Tecnologías de la Información,
Comunicaciones y Matemática Computacional

Universidad de Valencia, Spain

Joint Optimization of Sensor Selection and Routing for Distributed Estimation in Wireless Sensor Networks

Santosh Shah

Advisor

Prof. Dr. Baltasar Beferull-Lozano

Thesis submitted for the Degree of Doctor of Philosophy (Ph.D.)

Valencia, February 2014

Dedication

This thesis is dedicated to my parents for their endless love, support, and encouragement.

I lovingly dedicate this thesis to my wife, who supported me patiently throughout this endeavor and to my daughter, who inspired me.

Resumen

Avances recientes en redes inalámbricos de sensores (WSNs, Wireless Sensor Networks) han posibilitado que pequeños sensores, baratos y con recursos limitados tanto en sensado, comunicación, como en computación, sean desplegados a gran escala. En consecuencia, las WSNs pueden ofrecer diversos servicios en importantes aplicaciones para la sociedad. Entre las varias restricciones que aparecen en el diseño de WSNs, tales como la limitación en energía disponible, procesamiento y memoria, la limitación en energía es muy importante ya que en muchas aplicaciones (ej., monitorización remota de diferentes entornos, edificios administrativos, monitoreo del hábitat, los incendios forestales, la atención sanitaria, la vigilancia del tráfico, vigilancia del campo de batalla, las reservas de vida silvestre, etc.) los sensores están alimentados por baterías, pudiendo hacer uso también de captación de energía renovables. Dado que las comunicaciones son causantes del mayor consumo energético en un nodo, la transmisión y recepción de información deben optimizarse lo máximo posible. Estas limitaciones y el diseño específico de los sensores, hacen necesario el estudio de métodos eficientes energéticamente y que reduzcan la cantidad de información a transmitir.

Motivación y Objetivos

Aunque las WSNs necesitan cubrir en muchas ocasiones una importante área geográfica, muchos eventos necesitan ser detectados y tratados localmente. Algunos de estos ejemplos son la energía capturada por sensores de energía acústica donde existe una cierta fuente acústica localizada en el espacio, detección y verificación de un foco de fuego en un bosque, sensores de dirección de llegada para localización, u otra fuente

difusiva localmente generada (ej. radiación nuclear). Intuitivamente, en estos escenarios, los nodos que están localizados lejos de la fuente observarán medidas significativamente menos informativas que los nodos cercanos a la fuente. Por lo tanto, la vida útil de la red puede ser incrementada al considerar la activación de solo un subconjunto de sensores (los más informativos) cuya información es útil y por tanto debe ser recolectada. Además, la eficiencia energética puede ser mejorada aún más al elegir la mejor estructura de enrutamiento. Es importante resaltar que la técnica utilizada más tradicional es la transmisión directa inalámbrica de las medidas desde todos los nodos seleccionados al centro de fusión de datos (nodo solicitante de la estimación global), lo cual resultas en un ineficiente uso de los recursos energéticos. Una solución factible puede ser el uso de la naturaleza multisalto de la transmisión de datos, el cual puede significativamente reducir la potencia total de transmisión, y por tanto aumentar la vida de la red. La cuantificación de la información (fusión) puede también utilizarse en un procesado intrared para ahorrar energía, ya que reduce la cantidad de información a ser reenviada en dirección al nodo centro de fusión. La asignación dinámica de bit-rate (bits por muestra) en cada nodo puede también ser empleada para reducir también el consumo total de la red. De esta manera, se puede obtener un importante ahorro energético al realizar de manera distribuida una cierta tarea de estimación optimizando el conjunto de sensores activo; la estructura de enrutamiento, y los bits por muestra para cada sensor seleccionado.

