k}$ (number of bits allocated to sensor $k$ so far +1$)-\alpha_{k}$ (number of bits allocated to sensor $k$ so far). For instance, the green path in Fig. 4.1 is associated with the rate allocation $\left[r_{1}, r_{2}, r_{3}\right]=[3,2,0]$, and the corresponding objective function value is $\sum_{k=1}^{K} \alpha_{k}=w_{1}(1)+w_{1}(2)+w_{2}(0)+w_{2}(1)=$ $\alpha_{1}(3)+\alpha_{2}(2)$.


Figure 4.1: LRLP method

To solve ( $\mathbf{P} 1$ ), one needs to construct the associated binary tree with structures conforming to lemmas 2 and 3, then uses a search algorithm, such as Depth First Search (DFS) [151] to discover all possible "root to leaf" paths, and choose the path that results in the maximum objective function value. For $K=3$ and $B_{\text {tot }}=5$ bits, Fig. 4.1 shows there exist 5 "root to leaf" paths, all conforming to lemmas 2 and 3, corresponding to 5 distinct rate allocation among 3 sensors $\left[r_{1}, r_{2}, r_{3}\right] \in\{[5,0,0],[4,1,0],[3,2,0],[3,1,1],[2,2,1]\}$. We recognize these as different partitions of the integer number 5 , with 3 or fewer addends [152], i.e., the number of possible "root to

[^6]leaf" paths in a binary tree constructed as explained, conforming to lemmas 2 and 3, and characterizing ( $\mathbf{P} 1$ ), is equal to the number of solutions to the following equation:
\[

$$
\begin{align*}
& r_{1}+r_{2}+\ldots+r_{K}=B_{t o t}  \tag{4.11}\\
& \text { s.t. } r_{1} \geq r_{2} \geq \ldots \geq r_{K} \quad r_{k} \in \mathbb{Z}_{+} .
\end{align*}
$$
\]

Although the number of ways one can partition an integer number does not have a closed form formula, the literature [152] provides some useful asymptotic formulas or recurrence relations. Suppose $q_{k}(n)$ is the number of solutions to (4.11), then we have the recurrence relation $q_{k}(n)=$ $q_{k-1}(n)-q_{k}(n-k)$, with $q_{0}(n)=0, q_{k}(1)=1$ [152]. For $K=50$ and $B_{t o t}=60$ bits, $q_{K}\left(B_{t o t}\right)$ is in the order of $10^{6}$, which is much smaller than that of brute force $10^{31}$. The computational complexity of this method is still high for very large networks, e.g., $K \geq 100$, and hence its application is most beneficial for finding the optimal solution of small to moderate size networks.

### 4.3.2 Greedy Method

Recall from Lemma 1 that the maximum reduction in $\mathcal{D}$ corresponds to the maximum increase in $\sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right)$. Hence, our proposed greedy method in each iteration allocates one bit to the sensor that guarantees the maximum increase in $\sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right)$, i.e., in each iteration the algorithm loads one bit on sensor $k^{*}$ where $k^{*}=\underset{k}{\operatorname{argmax}} I_{k}\left(r_{k}\right)=\underset{k}{\operatorname{argmax}}\left(\alpha_{k}\left(r_{k}+1\right)-\alpha_{k}\left(r_{k}\right)\right)$. The iteration ends when all $B_{\text {tot }}$ bits are allocated to the sensors. Following algorithm illustrates the details:

For $K=3$ and $B_{t o t}=5$ bits, Fig. 4.2 shows the accepted decisions by the greedy method at each iteration/decision epoch with green arrows and the rejected decisions with red arrows. Note that the initial point is always $r_{k}=1, r_{k}=0$ for $k=2, \ldots, K$, since the first bit is always allocated to sensor 1 (for sorted sensors sensor 1 has the largest $\alpha_{k}$ or smallest $\sigma_{k}^{2}$ ).

Data: $B_{\text {tot }},\left\{\tau_{k}\right\}_{k=1}^{K},\left\{\sigma_{k}^{2}\right\}_{k=1}^{K}$
Result: rate allocation $\left\{r_{k}^{*}\right\}_{k=1}^{K}$
initialization;
$r_{1}=1, r_{k}=0$ for $k=2, \ldots, K, \mathcal{S}=\{1,2\}$
for $i=1: B_{t o t}$ do
$k^{*}=\underset{k \in \mathcal{S}}{\operatorname{argmax}}\left(\alpha_{k}\left(r_{k}+1\right)-\alpha_{k}\left(r_{k}\right)\right)$
$r_{k^{*}}=r_{k^{*}}+1$
$\mathcal{S}=\left\{k \mid r_{k}<r_{k-1}\right\} \cup\{1\}$
end
Algorithm: greedy method for rate allocation in (P1)


Figure 4.2: Greedy method

The second bit can be allocated to either sensor 1 or sensor 2, i.e., $k^{*}=\underset{k \in\{1,2\}}{\operatorname{argmax}}\left(\alpha_{k}\left(r_{k}+1\right)-\right.$ $\alpha_{k}\left(r_{k}\right)$ ), this is equivalent to making the decision $w_{1}(1) \underset{k^{*}=2}{\stackrel{k^{*}=1}{\gtrless}} w_{2}(0)$ (look at the weights on the edges in Fig. 4.2). The sequence of green arrows in Fig.4.2, $1 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 3$ is associated with the rate allocation $\left[r_{1}, r_{2}, r_{3}\right]=[3,1,1]$.

In the following, we look at the computational complexity of the greedy method in two cases: case (a) $B_{t o t} \leq K$, in this case the first bit has to be allocated to sensor 1, the second bit can be allocated to either sensor 1 or sensor 2 . In general the $i$ th bit, for $1 \leq i \leq B_{t o t}$, can be allocated to one of at most $i$ sensors (sensor $1, \ldots$, sensor $i$ ). In other words, in the $i$ th decision epoch, greedy method
should find the best sensor among eligible candidates in set $\mathcal{S}=\left\{k \mid k \leq i, r_{k}<r_{k-1}\right\} \cup\{1\}$, where $\mathcal{S}$ has a maximum of $i$ elements. Thus allocating $B_{\text {tot }}$ bits among $K$ sensors needs calculation of $I_{k}\left(r_{k}\right)=\alpha_{k}\left(r_{k}+1\right)-\alpha_{k}\left(r_{k}\right)$ for at most $2+3+\ldots+B_{t o t}=\left(\frac{B_{t o t}^{2}+B_{t o t}}{2}\right)-1$ times. Case $(b)$ $B_{t o t}>K$, adopting a reasoning similar to case ( $a$ ), in this case allocation of the first $K$ bits needs at most $\left(\frac{K^{2}+K}{2}\right)-1$ calculations. Each of the remaining $B_{t o t}-K$ bits can be allocated to one of at most $K$ sensors, leading into $\left(B_{t o t}-K\right) \times K$ number of calculations at most. Hence, the overall number of evaluations at most would be $\left(\frac{K^{2}+K}{2}\right)-1+\left(B_{t o t}-K\right) \times K \approx K\left(B_{t o t}-K / 2\right)$. For $K=50$ and $B_{t o t}=60$ bits, where $B_{t o t}>K$, the number of evaluations would be in order of $10^{3}$.

Remark 2: In the absence of a powerful FC, the proposed greedy algorithm can be implemented in a distributed way, assuming sensors can broadcast and hear the broadcast messages by other sensors. Sensor $k$ calculates the value $I_{k}\left(r_{k}\right)=\alpha_{k}\left(r_{k}+1\right)-\alpha_{k}\left(r_{k}\right)$, and broadcasts the value. Hearing all $I_{j}, j \neq k$, sensor $k$ increases $r_{k}$ by one if it has the largest $I_{k}$ among all sensors. Doing the mentioned process for $B_{t o t}$ times would complete the rate allocation.

### 4.3.3 Integer Relaxation Method

Convex relaxation for solving combinatorial optimization problems such as (P1) is a rather old technique, that has been widely used in research and applied to a variety of applications [153]. Relaxing the integer constraint on $r_{k}$ 's and letting them be positive numbers and using lemmas 1 and 2 , we consider the following relaxed problem:

$$
\begin{align*}
\left(\mathbf{P}^{\prime} \mathbf{1}\right) & \underset{r_{k} \forall k}{\operatorname{maximize}}  \tag{4.12}\\
& \sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right) \\
& \text { s.t. } \sum_{k=1}^{K} r_{k}=B_{t o t}, r_{k} \in \mathbb{R}_{+}, \forall k .
\end{align*}
$$

The Lagrangian for $\left(\mathbf{P}^{\prime} \mathbf{1}\right)$ is:

$$
\mathcal{L}\left(\left\{r_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)=\sum_{k=1}^{K} \alpha_{k}\left(r_{k}\right)-\mu_{k} r_{k}+\lambda\left(\sum_{k=1}^{K} r_{k}-B_{t o t}\right) .
$$

In the following we apply the first order KKT necessary optimality conditions for $\left(\mathbf{P}^{\prime} \mathbf{1}\right)$ which generate a closed-form solution for $r_{k}$ 's. Afterwards, we prove that the obtained solution satisfies the second order sufficient optimality conditions.

## - Necessary Optimality Conditions

After solving the KKT conditions corresponding to (4.12), we find:

$$
\begin{equation*}
r_{k}^{\dagger}=0.5\left[\log _{2}\left(\tau_{k}^{2}\left(\lambda^{\dagger}-\sigma_{k}^{2}-\sqrt{\lambda^{\dagger^{2}}-2 \lambda^{\dagger} \sigma_{k}^{2}}\right)^{-1}\right)-\log _{2} 3\right]^{+} \tag{4.13}
\end{equation*}
$$

where $[x]^{+}=\max (0, x)$ and $\lambda^{\dagger}$ in (4.13) is the solution to following equation:

$$
\begin{equation*}
g\left(\lambda,\left\{\sigma_{k}^{2}, \tau_{k}^{2}\right\}_{k=1}^{K^{\dagger}}\right)=\prod_{k=1}^{K^{\dagger}} \tau_{k}^{-2}\left(\lambda-\sigma_{k}^{2}-\sqrt{\lambda^{2}-2 \lambda \sigma_{k}^{2}}\right)=T \tag{4.14}
\end{equation*}
$$

in which $K^{\dagger}=\max \left\{k \mid \lambda^{\dagger}>2 \sigma_{K^{\dagger}}^{2}, r_{k}^{\dagger}>0\right\}$, and $T=4^{-B_{t o t} 3^{-K^{\dagger}} \text {. Consider a new equation }}$ which is obtained by replacing $K^{\dagger}$ in (4.14) with $M$. The new equation, which we refer to as (4.14'), does not necessarily have a real solution for $\lambda$, such that $\lambda>2 \sigma_{M}^{2}$ for any value of $M \in\{2,3, \ldots, K\}$. In order to find the requirements for $\left(4.14^{\prime}\right)$ to yield a real solution for $\lambda$, we present the following Lemma and ensuing discussion. For simplicity, we drop the parameters $\left\{\sigma_{k}^{2}, \tau_{k}^{2}\right\}_{k=1}^{M}$ in $g\left(\lambda,\left\{\sigma_{k}^{2}, \tau_{k}^{2}\right\}_{k=1}^{M}\right)$ and indicate it as $g(\lambda, M)$.

Lemma 4. The function $g(\lambda, M)$ is a decreasing function of $\lambda$, for $\lambda>2 \sigma_{M}^{2}$.

Proof. Consider $g(\lambda, M)=\prod_{k=1}^{M} g_{k}(\lambda)$, where $g_{k}(\lambda)=\tau_{k}^{-2}\left(\lambda-\sigma_{k}^{2}-\sqrt{\lambda^{2}-2 \lambda \sigma_{k}^{2}}\right)$. We can
verify that $g_{k}(\lambda)$ 's are strictly decreasing in $\lambda$, because $\frac{d g_{k}(\lambda)}{d \lambda}=1-\frac{\lambda-\sigma_{k}^{2}}{\sqrt{\lambda^{2}-2 \lambda \sigma_{k}^{2}}}<0$. Since all $g_{k}(\lambda)$ 's are strictly decreasing and positive, i.e., $g_{k}(\lambda)>0, \forall k$, we conclude that $g(\lambda, M)$ is a strictly decreasing function in $\lambda$.

Having Lemma 4, we consider two scenarios that occur when solving $g(\lambda, M)=T$ : case ( $i$ ) when $T \leq\left. g(\lambda)\right|_{\lambda=\sigma_{M}^{2}}$, in this case according to Lemma 4 we have a unique real solution for $\lambda$; case (ii) when $T>\left.g(\lambda)\right|_{\lambda=\sigma_{M}^{2}}$, in this case there is no real solution for $\lambda$. Hence we need to increase the value of $\left.g(\lambda)\right|_{\lambda=\sigma_{M}^{2}}$ to reach $T$. The only way to accomplish this is decreasing the number of active sensors that contribute to $\left.g(\lambda)\right|_{\lambda=\sigma_{M}^{2}}$ and deactivating sensors with largest $\sigma_{k}^{2}$ values, until we find a real solution for $\lambda$ or only one active sensor remains. In other words, solving (4.14') in case (ii) translates into obtaining the set of active sensors $\mathcal{A}=\left\{1,2, \ldots, K^{\dagger}\right\}$ and allocating $B_{\text {tot }}$ among these active sensors.

Remark 3: The solution in (4.13) can be implemented in a distributed fashion. FC solves (4.14) and broadcasts $\lambda^{\dagger}$. Each sensor calculates its own $r_{k}^{\dagger}$ using $\lambda^{\dagger}$ via (4.13). If a sensor finds its rate to be zero or a non-real value, it means that the sensor must be inactive. The integer relaxation method has a very low computational complexity, since it requires finding the root of the monotonic function in (4.14) once and, and then calculating the rates via (4.13) for a maximum of $K$ times.

We can find an approximate closed form solution for (4.14) under the special condition when $\left(\lambda-\sigma_{k}^{2}\right)^{2}$ is large compared to $\sigma_{k}^{4}$. Rewriting the function $g_{k}(\lambda)=\tau_{k}^{-2}\left[\lambda-\sigma_{k}^{2}-\left(\left(\lambda-\sigma_{k}^{2}\right)^{2}-\right.\right.$ $\left.\left.\sigma_{k}^{4}\right)^{1 / 2}\right]$ and keeping only the first two terms in the binomial expansion of the term $\left(\left(\lambda-\sigma_{k}^{2}\right)^{2}-\right.$ $\left.\sigma_{k}^{4}\right)^{1 / 2}$, we obtain $g_{k}(\lambda) \approx \frac{\sigma_{k}^{4}}{2 \tau_{k}^{2}\left(\lambda-\sigma_{k}^{2}\right)} \approx \frac{\sigma_{k}^{4}}{2 \tau_{k}^{2} \lambda}$. Substituting the approximation in (4.14), we reach $\prod_{k} g_{k}(\lambda) \approx \prod_{k} 2^{-1} \sigma_{k}^{4} \tau_{k}^{-2} \lambda^{-1}$, based on which the Lagrange multiplier can be approximated as $\lambda^{\dagger} \approx 1.5 \eta^{2} 4^{\frac{B_{\text {tot }}}{K^{\dagger}}}$, where $\eta=\prod_{k} \sigma_{k}^{2} \tau_{k}^{-1}$. Substituting the approximation for $\lambda^{\dagger}$ in (4.13) gives the following:

$$
\begin{equation*}
r_{k}^{\dagger} \approx\left[\frac{B_{t o t}}{K^{\dagger}}+\log _{2}\left(\eta \frac{\tau_{k}}{\sigma_{k}^{2}}\right)\right]^{+} . \tag{4.15}
\end{equation*}
$$

Examining (4.15), we note that first term inside the bracket is common among active sensors and can be perceived as average rate, whereas the second term (which depends on $\tau_{k}, \sigma_{k}^{2}$ ) differs among active sensors, such that an active sensor with a larger ratio $\frac{\tau_{k}}{\sigma_{k}^{2}}$ is allocated a larger $r_{k}^{\dagger}$. Consistent with the assumption in the proof of Lemma 3, suppose $\tau_{k}=\kappa \operatorname{var}\left(x_{k}\right)=\kappa\left(\sigma_{\theta}^{2}+\sigma_{k}^{2}\right)$. Interestingly, the second term in (4.15) takes the form $\log _{2}\left(\kappa \eta\left(1+\frac{\sigma_{\theta}^{2}}{\sigma_{k}^{2}}\right)\right)$, where the ratio $\frac{\sigma_{\theta}^{2}}{\sigma_{k}^{2}}$, can be viewed as the observation SNR in (5.1). We consider two scenarios: (i) high observation SNR: the quantization rates are large (fine quantization) and less sensors become active for a given $B_{\text {tot }}$. (ii) low observation SNR: the quantization rates are smaller (coarse quantization) and more sensors become active for the same $B_{\text {tot }}$ value, compared with that of scenario (i). Substituting (4.15) in (4.9) and after some simplifications we establish the bound $\mathcal{D} \leq \mathcal{D}^{c}\left(1+\frac{\sigma_{K^{\dagger}}^{2} 4^{-\frac{B_{\text {tot }}}{K \top}}}{3 \eta^{2}}\right)$.

## - Sufficient Optimality Conditions

The objective and equality constraint functions in (4.12) are twice differentiable. Hence, the second order sufficient optimality conditions for the solution in (4.13) and (4.14) to be strict minimum for $\left(\mathbf{P}^{\prime} \mathbf{1}\right)$ are [154, p.301, proposition 3.2.1]:

$$
\begin{array}{r}
-\boldsymbol{y}^{T}\left(\left.\nabla_{r}^{2} \mathcal{L}\left(\left\{r_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)\right|_{r_{k}=r_{k}^{\dagger} \forall k} ^{\lambda=\lambda^{\dagger}}\right) \boldsymbol{y}>0, \quad \forall \boldsymbol{y} \neq 0 \\
\text { with }\left[\nabla\left(\sum_{k \in \mathcal{A}} r_{k}^{\dagger}-B_{t o t}\right)\right]^{T} \boldsymbol{y}=0, \tag{4.16}
\end{array}
$$

where $\nabla_{r}^{2} \mathcal{L}\left(\left\{r_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)$ is the Hessian matrix of the Lagrangian in (4.13), and $\nabla\left(\sum_{k \in \mathcal{A}} r_{k}^{\dagger}-\right.$ $\left.B_{t o t}\right)$ is the gradient of the equality constraint in (4.12), both evaluated at the solution in (4.13) and (4.14). It is easy to verify that the Hessian matrix is diagonal with entries:

$$
\begin{equation*}
\left[\nabla_{r}^{2} \mathcal{L}\left(\left\{r_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)\right]_{k, k}=\sigma_{\epsilon_{k}}^{2}(\ln 4)^{2} \frac{\overbrace{\frac{\sigma_{\epsilon_{k}}^{2}-\sigma_{k}^{2}}{=\beta}}^{\left(\sigma_{k}^{2}+\sigma_{\epsilon_{k}}^{2}\right)^{3}}, \forall k . ~ . . ~ . ~}{=\beta} \tag{4.17}
\end{equation*}
$$

Noting that the denominator in (4.17) and $\sigma_{\epsilon_{k}}^{2}(\ln 4)^{2}$ are positive numbers we probe into $\beta$ evaluated at the solution in (4.13):

$$
\begin{equation*}
\beta^{\dagger}=\left(\sigma_{\epsilon_{k}}^{2}-\sigma_{k}^{2}\right)=\lambda^{\dagger}-2 \sigma_{k}^{2}-\sqrt{\lambda^{\dagger^{2}-2 \lambda^{\dagger} \sigma_{k}^{2}}}<0 \tag{4.18}
\end{equation*}
$$

The inequality in (4.18) is true, because $\beta$ (which is a function of $\lambda^{\dagger}$ ) is decreasing in $\lambda^{\dagger}$ and noting that $\lambda^{\dagger}>2 \sigma_{K^{\dagger}}^{2}$ in (4.18), we have $\operatorname{Sup}_{k, \lambda^{\dagger}}\left(\lambda^{\dagger}-2 \sigma_{k}^{2}-\sqrt{\lambda^{\dagger^{2}}-2 \lambda^{\dagger} \sigma_{k}^{2}}\right)=0$. Therefore $\left[\left.\nabla_{r}^{2} \mathcal{L}\left(\left\{r_{k}, \mu_{k}\right\}_{k=1}^{K}, \lambda\right)\right|_{r_{k}=r_{k}^{\dagger} \forall k} ^{\lambda=\lambda_{k}^{\dagger}}\right]_{k, k}<0 \forall k$, confirming that the sufficient optimality conditions in (4.16) are satisfied.

## - Migration to Integer Solution

We describe an approach for migrating from the continuous solution in (4.13) to an integer solution satisfying the integer constraint [155-157] in (4.2). We round the rates to nearest integers. In case the rounding violates the bandwidth constraint, we reduce the smallest rate by one, because this sensor is more likely to be the weakest player in the network (in the sense that it has the least contribution to $\mathcal{D}$ ) until the bandwidth constraint is satisfied. Although rounding the rates to nearest integers may sounds trivial [155-157], our simulation results corroborate that the performance loss is negligible, while at the same time it keeps the rate allocation scheme simple and easily implementable.

### 4.3.4 IRA Method

Examining (4.10) closely we realize that allocating $B_{\text {tot }}$ among sensors in order to minimize $\mathcal{D}$ presents a trade-off between the number of active sensors and quantization accuracy. If $B_{t o t}$ is distributed among only few sensors, we can have fine quantization, i.e., small $K$ and small $\Delta$. On the other hand, if $B_{t o t}$ is distributed among many sensors, we can only have coarse quantization,
i.e., large $K$ and large $\Delta$. Consider a network with homogeneous sensors $\sigma_{k}^{2}=\sigma^{2}, \forall k$. Given $B_{t o t}$, there exists an optimal number of active sensors $K^{\text {opt }}$, associated with an optimal quantization rate $r^{o p t}$, where $K^{o p t} r^{o p t}=B_{t o t}$. Thus the maximization of $\sum_{k=1}^{K} \alpha_{k}=K \alpha$, where we substitute $K=$ $B_{t o t} / r$, reduces to the following one dimensional simple search for $r^{o p t}$ in the set $\mathcal{S}_{h}=\left\{1, \ldots, B_{\text {tot }}\right\}$ :

$$
\begin{equation*}
r^{o p t}=\underset{r \in \mathcal{S}_{h}}{\operatorname{argmin}}\left\{r\left(\sigma^{2}+\tau^{2} 3^{-1}\left(2^{r}-1\right)^{-2}\right)\right\}, \tag{4.19}
\end{equation*}
$$

and consequently $K^{o p t}=\left\lfloor\frac{B_{t o t}}{r^{\text {opt }}}\right\rfloor$. Modifying the solution in (4.19) for heterogeneous networks, we reach the following:

$$
\begin{equation*}
r_{k}^{s o p t}=\underset{r \in \mathcal{S}_{k}}{\operatorname{argmin}}\left\{r\left(\sigma_{k}^{2}+\tau_{k}^{2} 3^{-1}\left(2^{r}-1\right)^{-2}\right)\right\}, \tag{4.20}
\end{equation*}
$$

in which $\mathcal{S}_{1}=\left\{1, \ldots, B_{t o t}\right\}, \mathcal{S}_{k}=\left\{1, \ldots, B_{t o t}-\sum_{i=1}^{k-1} r_{i}^{\text {sopt }}\right\}$ for $k=2, \ldots, K, K^{\text {sopt }}=\max \left\{k \mid \mathcal{S}_{k} \neq\right.$ $\{0\}\}, r_{k}^{\text {sopt }}$ and $K^{\text {sopt }}$ are the rates and number of active sensors, respectively. Note that the solution in (4.20) is integer and unique (since the objective function in (4.20) is convex for $r>0$ ). The drawback of the proposed rate allocation method is that for large $B_{t o t}$, all $B_{\text {tot }}$ bits may not be allocated to sensors, i.e., $\sum_{k=1}^{K^{\text {sopt }}} r_{k}^{\text {sopt }}<B_{t o t}$, causing the solution in (4.20) to deviate from the optimal solution according to Lemma 2. This method is similar to the one in [81], with the difference that starting from sensor 1 , we update and reduce the search domain, i.e., $\mathcal{S}_{k}$ for the next sensor. This accelerates the rate allocation process. Additionally, search domain reduction in some scenarios would help to use all $B_{t o t}$ bits by activating more sensors with coarse quantizers though. The proposed method exhibits a moderate computational complexity, since it only requires solving (4.20) for a maximum of $K$ times and it is almost fully distributed [81]. Note that efficient heuristic methods with low levels of complexity have been used in other engineering fields [158-160].

### 4.4 Solving Constrained Problem (P2)

Different from $(\mathbf{P} 1)$, satisfying the MSE constraint $\mathcal{D}\left(\left\{r_{k}\right\}_{k=1}^{K}\right) \leq \mathcal{D}_{0}$ in ( $\left.\mathbf{P} 2\right)$ enforces the number of active sensors to exceed a minimum number $K_{\min }$. Lemma 5 provides $K_{\text {min }}$.

Lemma 5. To satisfy the the constraint $\mathcal{D}\left(\left\{r_{k}\right\}_{k=1}^{K}\right) \leq \mathcal{D}_{0}$ we need at least $K_{\text {min }}$ active sensors, where $K_{\text {min }}=\min \left\{K \mid \sum_{k=1}^{K} \sigma_{k}^{-2}>\mathcal{D}_{0}^{\prime}=\mathcal{D}_{0}{ }^{-1}-\sigma_{\theta}^{-2}\right\}$.