En la literatura reciente, se ha demostrado claramente que la transmisión multisalto en WSNs es más eficiente energéticamente que la transmisión directa, donde cada medida es directamente transmitida al centro de fusión de datos (MT, Measure-and-Transmit). Además, transmisión mutlisalto, en general, permite el envío de las medidas al nodo fusión de dos formas: a) cada nodo reenviar directamente la información recibida, b) cada nodo reenviar la información agregada. Puede observarse que, el fusionar las medidas en sensores intermedios ofrece una mejora en la calidad global de la estimación con coste computacional limitado. Esto nos lleva a considerar los dos esquemas siguientes:

- *Medir-y-reenviar* (MF, Measure-and-Forward): En este esquema, los nodos sensores simplemente reenvían las medidas que reciben de sus nodos sensores hijos en dirección al nodo solicitante a lo largo de la estructura de enrutamiento elegida. El nodo solicitante obtendrá por tanto la estimación final, por lo tanto, no hay estimación agregadas incrementales en los sensores intermedios.

- *Estimar-y-reenviar* (EF, Estimate-and-Forward): En este esquema, se considera un enfoque con estimación agregada secuencial en los nodos intermedios de la ruta de encaminamiento. Dada una estructura de enrutamiento, cada sensor fusiona todas las otras medidas que son recibidas de sus nodos hijos junto con la suya propia, con el objetivo de obtener una estimación agregada, y luego enviar un único flujo de la información fusionada a su nodo sensor padre en la estructura de enrutamiento elegida.

El esquema EF tiene varias ventajas interesantes respecto al esquema MF. En primer lugar, el esquema EF es más eficiente energéticamente ya que un nodo sensor activo en una ruta solo tiene que reenviar la estimación fusionada (una único paquete de información transmitir), en vez de reenviar su propia medida además de las medidas de sus nodos hijos. Además, utilizando un esquema EF, los nodos intermedios en la ruta tienen una estimación del parámetro que es mejor conforme el nodo está más cercano al nodo solicitante. La otra principal desventaja del esquema MF es que los nodos cerca del nodo solicitante pueden sobrecargarse, lo crea un efecto de cuello de botella.

Por lo tanto, dada una WSN con una cierto grafo subyacente de conectividad de red, un cierto nodo solicitante, y una fuente localizada, esta tesis considera el problema de la estimación distribuida de un parámetro, donde la potencia total disponible esta limitada, por lo tanto, y donde utilizamos el esquema EF, optimizando conjuntamente el subconjunto de sensores activos, la asignación de bit-rate en cada sensor y la estructura de enrutamiento multisalto asociada hasta el nodo solicitante. Por lo tanto, la distorsión total en la estimación es minimizada para una cierta potencia total de transmisión. Un resultado importante de este trabajo el consiste en que el algoritmo Shortes Path Tree (SPT) basado solo en coste de comunicación (SPT-CC) no es la estructura óptima de enrutamiento en general cuando se busca alcanzar un compromiso óptimo entre la distorsión de la estimación y el coste total de comunicaciones, sin importar si uno usa el esquema MF o el EF.

En nuestra estimación distribuida multisalto, mientras nos dirigimos hacia el nodo solicitante, necesitamos asignar tasas mayores de bits ya que la precisión de la estimación mejora a medida que más información se fusiona en los nodos de sensores intermedios. Por lo tanto, la asignación de tasa de bits en un sensor depende del número de saltos que existe entre dicho nodo y el nodo solicitante, de tal manera que hay una necesidad de proporcionar mayores tasas de bits al ir acercándose al nodo solicitante en la ruta de multisalto escogido. Por otro parte, la localización de la fuente que determine fenómeno estimar también influencia la asignación de bit-rate para sensor. Por ejemplo, si un sensor está cerca de la fuente (relación Señal-Ruido alto), incluso aunque existe un gran número de saltos necesarios para llegar al nodo solicitante, necesitamos asignar un bit-rate razonablemente alto. En consecuencia, hay una clara necesidad de diseñar un cuantificador adaptativo en cada sensor con el objetivo de proporcionar un apropiado bit-rate, el cual depende del compromiso entre el número de saltos y la localización de la fuente. Además, el bit-rate también depende del coste de comunicación entre cada dos sensores.