Proof. Considering (4.9) and the definition of $\alpha_{k}$ 's, we find that $\mathcal{D}\left(\left\{r_{k}\right\}_{k=1}^{K}\right) \leq \mathcal{D}_{0}$ is equivalent to $\sum_{k=1}^{K} \alpha_{k} \geq \mathcal{D}_{0}^{\prime}=\mathcal{D}_{0}{ }^{-1}-\sigma_{\theta}^{-2}$. Thus ( $\mathbf{P} 2$ ) is equivalent to minimizing $\sum_{k=1}^{K} r_{k}$ such that $\sum_{k=1}^{K} \alpha_{k} \geq \mathcal{D}_{0}^{\prime}$ and $r_{k} \in \mathbb{Z}_{+}, \forall k$. Since $\alpha_{k}>0, \forall k$, we can increase $\sum_{k=1}^{K} \alpha_{k}=\sum_{k=1}^{K}\left(\frac{1}{\sigma_{k}^{2}+\sigma_{\epsilon_{k}}^{2}}\right)$, via increasing the number of active sensors, until the MSE constraint is satisfied. This implies that the minimum number of active sensors can be found by letting $\sigma_{\epsilon_{k}}^{2}=0, \forall k$, i.e., $K_{\text {min }}=$ $\min \left\{K \mid \sum_{k=1}^{K} \sigma_{k}^{-2}>\mathcal{D}_{0}^{\prime}\right\}$.

In the following we propose three methods for solving $(\mathbf{P} 2): A)$ greedy method, $B$ ) integer relaxation method, $C$ ) IRA method. We obtain these methods via applying some modifications to the proposed methods in section 4.3.

### 4.4.1 Greedy Method

According to Lemma 5, we need at least $K_{\min }$ active sensors. Therefore, we initiate the algorithm with $r_{k}=1$ for $k \in\left\{1, \ldots, K_{\min }\right\}$ and let $r_{k}=0$ otherwise, and go through the greedy method until the MSE constraint is satisfied.

Data: $\mathcal{D}_{0}, \sigma_{\theta}^{2},\left\{\tau_{k}\right\}_{k=1}^{K},\left\{\sigma_{k}^{2}\right\}_{k=1}^{K}$
Result: rate allocation $\left\{r_{k}^{*}\right\}_{k=1}^{K}$
initialization;

```
\(r_{k}=1\) for \(k \in\left\{1, \ldots, K_{\text {min }}\right\}\) o.w. \(r_{k}=0, \mathcal{S}=\left\{1, K_{\text {min }}+1\right\}, \quad \mathcal{D}_{0}^{\prime}=\mathcal{D}_{0}{ }^{-1}-\sigma_{\theta}^{-2}\),
    \(\mathbf{d}=\sum_{k=1}^{K_{\min }} \alpha_{k}\left(r_{k}\right)\)
while \(\mathbf{d}<\mathcal{D}_{0}^{\prime}\) do
    \(k^{*}=\underset{k \in \mathcal{S}}{\operatorname{argmax}}\left(\alpha_{k}\left(r_{k}+1\right)-\alpha_{k}\left(r_{k}\right)\right)\)
    \(r_{k^{*}}=r_{k^{*}}+1\)
    \(\mathbf{d}=\mathbf{d}-\alpha_{k^{*}}\left(r_{k^{*}}-1\right)+\alpha_{k^{*}}\left(r_{k^{*}}\right) \mathcal{S}=\left\{k \mid r_{k}<r_{k-1}\right\} \cup\{1\}\)
end
Algorithm: greedy method for rate allocation in (P2)
```


### 4.4.2 Integer Relaxation Method

Let $\left(\mathbf{P}^{\prime} \mathbf{2}\right)$ be the corresponding relaxed problem of $(\mathbf{P} 2)$. Solving the first order KKT necessary optimality conditions for $\left(\mathbf{P}^{\prime} \mathbf{2}\right)$ yields a similar solution to (4.13) as the following:

$$
\begin{equation*}
r_{k}^{\ddagger}=0.5\left[\log _{2}\left(\tau_{k}^{2}\left(\lambda^{\ddagger}-\sigma_{k}^{2}-\sqrt{\lambda^{\ddagger}-2 \lambda^{\ddagger} \sigma_{k}^{2}}\right)^{-1}\right)-\log _{2} 3\right]^{+} \tag{4.21}
\end{equation*}
$$

One can show that the objective function, i.e., $\sum_{k=1}^{K} r_{k}$ is a strictly increasing function of $\alpha_{k}$ 's. Hence, the optimal solution to ( $\left.\mathbf{P}^{\prime} \mathbf{2}\right)$ must satisfy the MSE constraint as equality, i.e., $\sum_{k=1}^{K} \alpha_{k}=$ $\mathcal{D}_{0}^{\prime}$. Using the MSE equality constraint we find that $\lambda^{\ddagger}$ in (4.21) is the solution to the following equation:

$$
\begin{equation*}
f\left(\lambda,\left\{\sigma_{k}^{2}\right\}_{k=1}^{K^{\ddagger}}\right)=\sum_{k=1}^{K^{\ddagger}} \frac{1}{\lambda-\sqrt{\lambda^{2}-2 \lambda \sigma_{k}^{2}}}=\mathcal{D}_{0}^{\prime} \tag{4.22}
\end{equation*}
$$

where $K^{\ddagger}=\max \left\{k \mid k \geq K_{\text {min }}, \lambda^{\ddagger}>2 \sigma_{K^{\ddagger}}^{2}, r_{k}^{\ddagger}>0\right\}$. Similar to what we did for the solution in (4.13), one can verify that the solution in (4.21) satisfies the second order sufficient optimality conditions in [154, p.301, proposition 3.2.1]. Note that (4.22) does not necessarily have a real
solution for $\lambda$. We first let $K^{\ddagger}=K$, i.e., the largest possible value for $K^{\ddagger}$ in the feasible set $\mathcal{F}=\left\{K_{\text {min }}, \ldots, K\right\}$ and solve (4.22). If there is no real solution for $\lambda$ we decrease the number of active sensors by one, i.e., $K^{\ddagger}=K-1$, and solve (4.22). We continue reducing the number of active sensors one by one until we reach a real solution for $\lambda^{\ddagger}$ or $K^{\ddagger}=K_{\text {min }}$ (the smallest possible value for $K^{\ddagger}$ in the feasible set). Even when $K^{\ddagger}=K_{\text {min }}$ it is still possible that solving (4.22) does not yield a real solution for $\lambda^{\ddagger}$. Since $f\left(\lambda,\left\{\sigma_{k}^{2}\right\}_{k=1}^{K_{\text {min }}}\right)$ is an increasing function of $\lambda$, this scenario would occur when $\mathcal{D}_{0}^{\prime}<\left.f\left(\lambda,\left\{\sigma_{k}^{2}\right\}_{k=1}^{K_{\min }}\right)\right|_{\lambda=2 \sigma_{K_{\text {min }}}^{2}}$. In this scenario we let $\lambda^{\ddagger}=2 \sigma_{K_{\text {min }}}^{2}$. Substituting $\lambda^{\ddagger}=2 \sigma_{K_{\text {min }}}^{2}$ in (4.21) and then $r_{k}^{\ddagger}$,s in $\sum_{k=1}^{K_{\text {min }}} \alpha_{k}$, we obtain:

$$
\sum_{k=1}^{K_{\min }} \alpha_{k}=\left.\left(\sum_{k=1}^{K_{\min }} \frac{1}{\lambda-\sqrt{\lambda^{2}-2 \lambda \sigma_{k}^{2}}}\right)\right|_{\lambda=2 \sigma_{K_{\min }}^{2}}>\mathcal{D}_{0}^{\prime}
$$

implying that the MSE constraint is met. Using similar approximation that led us to (4.15), we can approximate (4.21) as:

$$
\begin{equation*}
r_{k}^{\ddagger} \approx\left[\log _{2}\left(\frac{\tau_{k}}{\sigma_{k}^{2}}\right)+\log _{2}\left(\sqrt{\frac{K^{\ddagger}}{3\left(\sum_{k=1}^{K} \sigma_{k}^{-2}-\mathcal{D}_{0}^{\prime}\right)}}\right)\right]^{+} . \tag{4.23}
\end{equation*}
$$

Equation (4.23) shows as target MSE approaches its feasible minimum, i.e., as $\mathcal{D}_{0} \rightarrow \frac{1}{\sigma_{\theta}^{-2}+\sum_{k=1}^{K} \sigma_{k}^{-2}}$ and $\mathcal{D}_{0}^{\prime} \rightarrow \sum_{k=1}^{K} \sigma_{k}^{-2}$, the rates $r_{k}^{\ddagger}$ 's become very large, i.e., $r_{k}^{\ddagger} \rightarrow \infty$.

### 4.4.3 IRA Method

Following a similar reasoning to the one provided in Section 4.3.4 for a homogeneous network and recalling the discussion on satisfying the MSE constraint as equality in Section 4.4.2, we conclude that, given $\mathcal{D}_{0}^{\prime}$, there exists an optimal number of active sensors $K^{\text {opt }}$, associated with an optimal quantization rate $r^{o p t}$, where $K^{o p t} \alpha\left(r^{o p t}\right)=\mathcal{D}_{0}^{\prime}$ and our problem is to minimize $K^{o p t} r^{o p t}$ subject to
this MSE equality constraint. This optimization problem for a heterogeneous network, reduces to almost the same as in (4.20), with a difference that the search domain includes any positive integer number, i.e.,

$$
\begin{array}{r}
r_{k}^{\text {sopt }}=\underset{r \in \mathbb{Z}^{+}}{\operatorname{argmin}}\left\{r\left(\sigma_{k}^{2}+\tau_{k}^{2} 3^{-1}\left(2^{r}-1\right)^{-2}\right)\right\}, \\
K^{\text {sopt }}=\min \left\{k \mid K_{\min } \leq k \leq K, \sum_{i=1}^{k} \alpha_{i}\left(r_{i}\right) \geq \mathcal{D}_{0}^{\prime}\right\} . \tag{4.25}
\end{array}
$$

Note that there is no need to solve (4.24) for all sensors, since the rate allocation continues only till we find $K^{\text {sopt }}$ in (4.25).

### 4.5 BCRLB

We derive the BCRLB for any Bayesian estimator of $\theta$ based on quantized observations $\left\{m_{k}\right\}_{k=1}^{K}$. Assuming that the regularity condition is satisfied, i.e., $\mathbb{E}\left\{\frac{\partial \ln _{p(\mathbf{m}, \theta)}}{\partial \theta}\right\}=0$ [4] we write the Fisher information:

$$
\begin{equation*}
F=-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m}, \theta)}{\partial^{2} \theta}\right\}=-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta}\right\}-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\} \tag{4.26}
\end{equation*}
$$

Assuming that $m_{k}$ 's conditioned on $\theta$ are independent, i.e., $\ln p(\mathbf{m} \mid \theta)=\sum_{k=1}^{K} \ln p\left(m_{k} \mid \theta\right)$, the first and second derivatives of the log-likelihood function become:

$$
\begin{aligned}
& \frac{\partial \ln p(\mathbf{m} \mid \theta)}{\partial \theta}=\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta} \\
& \frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta}=\overbrace{\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial^{2} p\left(m_{k} \mid \theta\right)}{\partial^{2} \theta}}^{=F_{a}} \\
& -\underbrace{\sum_{k=1}^{K} \frac{1}{p^{2}\left(m_{k} \mid \theta\right)}\left(\frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta}\right)^{2}}_{=F_{b}}
\end{aligned}
$$

In the following, we find $\mathbb{E}\left\{F_{a}\right\}, \mathbb{E}\left\{F_{b}\right\}$. We have:

$$
\begin{aligned}
& \mathbb{E}\left\{F_{a}\right\}=\sum_{k=1}^{K} \int p(\theta) \frac{\overbrace{M_{k}}^{M_{k}} s_{k, i}(\theta))}{\partial^{2} \theta} d \theta=0, \\
& \mathbb{E}\left\{F_{b}\right\}=\sum_{k=1}^{K} \int p(\theta) \sum_{i=1}^{M_{k}} \frac{1}{s_{k, i}(\theta)}\left(\dot{s}_{k, i}(\theta)\right)^{2} d \theta,
\end{aligned}
$$

where $s_{k, i}(\theta)=p\left(m_{k}=m_{k, i} \mid \theta\right)=\mathbf{p}\left\{\left.m_{k, i}-\frac{\Delta_{k}}{2} \leq h_{k} \theta+n_{k} \leq m_{k, i}+\frac{\Delta_{k}}{2} \right\rvert\, \theta\right\}$ and $\dot{s}_{k, i}(\theta)=\frac{\partial s_{k, i}(\theta)}{\partial \theta}$. To complete the derivations of $F$ we need to characterize $s_{k, i}(\theta)$ and $\dot{s}_{k, i}(\theta)$. Combining all above and recalling $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$, we obtain:

$$
\begin{equation*}
F=\frac{1}{\sigma_{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{M_{k}} \int \frac{\left(\dot{s}_{k, i}(\theta)\right)^{2}}{s_{k, i}(\theta)} \phi\left(\frac{\theta}{\sigma_{\theta}}\right) d \theta+\frac{1}{\sigma_{\theta}^{2}}, \tag{4.27}
\end{equation*}
$$

where $\phi($.$) is the standard normal PDF. Equation (5.29) is true for arbitrarily distributed h_{k}$ 's with $\mathbb{E}\left\{h_{k}\right\}=1 \forall k$, and $\operatorname{var}\left(h_{k}\right)=\sigma_{h_{k}}^{2}$. When $h_{k}$ 's are Gaussian we have (for non-Gaussian $h_{k}$ 's see
appendix B.2):

$$
s_{k, i}^{G}(\theta)=\Phi\left(\frac{\zeta_{k, i+1}-\theta}{\sqrt{\theta^{2} \sigma_{h_{k}}^{2}+\sigma_{n_{k}}^{2}}}\right)-\Phi\left(\frac{\zeta_{k, i}-\theta}{\sqrt{\theta^{2} \sigma_{h_{k}}^{2}+\sigma_{n_{k}}^{2}}}\right),
$$

where $\zeta_{k, i}=m_{k, i}-\frac{\Delta_{k}}{2}, \zeta_{k, i+1}=m_{k, i}+\frac{\Delta_{k}}{2}$ are the quantizer boundaries, and $\Phi($.$) is the CDF of a$ standard normal random variable. Deriving $\dot{s}_{k, i}^{G}(\theta)$ is straightforward and reduces to subtraction of two scaled standard normal PDFs.

### 4.6 Extension to Erroneous Channels

To obtain our results so far we have focused on error-free communication channel model, i.e., the quantization bits from the sensors are available at the FC, to feed the LMMSE estimator. The results can be extended to independent BSCs with different error probabilities $p_{k}$. Suppose sensor $k$ uses Binary Natural Coding (BNC) to code its quantized message $m_{k}$, that is sent through a BSC with error probability $p_{k}$, and $\hat{m}_{k}$ is the corresponding recovered quantization level at the FC, where in general $\hat{m}_{k} \neq m_{k}$, due to channel errors.

### 4.6.1 LMMSE Estimator and its corresponding MSE

The LMMSE estimator and its MSE would have the same forms as in (4.4) and (4.5), with the difference that vector $\boldsymbol{m}$ is replaced with vector $\hat{\boldsymbol{m}}$. We characterize $\mathbb{E}\left\{\theta \hat{m}_{k}\right\}$ and $\mathbb{E}\left\{\hat{m}_{k} \hat{m}_{l}\right\}$ as the following:

$$
\begin{equation*}
\mathbb{E}\left\{\theta \hat{m}_{k}\right\}=\mathbb{E}\left\{\mathbb{E}\left\{\theta \hat{m}_{k} \mid \theta, m_{k}\right\}\right\}=\mathbb{E}\left\{\theta \mathbb{E}\left\{\hat{m}_{k} \mid m_{k}\right\}\right\} . \tag{4.28}
\end{equation*}
$$

With BNC of bit sequences and BSC model we have $\mathbb{E}\left\{\hat{m}_{k} \mid m_{k}\right\}=\left(1-2 p_{k}\right) m_{k}$ [161]. Thus (4.28) reduces to $\mathbb{E}\left\{\theta \hat{m}_{k}\right\}=\left(1-2 p_{k}\right) \mathbb{E}\left\{\theta m_{k}\right\}$, where $\mathbb{E}\left\{\theta m_{k}\right\}$ is characterized in (4.6). For $\mathbb{E}\left\{\hat{m}_{k} \hat{m}_{l}\right\}, k \neq l$ and $k=l$ we have:

$$
\begin{align*}
& \mathbb{E}\left\{\hat{m}_{k} \hat{m}_{l}\right\}=\mathbb{E}\left\{\mathbb{E}\left\{\hat{m}_{k} \hat{m}_{l} \mid \theta, m_{k}, m_{l}\right\}\right\} \stackrel{(a)}{=} \\
& \mathbb{E}\left\{\mathbb{E}\left\{\hat{m}_{k} \mid m_{k}\right\} \mathbb{E}\left\{\hat{m}_{l} \mid m_{l}\right\}\right\}=\left(1-2 p_{k}\right)\left(1-2 p_{l}\right) \mathbb{E}\left\{m_{k} m_{l}\right\}, \\
& \mathbb{E}\left\{\hat{m}_{k}^{2}\right\}=\mathbb{E}\left\{\mathbb{E}\left\{\hat{m}_{k}^{2} \mid m_{k}\right\}\right\} \stackrel{(b)}{=} g_{k} \mathbb{E}\left\{m_{k}^{2}\right\}+R_{k}, \tag{4.29}
\end{align*}
$$

where (a) in (4.29) is obtained using the facts that $(i)$ given $m_{k}, m_{l}$ then $\hat{m}_{k}, \hat{m}_{l}$ are independent, (ii) given $\theta$, then $m_{k}, m_{l}$ are uncorrelated (since $n_{k}, n_{l}, h_{k}, h_{l}$ are all uncorrelated). And $g_{k}=$ $\left(1-p_{k}\right)^{r_{k}-1}\left(1+p_{k}\left(r_{k}-5\right)\right)$ and $R_{k}=(4 / 3)\left(1-p_{k}\right)^{r_{k}-1} p_{k} \tau_{k}^{2}\left(2^{r_{k}}+1\right)\left(2^{r_{k}}-1\right)^{-1}$. To obtain (b) in (4.29) we assume at most one bit in a sequence of $r_{k}$ bits can be flipped due to the channel errors (roughly speaking $p_{k} \ll r_{k}^{-1}$ ). This is a reasonable assumption noting that for a poor channel with $p_{k} \approx 0.1$ and typical quantization rates of $r_{k} \leq 6$, flipping more than one bit in an $r_{k}$-bit sequence is unlikely [161]. Note that $\mathbb{E}\left\{m_{k} m_{l}\right\}$ and $\mathbb{E}\left\{m_{k}^{2}\right\}$ in (4.29) are characterized in (4.7). Having (4.28), (4.29), the LMMSE estimator and its corresponding MSE are characterized for BSC model.

### 4.6.2 BCRLB and Fisher Information Expressions

To find $F$ based on $\hat{m}_{k}$ 's, we need to find the counterpart of (5.28), where $\boldsymbol{m}$ is replaced with $\hat{\boldsymbol{m}}$. For independent BSC model, $\hat{m}_{k}$ 's conditioned on $\theta$ would be independent, leading to $\ln p(\hat{\boldsymbol{m}} \mid \theta)=$ $\sum_{k=1}^{K} \ln p\left(\hat{m}_{k} \mid \theta\right)$. Following similar steps as in Section 4.5, we find new $F_{a}$ to be zero. New $F_{b}$ can be found by replacing $p\left(m_{k}=m_{k, i} \mid \theta\right)$ with $p\left(\hat{m}_{k}=m_{k, i} \mid \theta\right)$ in the derivations. All that remains is
to characterize:

$$
p\left(\hat{m}_{k}=m_{k, i} \mid \theta\right)=\sum_{j=1}^{M_{k}} e_{k}^{j i} p\left(m_{k}=m_{k, j} \mid \theta\right)
$$

where $e_{k}^{i j}$ is the probability of receiving level $m_{k, i}$, while level $m_{k, j}$ is transmitted from sensor $k$. Note that $e_{k}^{i j}$ can be found in terms of $p_{k}$, i.e., $e_{k}^{j i}=\left(p_{k}\right)^{n\left(j, i, r_{k}\right)}\left(1-p_{k}\right)^{r_{k}-n\left(j, i, r_{k}\right)}$, where $n\left(j, i, r_{k}\right)$ is the Hamming distance between BNC representations of $m_{k, j}=\sum_{l=1}^{r_{k}} b_{k, j, l} 2^{r_{k}-l}$ and $m_{k, i}=\sum_{l=1}^{r_{k}} b_{k, i, l} 2^{r_{k}-l}$. To sum up, $F$ becomes:

$$
F=\sigma_{\theta}^{-1} \sum_{k=1}^{K} \sum_{i=1}^{M_{k}} \int \frac{\left(\sum_{j=1}^{M_{k}} e_{k}^{j i} \dot{s}_{k, j}(\theta)\right)^{2}}{\left(\sum_{j=1}^{M_{k}} e_{k}^{j i} s_{k, j}(\theta)\right)} \phi\left(\frac{\theta}{\sigma_{\theta}}\right) d \theta+\sigma_{\theta}^{-2}
$$

### 4.7 Numerical and Simulation Results

In this section, we corroborate our analytical results with numerical simulations. These results validate the accuracy of our analysis and illustrate the effectiveness and superiority of the proposed rate allocation schemes. We consider networks of sizes $K=5,10,50$ and conduct simulations for over $10^{5}$ observation channels with randomly generated $\left\{\sigma_{n_{k}}^{2}, \sigma_{h_{k}}^{2}\right\}_{k=1}^{K}$ and depict the average performance for all rate allocation methods (greedy, integer relaxation (relaxed), IRA, Order Aware (OA) uniform, and uniform). We generate $\sigma_{n_{k}}^{2}$ such that $\mathbb{E}\left\{\sigma_{n_{k}}^{2}\right\}=1$ or $\mathbb{E}\left\{\sigma_{n_{k}}^{2}\right\}=k_{n}$. To investigate the effect of multiplicative observation noise variance on the network dynamics and performance, we let $\mathbb{E}\left\{\sigma_{h_{k}}^{2}\right\}=k_{h}=0.1,1,2,4$ to indicate low, moderate, high, and very high multiplicative noise variance.


Figure 4.3: Analytical and simulated MSE of greedy vs. $B_{\text {tot }}$


Figure 4.4: MSE of different rate allocation methods vs. $B_{t o t}$ for $K=5$

Figs. 4.3a and 4.3b compare the analytical MSE in (4.9) and simulated MSE for $K=10,50$, when greedy method is employed ${ }^{4}$. The simulations are conducted for $h_{k}$ 's drawn from Gaussian, uniform, and Laplacian distributions. We observe that the analytical MSE is a good approximation of simulated MSE for almost all scenarios, and the approximation accuracy improves as $K$ increases and/or $k_{h}$ decreases. Also, except for small $K$ and very high $k_{h}$, the distribution of $h_{k}$ has negligible effect on the approximation accuracy.

[^7]

Figure 4.5: MSE performance for different allocation schemes for $K=50$

Figs. 4.4 and 4.5 compares the analytical MSE for different methods for $K=5,50$ and $k_{h}=$ $0.1,1,2$. We observe the MSE performance gap between uniform (including OA uniform), and greedy and integer relaxation are remarkable. As $B_{\text {tot }}$ increases, the performance of greedy, integer relaxation, uniform, OA uniform approaches that of the clairvoyant centralized estimation. However, there is a persistent gap between the performance of IRA and the clairvoyant case, even for large $B_{t o t}$. The performance of greedy and integer relaxation are almost the same for all scenarios. Similar observations are valid for $K=50$ and the plots are omitted due to lack of space. For large $K$ and small $k_{h}$ the performance of the individual rate allocation competes with greedy and integer relaxation methods, however, for small $K$ or high $k_{h}$ it loses the competition. On the other hand, when $B_{t o t}$ is relatively small compared to $K$, greedy, integer relaxation, and IRA have the same performance. As expected, we observe larger $k_{h}$ (larger $K$ ) leads to a larger (smaller) MSE for all methods.

Fig. 4.6 and 4.7 depicts the MSE performance of different methods and the associated BCRLB for $K=5,50$, where $h_{k}$ 's are drawn from Gaussian distribution for BCRLB. The $\sigma_{n_{k}}^{2}$ 's and $\sigma_{h_{k}}^{2}$ 's are independently generated with Chi-Square distribution $\sigma_{h_{k}}^{2} \sim \chi^{2}\left(k_{h}\right), \sigma_{n_{k}}^{2} \sim \chi^{2}(1)$.


Figure 4.6: MSE and CRLB vs. $B_{t o t}$ for $K=5$

For $k_{h}=1$ (moderate multiplicative noise), there is a noticeable gap between the MSE and the associated BCRLB for all methods, whereas for $k_{h}=0.1$ (low noise) and large $B_{\text {tot }}$, this gap tends to be very small. This is in agreement with the result that the MSE of MMSE estimator for a Gaussian linear observation model achieves the BCRLB [4]. In fact, for $k_{h}=0$ the observation model in (5.1) becomes the linear Gaussian model $x_{k}=\theta+n_{k}$ and when $B_{t o t} \rightarrow \infty$, LMMSE estimator in (4.4) becomes MMSE estimator, which achieves the BCRLB. Similar observations are valid for $K=50$ and the plots are omitted due to lack of space.


Figure 4.7: MSE and CRLB for proposed schemes for $K=50$

Figs. 4.8a and 4.8b depict the number of active sensors versus $B_{\text {tot }}$ for all methods and $K=10,50$. For $k_{h}=2$ (high noise) more sensors become active to reduce the noise effect, by averaging over observations coming from more sensors, leading to smaller quantization rates (coarser quantization). On the other hand, for $h_{k}=0.1$ (low noise) less sensors become active, leading to larger quantization rates (finer quantization). These observations illustrate the trade-off mentioned in explanations following (4.15) and in subsection 4.3.4. Note that greedy and integer relaxation methods activate fewer sensors, compared with those of IRA and uniform methods, and still provide better MSE performance (see also Fig. 4.4 and 4.5).


Figure 4.8: Number of active sensors vs. $B_{\text {tot }}$

Figs. 4.9 a and 4.9 b illustrate the average quantization rates of active sensors versus $B_{t o t}$ for all methods and $K=10,50$. For $k_{h}=2$ (high noise) the average quantization rates are smaller (more active sensors with coarser quantization). On the other hand, for $h_{k}=0.1$ (low noise) the average quantization rates are larger (less active sensors with finer quantization). These observations illustrate the trade-off mentioned in explanations following (4.15) and in subsection 4.3.4.


Figure 4.9: Average quantization rate of active sensors vs. $B_{t o t}$


Figure 4.10: required bandwidth versus a target MSE for different allocation schemes for $K=5$

Fig. 4.10 and 4.11 illustrates the required bandwidth, i.e., sum of quantization rates $\sum_{k=1}^{K} r_{k}$ versus a target MSE (to be satisfied), for all methods and $K=5,50$. Note that greedy and integer relaxation methods require much less bandwidth to achieve the target MSE, compared with those of IRA, uniform, and OA uniform. Similar observations are valid for $K=5$ and the plot is omitted due to lack of space. These figures (excluding IRA) show that, more bandwidth is required to $(i)$ satisfy a smaller target MSE, (ii) satisfy a fixed target MSE for larger $k_{h}$, (iii) satisfy a fixed target MSE for smaller $K$. In some sub-figures the required bandwidth for some target MSE values are left blank, since the target MSE is not achievable for that particular network setting. Note that IRA method is different from greedy and integer relaxation methods, since it is blind to target MSE and $K$ value (see (4.24),(4.25)), i.e., the assigned quantization rates are independent of the target MSE and $K$ value and the number of active sensors is kept at minimum, such that the target MSE is satisfied. For illustrative purposes, consider a large and easy-to-be-satisfied MSE target, such that it lies in the interval $[0.7,1]$. Such a target MSE most likely can be satisfied with one active sensor (see Figs. 4.12a, 4.12b), and one bit (see Fig. 4.13) in greedy and integer relaxation methods. However, since IRA is blind to the target MSE, it assigns a quantization rate to the only active sensor, according to the observation channel quality $\sigma_{1}^{2}$, that is likely to be larger than one bit (in fact, the smaller $k_{h}$ is, the larger $r_{k}$ is).


Figure 4.11: Required bandwidth of different rate allocation methods vs. target MSE for $K=50$


Figure 4.12: Number of active sensors vs. target MSE

Figs. 4.12a and 4.12 b depict the number of active sensors when the target MSE is met for all methods and $K=10,50$. Note that greedy and integer relaxation methods activate fewer sensors to satisfy the target MSE, compared with those of IRA and uniform and OA uniform. For $k_{h}=2$ (high noise), all methods require more active sensors to satisfy the target MSE (similar observations to those of Fig.4.8).


Figure 4.13: Quantization rate of active sensors vs. target MSE

Figs. 4.13a and 4.13b illustrate the average quantization rates of active sensors versus target MSE, for all methods and $K=10,50$. Similar conclusions to those for Fig. 4.11 can be made here. Note that IRA loses the competition to uniform methods for large target MSE values. These figures show that, larger average quantization rate is required to $(i)$ satisfy a smaller target MSE, (ii) satisfy a fixed target MSE for larger $k_{h}$, (iii) satisfy a fixed target MSE for smaller $K$ (compare the average rate of all algorithms except IRA in Fig. 4.13). In the figures the average quantization rates for some target MSE values are left blank, since the target MSE is not attainable for that particular network setting. Combining the observations from Figs. 4.11, 4.12, 4.13, we conclude that IRA method is not suitable to address (P2). To show the effect of erroneous communication channels [149, 162], Fig. 4.14 depicts the analytical and simulated MSE and compare them with BCRLB for $p=10^{-1}, 10^{-2}, 10^{-4}, K=50$, and $k_{h}=1$, when greedy method is employed. As expected, the analytical MSE is very accurate unless for large error probability $p=0.1$ (this is expected since to derive (4.29), we assume $p_{k} \ll r_{k}^{-1}$, which is not true for $p=0.1$ ).


Figure 4.14: MSE and CRLB of greedy vs. $B_{t o t}$ for $K=50$

- Comparison with Rate Distortion (R-D) Bound in [15]: As we mentioned in chapter 1, DES and the quadratic Gaussian CEO problem are different and hence the R-D bounds in [14-16] are less relevant to the problem in hand. Interestingly, our simulations show that in some scenarios, even the BCRLB based on multi-bit quantization can reach the R-D bound in [15]. Fig. 4.15 compares the BCRLB based on one-bit and multi-bit quantization with the R-D bound for different $K$ and $k_{n}$.

We use the R-D bound in [15] given below, which is for a heterogeneous network with limited number of agents:

$$
\begin{equation*}
R(D)=0.5\left[\log _{2}\left(\frac{\sigma_{\theta}^{2}}{D} \prod_{k=1}^{\tilde{K}}\left(\frac{\tilde{K}}{\sigma_{n_{k}}^{2}\left(\frac{1}{\tilde{D}(\tilde{K})}-\frac{1}{D}\right)}\right)\right)\right]^{+} \tag{4.30}
\end{equation*}
$$

where $\tilde{D}(\tilde{K})=\left(\sigma_{\theta}^{-2}+\sigma_{n_{1}}^{-2}+\ldots+\sigma_{n_{\tilde{K}}}^{-2}\right)^{-1}$ and $\tilde{K}$ is the largest value that satisfies $\frac{\tilde{K}}{\sigma_{n_{\tilde{K}}}^{2}}-\left(\frac{1}{\tilde{D}(\tilde{K})}-\frac{1}{D}\right) \geq 0$. Note that the gap between the R-D bound and the BCRLB based on multi-bit quantization is not persistent: as $B_{t o t}$ increases for a fixed $K$, the gap fades away and the latter approaches the former.


Figure 4.15: R-D and BCRLB for $K=5,50$ and $\sigma_{\theta}^{2}=1$

- One-bit MLE and MAP vs. Multi-bit LMMSE: As we mentioned earlier in literature review, there is a significant gap between the CRLB performance based on one-bit quantization and the clairvoyant benchmark (unquantized observations are available at the FC), when the dynamic range of $\theta$ is large with respect to $\sigma_{n_{k}}^{2}$ [33]. To illustrate this, Fig. 4.16 plots classical CRLB $=\left(\frac{\left(\dot{s}_{k, i}(\theta)\right)^{2}}{s_{k, i}(\theta)}\right)^{-1}$ versus $\theta$ in a homogeneous network with $\sigma_{n}^{2}=1$, assuming $\theta \in[-4,4]$. corresponding to one bit [121] and two bit quantization. We observe the classical CRLB corresponding to two bit quantization is significantly better than that of one bit quantization, when $\theta$ is larger than $\sigma_{n}^{2}$. One expects similar observation holds when we compare MSE of MLE corresponding to one bit and two bit quantization.


Figure 4.16: CRLB correspoding to one bit and two bit quantizers

In Appendix B. 1 we provide an analytical reason for the poor performance of CRLB and MLE based on one bit quantization. When PDF of $\theta$ is known a priori, MAP estimator can be used instead of MLE. Figs. 4.17 and 4.18 plot MSE of one-bit MLE, one-bit MAP, and the proposed multi-bit LMMSE estimators, as well as the associated (one-bit and multi-bit) CRLBs and R-D for a heterogeneous network of size $K=50$ versus $k_{n}$ and $B_{t o t}$, respectively. As expected, in all cases the proposed multi-bit LMMSE outperforms one-bit MLE and one-bit MAP. Fig. 4.17 also shows that for small $\sigma_{n}^{2}$ one-bit MLE and one-bit MAP perform poorly. Also, as $\sigma_{\theta}^{2}$ becomes larger the performance gap between one-bit MLE, one-bit MAP and the proposed multi-bit LMMSE increases significantly, as expected from our analysis in Section B.1.


Figure 4.17: Performance comparison of one-bit and multi-bit estimation vs. $k_{n}$, with $K=$ $50, B_{\text {tot }}=50, \theta_{\max }=2 \sigma_{\theta}$


Figure 4.18: Performance comparison of one bit and multi bit estimation vs. $B_{t o t}, K=50$, $\theta_{\max }=2 \sigma_{\theta}$

### 4.8 Conclusions

We considered DES of a Gaussian source in a heterogeneous bandwidth constrained WSN, where the source is corrupted by independent multiplicative and additive observation noises, with incomplete statistical knowledge of the multiplicative noise. For uniform multi-bit quantizers, we derived the closed-form MSE expression for the LMMSE estimator at the FC, and verified the accuracy of our derivations via simulations. For both error-free and erroneous communication channels (using BSC model) we proposed several rate allocation methods to (i) minimize the MSE given a network bandwidth constraint, and (ii) minimize the required network bandwidth given a target MSE. We also derived the BCRLB and compared the MSE performance of our proposed methods against the BCRLB. Our results corroborate that, for low power multiplicative observation noises and adequate network bandwidth, the gaps between the MSE of our proposed methods and the BCRLB are negligible, while the performance of other methods like IRA in [81], and uniform is not satisfactory. Through analysis and simulations, we showed that one-bit MLE and one-bit MAP in the literature have poor performance, when the realizations of unknown is large (compared with the observation noise variances), whereas our proposed multi-bit LMMSE significantly outperforms estimators based on one-bit quantization.

## CHAPTER 5: NOISE ENHANCEMENT IN BAYESIAN DISTRIBUTED ESTIMATION

In this chapter we investigate the problem of Bayesian DES of an unknown Gaussian source in a WSN with a FC. Gaussian multiplicative and additive noise environments are supposed to be present. Assuming that first and second order statistics of multiplicative noise is available and known, we derive the WWB and BCRLB for two cases: $i$ ) full precision observations of sensors are available in the FC , ii) only quantized version of observations are available in the FC. For both cases we report and characterize scenarios that according to the bounds, presence of multiplicative noise can enhance the estimation accuracy. We call this phenomena enhancement mode of multiplicative noise. We compare the MSE performance of MMSE estimator and MAP estimator with WWB and BCRLB bounds and characterize the enhancement modes according to MSE performance of the estimators. In sequel we consider the case that variance of multiplicative noise is unknown, considering the variance of the noise as a nuisance deterministic parameter, we derive three well-known bounds: HCRLB, NB and RUB to assess the performance limits of DES for this case. According to the bounds the enhancement modes are more scarce in comparison to the case that the variance of noise is known. Through simulation we demonstrate different examples of multiplicative noise enhancing the DES accuracy according to the bounds. We compare MSE performance of MMSE-ML and MAP-ML estimators with HCRLB, NB and RUB for the case that variance of multiplicative noise is unknown. According to the MSE of these estimators, there is no enhancement mode for this case, and multiplicative noise can not enhance the estimation when its variance is unknown.

### 5.1 System Model and Problem Statement

We consider a network with $K$ spatially distributed sensors and a FC, where the FC is interested in estimating a realization of a random unknown source $\theta \in \Theta=[-\infty, \infty]$ with Gaussian distribution $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$, via fusing the collective received data from all sensors. Note that $\theta$ is not directly observable. Instead each sensor independently observes a noisy version of $\theta$, where both multiplicative and additive observation noises are involved. Let $x_{k}$ denote the scalar noisy observation of $\theta$ at sensor $k$. We assume the following observation model ${ }^{1}$ :

$$
\begin{equation*}
x_{k}=h_{k} \theta+n_{k}, \quad \text { for } \quad k=1, \ldots, K \tag{5.1}
\end{equation*}
$$

where $h_{k}$ and $n_{k}$ are multiplicative and additive observation noises, respectively. Also, $\theta, h_{k}, n_{k}$ are all uncorrelated.

For sake of presentation, we assume multiplicative $h_{k}$ 's and additive $n_{k}$ 's noises are i.i.d, i.e., $n_{k} \sim \mathcal{N}\left(0, \sigma_{n}^{2}\right), h_{k} \sim \mathcal{N}\left(1, \sigma_{h}^{2}\right), \forall k^{2}$. We define vector $\mathbf{x}=\left[x_{1}, \ldots, x_{K}\right]^{T} \in \Omega=[-\infty,+\infty]^{K}$ that includes all sensors' observations. The sensors are not allowed to convene before transmitting to the FC. We focus on comprehending the effects of observation noises on the estimation accuracy and assume that the communication channels between the sensors and the FC are error-free. Error-free communication channel model has been used before in several classical works on DES, examples are [18-20,33, 81]. Having received data from all sensors, the FC employs a Bayesian

[^8]estimator to reconstruct $\theta$ and form $\hat{\theta}$. Let $\operatorname{MSE}=\mathbb{E}\left\{(\hat{\theta}-\theta)^{2}\right\}$ denote the corresponding MSE. We study the achievable lower bounds on the MSE and in particular we investigate the effect of multiplicative noise on the bounds. We unearth some scenarios in which the presence of multiplicative noise can enhance the DES accuracy, via decreasing the achievable lower bounds on the MSE. We characterize the mentioned scenarios as enhancement modes of multiplicative noise and compare those with the case that there is no multiplicative noise.

### 5.2 Lower Bounds on MSE

We investigate the effect of multiplicative noise on DES performance in two cases/frameworks: a) known variance for multiplicative noise, i.e., $\sigma_{h}^{2}$ is known $b$ ) unknown variance for multiplicative noise, i.e., $\sigma_{h}^{2}$ is unknown. In order to investigate the first case/framework and characterize the effect of multiplicative noise on achievable bounds we focus on deriving and probing into two well-known Bayesian bounds: $i$ ) WWB [125] which is known to be the tightest bound of the Weiss and Weinstein family and ii) BCRLB [129]. In order to investigate the second case/framework we focus on deriving and characterizing hybrid bounds such as well-known HCRLB (first introduced in [130] and build on [131]), NB [133, 134], and more recently proposed RUB [135].

### 5.2.1 Lower bound on the MSE for known $\sigma_{h}^{2}$

### 5.2.1.1 WWB

General form of WWB on the MSE of any Bayesian estimator is given in [125] as following:

$$
\begin{equation*}
W W B=\sup _{t, s} \frac{t^{2} e^{2 \mu(s, t)}}{e^{\mu(2 s, t)}+e^{\mu(2-2 s,-t)}-2 e^{\mu(s, 2 t)}}, \tag{5.2}
\end{equation*}
$$

where $s \in[0,1], t$ is chosen on the parameter support and $\mu(s, t)$ is given as:

$$
\begin{equation*}
\mu(s, t)=\ln \int_{\Theta} \int_{\Omega} \frac{p^{s}(\mathbf{x}, \theta+t)}{p^{s-1}(\mathbf{x}, \theta)} d \mathbf{x} d \theta \tag{5.3}
\end{equation*}
$$

The integral in (5.3) can be re written as following:

$$
\begin{equation*}
\mu(s, t)=\ln \int_{\Theta} \frac{p^{s}(\theta+t)}{p^{s-1}(\theta)} \underbrace{\int_{\Omega} \frac{p^{s}(\mathbf{x} \mid \theta+t)}{p^{s-1}(\mathbf{x} \mid \theta)} d \mathbf{x}}_{=J} d \theta \tag{5.4}
\end{equation*}
$$

$J$ in (5.4) is an $K$ dimensional integral, we delve into characterizing it. From now on we assume $s=0.5$ (since the optimum of (5.2) happens in $s=0.5$ [125]), and drop the $s$ argument in $\mu(s, t)$. Since $h_{k}$ 's and $n_{k}$ 's are uncorrelated Gaussian random variables, the $x_{k}$ 's conditioned on $\theta$ are independent Gaussian random variables, consequently we have:

$$
\begin{equation*}
p(\mathbf{x} \mid \theta)=\prod_{k=1}^{K} p\left(x_{k} \mid \theta\right)=\frac{e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)}}}{\left[2 \pi\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)\right]^{\frac{-K}{2}}}, \tag{5.5}
\end{equation*}
$$

where $1=[1, \ldots, 1]^{T}$. Substituting (5.5) in $J$ and doing tedious math, we obtain:

$$
\begin{equation*}
J=\frac{e^{-\frac{K t^{2}}{4\left(\sigma_{0}^{2}+\sigma_{t}^{2}\right)}}}{(2 \pi)^{\frac{K}{2}}\left(\sigma_{0}^{2} \sigma_{t}^{2}\right)^{\frac{K}{4}}} \underbrace{\int_{\Omega} e^{-a\|\mathbf{x}-1 b\|^{2}} d \mathbf{x}}_{=\left(\sqrt{\frac{\pi}{a}}\right)^{K}}, \tag{5.6}
\end{equation*}
$$

where $\sigma_{0}^{2}=\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}, \sigma_{t}^{2}=(\theta+t)^{2} \sigma_{h}^{2}+\sigma_{n}^{2}, a=\frac{\left(\sigma_{0}^{2}+\sigma_{t}^{2}\right)}{4 \sigma_{0}^{2} \sigma_{t}^{2}}$, and some $b \neq \pm \infty$, where the integral in (5.6) is independent of $b$. The definite integral in (5.6) is expressible in closed-form because it can be related to integral of an $K$ dimensional joint Gaussian pdf with mean vector of $1 b$ and covariance matrix of $(2 a)^{-1} \mathbf{I}$ ( $\mathbf{I}$ is identity matrix) over the support $\Omega$.

Substituting (5.6) in (5.4) and doing some clean up, we get:

$$
\begin{equation*}
\mu(t)=C(t)+\ln \left[\int_{\Theta}\left(\frac{\sqrt{\sigma_{0}^{2} \sigma_{t}^{2}}}{\sigma_{0}^{2}+\sigma_{t}^{2}}\right)^{\frac{K}{2}} e^{-\frac{(\theta+0.5 t)^{2}}{2 \sigma_{\theta}^{2}}-\frac{K t^{2}}{4\left(\sigma_{0}^{2}+\sigma_{t}^{2}\right)}} d \theta\right], \tag{5.7}
\end{equation*}
$$

where $C(t)=\ln \left(\frac{2^{\frac{K}{2}}}{\sqrt{2 \pi \sigma_{\theta}^{2}}}\right)-\frac{t^{2}}{8 \sigma_{\theta}^{2}}$. As can be seen we were able to reduce the $K+1$ dimensional integral in (5.3) into a one dimensional one in (5.7). Note that the integral in (5.7) can not be expressed in closed-form, however it can easily be calculated with numerical methods since the integrand is an smooth function with no singularities. After calculating $\mu(t)$ we need to do a one dimensional maximization over $t$ (look at (5.2)). Note that $\mu(t)$ is an even function (look at appendix C. 1 for the proof) and it also decays rapidly to zero, thus we confined the search domain to $t \in\left[0,4 \sigma_{\theta}\right]$ without any performance degradation.

### 5.2.1.2 BCRLB

In this subsection we derive the BCRLB for any estimator of $\theta$ in (5.1) based on observation vector $\mathbf{x}$. The Cramér Rao theorem [129] states that the MSE of any estimator is larger or equal to the inverse of the Fisher information. Assuming that regularity conditions is satisfied, i.e., $\mathbb{E}\left\{\frac{\partial \ln _{p(\mathbf{x}, \theta)}}{\partial \theta}\right\}=0[4,166]$, we write the Fisher information as:

$$
\begin{array}{r}
F=-\mathbb{E}_{\mathbf{x}, \theta}\left\{\frac{\partial^{2} \ln p(\mathbf{x}, \theta)}{\partial^{2} \theta}\right\}= \\
-\mathbb{E}_{\theta}\{\underbrace{\mathbb{E}_{\mathbf{x} \mid \theta}\left\{\frac{\partial^{2} \ln p(\mathbf{x} \mid \theta)}{\partial^{2} \theta}\right\}}_{*}\}-\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\} . \tag{5.8}
\end{array}
$$

The expression mentioned with $*$ in (5.8) can be perceived as negative of Fisher information for the case of estimating deterministic $\theta$ based on observation vector $\mathbf{x}$. Note that the observations in (5.1), conditioned on $\theta$ are jointly Gaussian random variables, so based on results for general

Gaussian case in $[4,3.9]$ we can write:

$$
\begin{align*}
\mathbb{E}_{\mathbf{x} \mid \theta}\left\{\frac{\partial^{2} \ln p(\mathbf{x} \mid \theta)}{\partial^{2} \theta}\right\}= & -\left[\frac{\partial u(\theta)}{\partial \theta}\right]^{T} \boldsymbol{\Sigma}^{-1}(\theta)\left[\frac{\partial u(\theta)}{\partial \theta}\right]- \\
& 0.5 \text { trace }\left[\left(\boldsymbol{\Sigma}^{-1}(\theta) \frac{\partial \boldsymbol{\Sigma}(\theta)}{\partial \theta}\right)^{2}\right] \tag{5.9}
\end{align*}
$$

where $u(\theta)=\mathbb{E}_{x \mid \theta}\{\mathbf{x} \mid \theta\}=\mathbf{1} \theta$ and $\boldsymbol{\Sigma}=\mathbb{E}_{x \mid \theta}\left\{\mathbf{x x}^{T} \mid \theta\right\}=\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right) \mathbf{I}$. Substituting these expressions in (5.9) and doing a tedious cleaning up, we obtain:

$$
\begin{equation*}
\mathbb{E}_{\mathbf{x} \mid \theta}\left\{\frac{\partial^{2} \ln p(\mathbf{x} \mid \theta)}{\partial^{2} \theta}\right\}=-K \frac{\sigma_{n}^{2}+\theta^{2} \sigma_{h}^{2}+2 \theta^{2} \sigma_{h}^{4}}{\left(\sigma_{n}^{2}+\sigma_{h}^{2} \theta^{2}\right)^{2}} . \tag{5.10}
\end{equation*}
$$

Next we need to take the expectation of expression in (5.10) with respect to distribution of $\theta$ :

$$
\begin{array}{r}
\mathbb{E}_{\theta}\left\{\mathbb{E}_{\mathbf{x} \mid \theta}\left\{\frac{\partial^{2} \ln p(\mathbf{x} \mid \theta)}{\partial^{2} \theta}\right\}\right\}= \\
\int_{-\infty}^{\infty}-K \frac{\sigma_{n}^{2}+\theta^{2} \sigma_{h}^{2}+2 \theta^{2} \sigma_{h}^{4}}{\sqrt{2 \pi \sigma_{\theta}^{2}}\left(\sigma_{n}^{2}+\sigma_{h}^{2} \theta^{2}\right)^{2}} e^{\frac{-\theta^{2}}{2 \sigma_{\theta}^{2}}} d \theta \tag{5.11}
\end{array}
$$

through ponderous and dull integral mathematics the definite integral in (5.11) can be expressed in closed-form. Substituting the closed-form expression into the (5.8) and replacing $\mathbb{E}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}=$ $\sigma_{\theta}^{-2}$, the complete expression for Bayesian Fisher information can be written as:

$$
\begin{equation*}
F=K\left(\sqrt{2 \pi} e^{\frac{\sigma_{n}^{2}}{\sigma_{\theta}^{2} \sigma_{h}^{2}}} \Phi\left(-\frac{\sigma_{n}}{\sigma_{\theta} \sigma_{h}}\right)\left(\frac{\sigma_{n}^{2}+\sigma_{\theta}^{2}+\sigma_{\theta}^{2} \sigma_{h}^{2}}{\sigma_{n} \sigma_{\theta}^{3} \sigma_{h}}\right)-\frac{1}{\sigma_{h}^{2}}\right)+\frac{1}{\sigma_{\theta}^{2}}, \tag{5.12}
\end{equation*}
$$

where $\Phi($.$) is the CDF of a standard normal random variable, i.e., \Phi(x)=\int_{-\infty}^{x} \frac{1}{\sqrt{2 \pi}} e^{-\frac{t^{2}}{2}} d t$. In order to delve into dynamics of Fisher information in (5.12) as a function of $\sigma_{h}$, we re ar-
range/formulate/write that as in (5.13), where $z \triangleq \sigma_{h}$ and $\alpha \triangleq \sigma_{n} / \sigma_{\theta}$. Existing the term $z / \alpha$ in Fisher expression (which is indicated with $*$ in (5.13)), can causes the Fisher to be increased with increasing $z$, which is totally unintuitive. We characterize such phenomena as enhancement mode of multiplicative noise and we will dig into/illustrate its details through numerical examples in section 5.5.

$$
\begin{equation*}
F=\frac{K \sqrt{2 \pi}}{\sigma_{\theta}^{2}} e^{\frac{\alpha^{2}}{2 z^{2}}} \Phi\left(-\frac{\alpha}{z}\right)(\frac{\alpha+\alpha^{-1}}{z}+\underbrace{\frac{z}{\alpha}}_{*})+\frac{1-K}{\sigma_{\theta}^{2}} . \tag{5.13}
\end{equation*}
$$

Remark 1: In case that there is no multiplicative noise, i.e., $z \rightarrow 0^{+}$, the Fisher information in (5.13) goes to $F^{c v}=\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}$, which is considered as the clairvoyant Fisher information for general Linear Gaussian model [129]. On other hand when multiplicative noise intensity goes to infinity, i.e., $z \rightarrow+\infty$, the Fisher information in (5.13) goes to infinity as well, meaning that multiplicative noise pushes the minimum achievable MSE to zero. Technically speaking when multiplicative noise is in enhancement mode, the more intense gets the noise the more enhancement we get till it reach the maximum of $\frac{1}{F^{c v}}-0=\left(F^{c v}\right)^{-1}$ enhancement in the MSE.