Metodología

En esta tesis, combinamos métodos de análisis teórico, diseño algoritmos iterativos inspirados en herramientas de optimización así como simulaciones por ordenador. En el caso del análisis teórico del problema mencionado anteriormente, hemos seguido la metodología estándar de estimación óptima lineal no sesgada; en otras palabras, Best Linear Unbiased Estimator (BLUE).

En particular, este trabajo de tesis se centra en el problema de optimizar conjuntamente la selección de sensores, la estructura de enrutamiento y la asignación de bit-rate para cada sensor seleccionado. En primer lugar, consideramos solamente la optimización conjunta de la selección de sensores y la estructura de enrutamiento, donde se asume una cuantificación fina, y por tanto se ignora la asignación óptima de bit-rate. En este caso, la función objetivo es lineal y las restricciones en el problema de optimización son no convexas, lo cual lleva a un problema a resolver que tiene una complejidad y alto.

En segundo lugar, tenemos en cuenta la asignación del bit-rate como una variable adicional en el primer problema, convirtiéndose en un problema de optimización no lineal no convexo. Por lo tanto, el problema de optimización conjunta se hace aún más difícil de resolver que el primera problema de optimización. La solución de este problema no convexo se aborda utilizando varios pasos de relajación convexa y resolviendo estos problemas relajados para las diferentes variables en tándem. El objetivo en ambos problemas anteriormente mencionadas es reducir al mínimo la distorsión total en la estimación bajo una cierta limitación de potencia total dada. También demostramos que nuestros problemas pertenecen a la clase de problemas NP-hard, realizando una reducción (de complejidad polinomial) de nuestro problema el problema Hamiltoniano no dirigido (UHP, Undirected Hamiltonian Path). Nuestros problemas de optimización relajados se pueden resolver a través de métodos de optimización convexa, tales como los métodos de punto interior.

Después de los análisis teóricos, los algoritmos propuestos considerados para ambos casos (cuantificación fina y cuantificación adaptativa), son simulados usando programación Matlab y el toolbox de CVX [10]. Los algoritmos propuestos son comparados, en cada caso, con los mejores algoritmos propuestos en la literatura para la asignación de recursos en WSN para estimación.

Conclusiones

En esta tesis, dada una WSN con un grafo subyacente de conectividad de red, un cierto nodo solicitante (sumidero) y una fuente localizada, hemos considerado el problema de la estimación distribuida de parámetros con donde la potencia total disponible esta limitada. Por lo tanto, para llevar a cabo un cierta tarea de estimación distribuida (por ejemplo, detección de fuego en un bosque, localización basada en dirección de llegada, estimación de cualquier otro fenómeno localizado, etc.), hemos considerado el problema, usando el esquema EF, de optimizar conjuntamente el subconjunto de sensores activas, la asignación de bit-rate y la asociada estructura de enrutamiento multisalto para enviar la información agregada hasta el nodo solicitante. De esta manera, la distorsión en la estimación total es minimizada una cierta potencia total.

La mayoría de las soluciones recientemente propuestas, intentan simplificar el problema considerando solamente la selección de un subconjunto de sensores, ignorando la optimización conjunta de la estructura de enrutamiento así como de la codificación. Sin embargo, optimizar la estructura de enrutamiento es una importante variable en el problema ya que, en general, transmitir información que está lejos del nodo solicitante es más costoso que desde un nodo cercano. La cuantificación de fuente también juega un papel importante ya que los sensores lejos de la fuente requieren menos niveles de cuantificación ya que reciben un SNR menor. A continuación resumimos nuestras principales contribuciones:

1. El problema de optimización conjunta de la selección de sensores, la estructura de enrutamiento multisalto y la asignación adaptativa de la tasa de bit (mediciones del sensor) para la estimación distribuida con un restricción en el coste total de comunicaciones, es formulado y analizado, tanto en términos de diseño de algoritmos como de análisis de complejidad, demostrando que es un problema NP-hard cuando se utiliza el esquema EF. También proporcionamos una cota inferior para la solución óptima del problema de optimización NP-hard original.

2. En primer lugar, consideramos el problema de optimización conjunta de la selección de los sensores y de la estructura de enrutamiento multisalto asumiendo que se dispone de una cuantificación fina[1] para cada medición de los sensores. A continuación, presentamos un Algoritmo que llamamos FTRA (Fixed-Tree Relaxation-based Algorithm) que consiste en una relajación de nuestro problema de optimización original, y que desacopla la elección de la estructura de enrutamiento de la selección de sensores activos.

3. A continuación, también diseñamos un nuevo y eficiente algoritmo iterativo distribuido que llamamos IDA (Iterative Distributed Algorithm), que optimiza de forma conjunta a nivel local y distribuida la selección de sensores y la estructura de enrutamiento de saltos múltiples. También demostramos experimentalmente que nuestro IDA genera una solución que está cerca de la solución óptima al problema NP-hard original, haciéndose uso de la cota anteriormente obtenida.

4. En segundo lugar, hemos considerado la asignación de tasa de bit como una variable adicional al anterior problema la optimización, en un problema de optimización no lineal y no convexo resultando

[1]En este contexto, la cuantificación fina significa que se asigna una cantidad de bits de representación de la medida tal que el error de la cuantificación en despreciable.

en un problema todavía mas complejo de resolver, y por tanto NP-Hard también.

5. Para este segundo problema de optimización, hemos desarrollado dos algoritmos: a) Algoritmo de Cuantificación Adaptativa basado en árbol Fijo (FTR-AQ, Fixed-Tree Relaxation-based Adaptive Quantization), y b) Algoritmo de Cuantificación Adaptativa basado en Optimización Local (LO-AQ, Local Optimization-based Adaptive Quantization). LO-AQ proporciona una estimación más precisa para la misma potencia total dada, aunque esto implica una complejidad computacional adicional en cada nodo.

6. Por último, comparamos nuestros algoritmos con los otros mejores trabajos relacionados presentados previamente en la literatura [47; 61; 84; 122], mostrando claramente un rendimiento superior en términos de distorsión en la estimación para la misma potencia total dada.

Acknowledgement

There are numerous people without whom this dissertation would not have been possible, and to whom I am greatly obligated. I would like to acknowledge their support, patience, and kind efforts.

First and foremost, I would like to thank my supervisor, Professor Baltasar Beferull-Lozano, who was abundantly helpful and offered invaluable assistance, support, and guidance throughout the tenure of my research work. I would like to thank him for patiently listening and providing advice on problems I occasionally brought before him as a student that were related to both non-academic as well as research matters. I also would like to thank him for his support in providing funding to make this research project possible. Without his guidance and persistent help this dissertation would not have been possible.

I am extremely grateful to my parents for their love, prayers, and caring for educating and preparing me for my future. Then, I wish to express my love and gratitude to my adored wife Vandana. Without her understanding, inspiration, and endless love, throughout the tenure of my studies, I would not be able to achieve what I have accomplished. I wish to acknowledge most importantly her sacrificing work as the mother of my daughter Saesha that I am most grateful for. There is no better cure for a tough day at work than the smiles and cheers I get from Saesha when I come home from work.

In addition, a thank you to Professor Carmen Botella Mascarell, she has been a valuable interdisciplinary resource to me.

Loving thanks to my colleagues and friends in the Group of Information and Communication Systems: Eugenio Celada, César Asensio, Huang Xiaoge, Ibrahim Khalife, Gustavo Hernández, Fernando Camaró Nogues, Daniel Alonso, and Vinay Prasad. I would like to thank all my colleagues for the valuable group/personal discussion and for all the fun we have had throughout the tenure of my research work. I would like to say a special thanks to Eugenio, Ibrahim, and Fernando, who helped me to face the linguistic problem upon my first arrival in Spain.

I would like to express special thanks to Professor Juan José Pérez Solano and to Professor Santiago Felici for being such good company and for all I have learnt from them.