Lemma 6. For the observation model in (5.1) with no multiplicative noise, i.e., $\sigma_{h}^{2}=0$, (which reduces to general linear Gaussian model, i.e., $\left.x_{k}=\theta+n_{k}, \quad \forall k\right)$, the WWB bound reduces to clairvoyant BCRLB , i.e., $W W B=\left(F^{c v}\right)^{-1}$. In other word for the case that there is no multiplicative noise there is no need to calculate WWB bound because it is as tight as BCRLB.

Proof. For $\sigma_{h}^{2}=0$ we have $\sigma_{0}^{2}=\sigma_{t}^{2}=\sigma_{n}^{2}$, and it is easy to verify that the expression in (5.7) reduces to:

$$
\mu(t)=-\frac{t^{2}}{8}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right),
$$

and the WWB for this case is as following:

$$
\begin{equation*}
W W B=\sup _{t} \frac{t^{2} e^{-\frac{t^{2}}{4}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}}{2\left(1-e^{-\frac{t^{2}}{2}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}\right)} . \tag{5.14}
\end{equation*}
$$

Taking the derivative of the function in (5.14) with respect to $t$ and finding its zeros, easily reveals that it has a unique zero/solution in $t=0$, thus we have:

$$
\begin{array}{r}
W W B=\sup _{t} \frac{t^{2} e^{-\frac{t^{2}}{4}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}}{2\left(1-e^{-\frac{t^{2}}{2}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}\right)} \\
=\lim _{t \rightarrow 0^{+}} \frac{t^{2} e^{-\frac{t^{2}}{4}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}}{2\left(1-e^{-\frac{t^{2}}{2}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)}\right)} \stackrel{(a)}{=}\left(\frac{K}{\sigma_{n}^{2}}+\frac{1}{\sigma_{\theta}^{2}}\right)^{-1}=\left(F^{c v}\right)^{-1},
\end{array}
$$

the calculation of the limit in $(a)$ is easy to verify with $L^{\prime} H \hat{o p}$ otal's rule [167].

Lemma 7. For the observation model in (5.1) with multiplicative noise, i.e., $\sigma_{h}^{2}>0$, the WWB is strictly tighter than the BCRLB, i.e., $W W B>F^{-1}$.

Proof. Since $W W B \geq F^{-1}$ [125], and $W W B \neq F^{-1}$, so $W W B>F^{-1}$.

### 5.2.2 Lower bounds on the MSE for unknown $\sigma_{h}^{2}$

For the case that the value of $\sigma_{h}^{2}$ is not available, we assume that to be as a deterministic ${ }^{3}$ nuisance parameter, while we are interested in estimating random $\theta$. It is dealing with the situation where we have an unknown vector to be estimated defined as $\boldsymbol{\eta} \triangleq\left[\theta, \sigma_{h}^{2}\right]^{T}$, where the unknown vector consist of both random and deterministic parameters. To address such an estimation problem, the HCRLB first has been presented in context of a specific application in [130] and formulated as a general tool in [131]. The attractiveness of HCRLB stems from the fact that it provides a matrix lower bound on the MSE of both the deterministic and random parameters. Other bounds that address the same issue are YB [132], NB [133, 134] and RUB [135]. We will argue that the RUB is tighter than former ones for the problem in hand.

### 5.2.2.1 HCRLB

According to [131], the HCRLB matrix for estimating $\boldsymbol{\eta}$ can be formulated as in (5.15), where $\hat{\sigma_{h}^{2}}$ and $\hat{\theta}$ are estimators for $\sigma_{h}^{2}$ and $\theta$. Note that conditioned on $\boldsymbol{\eta}$ the vector of observations $\mathbf{x}$ form a jointly Gaussian vector, consequently after some tedious mathematical operations the elements of matrix $\mathbf{G} \triangleq\left[g_{11}, g_{12} ; g_{21}, g_{22}\right]$ can be written (refer to results for general Gaussian case in $[4,3.9]$ )

[^9]as following:
\[

$$
\begin{align*}
& g_{11}=\frac{-K \theta^{4}}{2\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)^{2}}, \quad g_{22}=\frac{-K\left(\sigma_{n}^{2}+\theta^{2} \sigma_{h}^{2}+2 \theta^{2} \sigma_{h}^{4}\right)}{\left(\sigma_{n}^{2}+\sigma_{h}^{2} \theta^{2}\right)^{2}}, \\
& g_{12}=g_{21}=\frac{-K \theta^{3} \sigma_{h}^{2}}{\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)^{2}}, \tag{5.16}
\end{align*}
$$
\]

accordingly the HCRLB in (5.15) can be written as following:

$$
\begin{array}{r}
H C R L B=\left(\left[\begin{array}{cc}
\mathbb{E}_{\theta}\left\{g_{11}\right\} & \mathbb{E}_{\theta}\left\{g_{12}\right\} \\
\mathbb{E}_{\theta}\left\{g_{21}\right\} & \mathbb{E}_{\theta}\left\{g_{22}\right\}+\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}
\end{array}\right]\right)^{-1} \\
=\frac{1}{\mathbb{E}_{\theta}\left\{g_{11}\right\}\left(\mathbb{E}_{\theta}\left\{g_{22}\right\}+\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}\right)-\left(\mathbb{E}_{\theta}\left\{g_{12}\right\}\right)^{2}} \times \\
{\left[\begin{array}{cc}
\mathbb{E}_{\theta}\left\{g_{22}\right\}+\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\} & -\mathbb{E}_{\theta}\left\{g_{12}\right\} \\
-\mathbb{E}_{\theta}\left\{g_{21}\right\} & \left.\mathbb{E}_{\theta}\left\{g_{11}\right\}\right\}
\end{array}\right] .}
\end{array}
$$

Thus we conclude that the HCRLB bound for estimator $\hat{\theta}$ is obtained as following:

$$
\begin{array}{r}
\mathbb{E}_{\mathbf{x}, \theta}\{\hat{\theta}-\theta\}^{2} \geq H C R L B= \\
\mathbb{E}_{\theta}\left\{g_{11}\right\}  \tag{5.17}\\
\mathbb{E}_{\theta}\left\{g_{11}\right\}\left(\mathbb{E}_{\theta}\left\{g_{22}\right\}+\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}\right)-\left(\mathbb{E}_{\theta}\left\{g_{12}\right\}\right)^{2}
\end{array}
$$

According to (5.16) one can easily verify that $\mathbb{E}_{\theta}\left\{g_{12}\right\}=0$, because $g_{12}$ is an odd function of $\theta$ so the expectation of that with respect to Gaussian distribution with zero mean is always zero. Thus HCRLB in (5.17) reduces to:

$$
\begin{equation*}
H C R L B=\frac{1}{\mathbb{E}_{\theta}\left\{g_{22}\right\}+\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}} \tag{5.18}
\end{equation*}
$$

According to definition of $g_{22}$ in (5.16) and comparing it to (5.10), (5.11) and noting $\mathbb{E}_{\theta}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\}=$ $\sigma_{\theta}^{-2}$, we conclude that expression of HCRLB in (5.18) is same as the expression for BCRLB in (5.12).

Remark 2: For the observation model of (5.1), the HCRLB for estimating $\theta$ for the case of unknown multiplicative noise variance $\sigma_{h}^{2}$, is equal to the BCRLB for the case of known $\sigma_{h}^{2}$. Thus the well known HCRLB does not yields any new information on estimation performance for the model in (5.1) in comparison to BCRLB ${ }^{4}$.

### 5.2.2.2 $Y B$

According to [132] the YB for the model in (5.1) can be established as following:

$$
\begin{array}{r}
\mathbb{E}_{\mathbf{x}, \theta}\left\{(\hat{\theta}-\theta)^{2}\right\} \geq Y B= \\
{\left[\mathbb{E}_{\mathbf{x}}\left\{\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}\right]^{2} H^{-1}} \tag{5.19}
\end{array}
$$

where $\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)$ is the MMSE estimator of unknown $\theta$ (assuming $\sigma_{h}^{2}$ is known), i.e., $\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)=$ $\mathbb{E}\left\{\theta \mid \mathbf{x} ; \sigma_{h}^{2}\right\}$, and $H$ is the Fisher information with respect to deterministic unknown $\sigma_{h}^{2}$.
${ }^{4}$ Note that this is true for zero mean $\theta$. On the other hand for the case of $\mathbb{E}\{\theta\} \neq 0$, we have $\mathbb{E}\left\{g_{12}\right\} \neq 0$ so (5.17) does not reduce to (5.18).

Lemma 8. For the observation model in (5.1) we have following equality:

$$
\mathbb{E}_{\mathbf{x}}\left\{\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}=0
$$

Proof. Refer to appendix C.2.

Having the result in lemma 8 and without digging into deriving $H$ in (5.19) we conclude that YB is always zero for the system model in (5.1). Thus the YB is not a useful bound for the problem in hand.

### 5.2.2.3 $N B$

According to $[133,134]$ the NB for the model in (5.1) can be formulated as following:

$$
\begin{array}{r}
\mathbb{E}_{\mathbf{x}, \theta}\left\{(\hat{\theta}-\theta)^{2}\right\} \geq N B= \\
\mathbb{E}_{\mathbf{x}, \theta}\left\{\left(\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)-\theta\right)^{2}\right\}+Y B
\end{array}
$$

As stated in lemma 8 the YB is equal to zero, so the NB reduces to the MSE of optimal MMSE estimator assuming that $\sigma_{h}^{2}$ is known. In other word NB states that the lower bound on MSE for observation model in (5.1) with unknown $\sigma_{h}^{2}$ is equal to the MSE of the MMSE estimator of $\theta$, with known $\sigma_{h}^{2}$. In numerical examples we will see that for adequate number of sensors NB will decrease as multiplicative noise gets stronger.

### 5.2.2.4 RUB

According to [135] the RUB for the model in (5.1) can be written as in (5.20) or as in (5.21).
$R U B=\mathbb{E}_{\mathbf{x}, \theta}\left\{\left(\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)-\theta\right)^{2}\right\}+$
$\left(\mathbb{E}_{\left.\mathbf{x}\left\{\left(\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right)^{2}\right\}\right)^{3}} \begin{array}{r}\mathbb{E}_{\mathbf{x}}\left\{\left(\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right)^{2}\right\} \mathbb{E}_{\mathbf{x}}\left\{\left(\frac{\partial^{2} \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial^{2} \sigma_{h}^{2}}+\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} \frac{\partial \ln p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right)^{2}\right\}-\left(\mathbb{E}_{\mathbf{x}}\left\{\left(\frac{\partial^{2} \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial^{2} \sigma_{h}^{2}}+\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} \frac{\partial \ln p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right) \frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}\right)^{2}\end{array}\right.$

$$
\frac{R U B=N B+}{} \frac{\left(\mathbb{E}_{\mathbf{x}}\left\{d^{2}\left(\mathbf{x}, \sigma_{h}^{2}\right)\right\}\right)^{3}}{\mathbb{E}_{\mathbf{x}}\left\{d^{2}\left(\mathbf{x}, \sigma_{h}^{2}\right)\right\} \mathbb{E}_{\mathbf{x}}\left\{e^{2}\left(\mathbf{x}, \sigma_{h}^{2}\right)\right\}-\left(\mathbb{E}_{\mathbf{x}}\left\{d\left(\mathbf{x}, \sigma_{h}^{2}\right) e\left(\mathbf{x}, \sigma_{h}^{2}\right)\right\}\right)^{2}},
$$

where

$$
\begin{array}{r}
d\left(\mathbf{x}, \sigma_{h}^{2}\right)=\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} \\
e\left(\mathbf{x}, \sigma_{h}^{2}\right)=\frac{\partial^{2} \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial^{2} \sigma_{h}^{2}}+d\left(\mathbf{x}, \sigma_{h}^{2}\right) \frac{\partial \ln p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}
\end{array}
$$

for derivations of $d\left(\mathbf{x}, \sigma_{h}^{2}\right)$ see appendix C.2, we omitted the derivations for $e\left(\mathbf{x}, \sigma_{h}^{2}\right)$ because it is very straight forward and similar to what we did in appendix C.2.

Note that the RUB is valid for Risk Unbiased (RU) estimators [135]. An estimator $\hat{\theta}$, for the model in 5.1 is considered to be RU provided that it satisfies following:

$$
\mathbb{E}_{\mathbf{x}, \theta}\left\{\left(\hat{\theta}-\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)\right) \frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}=0
$$

Lemma 9. RUB is always tighter than NB.

Proof. the second term in (5.21) is always positive because ( $i$ ) the term in nominator is positive
since it is expectation of a positive function. (ii) the term in denominator is positive according to Cauchy- Schwartz inequality $\mathbb{E}\left\{d^{2}\right\} \mathbb{E}\left\{e^{2}\right\} \geq(\mathbb{E}\{d e\})^{2}$.

### 5.3 Estimation Schemes

### 5.3.1 Estimation Schemes for known $\sigma_{h}^{2}$

### 5.3.1.1 MMSE estimator

For the case of known $\sigma_{h}^{2}$ we can employ MMSE estimator. This estimator can be derived by calculating two one dimensional integrals as following (see Appendix C. 2 for details):

$$
\begin{align*}
& \hat{\theta}_{M M S E}=\mathbb{E}\{\theta \mid \mathbf{x}\}= \\
& \frac{\int_{-\infty}^{+\infty} \frac{e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}}{\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2} \frac{K}{2}\right.} d \theta}{\int_{-\infty}^{+\infty} \frac{\theta e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}}{\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)^{\frac{K}{2}}}} . \tag{5.22}
\end{align*}
$$

The integrals in (5.22) do not have closed-form solutions, however those are easily calculated with numerical methods.

### 5.3.1.2 MAP estimator

The MAP estimator for the model in (5.1) can be written as following:

$$
\begin{array}{r}
\hat{\theta}_{M A P}=\underset{\theta}{\operatorname{argmax}}[\ln p(\mathbf{x} \mid \theta)+\ln p(\theta)] \stackrel{(a)}{=} \\
\underset{\theta}{\operatorname{argmax}}\left[K \ln \left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)+\frac{\|\mathbf{x}-\mathbf{1} \theta\|^{2}}{\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}}+\frac{\theta^{2}}{\sigma_{\theta}^{2}}\right], \tag{5.23}
\end{array}
$$

where (a) can be verified noting that the vector $\mathbf{x}$ conditioned on unknown $\theta$, has a joint Gaussian distribution with mean vector of $\theta \mathbf{1}$, and covariance matrix of $\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right) \mathbf{I}$. Taking the derivative of a posteriori distribution in (5.23) and finding its zeros, provide us with following equation for MAP estimator:

$$
\begin{align*}
& \left(\frac{\sigma_{h}^{4}}{\sigma_{\theta}^{2} K}\right) \theta^{5}+\left(\sigma_{h}^{4}+\frac{2 \sigma_{h}^{2} \sigma_{n}^{2}}{\sigma_{\theta}^{2} K}\right) \theta^{3}+\left(\bar{S} \sigma_{h}^{2}\right) \theta^{2}+ \\
& \left(\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}\right)-\bar{V} \sigma_{h}^{2}+\frac{\sigma_{n}^{4}}{\sigma_{\theta}^{2} K}\right) \theta-\bar{S} \sigma_{n}^{2}=0 \tag{5.24}
\end{align*}
$$

where $\bar{S}=\left(\mathbf{x}^{\prime} \mathbf{1}\right) / K=\left(\sum_{k=1}^{K} x_{k}\right) / K, \bar{V}=\left(\|\mathbf{x}\|^{2}\right) / K$. The zero of the equation in (5.24) which maximizes the posteriori distribution in (5.23), is the MAP estimator. Note that for large $K$ the MAP estimator reduces to Bayesian MLE estimator with ML derivative as following:

$$
\sigma_{h}^{4} \theta^{3}+\bar{S} \sigma_{h}^{2} \theta^{2}+\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}-\bar{V} \sigma_{h}^{2}\right) \theta-\bar{S} \sigma_{n}^{2}=0
$$

### 5.3.2 Estimation Schemes for unknown $\sigma_{h}^{2}$

For the case that $\sigma_{h}^{2}$ is unknown, we employ the well-known MMSE-ML and MAP-ML estimators [171]. For these estimators, first we find the ML estimate of unknown $\sigma_{h}^{2}$, afterwards having the estimate of $\sigma_{h}^{2}$ we use the MMSE and MAP estimators in (5.22) and (5.23) respectively to form the MMSE-ML and MAP-ML estimates for $\theta$ for case of unknown $\sigma_{h}^{2}$. It is proved in [171] that MMSE-ML and MAP-ML estimators are RU and asymptotically converge to MMSE estimator in probability, i.e.,

$$
\begin{array}{r}
\lim _{K \rightarrow+\infty} \mathbb{E}\left\{\left(\hat{\theta}_{M M S E-M L}-\hat{\theta}_{M M S E}\right)\left(\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right)\right\}=0 \\
\lim _{K \rightarrow+\infty} \mathbb{E}\left\{\left|\hat{\theta}_{M M S E-M L}-\hat{\theta}_{M M S E}\right|\right\}=0 .
\end{array}
$$

The results in above equalities are valid for $\hat{\theta}_{M A P-M L}$ estimator as well [171]. Having observation vector $\mathbf{x}$, the ML equation for estimating $\sigma_{h}^{2}$, is as following (luckily the zeros of a cubic function can be derived in closed form):

$$
\begin{equation*}
\theta^{3}-3 \bar{S} \theta^{2}+\left(2 \bar{S}^{2}+\bar{V}\right) \theta-\bar{S} \times \bar{V}=0 \tag{5.25}
\end{equation*}
$$

for details of derivation of (5.25) look at appendix C.3.

### 5.4 Extension of Derivations of WWB and BCRLB to the Case of Quantized Observations

Suppose the scalar observations $x_{k}$ 's in (5.1) are separately quantized via uniform quantizers in sensors with fixed step size of $\Delta$. The quantizers have $M$ quantization levels with boundaries $\left\{\zeta_{1}, \ldots, \zeta_{M+1}\right\}$. The observation $x_{k}$ is mapped into the quantization level $m_{k}$, where it belongs to one of $M$ quantization levels, i.e., $m_{k}=\left\{l_{1}, \ldots, l_{M}\right\}$. The mapping function is like $x_{k} \in$ $\left[\zeta_{i}, \zeta_{i+1}\right) \rightarrow m_{k}=l_{i}$. Assuming that with very high probability $x_{k}$ lies in the interval $[-\tau, \tau]$ for a reasonably large value of $\tau$, we consider the step size of the quantizers to be $\Delta=\frac{2 \tau}{M-1}$. The boundaries are set as $\zeta_{1+(M / 2)}=0$ and $\zeta_{i+1}-\zeta_{i}=\Delta$ for $i \in\{1, \ldots, M\}$. The quantization levels are set as $l_{i}=\frac{\zeta_{i}+\zeta_{i+1}}{2}$. The quantized observations can be mapped into a binary sequence of length $r=\log _{2} M$ (bits) and transmitted to the FC. We refer to $r$ as quantization rate. We define $\mathbf{m}=$ $\left[m_{1}, \ldots, m_{K}\right]^{T}$ as vector of received quantized observations in the FC, where $m_{k} \in \omega=\left\{l_{1}, \ldots, l_{m}\right\}$ , and $\mathbf{m} \in \Omega^{\prime}=\omega^{K}$.

### 5.4.1 WWB derivation for quantized observations

The counterpart of (5.3) for the case of quantized observations named as $\mu_{q}(t)$ can be formulated as following:

$$
\begin{gather*}
\mu_{q}(t)=\ln \int_{\Theta} \sum_{\mathbf{m} \in \Omega^{\prime}} \frac{p^{0.5}(\mathbf{m}, \theta+t)}{p^{-0.5}(\mathbf{m}, \theta)} d \theta \\
\mu_{q}(t)=\ln \int_{\Theta} \frac{p^{0.5}(\theta+t)}{p^{-0.5}(\theta)} \underbrace{\sum_{\mathbf{m} \in \Omega^{\prime}} \frac{p^{0.5}(\mathbf{m} \mid \theta+t)}{p^{-0.5}(\mathbf{m} \mid \theta)}}_{=J^{\prime}} d \theta . \tag{5.26}
\end{gather*}
$$

Since $h_{k}$ 's and $n_{k}$ 's are uncorrelated Gaussian random variables, and quantizers in sensors are independent, the $m_{k}$ 's conditioned on $\theta$ are independent, so $p(\mathbf{m} \mid \theta)=\prod_{k} p\left(m_{k} \mid \theta\right)$, where:

$$
\begin{aligned}
& p\left(m_{k}=\right.\left.l_{i} \mid \theta\right)=p\left(\zeta_{i} \leq h_{k} \theta+n_{k}<\zeta_{i+1} \mid \theta\right)= \\
& \underbrace{\Phi\left(\frac{\zeta_{i+1}-\theta}{\sqrt{\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}}}\right)-\Phi\left(\frac{\zeta_{i}-\theta}{\sqrt{\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}}}\right)}_{\triangleq P_{i}(\theta)} .
\end{aligned}
$$

For brevity we drop the $\theta$ argument in $P_{i}(\theta)$ and represent it as $P_{i}$. We also define $P_{i}^{\prime} \triangleq P_{i}(\theta+t)$. Then $p\left(m_{k}=l_{i}\right)$ can be written as following:

$$
\begin{array}{r}
p\left(m_{k} \mid \theta\right)=P_{1}^{\delta\left[m_{k}-l_{1}\right]} \times \ldots \times P_{M}^{\delta\left[m_{k}-l_{M}\right]} \\
p(\mathbf{m} \mid \theta)=P_{1}^{\sum_{k} \delta\left[m_{k}-l_{1}\right]} \times \ldots \times P_{M}^{\sum_{k} \delta\left[m_{k}-l_{M}\right]},
\end{array}
$$

where $\delta[$.$] is the discrete delta function. Accordingly we can write:$

$$
\begin{array}{r}
\frac{p^{0.5}(\mathbf{m} \mid \theta+t)}{p^{-0.5}(\mathbf{m} \mid \theta)}= \\
{\left[\left\{P_{1} P_{1}^{\prime}\right\}^{\sum_{k} \delta\left[m_{k}-l_{1}\right]} \ldots\left\{P_{M} P_{M}^{\prime}\right\}^{\sum_{k} \delta\left[m_{k}-l_{M}\right]}\right]^{1 / 2}} \\
=\left[\prod_{k=1}^{K} \prod_{i=1}^{M}\left\{P_{i} P_{i}^{\prime}\right\}^{\delta\left[m_{k}-l_{i}\right]}\right]^{1 / 2} . \tag{5.27}
\end{array}
$$

Having (5.27), it is easy to verify that the $K$ dimensional sum in $J^{\prime}$ can be decoupled into multiplication of $K$ independent one dimensional sum as following:

$$
J^{\prime}=\underbrace{\sum_{m_{1} \in \omega} \prod_{i=1}^{M}\left\{P_{i} P_{i}^{\prime}\right\}^{\frac{\delta\left[m_{1}-l_{i}\right]}{2}}}_{=\sum_{i=1}^{M} \sqrt{P_{i} P_{i}^{\prime}}} \times \ldots \times \underbrace{\sum_{m_{K} \in \omega} \prod_{i=1}^{M}\left\{P_{i} P_{i}^{\prime}\right\}^{\frac{\delta\left[m_{M}-l_{i}\right]}{2}}}_{=\sum_{i=1}^{M} \sqrt{P_{i} P_{i}^{\prime}}} .
$$

One can verify that each summation in above formula reduces to $\sum_{i=1}^{M} \sqrt{P_{i} P_{i}^{\prime}}$, thus we have $J^{\prime}=\left[\sum_{i=1}^{M} \sqrt{P_{i} P_{i}^{\prime}}\right]^{K}$, consequently the $K+1$ dimensional sum-integral in (5.26) reduces to following one dimensional integral:

$$
\mu_{q}(t)=C_{q}(t)+\ln \int_{-\infty}^{+\infty} e^{-\frac{(\theta+0.5 t)^{2}}{2 \sigma_{\theta}^{2}}}\left[\sum_{i=1}^{M} \sqrt{P_{i} P_{i}^{\prime}}\right]^{K} d \theta
$$

where $C_{q}(t)=\ln \left(\frac{1}{\sqrt{2 \sigma_{\theta}^{2}}}\right)-\frac{t^{2}}{8 \sigma_{\theta}^{2}}$. The function $\mu_{q}(t)$ is an even function w.r.s to $t$ (see Appendix C. 1 for details), thus the WWB can be found by a one dimensional search for maximum of $\mu_{q}(t)$ in the interval $\left[0,4 \sigma_{\theta}^{2}\right]$.