I am thankful to the non-teaching staff for their cooperation and timely help. I wish to thank the Universidad de Valencia for providing the necessary infrastructure and support to undertake this research work.

Finally, I thank each and everyone who supported directly or indirectly in completing this thesis.

This work was supported by the Spanish MEC grants TEC2010-19545-C04-04 "COSIMA", CONSOLIDER-INGENIO 2010 CSD2008-00010, "COMONSENS", the European STREP project "HYDROBION"-'ETS" grant no. 287613 within the FP7 Framework Programme, and by a Telefonica Chair.

SANTOSH SHAH

Publications

Journal Articles

1. Shah, S.; Beferull-Lozano, B., "Joint Sensor Selection and Multi-hop Routing for Distributed Estimation in Ad-hoc Wireless Sensor Networks," *Signal Processing, IEEE Transactions on*, vol. 61, no. 24, pp. 6355-6370, December 15, 2013. (Impact factor: 2.813, Q1, JCR 2013). doi: 10.1109/TSP.2013.2284486

2. Shah, S.; Beferull-Lozano, B., "Energy-efficient Multihop Progressive Estimation and Adaptive Quantization for Ad-hoc Wireless Sensor Networks," *Signal Processing, IEEE Transactions on*, 2014. Submitted, (Impact factor: 2.813, Q1, JCR 2013).

3. Shah, S.; Beferull-Lozano, B., "Joint Sensor Selection and Routing for Power-efficient Tracking and Estimation of Moving Target Sources," *Signal Processing, IEEE Transactions on*, 2014. To be submitted, (Impact factor: 2.813, Q1, JCR 2013).

CORE-Conference Proceedings

1. Shah, S.; Beferull-Lozano, B., "Adaptive Quantization for Multihop Progressive Estimation in Wireless Sensor Networks", *21st European Signal Processing Conference 2013*, EUSIPCO-2013, Marrakech, Morocco, September 2013.

2. Shah, S.; Beferull-Lozano, B., "In-Network Iterative Distributed Estimation for Power-Constrained Wireless Sensor Networks," *Distributed Computing in Sensor Systems (DCOSS), 2012 IEEE 8th International Conference on*, pp. 239-246, 16-18 May 2012. (*Best Paper Award*). doi: 10.1109/DCOSS.2012.18

3. Shah, S.; Beferull-Lozano, B., "Power-Aware Joint Sensor Selection and Routing for Distributed Estimation: A Convex Optimization Approach," *Distributed Computing in Sensor Systems (DCOSS), 2012 IEEE 8th International Conference on*, pp. 230-238, 16-18 May 2012. doi: 10.1109/DCOSS.2012.19

4. Shah, S.; Beferull-Lozano, B., "In-Network Local Distributed Estimation for Power-Constrained Wireless Sensor Networks," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1-5, 6-9 May 2012. doi: 10.1109/VETECS.2012.6240130

Notice that all these conference papers are classified as CORE conferences (http://www.core.edu.au/coreportal).

Abstract

In this PhD thesis, we consider the problem of power efficient distributed estimation of a deterministic parameter related to a localized *phenomena* in a Wireless Sensor Network (WSN), where due to the power constraints, we propose to jointly optimize (*i*) selection of a subset of active sensors, (*ii*) multihop routing structure and (*iii*) bit-rate allocation for all active sensor measurements. Thus, our goal is to obtain the best possible estimation performance at a given querying (sink) node, for a given total power budget in the WSN. Furthermore, because of the power constraints, each selected sensor fuses all other measurements that are received from its child sensors on the chosen multihop routing tree structure together with its own measurement to perform an aggregated parameter estimation, and then it sends only one flow of fused data to its parent sensor on the tree. We call this scheme as an Estimate-and-Forward (EF).