### 5.4.2 BCRLB derivation for quantized observations

Parallel to subsection 5.2.1.2 we can derive the BCRLB based on quantized observation vector $\mathbf{m}$. Assuming that the regularity condition is satisfied, i.e., $\mathbb{E}\left\{\frac{\partial \ln _{p(\mathbf{m}, \theta)}}{\partial \theta}\right\}=0$ [4], we write the Fisher information for quantized observations as following:

$$
\begin{array}{r}
F_{q}=-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m}, \theta)}{\partial^{2} \theta}\right\}= \\
-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta}\right\}-\mathbb{E}\left\{\frac{\partial^{2} \ln p(\theta)}{\partial^{2} \theta}\right\} . \tag{5.28}
\end{array}
$$

Assuming that $m_{k}$ 's conditioned on $\theta$ are independent, i.e., $\ln p(\mathbf{m} \mid \theta)=\sum_{k=1}^{K} \ln p\left(m_{k} \mid \theta\right)$, the first and second derivatives of the log-likelihood function become:

$$
\begin{aligned}
& \frac{\partial \ln p(\mathbf{m} \mid \theta)}{\partial \theta}=\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta} \\
& \frac{\partial^{2} \ln p(\mathbf{m} \mid \theta)}{\partial^{2} \theta}=\overbrace{\sum_{k=1}^{K} \frac{1}{p\left(m_{k} \mid \theta\right)} \frac{\partial^{2} p\left(m_{k} \mid \theta\right)}{\partial^{2} \theta}}^{=F_{a}} \\
& -\underbrace{\sum_{k=1}^{K} \frac{1}{p^{2}\left(m_{k} \mid \theta\right)}\left(\frac{\partial p\left(m_{k} \mid \theta\right)}{\partial \theta}\right)^{2}}_{=F_{b}}
\end{aligned}
$$

In the following, we find $\mathbb{E}\left\{F_{a}\right\}, \mathbb{E}\left\{F_{b}\right\}$. We have:

$$
\begin{aligned}
& \mathbb{E}\left\{F_{a}\right\}=K \int p(\theta) \frac{\overbrace{}^{2}(\overbrace{i=1}^{M} P_{i}(\theta))}{\partial^{2} \theta} d \theta=0 \\
& \mathbb{E}\left\{F_{b}\right\}=K \int p(\theta) \sum_{i=1}^{M} \frac{1}{P_{i}(\theta)}\left(\dot{P}_{i}(\theta)\right)^{2} d \theta
\end{aligned}
$$

where $\dot{P}_{i}(\theta)=\frac{\partial P_{i}(\theta)}{\partial \theta}$ (Deriving $\dot{P}_{i}(\theta)$ is straightforward and reduces to subtraction of two scaled standard normal PDFs). Combining all above and recalling $\theta \sim \mathcal{N}\left(0, \sigma_{\theta}^{2}\right)$, we obtain:

$$
\begin{equation*}
F_{q}=\frac{K}{\sqrt{2 \pi \sigma_{\theta}^{2}}} \sum_{i=1}^{M} \int \frac{\left(\dot{P}_{i}(\theta)\right)^{2}}{P_{i}(\theta)} e^{-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}} d \theta+\frac{1}{\sigma_{\theta}^{2}} \tag{5.29}
\end{equation*}
$$

Note that we managed to reduce the $K+1$ dimensional sum-integral in (5.28) to sum of $M$ one dimensional integral in (5.29).

### 5.5 Simulation Results

In this section we illustrate the behavior of various lower bounds derived in previous sections with numerous numerical simulations. We demonstrate the existence of the so called enhancement mode of multiplicative noise and characterize the scenarios that it can enhance the DES performance. We also probe into effect of enhancement mode on performance of well-known estimators such as MMSE, MAP, MMSE-ML, MAP-ML and compare their performance's with associated lower bounds. We consider networks of different sizes i.e., $K=10,20,50,100$, and different additive noise variances i.e., $\sigma_{n}^{2}=0.5,1,2$, to investigate the effect of those on enhancement modes. The Monte-Carlo simulations were executed on $10^{6}$ unknown and noise realizations to produce the MSE performance of different estimators.

Fig. 5.1 depicts the WWB and BCRLB vs known $\sigma_{h}^{2}$ for different values of $\sigma_{n}^{2}$. Apparently for $\sigma_{h}^{2}=0$ the WWB and BCRLB are equal (see lemma 6), and as $\sigma_{h}^{2}$ increases WWB tends to be a tighter bound (see lemma 7). The enhancement modes are marked with arrows.


Figure 5.1: WWB and BCRLB vs known $\sigma_{h}^{2}$

As can be seen the enhancement modes tend to start with smaller $\sigma_{h}^{2}$ 's with larger $\sigma_{n}^{2}$,s. It is also observable that for a fixed value of $\sigma_{h}^{2}$, more enhancement is happening for larger $\sigma_{n}^{2}$.

Fig. 5.2 depicts the WWB and BCRLB vs known $\sigma_{h}^{2}$ for different values of quantization rate $(r)$, and different additive noise variance $\sigma_{n}^{2}$. For small quantization rates, i.e., $r=1,2$, the gap between WWB and BCRLB is more than the case of $r=3,4$. Note that there is no enhancement modes for some scenarios, for instance for small quantization rates ( $r=1,2$-Fig. $5.2 \mathrm{a}, \mathrm{b}$ ).


Figure 5.2: WWB and BCRLB vs $\sigma_{h}^{2}$ for different quantization rates $(r)$ and different $\sigma_{n}^{2}$

For $r=3$, there is no enhancement mode for $\sigma_{n}^{2}=0.5$ (this is in agreement to the observation we made in Fig. 5.1, where the enhancement modes were bolder for larger $\sigma_{n}^{2}$ and weaker for smaller $\sigma_{n}^{2}$ ), but for $\sigma_{n}^{2}=1,2$ in the same case we notice enhancement modes. As $r$ increases the behaviors of bounds get alike the ones in Fig. 5.1, for instance for $r=4$ we notice the enhancement modes for all values of $\sigma_{n}^{2}$ in the figure.

Fig. 5.3 compares the behavior of WWB and BCRLB vs known $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different quantization rate values ( $r=1,2,3,4$ ) with clairvoyant benchmarks. It can be seen that how the increase in quantization rate can change the behavior of the bounds turning the bounds from increasing w.r.t $\sigma_{h}^{2}$ to decreasing w.r.t $\sigma_{h}^{2}$ (for instance compare the behavior of cases $r=1$ and $r=4)$. As $r$ increases the gap between bounds for quantized observation cases and the clairvoyant benchmark decrease (compare the case of $r=4$ and clairvoyant benchmarks). Clairvoyant bounds are indicated with (cv) in the legends.


Figure 5.3: Comparison of quantized bounds and clairvoyant benchmark, $\sigma_{n}^{2}=1$

Fig. 5.4 compares the MSE performance of MMSE and MAP estimators with WWB vs known $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network sizes of of $K=20,50,100$. For performance comparison we just kept the tighter bound (WWB) and omitted the BCRLB (see Lemma 7). As can be seen provided that number of sensors $K$, is large enough, the enhancement mode happens for these estimators. For instance for $K=100$ the enhancement mode is apparent in the figure, on the other hand there are no enhancement modes for same estimators for case of $K=20$, despite the fact that WWB suggest that there is an enhancement mode. Thus in practice we need more sensors than that of suggested by lower bounds like WWB to see the enhancement effect of multiplicative noise in estimators. Note that for $\sigma_{h}^{2}=0$ the MSE of MMSE and MAP estimators both reach the WWB bound.


Figure 5.4: Performance of MMSE and MAP estimator

For the case of unknown $\sigma_{h}^{2}$, Fig. 5.5 compares the three lower bounds of HCRLB, NB and RUB vs unknown $\sigma_{h}^{2}$ for a small network of size $K=10$ and different values of $\sigma_{n}^{2}=0.5,1,2$. Obviously RUB is tighter than NB (see Lemma 9). As can be seen HCRLB suggests existence of enhancement modes however it is not trustworthy and realistic since according to Remark 2 HCRLB is equal to BCRLB for the problem (BCRLB is representing bound for the case of known $\sigma_{h}^{2}$ ), thus for realistic assessment, we need to focus on NB and RUB which are tighter and reflect better on the case of unknown $\sigma_{h}^{2}$.


Figure 5.5: comparison of HCRLB and NB and RUB vs unknown $\sigma_{h}^{2}$, for $K=10$ and different $\sigma_{n}^{2}$ values

Fig. 5.6 compares the three lower bounds of HCRLB, NB and RUB vs unknown $\sigma_{h}^{2}$ and $\sigma_{n}^{2}=1$, for different network sizes of $K=10,50,100$. As can be seen the enhancement mode in this case happens only for large network size of $K=100$. This in contrast with the case of known $\sigma_{h}^{2}$ in Fig. 5.1 where enhancement mode happens for all network sizes (Fig. 5.7 demonstrates this phenomena with more details). Another observation in Fig. 5.6 is that for small network sizes like $K=10$ the RUB is apparently tighter than NB, however as network size increase (for instance $K=50,100$ in the Figure), the gap between RUB and NB almost vanishes.


Figure 5.6: HCRLB, NB and RUB vs unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different $K$ 's


Figure 5.7: comparison of WWB and RUB vs known and unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network size $K$

Fig. 5.7 compares the tightest bounds associated with two cases of known and unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network sizes of $K=20,50,100$. As can be seen the bounds suggest that the enhancement for the case of known $\sigma_{h}^{2}$ happens for all network sizes, however for the case of
unknown $\sigma_{h}^{2}$ the network size must be large enough to give rise to enhancement phenomena.

Fig. 5.8 compares the MSE performance of two estimation schemes of MMSE-ML and MAPML with the RUB (RUB is tighter than other two bounds HCRLB and NB, so omitted those in this figure and just kept the tightest one, see lemma 9) for a network of size $K=20$ and different additive noise variances $\sigma_{n}^{2}=0.5,1,2$. The plots are vs unknown $\sigma_{h}^{2}$. As can be seen there is no enhancement mode for this network size according to the RUB and MSE of estimators.


Figure 5.8: Performance of MMSE-ML and MAP-ML estimator, for network size $K=20$

Fig. 5.9 compares the MSE performance of MMSE-ML and MAP-ML with RUB for $\sigma_{n}^{2}=1$ and different network sizes of $K=10,50,100$. The plots are vs unknown $\sigma_{h}^{2}$. As can be seen according to the MSE performance of estimators, we notice no enhancement modes, despite the fact that according to the RUB there is enhancement mode for large networks like $K=100$. In other words for the case of unknown $\sigma_{h}^{2}$ we can not enjoy the enhancement modes with practical estimators like MMSE-ML and MAP-ML even with large network sizes. This is in contrast with the case of known $\sigma_{n}^{2}$ (look at Fig. 5.4), where provided that network size is large enough we can enjoy the enhancement mode for estimators like MMSE and MAP. Fig.5.10 clarifies this
phenomena by comparing the MSE of mentioned estimators in two cases of known and unknown $\sigma_{h}^{2}$. As can be seen according to MSE of estimators, there is no enhancement mode for case of unknown $\sigma_{h}^{2}$ (MMSE-ML, MAP-ML) even for large network sizes like $K=100$. On the other hand for case of known $\sigma_{h}^{2}$, provided that network size is large enough, the estimators can enjoy the enhancement mode of multiplicative noise, for instance look at MSE of MMSE and MAP for $K=100$.


Figure 5.9: Performance of MMSE-ML and MAP-ML estimator, for $\sigma_{n}^{2}=1$ and different network size $K=10,50,100$


Figure 5.10: comparison of MSE of various estimators for case of known and unknown $\sigma_{h}^{2}$, for $\sigma_{n}^{2}=1$ and different network size $K=10,50,100$

### 5.6 Conclusions

We considered DES of a Gaussian unknown parameter with zero mean and known variance, in a WSN where an FC is tasked with estimating the unknown parameter. Assuming that the variance of multiplicative noise is known, we derived the WWB and BCRLB for the case that full precision observations are available in the FC and also the case that only quantized version of observations are available in the FC. We analyzed the bounds and studied the effect of multiplicative noise on DES performance. The analytical and simulation results led us to the conclusion that presence of multiplicative noise in some scenarios can aid the estimation and decreases the MSE. In addition, simulation examples verifies existence of enhancement modes according to MSE metric of MMSE and MAP estimators. In sequel we dug into a situation that variance of multiplicative noise is unknown. Considering the unknown variance as a deterministic nuisance parameter, we derived the HCRLB, NB and RUB to investigate the effect of multiplicative noise. The bounds predicts
existence of enhancement modes for some rare cases, however comparing the MSE performance of MMSE-ML and MAP-ML and the bounds reveals that there is no enhancement modes of multiplicative noise with unknown variance.

## CHAPTER 6: CONCLUSION AND FUTURE WORK DIRECTIONS

In this thesis we studied the effects of several factors, including correlated additive observation noises, multiplicative observation noises, quantization schemes, communication channel qualities, network bandwidth and power constraints and fusion rules on the performance of DES in WSNs. We characterized the effects of two main contributing factors on the MSE: i) observation quality (dependent on observation noises and quantization errors) ii) communication channel quality (dependent on channel fading and additive noise).

In the following, we summarize chapters 2-5, present our main contributions and provide some ideas for future works.

### 6.1 Conclusions

In Chapter 2, we considered the DES of a Gaussian vector with a known covariance matrix and linear observation model, in which the FC is tasked with reconstructing the unknowns vector using a linear estimator. Sensors employ uniform multi-bit quantizers and BPSK modulation, and communicate with the FC over power- and bandwidth-constrained channels. We derived two closedform upper bounds on the MSE in terms of the optimization parameters (i.e., transmit power and quantization rate per sensor). Each bound consists of two terms: the first term is the MSE due to observation noises and quantization errors, and the second term is the MSE due to communication channel errors. We proposed "coupled" and "decoupled" resource allocation schemes that minimize these bounds. The "coupled" schemes utilize the iterative modified ellipsoid method to conduct a $K$-dimensional search and find the quantization rate vector, whereas the "decoupled" ones rely on a one-dimensional search to find the quantization rates. Our simulations show that
when $P_{t o t}$ and $B_{t o t}$ are not too scarce, the bounds are good approximations of the actual MSE. Through simulations, we verified the effectiveness of the proposed schemes and confirmed that their performance approaches the clairvoyant CE for large $P_{t o t}$ and $B_{t o t}\left(P_{t o t} \approx 25 \mathrm{~dB}, B_{t o t} \approx 30\right.$ bits). Our results indicate that resource allocation is affected by the sensors' observation qualities, channel gains, and by $P_{t o t}$ and $B_{t o t}$, e.g., two WSNs with identical conditions and $P_{t o t}\left(B_{t o t}\right)$ and different $B_{t o t}\left(P_{t o t}\right)$ require two different power (rate) allocations. Additionally, more quantization rate and transmit power are allotted to sensors with better observation qualities.

In Chapter 3, we derived the BCRLB for DES of a Gaussian random variable where the individual observations in sensors are separately quantized with uniform quantizers and sensors have limited sensing dynamic range. The observation model is assumed to be linear with additive Gaussian noise. In sequel we provided closed-form approximations for the BCRLB and studied the behavior of the BCRLB and corresponding approximations as quantization rates, variances of additive observation noises and sensing dynamic range of sensors vary. The simulation results corroborate the accuracy of the proposed approximations and verify that increasing the variance of the additive observation noise always degrades the estimation accuracy and increasing the quantization rates always improves the estimation accuracy. Our simulation results illustrate that for a Gaussian linear model provided that the dynamic sensing range of sensors stays larger than 3.5 times the standard deviation of the sensors' observations the information loss (in terms of increasing MSE) is negligible, in other word we can limit observations larger than 3.5 times the standard deviation of observations into a confined limited range, without noticeable degradation in the estimation performance.

In Chapter 4, we considered DES of a Gaussian source in a heterogeneous bandwidth constrained WSN, where the source is corrupted by independent multiplicative and additive observation noises, with incomplete statistical knowledge of the multiplicative noise. For uniform multi-bit quantizers, we derived the closed-form MSE expression for the LMMSE estimator at the FC, and verified
the accuracy of our derivations via simulations. For both error-free and erroneous communication channels (modeled as BSC) we proposed several rate allocation methods to (i) minimize the MSE given a network-wide bandwidth constraint, and (ii) minimize the required network-wide bandwidth given a target MSE. We also derived the BCRLB and compared the MSE performance of our proposed methods against the bound. Our results corroborate that, for low power multiplicative observation noises and adequate network-wide bandwidth, the gaps between the MSE of our proposed methods and the BCRLB are negligible, while the performance of other methods like IRA [81], and uniform is not satisfactory. Through analysis and simulations, we showed that one-bit MLE and one-bit MAP in the literature have poor performance, when the realizations of unknown parameter is large compared with the additive observation noise variances, whereas our proposed multi-bit LMMSE significantly outperforms the estimators based on one-bit quantization.

In Chapter 5, we considered DES of a Gaussian unknown parameter with zero mean and known variance in a WSN where the FC is tasked with estimating the unknown parameter. Assuming that the variance of multiplicative noise is known, we derived the WWB and BCRLB for the case that full precision observations are available in the FC and also the case that only quantized version of observations are available in the FC. We analyzed the bounds and studied the effect of multiplicative noise on the DES performance. The analytical and simulation results led us to the conclusion that the presence of multiplicative observation noise in some scenarios can improve the estimation accuracy and decrease the MSE. In addition, simulation examples based on the MSE of MMSE and MAP estimators verify the existence of enhancement modes. In sequel we studied a situation where the variance of the multiplicative observation noise is unknown. Modeling the unknown variance as a deterministic nuisance parameter, we derived the HCRLB, NB and RUB to investigate the effect of the multiplicative observation noise. The bounds predict the existence of enhancement modes for some special cases, however comparing the MSE performance of the

MMSE-ML and MAP-ML estimators and the bounds reveals that there is no enhancement modes due to multiplicative noise when its variance is unknown.

### 6.2 Future Work Directions

Based on our contributions in this thesis, one may extend the work in the following directions:

### 6.2.1 Fusion Rules based on channel outputs

In this thesis we assumed that the FC first recovers the transmitted quantization levels and then feeds these levels into the estimator to reconstruct the unknown parameter. One can extend the work to the case where the FC feeds the channel outputs directly to the estimator ${ }^{1}$. That would be interesting to quantify the MSE distortion reduction provided by using channel outputs directly to feed the estimators, as opposed to using the recovered quantization levels..

### 6.2.2 Effect of Correlated Multiplicative Noise on DES Design and Performance

In Chapters 4 and 5, we assumed the multiplicative observation noises across sensors are independent. The assumption helped us to provide a closed-form expression for the MSE of the corresponding LMMSE estimator and closed-form solutions for optimal rate allocation. That would be interesting to extend the results to the case of correlated multiplicative observation noises to assess the effect of correlation on the system performance and rate allocation scheme.

[^10]
## APPENDIX A: APPENDIX FOR CHAPTER 2

## A. 1 Finding Upper Bound on $\mathbb{E}\left\{\left(\hat{m}_{k}-m_{k}\right)^{2}\right\}$

Suppose that the bit sequence representations of $m_{k}$ and $\hat{m}_{k}$ are $\left(b_{k, 1}, \ldots, b_{k, L_{k}}\right)$ and $\left(\hat{b}_{k, 1}, \ldots, \hat{b}_{k, L_{k}}\right)$, respectively, i.e., $m_{k}=\Delta_{k}\left(0.5-2^{L_{k}-1}+\sum_{j=1}^{L_{k}} b_{k, j} 2^{L_{k}-j}\right)$ and $\hat{m}_{k}=\Delta_{k}\left(0.5-2^{L_{k}-1}+\sum_{j=1}^{L_{k}} \hat{b}_{k, j} 2^{L_{k}-j}\right)$. Therefore,

$$
\begin{aligned}
\mathbb{E}\left\{\left(\hat{m}_{k}-m_{k}\right)^{2}\right\} & =\Delta_{k}^{2} \mathbb{E}\left\{\left(\sum_{j=1}^{L_{k}} 2^{L_{k}-j}\left(b_{k, j}-\hat{b}_{k, j}\right)\right)^{2}\right\} \\
& \stackrel{(a)}{\leq} \Delta_{k}^{2} L_{k}\left(4^{L_{k}}\right) \sum_{j=1}^{L_{k}} 4^{-j} \underbrace{\mathbb{E}\left\{\left(b_{k, j}-\hat{b}_{k, j}\right)^{2}\right\}}_{=p_{e}} \\
& \stackrel{(b)}{=} p_{e} \Delta_{k}^{2} L_{k}\left(4^{L_{k}}\right) \frac{1-(1 / 4)^{L_{k}}}{3} \\
& \stackrel{(c)}{<} \frac{4 p_{e} L_{k} \tau_{k}^{2}}{3} \stackrel{(d)}{\leq} \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right) \frac{4 L_{k} \tau_{k}^{2}}{3}
\end{aligned}
$$

where ( $a$ ) comes from Cauchy's inequality $\left(\sum_{j} \alpha_{j} \beta_{j}\right)^{2} \leq\left(\sum_{j} \alpha_{j}^{2}\right)\left(\sum_{j} \beta_{j}^{2}\right)$ for arbitrary $\alpha_{j}, \beta_{j} \mathrm{~s}$ and the fact that $\left(b_{k, j}-\hat{b}_{k, j}\right)^{2}$ is a Bernoulli random variable with success probability $p_{e},(b)$ is due to the sum of a geometric series, $(c)$ is found using the definition of $\Delta_{k}$ and $1-(1 / 4)^{L_{k}}<1$, and $(d)$ is obtained because $p_{e}=\mathcal{Q}\left(\sqrt{\frac{2 \gamma_{k} P_{k}}{L_{k}}}\right)$ and $\mathcal{Q}(x) \leq \exp \left(-\frac{x^{2}}{2}\right)$.

## A. 2 Properties of $\mathcal{D}_{2}^{u p b}$ and $\mathcal{D}_{2}^{u u p b}$

One can verify the following:

$$
\begin{aligned}
\frac{\partial \mathcal{D}_{2}^{u p b}}{\partial P_{k}} & =-\alpha_{k} \gamma_{k} \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right) \leq 0, \\
\frac{\partial^{2} \mathcal{D}_{2}^{u p b}}{\partial P_{k} \partial P_{l}} & = \begin{cases}0, & \text { if } k \neq l \\
\frac{\alpha_{k} \gamma_{k}^{2}}{L_{k}} \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right) \geq 0, & \text { if } k=l\end{cases}
\end{aligned}
$$

These facts imply that $\mathcal{D}_{2}^{u p b}$ is a decreasing function of $P_{k} \mathrm{~s}$ and $P_{t o t}$ as well as a jointly convex function of $P_{k} \mathrm{~s}$. Similarly,

$$
\begin{aligned}
\frac{\partial \mathcal{D}_{2}^{u u p b}}{\partial P_{k}} & =\tilde{\lambda}\left[\frac{\partial \boldsymbol{\mathcal { M }}^{\prime}}{\partial P_{k}}\right]_{k, k}=-\tilde{\lambda}\left(\frac{4 \tau_{k}^{2} \gamma_{k}}{3}\right) \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right) \leq 0, \forall k \\
\frac{\partial^{2} \mathcal{D}_{2}^{u u p b}}{\partial P_{k} \partial P_{l}} & = \begin{cases}0, & \text { if } \\
0 \neq l \\
\tilde{\lambda}\left(\frac{4 \tau_{k}^{2} \gamma_{k}^{2}}{3 L_{k}}\right) \exp \left(-\frac{\gamma_{k} P_{k}}{L_{k}}\right) \geq 0, & \text { if } k=l\end{cases}
\end{aligned}
$$

These facts imply that $\mathcal{D}_{2}^{\text {uupb }}$ is a decreasing function of $P_{k} \mathrm{~s}$ as well as a jointly convex function of $P_{k} \mathrm{~s}$.
A. 3 Properties of $\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$

One can verify the following:

$$
\begin{aligned}
& \frac{\partial\left(\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}\right)}{\partial L_{k}}=-\frac{2 \ln 2 \delta_{k} \tau_{k}^{2} 2^{L_{k}}}{3\left(2^{L_{k}}-1\right)^{3}} \leq 0, \forall k \\
& \frac{\partial^{2}\left(\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}\right)}{\partial L_{k} \partial L_{l}}= \begin{cases}0, & \text { if } \quad k \neq l \\
\frac{(\ln 2)^{2} \delta_{k} \tau_{k}^{2} 2^{L_{k}+1}\left(1+2^{L_{k}+1}\right)}{3\left(2^{L_{k-1}}\right)^{4}} \geq 0, & \text { if } \quad k=l\end{cases}
\end{aligned}
$$

These facts imply that $\sum_{k=1}^{K} \delta_{k} \sigma_{\epsilon_{k}}^{2}$ is a decreasing function of $L_{k} \mathrm{~S}$ as well as a jointly convex function of $L_{k} \mathrm{~s}$.

## A. 4 More Properties on $\mathcal{D}_{1}$ and $\mathcal{D}_{2}^{u p b}$

- $\mathcal{D}_{1}$ is a decreasing function of $B_{t o t}$ : after some mathematical manipulations, one can show that

$$
\frac{\partial \mathcal{D}_{1}}{\partial B_{t o t}}=\operatorname{tr}\left(\boldsymbol{G} \frac{\partial \boldsymbol{Q}}{\partial B_{t o t}} \boldsymbol{G}^{T}\right)=\sum_{k=1}^{K}\left[\frac{\partial \boldsymbol{Q}}{\partial B_{t o t}}\right]_{k, k}\left\|\boldsymbol{g}_{k}\right\|^{2} \leq 0
$$

where after substituting (2.33) into $\boldsymbol{Q}$, we find that $\left[\frac{\partial \boldsymbol{Q}}{\partial B_{t o t}}\right]_{k, k}=-\frac{2 \ln 2 \tau_{k}^{2} 2^{L_{k}}}{3\left(2^{L_{k}}-1\right)^{3} K} \leq 0$. Similarly, one can show that $\frac{\partial \mathcal{D}_{p}^{u p b}}{\partial B_{t o t}} \leq 0$. As $B_{\text {tot }} \rightarrow 0,[\boldsymbol{Q}]_{k, k} \rightarrow \infty$, i.e., all eigenvalues of $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)$ go to infinity. Consequently, due to Weyl's inequality [140], all eigenvalues of $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ go to zero. Because $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ is a diagonalizable matrix, this means that $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ goes to an all-zero matrix and $\mathcal{D}_{1} \rightarrow \operatorname{tr}\left(\mathcal{C}_{\theta}\right)$. However, as $B_{\text {tot }} \rightarrow \infty,[\boldsymbol{Q}]_{k, k} \rightarrow 0$, and thus, $\mathcal{D}_{1} \rightarrow d_{0}=\operatorname{tr}\left(\mathcal{C}_{\theta}\right)-\operatorname{tr}\left(\mathcal{C}_{\boldsymbol{x} \theta}^{T} \mathcal{C}_{\boldsymbol{x}}^{-1} \mathcal{C}_{\boldsymbol{x} \theta}\right)$.

- $\mathcal{D}_{2}^{u p b}$ is an increasing function of $B_{t o t}$ : after some mathematical manipulations, one can verify that:

$$
\begin{equation*}
\frac{\partial \mathcal{D}_{2}^{u p b}}{\partial B_{t o t}}=\operatorname{tr}\left(\boldsymbol{G} \frac{\partial \boldsymbol{\mathcal { M }}^{\prime}}{\partial B_{t o t}} \boldsymbol{G}^{T}\right)+\operatorname{tr}\left(\boldsymbol{G}\left(\mathbf{B}+\mathbf{B}^{T}\right) \boldsymbol{G}^{T}\right), \tag{A.1}
\end{equation*}
$$

where $\mathbf{B}=-\left(\frac{\partial \boldsymbol{Q}}{\partial B_{\text {tot }}}\right)\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1} \boldsymbol{\mathcal { M }}^{\prime}$. Substituting (2.33) into $\boldsymbol{\mathcal { M }}^{\prime}$, we find that $\left[\frac{\partial \boldsymbol{\mathcal { M }}^{\prime}}{\partial B_{\text {tot }}}\right]_{k, k}=\frac{\partial u_{k}}{\partial B_{\text {tot }}}=$ $\frac{4 \tau_{k}^{2}}{3 K} \exp \left(\frac{-\gamma_{k} P_{k}}{L_{k}}\right)\left[1+\frac{\gamma_{k} P_{k}}{L_{k}}\right] \geq 0$. Hence, $\boldsymbol{G} \frac{\partial \mathcal{M}^{\prime}}{\partial B_{t o t}} \boldsymbol{G}^{T} \succeq 0$, and the first term in (A.1) is non-negative. Next, we show that the second term in (A.1) is also non-negative, and hence, $\frac{\partial \mathcal{D}_{2}^{u p b}}{\partial B_{t o t}} \geq 0$. Because $\boldsymbol{G}^{T} \boldsymbol{G}$ and $\mathrm{B}+\mathbf{B}^{T}$ are symmetric matrices and $\boldsymbol{G}^{T} \boldsymbol{G} \succeq 0$, by using inequality (2) of [172] we obtain

$$
\operatorname{tr}\left(\boldsymbol{G}\left(\mathbf{B}+\mathbf{B}^{T}\right) \boldsymbol{G}^{T}\right)=\operatorname{tr}\left(\boldsymbol{G} \boldsymbol{G}^{T}\left(\mathbf{B}+\mathbf{B}^{T}\right)\right) \geq \lambda_{\min }\left(\boldsymbol{G} \boldsymbol{G}^{T}\right) \operatorname{tr}\left(\mathbf{B}+\mathbf{B}^{T}\right)
$$

where $\lambda_{\min }\left(\boldsymbol{G} \boldsymbol{G}^{T}\right) \geq 0$. Furthermore,

$$
\begin{aligned}
\operatorname{tr}\left(\mathbf{B}+\mathbf{B}^{T}\right) & =2 \operatorname{tr}(\mathbf{B})=2 \operatorname{tr}\left(\boldsymbol{\mathcal { M }}^{\prime}\left(-\frac{\partial \boldsymbol{Q}}{\partial B_{t o t}}\right)\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}\right) \\
& \stackrel{(e)}{\geq} 2 \lambda_{\min }\left(\boldsymbol{\mathcal { M }}^{\prime}\left(-\frac{\partial \boldsymbol{Q}}{\partial B_{\text {tot }}}\right)\right) \operatorname{tr}\left(\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}\right) \stackrel{(f)}{\geq} 0
\end{aligned}
$$

in which $(e),(f)$ above are found because $\boldsymbol{\mathcal { M }}^{\prime}\left(-\frac{\partial \boldsymbol{Q}}{\partial B_{t o t}}\right)$ and $\left(\mathcal{C}_{\boldsymbol{x}}+\boldsymbol{Q}\right)^{-1}$ are symmetric and positive definite matrices. Similarly, one can show that $\frac{\partial \mathcal{D}_{2}^{u u p b}}{\partial B_{\text {tot }}} \geq 0$. As $B_{\text {tot }} \rightarrow 0,\left[\mathcal{M}^{\prime}\right]_{k, k} \rightarrow 0$ and $\boldsymbol{G}, \boldsymbol{G}^{T}$ go to all-zero matrices, and therefore, $\mathcal{D}_{2}^{u p b} \rightarrow 0$. However, as $B_{\text {tot }} \rightarrow \infty,\left[\mathcal{M}^{\prime}\right]_{k, k} \rightarrow \infty$ and $\boldsymbol{G} \rightarrow \mathcal{C}_{x \theta}^{T} \mathcal{C}_{x}^{-1}$, and therefore, $\mathcal{D}_{2}^{u p b} \rightarrow \infty$.

## A. 5 Rate Discretization Algorithm

Data: $\mathbf{L}^{c}$ \& all the data required for running Algorithm "a-coupled"
Result: Discretized Quantization Rates and Ascociated Power Allocation initialization;
$\mathbf{L}=\mathbf{L}^{c}$
$\mathcal{U}=\{1, \ldots, K\}$
$\mathcal{X}=\emptyset \%$ set of sensor indices with discretized rates
while $|\mathcal{X}|<|\mathcal{U}|$ do

$$
\begin{aligned}
& \text { if } \sum_{j=1}^{K} L_{j}<B_{t o t} \text { then } \\
& m=\underset{j \notin \mathcal{X}}{\operatorname{argmin}}\left(L_{j}\right) ; \\
& \text { else } \\
& m=\underset{j \notin \mathcal{X}}{\operatorname{argmax}}\left(L_{j}\right) ;
\end{aligned}
$$

$\mathbf{L}^{u r}=\mathbf{L}+[0, \ldots, \overbrace{\left\lceil L_{m}\right\rceil-L_{m}}^{m \text { th place }}, \ldots, 0]^{T}$
$\mathbf{L}^{l r}=\mathbf{L}+[0, \ldots, \underbrace{\left\lfloor L_{m}\right\rfloor-L_{m}}_{m \text { th place }}, \ldots, 0]^{T}$
if $\mathcal{D}^{u p b}\left(\mathbf{P}, \mathbf{L}^{u r}\right) \geq \mathcal{D}^{u p b}\left(\mathbf{P}, \mathbf{L}^{l r}\right)$ then
$m$ th place(fixed)
$\mathbf{L}=\mathbf{L}+[0, \ldots, \overbrace{\left\lceil L_{m}\right\rceil-L_{m}}, \ldots, 0]^{T} ;$
else

$$
\mathbf{L}=\mathbf{L}+[0, \ldots, \underbrace{\left\lfloor L_{m}\right\rfloor-L_{m}}_{m \text { th place }(\text { fixed })}, \ldots, 0]^{T} ;
$$

end
$\mathcal{X} \leftarrow(\mathcal{X} \cup\{m\})$ $B_{t o t} \leftarrow\left(B_{t o t}-L_{m}\right)$

Apply the Algorithm "a-coupled" to reallocate power to all sensors and reallocate rates to sensors with indices in the set $(\mathcal{U}-\mathcal{X})$
end
Algorithm: Discretizing Quantization Rates

## A. 6 "a-decoupled" algorithm

Data: System Parameters defined in system model
Result: Continuoes solution for quantization rates and transmission powers initialization;
$B^{o p t}=1$

```
for \(i=1: B_{t o t}\) do
    \(B^{(i)}=i\)
    \(L_{k}=\frac{B^{(i)}}{K}+\log _{2}\left(\sqrt{\frac{\delta_{k} \tau_{k}^{2}}{\prod_{j=1}^{K} \delta_{j} \tau_{j}^{2}}}\right)\)
    \(\gamma_{k}=\frac{\left|h_{k}\right|^{2}}{2 \sigma_{\omega_{k}}^{L_{k}}}, \quad \alpha_{k}=\tau_{k}^{2}\left(4 L_{k} / 3\right) \sum_{i=1}^{q}\left([\boldsymbol{G}]_{i k}\right)^{2}\)
    \(\left.\lambda=\left(\frac{\prod_{k \neq I_{k=1}^{K}}^{{ }^{K}}\left(\gamma_{k} \alpha_{k}\right)^{\left(1 / \gamma_{k}\right)}}{e^{P_{t o t}}}\right)\right)^{\left(\frac{1}{\sum_{k \neq I_{k=1}}^{K} \frac{I}{\gamma_{k}}}\right)}\)
    \(P_{k}=\left[\frac{1}{\gamma_{k}} \log \left(\frac{\gamma_{k} \alpha_{k}}{\lambda}\right)\right]^{+}, \quad k=1, \ldots, K\)
    \(\mathcal{D}^{u p b^{(i)}}=2\left(\mathcal{D}_{1}(\mathbf{L}, \mathbf{P})+\mathcal{D}_{2}^{u p b}(\mathbf{L}, \mathbf{P})\right)\)
    if \(i>1 \quad \& \mathcal{D}^{\text {upb } b^{(i)}}<\mathcal{D}^{u p b^{(i-1)}}\) then
        \(B^{o p t}=B^{(i)}\)
    end
```

    Final rate and power allocation based on \(B^{o p t}\) :
    \(L_{k}^{o p t}=\frac{B^{o p t}}{K}+\log _{2}\left(\sqrt{\frac{\delta_{k} \tau_{k}^{2}}{\prod_{j=1}^{K} \delta_{j} \tau_{j}^{2}}}\right)\)
    \(\gamma_{k}^{o p t}=\frac{\left|h_{k}\right|^{2}}{2 \sigma_{\omega_{k}}^{2} L_{k}^{\text {opt }}}, \quad \alpha_{k}^{o p t}=\tau_{k}^{2}\left(4 L_{k}^{o p t} / 3\right) \sum_{i=1}^{q}\left([\boldsymbol{G}]_{i k}\right)^{2}\)
    \(\lambda^{\text {opt }}=\left(\frac{\prod_{k \neq \mathcal{I}_{k=1}}^{K}\left(\gamma_{k}^{o p t} \alpha_{k}^{o p t}\right)^{\left(1 / \gamma_{k}^{o p t}\right)}}{e^{P_{\text {Pot }}}}\right)\left(\frac{1}{\sum_{k \notin \mathcal{I}_{k=1}^{K}} \frac{1}{\gamma_{k}^{o p t}}}\right)\)
    \(P_{k}^{o p t}=\left[\frac{1}{\gamma_{k}^{\text {opt }}} \log \left(\frac{\gamma_{k}^{\text {opt }} \alpha_{k}^{\text {opt }}}{\lambda^{\text {opt }}}\right)\right]^{+}, \quad k=1, \ldots, K\)
    end

Algorithm: A-Decoupled

## A. 7 Ellipsoid Method is Guaranteed to Converge

We consider Ellipsoid method iterative updating formula (2.24) in general form of $\mathbf{x}^{(i+1)}=\mathbf{x}^{(i)}+$ $p S^{(i)}$, where $\mathbf{x}^{(i)}$ is the optimization variable in iteration $i, p$ is the step size and $S^{(i)}$ is the 'update direction' of algorithm. We also assume that the objective function [of optimization] is $F(\mathbf{x})$. Ac-
cording to (prop .. ref bertsekas) if objective $F(\mathbf{x})$ and $S^{(i)}$ satisfy following conditions, then the Ellipsoid method will converge to KKT conditions/stationary point of problem.

1) $F(\mathrm{x}) \geq 0$
2) $\nabla F$ must be Lipschitz Continuous. i.e $\|\nabla F(\mathbf{x})-\nabla F(\mathbf{y})\| \leq K_{1}\|\mathbf{x}-\mathbf{y}\| \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n}$
3) $\left\|S^{(i)}\right\| \geq K_{2}\left\|\nabla F\left(\mathbf{x}^{(i)}\right)\right\|$
4) $S^{(i)^{T}} \nabla F\left(\mathbf{x}^{(i)}\right) \leq-K_{3}\left\|S^{(i)}\right\|^{2}$

For condition 1, it is obvious that our objective function $\mathcal{D}^{u p b}$ is always nonnegative because it is an upper bound on MSE metric. For the Ellipsoid method we have $S^{(i)}=-\mathbf{A}^{(i)} \tilde{\nabla}^{(i)}=$ $-\left(\frac{\mathbf{A}^{(i)}}{\sqrt{\nabla^{(i) T} \mathbf{A}^{(i)} \nabla^{(i)}}}\right) \nabla^{(i)}$. For ease of representation we introduce $z^{(i)}=\sqrt{\nabla^{(i) T} \mathbf{A}^{(i)} \nabla^{(i)}}$. The matrix $\mathbf{A}^{(i)}$ which describes the ellipsoid in each iteration is a positive definite (p.d) matrix, so $\left(\mathbf{A}^{(i)}\right)^{-1}$ is also a p.d matrix. Suppose that we choose a positive $K_{3}$ in a way such that $K_{3} \leq \lambda_{\min }\left(\left(\mathbf{A}^{(i)}\right)^{-1}\right) \forall i$ then we have $\left(\mathbf{A}^{(i)}\right)^{-1}-K_{3} \mathbf{I} \geq \mathbf{0} \forall i$, having this we can follow:

$$
\begin{equation*}
K_{3}\left\|S^{(i)}\right\|^{2} \leq\left(S^{(i)}\right)^{T}\left(\mathbf{A}^{(i)}\right)^{-1} S^{(i)}=\frac{-\left(S^{(i)}\right)^{T} \nabla F}{z^{(i)}} \tag{A.2}
\end{equation*}
$$

If we choose $K_{3}^{\prime}=K_{3} \inf \left(z^{(i)}\right),\left(K_{3}^{\prime} \geq 0\right.$ because $z^{(i)} \geq 0$ and $\left.K_{3} \geq 0\right)$ then we have $\left(S^{(i)}\right)^{T} \nabla F \leq-K_{3}^{\prime}\left\|S^{(i)}\right\|^{2}$, so condition 4 is satisfied.
For condition 5 we can easily follow that $z^{(i)}\left(\mathbf{A}^{(i)}\right)^{-1} S^{(i)}=-\nabla F$ and taking norm from both sides we have the inequality $\left.\left|z^{(i)}\right| \| \mathbf{A}^{(i)}\right)^{-1}\| \| S^{(i)}\|\geq\| \nabla F \|$, having the inequality and choos$\operatorname{ing} K_{2}=\frac{1}{\left.\operatorname{Sup}\left(\left|z^{(i)}\right|\right) \| \mathbf{A}^{(i)}\right)^{-1} \|}$, it can easily followed $\left\|S^{(i)}\right\| \geq K_{2}\|\nabla F\|$.
on condition 2) : For a twice continuously differentiable function (It is easy to verify that the objective function in (2.18) is twice continuously differentiable function.), Lipschitz Continuity condition can be replaced with following simple condition: The set $\{\mathbf{x} \mid F(\mathbf{x}) \leq c\}$ must be bounded for every $c \in \mathbb{R}$. If we assume that $D_{\text {min }}=\operatorname{minimum}\left(\mathcal{D}^{u p b}\left(\left\{L_{k}\right\}_{k=1}^{K}\right)\right)$ given $\left\{P_{k}\right\}_{k=1}^{K}$. Then the set $\left\{\mathbf{L} \mid \mathcal{D}(\mathbf{L})_{1}+\mathcal{D}(\mathbf{L})_{2}^{u p b} \leq c, D_{\text {min }} \leq c \leq \infty\right\}$ is bounded. For proving this lets assume that the
set is not bounded, so at least one of $L_{i}$ goes to infinity ( $L_{i} \rightarrow \pm \infty$ ), having this we can easily show that $\mathcal{D}_{2}^{u p b} \rightarrow \pm \infty$ (because $\mathcal{D}_{2}{ }^{u p b}=\sum_{k=1}^{K} u_{k}\left\|g_{k}\right\|^{2}$ and $u_{i} \rightarrow \pm \infty$ when $L_{i} \rightarrow \pm \infty$, while $\left.\left\|g_{k}\right\|^{2}>0\right)$ and that is a contradiction which leaves us with the result that the set is bounded.

## APPENDIX B: APPENDIX FOR CHAPTER 4

## B. 1 On one-bit MLE

For a homogeneous network, the likelihood function corresponding to one bit quantization is $L(\theta)=\sum_{k=1}^{K} \ln \Phi\left(y_{k} \frac{\theta}{v(\theta)}\right)$, where $v(\theta)=\sqrt{\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}}$ and $y_{k}=\operatorname{sign}\left(h_{k} \theta+n_{k}\right)$. Therefore, the one-bit MLE is $\hat{\theta}_{M L}=\underset{\theta \in\left[-\theta_{\max }, \theta_{\max }\right]}{\operatorname{argmax}} L(\theta)$. We have:

$$
\begin{equation*}
\dot{L}(\theta)=\frac{\partial L}{\partial \theta}=\overbrace{\frac{\sigma_{n}^{2}}{v^{2}(\theta)} \phi\left(\frac{\theta}{v(\theta)}\right)}^{=l_{a}(\theta)} \overbrace{\left[\frac{N^{+}}{\Phi\left(\frac{\theta}{v(\theta)}\right)}-\frac{N^{-}}{\Phi\left(\frac{-\theta}{v(\theta)}\right)}\right]}^{=l_{b}(\theta)}, \tag{B.1}
\end{equation*}
$$

where $N^{+}$and $N^{-}$are number of +1 's and -1 's in all $y_{k}$ 's, respectively, and $N^{+}+N^{-}=K$. Since $l_{a}(\theta)>0$, the solution of $l_{b}(\theta)=0$ is $\hat{\theta}_{M L}$. However, in some cases this equation has no solution in $\left[-\theta_{\max }, \theta_{\max }\right]$ and $L(\theta)$ is strictly monotonic. One instance is when both multiplicative and additive noises are very small, such that all $y_{k}$ 's are equal to +1 almost surely, i.e., $N^{+}=K, N^{-}=$ 0 , and hence $\hat{\theta}_{M L}=\theta_{\max }$. This poor performance of MLE for low power noises is unintuitive. Below, we provide an analytical explanation for this. Let $P_{s}$ denote the probability of $L(\theta)$ being strictly monotonic (i.e., the probability of $l_{b}(\theta)>0 \vee l_{b}(\theta)<0$ ). We find:

$$
\begin{align*}
P_{s} & =\left[\Phi^{K-q}\left(\frac{\theta}{v(\theta)}\right)+\Phi^{K-q}\left(\frac{-\theta}{v(\theta)}\right)\right]  \tag{B.2}\\
& \times \sum_{i=0}^{q}\left(\Phi^{i}\left(\frac{\theta}{v(\theta)}\right) \Phi^{q-i}\left(\frac{-\theta}{v(\theta)}\right)\right),
\end{align*}
$$

where $q=\left\lfloor\Phi\left(\frac{-\theta_{\max }}{v\left(\theta_{\text {max }}\right)}\right)\right\rfloor$. Interestingly, $P_{s}$ is an even function of $\frac{|\theta|}{\sigma_{n}}$ and can get very close to one, plot of $P_{s}$ versus $\frac{|\theta|}{\sigma_{n}}$ is depicted in Fig. B.1. In fact, when $\frac{|\theta|}{\sigma_{n}}$ is large enough MLE becomes independent of $y_{k}$ 's and realization of $\theta$, i.e., $\hat{\theta}_{M L}= \pm \theta_{\max }$, leading to poor estimation performance. The larger is the dynamic range of $\theta$ (compared with $\sigma_{n}^{2}$ ), the larger is the estimation error of one-bit MLE.


Figure B.1: $P_{s}$ with $\sigma_{n}^{2}=1, \theta_{\max }=5$

When PDF of $\theta$ is known a priori, MAP can be used instead of MLE, where $\hat{\theta}_{M A P}=\operatorname{argmax}[L(\theta)+$ $\left.\ln \left(\sigma_{\theta}^{-1} \phi\left(\frac{\theta}{\sigma_{\theta}}\right)\right)\right]$, i.e., $\hat{\theta}_{M A P}$ is the solution of $\dot{L}(\theta)=\frac{\theta}{\sigma_{\theta}^{2}}$. As we discussed before, in some cases $\dot{L}(\theta)$ becomes almost independent of $y_{k}$ 's and realization of $\theta$, leading to poor estimation performance. Figs. B.2.a and B.2.b depict Mont Carlo averages of $\hat{\theta}_{M L}$ and $\hat{\theta}_{M A P}$ versus realizations of $\theta$ for $\sigma_{n}^{2}=1, \sigma_{h}^{2}=0.1,1, K=5,50$. These figures show that as the realization of $\theta$ becomes large enough, the estimates become more inaccurate, i.e., $\hat{\theta}_{M L}$ approaches $\pm \theta_{\max }$ and $\hat{\theta}_{M A P}$ gets independent of $\theta$ realization, and is determined ${ }^{1}$ by $\sigma_{\theta}^{2}, \sigma_{h}^{2}, K$, causing severe estimation errors.

[^11]
(a) $\hat{\theta}_{M L}$ vs. realization of $\theta$ with $\theta_{\max }=5$

(b) $\hat{\theta}_{M A P}$ vs. realization of $\theta$ with $\theta_{\max }=5$

Figure B.2: $\hat{\theta}_{M L}$ and $\hat{\theta}_{M A P}$ vs. realizations of $\theta$

## B. 2 Fisher information for non Gaussian Multiplicative Noise

- Uniform distribution: Suppose $h_{k}$ 's are uniformly distributed with unit mean and variance $\sigma_{h_{k}}^{2}$, i.e $h_{k} \sim \mathcal{U}\left(a_{k}, b_{k}\right)$ with $a_{k}+b_{k}=2$ and $\left(b_{k}-a_{k}\right)^{2}=12 \sigma_{h_{k}}^{2}$. Consider $x_{k}=h_{k} \theta+n_{k}$, conditioned on $\theta$, has the following PDF:

$$
f_{x_{k} \mid \theta}(x \mid \theta)= \begin{cases}\frac{\Phi\left(\frac{x-a_{k} \theta}{\sigma n_{k}}\right)-\Phi\left(\frac{x-b_{k} \theta}{\sigma n_{k}}\right)}{\theta\left(b_{k}-a_{k}\right)}, & \text { if } \theta \neq 0 \\ \frac{1}{\sigma_{n_{k}}} \phi\left(\frac{x}{\sigma_{n_{k}}}\right), & \text { if } \theta=0\end{cases}
$$

Thus we can follow:

$$
s_{k, i}^{U}(\theta)=\mathrm{p}\left\{\zeta_{k, i} \leq x_{k} \leq \zeta_{k, i+1} \mid \theta\right\}=\int_{\zeta_{k, i}}^{\zeta_{k, i+1}} f\left(x_{k} \mid \theta\right) d x_{k}
$$

where $\zeta_{k, i}=m_{k, i}-\frac{\Delta_{k}}{2}$ and $\zeta_{k, i+1}=m_{k, i}+\frac{\Delta_{k}}{2}$. Using normal integrals in [173] we can write $s_{k, i}^{U}(\theta)$ as following:

$$
s_{k, i}^{U}(\theta)= \begin{cases}\frac{\sigma_{n_{k}}}{\theta\left(b_{k}-a_{k}\right)}\left[\Psi\left(\zeta_{k, i+1}^{a}, \zeta_{k, i}^{a}\right)-\Psi\left(\zeta_{k, i+1}^{b}, \zeta_{k, i}^{b}\right)+\right. \\ \left.\psi\left(\zeta_{k, i+1}^{a}, \zeta_{k, i+1}^{b}\right)-\psi\left(\zeta_{k, i}^{a}, \zeta_{k, i}^{b}\right)\right], & \text { if } \theta \neq 0 \\ \Phi\left(\frac{\zeta_{k, i+1}}{\sigma_{n_{k}}}\right)-\Phi\left(\frac{\zeta_{k, i}}{\sigma_{n_{k}}}\right), & \text { if } \theta=0\end{cases}
$$

where $\zeta_{k, i}^{a}(\theta)=\zeta_{k, i}^{a}=\frac{\zeta_{k, i}-a_{k} \theta}{\sigma_{n_{k}}}, \zeta_{k, i}^{b}(\theta)=\zeta_{k, i}^{b}=\frac{\zeta_{k, i}-b_{k} \theta}{\sigma_{n_{k}}}$ and $\Psi(y, z)=y \Phi(y)-z \Phi(z)$ and $\psi(y, z)=\phi(y)-\phi(z)$.

- Laplace distribution: Suppose $h_{k}$ 's have Laplace distributions with unit mean and scale parameter of $b$ (variance of $h_{k}$ is $\sigma_{h_{k}}^{2}=2 b^{2}$ ), i.e $f_{h_{k}}(x)=\frac{1}{2 b} e^{\frac{-|x-1|}{b}}$, consider $x_{k}=h_{k} \theta+n_{k}$, conditioned on
$\theta$, has the following PDF:

$$
f_{x_{k} \mid \theta}(x \mid \theta)=\left\{\begin{array}{ll}
\frac{e^{\frac{x-\theta}{\theta \mid b}} \Phi\left(-\frac{x-\theta}{\sigma_{n_{k}}}+\nu_{k}\right)+e^{-\frac{x-\theta}{\theta \mid b}} \Phi\left(\frac{x-\theta}{\sigma_{n_{k}}}+\nu_{k}\right)}{2|\theta| b e^{-\frac{\nu_{k}^{2}}{2}},}, & \text { if } \theta \neq 0 \\
\frac{1}{\sigma_{n_{k}}} \phi\left(\frac{x}{\sigma_{n_{k}}}\right), & \text { if } \theta=0
\end{array},\right.
$$

where $\nu_{k}=-\frac{\sigma_{n_{k}}}{|\theta| b}$. Using the integrals in [173] we can obtain:

$$
\begin{gathered}
s_{k, i}^{L}(\theta)=\mathrm{p}\left\{\zeta_{k, i} \leq x_{k} \leq \zeta_{k, i+1} \mid \theta\right\}= \\
\begin{cases}\frac{\Omega\left(\zeta_{k, i}^{\prime}, \zeta_{k, i+1}^{\prime}, \nu_{k}\right)-\Omega\left(\zeta_{k, i}^{\prime},-\zeta_{k, i+1}^{\prime}, \nu_{k}\right)}{2 e^{-\frac{\nu_{k}^{2}}{2}}+} \\
\Phi\left(\zeta_{k, i+1}^{\prime}\right)-\Phi\left(\zeta_{k, i}^{\prime}\right), & \text { if } \quad \theta \neq 0 \\
\Phi\left(\frac{\zeta_{k, i+1}}{\sigma_{n_{k}}}\right)-\Phi\left(\frac{\zeta_{k, i}}{\sigma_{n_{k}}}\right), & \text { if } \theta=0\end{cases}
\end{gathered}
$$

where $\Omega(y, z, t)=e^{t y} \Phi(y+t)-e^{t z} \Phi(z+t)$, and $\zeta_{k, i}^{\prime}=\frac{\zeta_{k, i}-\theta}{\sigma_{n_{k}}}$.

## APPENDIX C: APPENDIX FOR CHAPTER 5

C. $1 \quad \mu(t)$ and $\mu_{q}(t)$ are even functions of $t$

- on $\mu(t)$ :

Note that $C(t)$ in (5.7) is an even function. For the integral in (5.7) we substitute $\theta=-\theta^{\prime}$ and replace $t$ with $-t$, then we have: $\sigma_{0}^{\prime 2}=\theta^{\prime 2} \sigma_{h}^{2}+\sigma_{n}^{2}, \quad \sigma_{t}^{\prime 2}=\left(-\theta^{\prime}-t\right)^{2} \sigma_{h}^{2}+\sigma_{n}^{2}=\left(\theta^{\prime}+t\right)^{2} \sigma_{h}^{2}+\sigma_{n}^{2}$ and

$$
\begin{align*}
& \int_{-\infty}^{\infty}\left(\frac{\sqrt{\sigma_{0}^{\prime 2} \sigma_{t}^{2}}}{\sigma_{0}^{\prime 2}+\sigma_{t}^{\prime 2}}\right)^{\frac{K}{2}} e^{-\frac{\left(-\theta^{\prime}-0.5 t\right)^{2}}{2 \sigma_{\theta}^{2}}-\frac{K t^{2}}{4\left(\sigma_{0}^{\prime 2}+\sigma_{t}^{\prime 2}\right)}} d \theta^{\prime}= \\
& \int_{-\infty}^{\infty}\left(\frac{\sqrt{\sigma_{0}^{2} \sigma_{t}^{2}}}{\sigma_{0}^{2}+\sigma_{t}^{2}}\right)^{\frac{K}{2}} e^{-\frac{(\theta+0.5 t)^{2}}{2 \sigma_{\theta}^{2}}-\frac{K t^{2}}{4\left(\sigma_{0}^{2}+\sigma_{t}^{2}\right)}} d \theta \tag{C.1}
\end{align*}
$$

the equality in (C.1) is axiomatic with substituting $\theta^{\prime}=\theta$.

- on $\mu_{q}(t)$ :

Noting that: $(i) \Phi(-\alpha)=1-\Phi(\alpha)$, (ii) quantizer boundaries are symmetric around zero, i.e., $\zeta_{i}=-\zeta_{M+2-i}$ for $i \in\{1, \ldots, M+1\}$. it is easy to verify following:

$$
\left.\left.\left[P_{i}(\theta) P_{i}^{\prime}\right]\right]_{\theta=-\theta^{\prime}}^{t=-t}=\left[P_{i}(\theta) P_{i}(\theta+t)\right]\right]_{\theta=-\theta^{\prime}}^{t=-t}=P_{i}\left(\theta^{\prime}\right) P_{i}^{\prime},
$$

thus replacing $t$ with $-t$ and changing the integration variable from $\theta$ to $-\theta^{\prime}$ in (5.28) one can verify $\mu_{q}(-t)=\mu_{q}(t)$.

## C. 2 Proof for lemma 8

The MMSE estimator of $\theta$ can be written as following:

$$
\begin{equation*}
\hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)=\frac{q\left(\mathbf{x}, \sigma_{h}^{2}\right)}{p\left(\mathbf{x}, \sigma_{h}^{2}\right)} \tag{C.2}
\end{equation*}
$$

where $q\left(\mathbf{x}, \sigma_{h}^{2}\right)$ and $p\left(\mathbf{x}, \sigma_{h}^{2}\right)$ (the joint distribution observation vector $\mathbf{x}$ ), are formulated as following:

$$
\begin{gathered}
p\left(\mathbf{x}, \sigma_{h}^{2}\right)=\int_{-\infty}^{+\infty} p(\mathbf{x} \mid \theta) p(\theta) d \theta= \\
\frac{1}{\sqrt{(2 \pi)^{K+1} \sigma_{\theta}^{2}}} \int_{-\infty}^{+\infty} \frac{e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}}{v\left(\theta^{2}\right)^{\frac{K}{2}}} d \theta,
\end{gathered}
$$

$$
\begin{gathered}
q\left(\mathbf{x}, \sigma_{h}^{2}\right)=\int_{-\infty}^{+\infty} \theta p(\mathbf{x} \mid \theta) p(\theta) d \theta= \\
\frac{1}{\sqrt{(2 \pi)^{K+1} \sigma_{\theta}^{2}}} \int_{-\infty}^{+\infty} \frac{\theta e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}}{v\left(\theta^{2}\right)^{\frac{K}{2}}} d \theta,
\end{gathered}
$$

where $v\left(\theta^{2}\right)=\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}$. The partial derivative of the estimator in (C.2) with respect to $\sigma_{h}^{2}$ can be written as following:

$$
\begin{equation*}
\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}=\frac{\frac{\partial q\left(\mathbf{x}, \sigma_{2}^{2}\right)}{\partial \sigma_{h}^{2}} p\left(\mathbf{x}, \sigma_{h}^{2}\right)-\frac{\partial p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} q\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\left(p\left(\mathbf{x}, \sigma_{h}^{2}\right)\right)^{2}} . \tag{C.3}
\end{equation*}
$$

It is easy to derive the partial derivatives in (C.3) as following:

$$
\begin{gather*}
\frac{\partial p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}=\frac{1}{\sqrt{(2 \pi)^{K+1} \sigma_{\theta}^{2}}} \times \\
\int_{-\infty}^{+\infty} \frac{\theta^{2} e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}\left(\|\mathbf{x}-\mathbf{1} \theta\| \|^{2}-K v(\theta)\right)}{v\left(\theta^{2}\right)^{\frac{K}{2}}} d \theta \\
\frac{\partial q\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}=\frac{1}{\sqrt{(2 \pi)^{K+1} \sigma_{\theta}^{2}}} \times \\
\int_{-\infty}^{+\infty} \frac{\theta^{3} e^{-\frac{\|\mathbf{x}-1 \theta\|^{2}}{2 v\left(\theta^{2}\right)}-\frac{\theta^{2}}{2 \sigma_{\theta}^{2}}}\left(\|\mathbf{x}-\mathbf{1} \theta\|^{2}-K v(\theta)\right)}{v\left(\theta^{2}\right)^{\frac{K}{2}}} d \theta \tag{C.4}
\end{gather*}
$$

Changing the variables for integration in (C.4) as $\theta=-\theta^{\prime}$ and replacing $\mathrm{x}=-\mathrm{x}$ one can verify that :

$$
\begin{array}{r}
p\left(-\mathbf{x}, \sigma_{h}^{2}\right)=p\left(\mathbf{x}, \sigma_{h}^{2}\right), q\left(-\mathbf{x}, \sigma_{h}^{2}\right)=-q\left(\mathbf{x}, \sigma_{h}^{2}\right), \\
\frac{\partial p\left(-\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}=\frac{\partial p\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}, \quad \frac{\partial q\left(-\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}=-\frac{\partial q\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} . \tag{C.5}
\end{array}
$$

Based on (C.5) it is easy to verify that:

$$
\frac{\partial \hat{\theta}_{M M S E}\left(-\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} p\left(-\mathbf{x}, \sigma_{h}^{2}\right)=-\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} p\left(\mathbf{x}, \sigma_{h}^{2}\right)
$$

then we can continue:

$$
\begin{array}{r}
\mathbb{E}_{\mathbf{x}}\left\{\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}=\int_{\Omega} \frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} p\left(\mathbf{x}, \sigma_{h}^{2}\right) d \mathbf{x} \\
\overline{(a)}-\int_{\Omega}(-1)^{2 K} \frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}} p\left(\mathbf{x}, \sigma_{h}^{2}\right) d \mathbf{x} \tag{C.6}
\end{array}
$$

where $(a)$ is established with changing the integration variable as $\mathbf{x}=-\mathbf{x}$. Based on the equality in (C.6) we can easily conclude that $\mathbb{E}_{\mathbf{x}}\left\{\frac{\partial \hat{\theta}_{M M S E}\left(\mathbf{x}, \sigma_{h}^{2}\right)}{\partial \sigma_{h}^{2}}\right\}=0$.

## C. 3 Derivation of MLE for $\sigma_{h}^{2}$ in (5.25)

Writing the ML function of $\sigma_{h}^{2}$ for model in (5.1) and taking the derivative of that w.r.s to $\sigma_{h}^{2}$ and equaling that to zero yields following solution for MLE of $\sigma_{h}^{2}$ :

$$
\begin{equation*}
\hat{\sigma_{h M L}^{2}}=\frac{\bar{V}-\sigma_{n}^{2}}{\theta^{2}}-\frac{2 \bar{S}}{\theta}+1 \tag{C.7}
\end{equation*}
$$

doing the same for finding the MLE of $\theta$ we reach following equation, where its zero that maximize the likelihood function is the MLE of $\theta$ :

$$
\begin{equation*}
\sigma_{h}^{4} \theta^{3}+\bar{S} \sigma_{h}^{2} \theta^{2}+\left(\theta^{2} \sigma_{h}^{2}+\sigma_{n}^{2}-\bar{V} \sigma_{h}^{2}\right) \theta-\bar{S} \sigma_{n}^{2}=0 \tag{C.8}
\end{equation*}
$$

Jointly solving (C.7) and (C.8) we obtain following equation (the equation in (5.25)), where it's positive zero is the MLE for $\sigma_{h}^{2}$ :

$$
\theta^{3}-3 \bar{S} \theta^{2}+\left(2 \bar{S}^{2}+\bar{V}\right) \theta-\bar{S} \times \bar{V}=0
$$

## APPENDIX D: BIOGRAPHICAL SKETCH

The author was born in Tabriz, Iran on December 10, 1985. He received his Bachelor of Science degree from University of Tabriz, Tabriz, Iran in 2009 and his Master of Science degree from University of Tehran, Tehran, Iran in 2012; both in Electrical Engineering. In his master's program, he worked on resource allocation schemes for Orthogonal Frequency Division Multiple Access (OFDMA) networks with Quality of Service (QoS) considerations for users. In August 2013, he joined Department of Electrical Engineering and Computer Science in University of Central Florida as a PhD student under direction of Dr. Azadeh Vosoughi. His current research interests include adaptive and statistical signal processing, distributed estimation and detection.

## APPENDIX E: PUBLICATIONS

## Journal papers

- A. Sani and A. Vosoughi, "Distributed Vector Estimation for Power- and Bandwidth-Constrained Wireless Sensor Networks," in IEEE Transactions on Signal Processing, vol. 64, no. 15, pp. 38793894, Aug.1, 1 2016. doi: 10.1109/TSP.2016.2552504
- A. Sani and A. Vosoughi, "On Distributed Linear Estimation in Multiplicative Noise Environments and Observation Model Uncertainties," Submitted to IEEE Transactions on Signal Processing
- A. Sani and A. Vosoughi, "Noise Enhancement in Bayesian Distributed Estimation," Submitted to IEEE Transactions on Signal Processing
- A. Sani and A. Vosoughi, "Closed-form Approximation for Cramér-Rao Lower Bound for Distributed Estimation with Quantized Observations," Submitted to IEEE Signal Processing Letters.


## Conference papers

- A. Sani and A. Vosoughi, "Resource allocation optimization for distributed vector estimation with digital transmission," 2014 48th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2014, pp. 1463-1467. doi: 10.1109/ACSSC.2014.7094705
- A. Sani and A. Vosoughi, "Bandwidth and power constrained distributed vector estimation in wireless sensor networks," MILCOM 2015-2015 IEEE Military Communications Conference, Tampa, FL, 2015, pp. 1164-1169. doi: 10.1109/MILCOM.2015.7357603
- A. Sani and A. Vosoughi, "On Quantizer Design for Distributed Estimation in Bandwidth Constrained Networks," 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, MA, 2015, pp. 1-2. doi: 10.1109/VTCFall.2015.7391042
- A. Sani and A. Vosoughi, "Noise enhanced distributed Bayesian estimation," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 4217-4221. doi: 10.1109/ICASSP.2017.7952951


## LIST OF REFERENCES

[1] M. F. Othman and K. Shazali, "Wireless sensor network applications: A study in environment monitoring system," Procedia Engineering, vol. 41, no. Supplement C, pp. 1204 - 1210, 2012, international Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012). [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S1877705812027026
[2] M. Bhende, S. J. Wagh, and A. Utpat, "A quick survey on wireless sensor networks," in 2014 Fourth International Conference on Communication Systems and Network Technologies, April 2014, pp. 160-167.
[3] S. Manaffam and A. Jabalameli, "Rf-localize: An rfid-based localization algorithm for internet-of-things," in 2016 Annual IEEE Systems Conference (SysCon), April 2016, pp. $1-5$.
[4] S. Kay, Fundamentals of statistical signal processing: estimation theory. Prentice Hall,Upper Saddle River,NJ, 1993.
[5] Z. Zhang and T. Berger, "Estimation via compressed information," IEEE Transactions on Information Theory, vol. 34, no. 2, pp. 198-211, Mar 1988.
[6] T. S. Han and S. Amari, "Parameter estimation with multiterminal data compression," IEEE Transactions on Information Theory, vol. 41, no. 6, pp. 1802-1833, Nov 1995.
[7] __, "Statistical inference under multiterminal data compression," IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2300-2324, Oct 1998.
[8] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," IECE Trans of Japan, vol. 63, no. 1, pp. 700-706, Jan 1980.
[9] T. Flynn and R. Gray, "Encoding of correlated observations," IEEE Transactions on Information Theory, vol. 33, no. 6, pp. 773-787, Nov 1987.
[10] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," IEEE Transactions on Information Theory, vol. 45, no. 1, pp. 106-117, Jan 1999.
[11] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," IEEE Journal on Selected Areas in Communications, vol. 23, no. 4, pp. 765-775, April 2005.
[12] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," IEEE Transactions on Information Theory, vol. 42, no. 3, pp. 887-902, May 1996.
[13] H. Viswanathan and T. Berger, "The quadratic gaussian ceo problem," IEEE Transactions on Information Theory, vol. 43, no. 5, pp. 1549-1559, Sep 1997.
[14] Y. Oohama, "Rate-distortion theory for gaussian multiterminal source coding systems with several side informations at the decoder," IEEE Transactions on Information Theory, vol. 51, no. 7, pp. 2577-2593, July 2005.
[15] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the ceo problem," IEEE Journal on Selected Areas in Communications, vol. 22, no. 6, pp. 977-987, Aug 2004.
[16] H. Behroozi and M. R. Soleymani, "Distortion sum-rate performance of successive coding strategy in quadratic gaussian ceo problem," IEEE Transactions on Wireless Communications, vol. 6, no. 12, pp. 4361-4365, December 2007.
[17] J. Chen, X. Zhang, T. Berger, and S. Wicker, "Rate allocation in distributed sensor networks," in 41 st Annual Allerton Conference Communication, Control and Computing, 2003. . Proceedings., 2003.
[18] A. Vempaty, H. He, B. Chen, and P. Varshney, "On quantizer design for distributed bayesian estimation in sensor networks," Signal Processing, IEEE Transactions on, vol. 62, no. 20, pp. 5359-5369, Oct 2014.
[19] H. Chen and P. Varshney, "Performance limit for distributed estimation systems with identical one-bit quantizers," Signal Processing, IEEE Transactions on, vol. 58, no. 1, pp. 466471, Jan 2010.
[20] S. Kar, H. Chen, and P. Varshney, "Optimal identical binary quantizer design for distributed estimation," Signal Processing, IEEE Transactions on, vol. 60, no. 7, pp. 3896-3901, July 2012.
[21] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic gaussian ceo problem," in International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings., June 2004, pp. 119-.
[22] W. M. Lam and A. R. Reibman, "Source coding for decentralized estimation systems," in 29th IEEE Conference on Decision and Control, Dec 1990, pp. 2686-2691 vol.5.
[23] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," IEEE Transactions on Communications, vol. 41, no. 11, pp. 1602-1605, Nov 1993.
[24] S. Marano, V. Matta, and P. Willett, "Quantizer precision for distributed estimation in a large sensor network," Signal Processing, IEEE Transactions on, vol. 54, no. 10, pp. 4073-4078, Oct 2006.
[25] J. A. Gubner, "Distributed estimation and quantization," IEEE Transactions on Information Theory, vol. 39, no. 4, pp. 1456-1459, Jul 1993.
[26] K. Zhang and X. R. Li, "Optimal sensor data quantization for best linear unbiased estimation fusion," in 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601), vol. 3, Dec 2004, pp. 2656-2661 Vol.3.
[27] M. L. Fowler, M. Chen, and S. Binghamton, "Fisher-information-based data compression for estimation using two sensors," IEEE Transactions on Aerospace and Electronic Systems, vol. 41, no. 3, pp. 1131-1137, July 2005.
[28] M. L. Fowler and M. Chen, "Evaluating fisher information from data for task-driven data compression," in 2006 40th Annual Conference on Information Sciences and Systems, March 2006, pp. 967-972.
[29] M. Chen, M. L. Fowler, and A. Noga, "Data compression for simultaneous/sequential inference tasks in sensor networks," in 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 5, May 2006, pp. V-V.
[30] P. Venkitasubramaniam, G. Mergen, L. Tong, and A. Swami, "Quantization for distributed estimation in large scale sensor networks," in 2005 3rd International Conference on Intelligent Sensing and Information Processing, Dec 2005, pp. 121-127.
[31] P. Venkitasubramaniam, L. Tong, and A. Swami, "Score-function quantization for distributed estimation," in 2006 40th Annual Conference on Information Sciences and Systems, March 2006, pp. 369-374.
[32] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," IEEE Transactions on Information Theory, vol. 47, no. 3, pp. 978-1002, Mar 2001.
[33] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part i: Gaussian case," Signal Processing, IEEE Transactions on, vol. 54, no. 3, pp. 1131-1143, March 2006.
[34] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," Signal Processing Magazine, IEEE, vol. 23, no. 4, pp. 27-41, July 2006.
[35] H. Li and J. Fang, "Distributed adaptive quantization and estimation for wireless sensor networks," IEEE Signal Processing Letters, vol. 14, no. 10, pp. 669-672, Oct 2007.
[36] J. Fang and H. Li, "Adaptive distributed estimation of signal power from one-bit quantized data," Aerospace and Electronic Systems, IEEE Transactions on, vol. 46, no. 4, pp. 18931905, Oct 2010.
[37] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," Information Theory, IEEE Transactions on, vol. 51, no. 6, pp. 2210-2219, June 2005.
[38] ——, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," IEEE Journal on Selected Areas in Communications, vol. 23, no. 4, pp. 735-744, April 2005.
[39] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," Information Theory, IEEE Transactions on, vol. 51, no. 10, pp. 3564-3575, Oct 2005.
[40] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part ii: unknown probability density function," Signal Processing, IEEE Transactions on, vol. 54, no. 7, pp. 2784-2796, July 2006.
[41] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduceddimensionality sensor observations," IEEE Transactions on Signal Processing, vol. 55, no. 8, pp. 4284-4299, Aug 2007.
[42] O. Roy and M. Vetterli, "Dimensionality reduction for distributed estimation in the infinite dimensional regime," IEEE Transactions on Information Theory, vol. 54, no. 4, pp. 16551669, April 2008.
[43] J. Fang and H. Li, "Joint dimension assignment and compression for distributed multisensor estimation," IEEE Signal Processing Letters, vol. 15, pp. 174-177, 2008.
[44] _-, "Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios," IEEE Transactions on Signal Processing, vol. 58, no. 8, pp. 4339-4353, Aug 2010.
[45] M. Joneidi, A. Zaeemzadeh, S. Rezaeifar, M. Abavisani, and N. Rahnavard, "Lfm signal detection and estimation based on sparse representation," in 2015 49th Annual Conference on Information Sciences and Systems (CISS), March 2015, pp. 1-5.
[46] A. Sani and A. Vosoughi, "On distributed linear estimation with observation model uncertainties," CoRR, vol. abs/1709.02040, 2017. [Online]. Available: http: //arxiv.org/abs/1709.02040
[47] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: A survey," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 1, pp. 5-23, Feb 2011.
[48] K. Kotobi and S. G. Bilén, "Spectrum sharing via hybrid cognitive players evaluated by an m/d/1 queuing model," EURASIP Journal on Wireless Communications and Networking, vol. 2017, no. 1, p. 85, May 2017. [Online]. Available: https: //doi.org/10.1186/s13638-017-0871-x
[49] K. Kotobi, P. B. Mainwaring, and S. G. Bilen, "Puzzle-based auction mechanism for spectrum sharing in cognitive radio networks," in 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct 2016, pp. $1-6$.
[50] H. Yazdani and A. Vosoughi, "On cognitive radio systems with directional antennas and imperfect spectrum sensing," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 3589-3593.
[51] H. Yazdani, A. M. Rabiei, and N. C. Beaulieu, "On the benefits of mai-plus-noise whitening in th-bpsk ir-uwb systems," IEEE Transactions on Wireless Communications, vol. 13, no. 7, pp. 3690-3700, July 2014.
[52] ——, "Multiaccess interference whitening for th-bpsk ir-uwb systems," in 2012 26th Biennial Symposium on Communications (QBSC), May 2012, pp. 111-114.
[53] D. M. Najafabadi, A. A. Tadaion, and M. R. A. Sahaf, "Wideband spectrum sensing by compressed measurements," in 2012 IEEE Symposium on Computers and Communications (ISCC), July 2012, pp. 000 667-000 671.
[54] - , "Wideband spectrum sensing using compressive sampling based energy reconstruction," in 2012 35th International Conference on Telecommunications and Signal Processing (TSP), July 2012, pp. 667-670.
[55] D. M. NAJAFABADI, M. R. A. SAHAF, and A. A. TADAION, "Sparsity and block-sparsity concepts based wideband spectrum sensing," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E96.A, no. 2, pp. 573-583, 2013.
[56] S. Vahidian, E. Soleimani-Nasab, S. Assa, and M. Ahmadian-Attari, "Bidirectional af relaying with underlay spectrum sharing in cognitive radio networks," IEEE Transactions on Vehicular Technology, vol. 66, no. 3, pp. 2367-2381, March 2017.
[57] S. Vahidian, M. Najafi, M. Najafi, and F. S. Al-Qahtani, "Power allocation and cooperative diversity in two-way non-regenerative cognitive radio networks," $\operatorname{CoRR}$, vol. abs/1705.02242, 2017. [Online]. Available: http://arxiv.org/abs/1705.02242
[58] S. Hatamnia, S. Vahidian, S. Assa, B. Champagne, and M. Ahmadian-Attari, "Networkcoded two-way relaying in spectrum-sharing systems with quality-of-service requirements," IEEE Transactions on Vehicular Technology, vol. 66, no. 2, pp. 1299-1312, Feb 2017.
[59] A. Zaeemzadeh, M. Joneidi, B. Shahrasbi, and N. Rahnavard, "Missing spectrum-data recovery in cognitive radio networks using piecewise constant nonnegative matrix factorization," in MILCOM 2015-2015 IEEE Military Communications Conference, Oct 2015, pp. 238-243.
[60] M. K. Kiskani and B. H. Khalaj, "Novel power control algorithms for underlay cognitive radio networks," in 2011 21st International Conference on Systems Engineering, Aug 2011, pp. 206-211.
[61] M. Fozooni, A. Abbasfar, and A. Sani, "Spectrum sharing in heavy loaded networks," in 2012 15th International Telecommunications Network Strategy and Planning Symposium (NETWORKS), Oct 2012, pp. 1-5.
[62] A. D. Yaghjian and S. R. Best, "Impedance, bandwidth, and q of antennas," IEEE Transactions on Antennas and Propagation, vol. 53, no. 4, pp. 1298-1324, April 2005.
[63] M. Shirazi, T. Li, and X. Gong, "Reconfigurable double slot-ring antenna for bandwidth enhancement," in 2016 International Workshop on Antenna Technology (iWAT), Feb 2016, pp. 15-17.
[64] M. Shirazi and X. Gong, "Singly-polarized reconfigurable slot-ring antenna/array with fractal shapes," in 2016 IEEE International Symposium on Antennas and Propagation (APSURSI), June 2016, pp. 843-844.
[65] J.-Y. Sze and K.-L. Wong, "Slotted rectangular microstrip antenna for bandwidth enhancement," IEEE Transactions on Antennas and Propagation, vol. 48, no. 8, pp. 1149-1152, Aug 2000.
[66] Jabalameli, Amirhossein, "Characterization of a spiking neuron model via a linear approach," (2015). Electronics Theses and Dissertations. 1458.
[67] S. Marano, V. Matta, and P. Willett, "Quantizer precision for distributed estimation in a large sensor network," Signal Processing, IEEE Transactions on, vol. 54, no. 10, pp. 4073-4078, Oct 2006.
[68] S. Talarico, N. Schmid, M. Alkhweldi, and M. Valenti, "Distributed estimation of a parametric field: Algorithms and performance analysis," Signal Processing, IEEE Transactions on, vol. 62, no. 5, pp. 1041-1053, March 2014.
[69] I. Nevat, G. Peters, and I. Collings, "Random field reconstruction with quantization in wireless sensor networks," Signal Processing, IEEE Transactions on, vol. 61, no. 23, pp. 60206033, Dec 2013.
[70] I. Bahceci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," Communications, IEEE Transactions on, vol. 56, no. 7, pp. 1146-1156, July 2008.
[71] A. Behbahani, A. Eltawil, and H. Jafarkhani, "Linear decentralized estimation of correlated data for power-constrained wireless sensor networks," Signal Processing, IEEE Transactions on, vol. 60, no. 11, pp. 6003-6016, Nov 2012.
[72] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," Information Theory, IEEE Transactions on, vol. 55, no. 12, pp. 56825699, Dec 2009.
[73] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. Poor, "Estimation diversity and energy efficiency in distributed sensing," Signal Processing, IEEE Transactions on, vol. 55, no. 9, pp. 4683-4695, Sept 2007.
[74] M. Gastpar, "Uncoded transmission is exactly optimal for a simple gaussian "sensor" network," in 2007 Information Theory and Applications Workshop, Jan 2007, pp. 177-182.
[75] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," IEEE Wireless Communications, vol. 9, no. 4, pp. 8-27, Aug 2002.
[76] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Linear coherent decentralized estimation," Signal Processing, IEEE Transactions on, vol. 56, no. 2, pp. 757-770, Feb 2008.
[77] A. Behbahani, A. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise," Signal Processing, IEEE Transactions on, vol. 62, no. 21, pp. 5603-5614, Nov 2014.
[78] M. Ahmed, T. Al-Naffouri, M.-S. Alouini, and G. Turkiyyah, "The effect of correlated observations on the performance of distributed estimation," Signal Processing, IEEE Transactions on, vol. 61, no. 24, pp. 6264-6275, Dec 2013.
[79] X. Luo and G. Giannakis, "Energy-constrained optimal quantization for wireless sensor networks,", EURASIP J. Adv. Signal Process., p. 12, 2008.
[80] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," IEEE Transactions on Signal Processing, vol. 56, no. 4, pp. 1398-1410, April 2008.
[81] J. Li and G. AlRegib, "Rate-constrained distributed estimation in wireless sensor networks," Signal Processing, IEEE Transactions on, vol. 55, no. 5, pp. 1634-1643, May 2007.
[82] ——, "Distributed estimation in energy-constrained wireless sensor networks," Signal Processing, IEEE Transactions on, vol. 57, no. 10, pp. 3746-3758, Oct 2009.
[83] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," Signal Processing, IEEE Transactions on, vol. 54, no. 2, pp. 413-422, Feb 2006.
[84] M. Chaudhary and L. Vandendorpe, "Power constrained linear estimation in wireless sensor networks with correlated data and digital modulation," Signal Processing, IEEE Transactions on, vol. 60, no. 2, pp. 570-584, Feb 2012.
[85] M. Shirazi and A. Vosoughi, "Fisher information maximization for distributed vector estimation in wireless sensor networks," CoRR, vol. abs/1705.00803, 2017. [Online]. Available: http://arxiv.org/abs/1705.00803
[86] ——, "Bayesian cramer rao bound for distributed estimation of correlated data with nonlinear observation model," in 2014 48th Asilomar Conference on Signals, Systems and Computers, Nov 2014, pp. 1484-1488.
[87] _-, "Bayesian cramer rao bound for distributed vector estimation with linear observation model," in 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), Sept 2014, pp. 712-716.
[88] V. Naghshin, B. Maham, W. Saad, and M. Nickray, "Lifetime maximization in cluster-based outage-restricted wireless sensor networks," in 2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2012, pp. 369373.
[89] Hajibabaei Najafabadi, Zahra, "Impact of wireless channel uncertainty upon m-ary distributed detection systems." (2016). Electronics Theses and Dissertations. 4901.
[90] A. Carleial, "Multiple-access channels with different generalized feedback signals," IEEE Transactions on Information Theory, vol. 28, no. 6, pp. 841-850, Nov 1982.
[91] S. Hajizadeh, M. Monemizadeh, and G. A. Hodtani, "A coding theorem for the discrete memoryless compound multiple access channel with common message and generalized feedback," in 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Oct 2012, pp. 1991-1996.
[92] S. Hajizadeh and N. Devroye, "Dependence balance outer bounds for the discrete memoryless two-way multiple access broadcast channel," in 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Sept 2014, pp. 996-1003.
[93] M. Gastpar, "Uncoded transmission is exactly optimal for a simple gaussian 201c;sensor 201d; network," IEEE Transactions on Information Theory, vol. 54, no. 11, pp. 5247-5251, Nov 2008.
[94] S. Talarico, N. Schmid, M. Alkhweldi, and M. Valenti, "Distributed estimation of a parametric field: Algorithms and performance analysis," Signal Processing, IEEE Transactions on, vol. 62, no. 5, pp. 1041-1053, March 2014.
[95] G. Ardeshiri, A. Jamshidi, and A. Keshavarz-Haddad, "Performance analysis of decode and forward relay network in diffusion based molecular communication," in 2017 Iranian Conference on Electrical Engineering (ICEE), May 2017, pp. 1992-1997.
[96] - , "Optimal relay location in diffusion based molecular communications," in 2017 Iranian Conference on Communication Engineering (ICEE), May 2017, pp. 1992-1997.
[97] S. Vahidian, S. Assa, and S. Hatamnia, "Relay selection for security-constrained cooperative communication in the presence of eavesdropper's overhearing and interference," IEEE Wireless Communications Letters, vol. 4, no. 6, pp. 577-580, Dec 2015.
[98] S. Hataminia, "Performance analysis of two-way decode-and-forward relaying in the presence of co-channel interferences," IET Communications, vol. 8, pp. 3349-3356(7), December 2014. [Online]. Available: http://digital-library.theiet.org/content/journals/10. 1049/iet-com. 2014.0592
[99] J. J. Xiao and Z. Q. Luo, "Multiterminal source 2013;channel communication over an orthogonal multiple-access channel," IEEE Transactions on Information Theory, vol. 53, no. 9, pp. 3255-3264, Sept 2007.
[100] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," IEEE Transactions on Information Theory, vol. 58, no. 6, pp. 3575-3605, June 2012.
[101] S. Zhang, C. Tepedelenliolu, M. K. Banavar, and A. Spanias, "Max consensus in sensor networks: Non-linear bounded transmission and additive noise," IEEE Sensors Journal, vol. 16, no. 24, pp. 9089-9098, Dec 2016.
[102] S. Dasarathan, C. Tepedelenliolu, M. K. Banavar, and A. Spanias, "Robust consensus in the presence of impulsive channel noise," IEEE Transactions on Signal Processing, vol. 63, no. 8, pp. 2118-2129, April 2015.
[103] J. He, L. Duan, F. Hou, P. Cheng, and J. Chen, "Multiperiod scheduling for wireless sensor networks: A distributed consensus approach," IEEE Transactions on Signal Processing, vol. 63, no. 7, pp. 1651-1663, April 2015.
[104] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," IEEE Transactions on Wireless Communications, vol. 8, no. 7, pp. 3822-3832, July 2009.
[105] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," IEEE Transactions on Signal Processing, vol. 56, no. 7, pp. 2649-2661, July 2008.
[106] W. Li and C. G. Cassandras, "A minimum-power wireless sensor network self-deployment scheme," in IEEE Wireless Communications and Networking Conference, 2005, vol. 3, March 2005, pp. 1897-1902 Vol. 3.
[107] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood," IEEE Transactions on Signal Processing, vol. 56, no. 10 , pp. 5246-5257, Oct 2008.
[108] O. Dabeer and A. Karnik, "Signal parameter estimation using 1-bit dithered quantization," IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5389-5405, Dec 2006.
[109] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," IEEE Transactions on Information Theory, vol. 51, no. 6, pp. 2210-2219, June 2005.
[110] A. Sani and A. Vosoughi, "On distributed vector estimation for power and bandwidth constrained wireless sensor networks," IEEE Transactions on Signal Processing, vol. PP, no. 99, pp. 1-1, 2016.
[111] _—, "Bandwidth and power constrained distributed vector estimation in wireless sensor networks," in Military Communications Conference, MILCOM 2015-2015 IEEE, Oct 2015, pp. 1164-1169.
[112] -_, "Resource allocation optimization for distributed vector estimation with digital transmission," in Signals, Systems and Computers, 2014 48th Asilomar Conference on, Nov 2014, pp. 1463-1467.
[113] Z. Hajibabaei, A. Sani, and A. Vosoughi, "M-ary distributed detection in the presence of channel estimation error," in 2014 48th Asilomar Conference on Signals, Systems and Computers, Nov 2014, pp. 2147-2151.
[114] O. Besson, F. Vincent, P. Stoica, and A. B. Gershman, "Approximate maximum likelihood estimators for array processing in multiplicative noise environments," IEEE Transactions on Signal Processing, vol. 48, no. 9, pp. 2506-2518, Sep 2000.
[115] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," IEEE Transactions on Signal Processing, vol. 53, no. 1, pp. 168-181, Jan 2005.
[116] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," IEEE Transactions on Signal Processing, vol. 54, no. 2, pp. 601-612, Feb 2006.
[117] A. Wiesel, Y. C. Eldar, and A. Beck, "Maximum likelihood estimation in linear models with a gaussian model matrix," IEEE Signal Processing Letters, vol. 13, no. 5, pp. 292-295, May 2006.
[118] A. Wiesel, Y. C. Eldar, and A. Yeredor, "Linear regression with gaussian model uncertainty: Algorithms and bounds," IEEE Transactions on Signal Processing, vol. 56, no. 6, pp. 21942205, June 2008.
[119] Y. Tang, L. Chen, and Y. Gu, "On the performance bound of sparse estimation with sensing matrix perturbation," IEEE Transactions on Signal Processing, vol. 61, no. 17, pp. 43724386, Sept 2013.
[120] J. Zhu, X. Wang, X. Lin, and Y. Gu, "Maximum likelihood estimation from sign measurements with sensing matrix perturbation," IEEE Transactions on Signal Processing, vol. 62, no. 15, pp. 3741-3753, Aug 2014.
[121] J. Zhu, X. Lin, R. S. Blum, and Y. Gu, "Parameter estimation from quantized observations in multiplicative noise environments," IEEE Transactions on Signal Processing, vol. 63, no. 15, pp. 4037-4050, Aug 2015.
[122] F. m. c. Chapeau-Blondeau, "Noise-aided nonlinear bayesian estimation," Phys. Rev. E, vol. 66, p. 032101, Sep 2002. [Online]. Available: http://link.aps.org/doi/10.1103/ PhysRevE.66.032101
[123] G. O. Balkan and S. Gezici, "Crlb based optimal noise enhanced parameter estimation using quantized observations," IEEE Signal Processing Letters, vol. 17, no. 5, pp. 477-480, May 2010.
[124] J. Zhu, X. Wang, X. Lin, and Y. Gu, "Maximum likelihood estimation from sign measurements with sensing matrix perturbation," IEEE Transactions on Signal Processing, vol. 62, no. 15, pp. 3741-3753, Aug 2014.
[125] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," IEEE Transactions on Information Theory, vol. 34, no. 2, pp. 338-342, Mar 1988.
[126] F. Bellili, C. Elguet, S. B. Amor, S. Affes, and A. Stphenne, "Closed-form cramer-rao lower bounds for doa estimation from turbo-coded square-qam-modulated transmissions," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April 2015, pp. 3492-3496.
[127] A. Methenni and S. Affes, "Closed-form crlbs for the snr estimates from turbo-coded pamand rectangular-qam-modulated signals," in 2016 IEEE Global Communications Conference (GLOBECOM), Dec 2016, pp. 1-7.
[128] Y. Huang and Y. Hua, "Energy planning for progressive estimation in multihop sensor networks," IEEE Transactions on Signal Processing, vol. 57, no. 10, pp. 4052-4065, Oct 2009.
[129] H. L. V. Trees, Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise. Melbourne, FL, USA: Krieger Publishing Co., Inc., 1992.
[130] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locationspart i: Far-field sources," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 35, no. 3, pp. 286-299, Mar 1987.
[131] H. Messer, "The hybrid cramer-rao lower bound - from practice to theory," in Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006., July 2006, pp. 304-307.
[132] Y. G. Yatracos, "On prediction and mean squared error," Canadian Journal of Statistics, vol. 20, no. 2, pp. 187-200, 1992. [Online]. Available: http://dx.doi.org/10.2307/3315467
[133] T. K. Nayak, "On best unbiased prediction and its relationships to unbiased estimation," Journal of Statistical Planning and Inference, vol. 84, no. 1, pp. 171 - 189, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378375899001524
[134] ——, "Rao-cramer type inequalities for mean squared error of prediction," The American Statistician, vol. 56, no. 2, pp. 102-106, 2002. [Online]. Available: http://www.jstor.org/stable/3087379
[135] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters;part i: Performance bounds," IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6632-6646, Dec 2015.
[136] T. Kailath, A. Sayed, and B. Hassibi, Linear Estimation, ser. Prentice Hall Information and. Prentice Hall, 2000. [Online]. Available: https://books.google.com/books?id= zNJFAQAAIAAJ
[137] B. Widrow and I. Kollar, Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control and Communications. Cambridge Univ. Press, 2008.
[138] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the CRB," Signal Processing, IEEE Transactions on, vol. 54, no. 3, pp. 940-954, March 2006.
[139] F. Zhang and Q. Zhang, "Eigenvalue inequalities for matrix product," Automatic Control, IEEE Transactions on, vol. 51, no. 9, pp. 1506-1509, Sept 2006.
[140] R. Bhatia, Matrix Analysis, Graduate Texts in Mathematics. Springer, 1997.
[141] D. Li and X. Sun, Nonlinear Integer Programming. Springer, 2006.
[142] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," Journal of Optimization Theory and Applications, vol. 109, pp. 475-494, 42001.
[143] S. Boyd, "Convex optimization," University Lecture, 2008.
[144] A. Sani, M. M. Feghhi, and A. Abbasfar, "Discrete bit loading and power allocation for ofdma downlink with minimum rate guarantee for users," AEU - International Journal of Electronics and Communications, vol. 68, no. 7, pp. 602 - 610, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1434841114000077
[145] A. M. Davie and A. J. Stothers, "Improved bound for complexity of matrix multiplication," Proceedings of the Royal Society of Edinburgh, Section: A Mathematics, vol. 143, pp. 351369, 42013.
[146] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," Comput. Netw., vol. 45, no. 3, pp. 245-259, Jun. 2004. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2004.03.007
[147] J. O. Berger, V. D. Oliveira, and B. Sans, "Objective bayesian analysis of spatially correlated data," Journal of the American Statistical Association, vol. 96, no. 456, pp. 1361-1374, 2001. [Online]. Available: http://dx.doi.org/10.1198/016214501753382282
[148] A. Vempaty, H. He, B. Chen, and P. Varshney, "On quantizer design for distributed bayesian estimation in sensor networks," Signal Processing, IEEE Transactions on, vol. 62, no. 20, pp. 5359-5369, Oct 2014.
[149] J. . G. Proakis, Digital Communications. New York, NY, USA: Mc-Graw-Hill, 2001.
[150] A. Sani and A. Vosoughi, "On quantizer design for distributed estimation in bandwidth constrained networks," in 2015 IEEE 82nd Vehicular Technology Conference (VTC2015Fall), Sept 2015, pp. 1-2.
[151] R. Garnier and J. Taylor, Discrete Mathematics: Proofs, Structures and Applications, Third Edition, 3rd ed. Bristol, PA, USA: Taylor \& Francis, Inc., 2009.
[152] G. E. Andrews, The Theory of Partitions. Cambridge University Press, 1984, cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511608650
[153] S. Joshi and S. Boyd, "Sensor selection via convex optimization," Signal Processing, IEEE Transactions on, vol. 57, no. 2, pp. 451-462, Feb 2009.
[154] D. P. Bertsekas, Nonlinear Programming. Belmont, MA: Athena Scientific, 1999.
[155] A. Sani and A. Abbasfar, "Discrete bit-loading for ofdma downlink with qos consideration," in 2012 15th International Telecommunications Network Strategy and Planning Symposium (NETWORKS), Oct 2012, pp. 1-6.
[156] ——, "Qos-aware discrete bit loading for ofdma networks," in 2013 International Conference on Computing, Networking and Communications (ICNC), Jan 2013, pp. 994-998.
[157] A. Sani, A. Abbasfar, and M. Fozooni, "Resource allocation for ofdma downlink with qos consideration," in 2012 Fourth International Conference on Ubiquitous and Future Networks (ICUFN), July 2012, pp. 261-265.
[158] V. Pasias, D. A. Karras, and R. C. Papademetriou, "Heuristic algorithms for efficient wireless multimedia network design," in 32nd EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO'06), Aug 2006, pp. 372-378.
[159] A. Jabalameli and A. Behal, "A constrained linear approach to identify a multi-timescale adaptive threshold neuronal model," in 2015 IEEE 5th International Conference on Compиtational Advances in Bio and Medical Sciences (ICCABS), Oct 2015, pp. 1-6.
[160] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor, "A novel method for identification of copd in inspiratory and expiratory states of ct images," in 2011 1st Middle East Conference on Biomedical Engineering, Feb 2011, pp. 235-238.
[161] H. Leung, C. Seneviratne, and M. Xu, "A novel statistical model for distributed estimation in wireless sensor networks," IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3154-3164, June 2015.
[162] 2015 IEEE International Conference on Communications, ICC 2015, London, United Kingdom, June 8-12, 2015. IEEE, 2015. [Online]. Available: http://ieeexplore.ieee.org/ $\mathrm{xpl} /$ mostRecentIssue.jsp?punumber=7225357
[163] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," IEEE Journal on Selected Areas in Communications, vol. 21, no. 5, pp. 684-702, June 2003.
[164] M. K. Kiskani, Z. Wang, H. R. Sadjadpour, J. A. Oviedo, and J. J. Garcia-Luna-Aceves, "Opportunistic interference management: a new approach for multiantenna downlink cellular networks," Wireless Communications and Mobile Computing, vol. 15, no. 14, pp. 1837-1850, 2015. [Online]. Available: http://dx.doi.org/10.1002/wcm. 2465
[165] B. S. M. Mousaie, M. Soltanian, "Comsens: Exploiting pilot diversity for pervasive integration of communication and sensing in mimo-tdd-frameworks," in MILCOM, 2017.
[166] A. Sani and A. Vosoughi, "Noise enhanced distributed bayesian estimation," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 4217-4221.
[167] A. E. Taylor, "L'hospital's rule," The American Mathematical Monthly, vol. 59, no. 1, pp. 20-24, 1952. [Online]. Available: http://www.jstor.org/stable/2307183
[168] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in ofdm systems," in 1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century, vol. 2, Jul 1995, pp. 815-819 vol.2.
[169] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance," IEEE Transactions on Signal Processing, vol. 58, no. 3, pp. 1807-1820, March 2010.
[170] A. N. Mody and G. L. Stuber, "Parameter estimation for ofdm with transmit receive diversity," in IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202), vol. 2, 2001, pp. 820-824 vol.2.
[171] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters ;part ii: Estimation methods," IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6647-6658, Dec 2015.
[172] J. Baksalary and S. Puntanen, "An inequality for the trace of matrix product," Automatic Control, IEEE Transactions on, vol. 37, no. 2, pp. 239-240, Feb 1992.
[173] D. B. Owen, "A table of normal integrals," Communications in Statistics - Simulation and Computation, vol. 9, no. 4, pp. 389-419, 1980. [Online]. Available: http: //dx.doi.org/10.1080/03610918008812164


[^0]:    ${ }^{1}$ Note that wireless spectrum is a notoriously scarce resource and designing bandwidth constrained systems has always been of great interest in both system [47-61] and antenna level [62-66].

[^1]:    ${ }^{2}$ Because it is attainable.

[^2]:    ${ }^{1}$ The function $f(x)=\left(a+b x^{2}\right) \phi(c x)$ for $a, b \neq \pm \infty$ and $c>0$ is a Riemann-Integrable function.

[^3]:    ${ }^{2}$ Since $\theta$ and the additive noises are Gaussian, we can assume $x_{k}$ lies in a bounded interval with a high probability, i.e., $x_{k} \in\left[-\tau_{k}, \tau_{k}\right]$ for a reasonably large value of $\tau_{k}$, thus intuitively if sensing range of sensor is large enough there is almost no information loss due to limited/definite sensing range, we will demonstrate this intuition by analytical and simulation results.

[^4]:    ${ }^{1}$ For the general case $\mathbb{E}\left\{h_{k}\right\}=\mu_{k}$ we can scale $x_{k}$ and obtain $x_{k}^{\prime}=h_{k}^{\prime} \theta+n_{k}^{\prime}$, where $x_{k}^{\prime}=x_{k} / \mu_{k}, h_{k}^{\prime}=h_{k} / \mu_{k}$, $n_{k}^{\prime}=n_{k} / \mu_{k}, \mathbb{E}\left\{h_{k}^{\prime}\right\}=1, \operatorname{var}\left(h_{k}^{\prime}\right)=\sigma_{h_{k}}^{2} / \mu_{k}^{2}, n_{k}^{\prime} \sim \mathcal{N}\left(0, \sigma_{n_{k}}^{2} / \mu_{k}^{2}\right)$. Thus without loss of generality, we assume $\mathbb{E}\left\{h_{k}\right\}=1, \forall k$.

[^5]:    ${ }^{2}$ We assume sorted sensors throughout this work for all scenarios.

[^6]:    ${ }^{3}$ For definition of weight $w_{k}(0)$, we consider $\alpha_{k}(0)=0$.

[^7]:    ${ }^{4}$ Integer relaxation and IRA exhibit similar results, and the plots are omitted, for the sake of saving space.

[^8]:    ${ }^{1}$ The observation model in (5.1) is general and can be used for wireless communication systems with fading channel model and multiple receiver antennas [163-165], as well as sensing systems which observation model has a multiplicative sensing matrix with random perturbations.
    ${ }^{2}$ All derivations in this work can be easily re-derived for the case which the variance of additive and multiplicative noises are different across the sensors. For the general case $\mathbb{E}\left\{h_{k}\right\}=\mu_{k}$, we can scale $x_{k}$ and obtain $x_{k}^{\prime}=h_{k}^{\prime} \theta+n_{k}^{\prime}$, where $x_{k}^{\prime}=x_{k} / \mu_{k}, h_{k}^{\prime}=h_{k} / \mu_{k}, n_{k}^{\prime}=n_{k} / \mu_{k}$, with $\mathbb{E}\left\{h_{k}^{\prime}\right\}=1, \operatorname{var}\left\{h_{k}^{\prime}\right\}=\sigma_{h}^{2} / \mu_{k}^{2}, n_{k}^{\prime} \sim \mathcal{N}\left(0, \sigma_{n}^{2} / \mu_{k}^{2}\right)$. Thus without loss of generality, we assume $\mathbb{E}\left\{h_{k}\right\}=1, \forall k$

[^9]:    ${ }^{3}$ It is prevalent/common to consider the channel gain to be stochastic with deterministic statistics [135, 168, 169]. An example of modeling channel as stochastic with unknown statistics can be found in [170]. Note that modeling the unknown $\sigma_{h}^{2}$ as a random variable is another way of looking into problem, however it requires a prior full information on distribution of $\sigma_{h}^{2}$, which does not sound a rationale assumption.

[^10]:    ${ }^{1}$ For example suppose $u_{k}$ is a Pulse Amplitude Modulated (PAM) symbol, sent from sensor $k$ to the FC, where $u_{k}=m_{k} \sqrt{P_{k}}$. The received signal at the FC would be $y_{k}=h_{k} u_{k}+w_{K}$. Assuming perfect Channel State Information (CSI), the FC finds the messages as $z_{k}=\frac{y_{k}}{h_{k} \sqrt{P_{k}}}$.

[^11]:    ${ }^{1}$ In contrast to $\hat{\theta}_{M L}$, our simulations suggest, provided that $\theta_{\max }>3 \sigma_{\theta}, \hat{\theta}_{M A P}$ does not depend on $\theta_{\text {max }}$ in these cases.