The thesis is divided in two parts. In the first part, an optimization problem is formulated where fine quantization (high bit-rates) is assumed to be provided at all the sensor measurements, that is, ignoring the bit-rate optimization problem. Then, only the sensor selection and multihop routing structure are jointly optimized in order to minimize the total distortion in estimation (estimation error) under a constraint on the total multihop communication cost. The resulting problem is non-convex, and we show that, in fact, it is an NP-Hard problem. Thus, first we propose an algorithm based on a relaxation of our original optimization problem, where the choice of the sensor selection is decoupled from the choice of the multihop routing structure. In this case, the routing structure is taken from the Shortest Path Tree, that is, it's based only on the Communication Cost (SPT-CC). Furthermore, we also design an efficient iterative distributed algorithm that jointly optimizes the sensor selection and multihop routing structure. Then, we also pro-

vide a lower bound for the optimal solution of our original NP-Hard optimization problem and show experimentally that our iterative distributed algorithm generates a solution that is close to this lower bound, thus approaching optimality. Although there is no strict guarantee that the gap between this lower bound and the optimal solution of the main problem is always small, our numerical experiments support that this gap is actually very small in many cases.

In the second part, the bit-rate allocation is also considered in the optimization problem along with the sensor selection and multihop routing structure. In this case, the problem becomes a nonlinear non-convex optimization problem. Note that in the first part, the objective function was linear, but the constraints were non-convex. Since the problem in the second part is a nonlinear non-convex optimization problem, very interestingly, we address this nonlinear non-convex optimization problem using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of linear (convex) subproblems that can be solved efficiently. Then, we propose an algorithm using the EF scheme and an adaptive uniform dithered quantizer to solve this problem. First, by assuming a certain fixed routing structure and high bit-rates to each sensor measurement are available, we optimize the sensor selection. Then, given the subset of sensors and associated routing structure, we optimize the bit-rate allocation only for the selected sensors for a given total power budget, in order to minimize the total distortion in estimation. In addition, we also show that the total distortion in estimation can be further minimized by allowing interplay between the edges of the selected routing structure and other available smaller communication cost edges, while keeping the routing tree routed at the sink node.

An important result from our work is that because of the interplay between the communication cost over the links and the gain in estimation accuracy obtained by choosing certain sensors and fusing their measurements on the routing tree, the traditional SPT routing structure, widely used in practice, is no longer optimal. To be more specific, our routing structures provide a better trade-off between the overall power consumption and the final estimation accuracy obtained at the sink node. Comparing to more conventional sensor selection, adaptive quantization and fixed routing algorithms, our proposed joint optimization algorithms yield a significant amount of energy saving for the same estimation accuracy.

Contents

List of Figures

Nomenclature

Roman Symbols

a_{jm}	scalar element associated to h_{jm}
$\boldsymbol{A}$	incidence matrix of the graph G
$A_{i,j}$	status whether a link between node i and node j is present, $A_{i,j} \in \{0,1\}$
$\mathcal{A}$	maximum value of $\hat{\theta}_j$
b_j	status of the sensor (active or inactive), $b_j \in \{0,1\}$
b_j^r	relaxed version of the variable b_j
$\widehat{b_j^r}$	binary form of the variable b_j^r
B_j	bit-rate used to quantize samples at sensor j
$c_{i,j}$	communication cost of the edge $e_{i,j}$
c_{v_j}	characteristic function of the dither signal v_j
$Comm\_cost$	total communication cost associated to the subtree T
$\boldsymbol{C}$	covariance matrix
$\mathcal{C}_j$	set of child nodes of node j
$d_{i,j}$	distance between node i and node j
d_{norm}	normalized maximum distance reachable by each sensor
$d_{j,t}$	distance from a particular sensor node j to the source target t
$Distortion$	the average estimation error
$e_{i,j}$	edge between node i and node j
$\mathbb{E}[x]$	expected value of x
E	set of one-to-one orthogonal communication links, set of edges
E_T	set of edges in the nonspanning subtree T
f_c	communication cost function
$f_{i,j}$	signal flow going from node i to node j
$f_c(d_{i,j})$	communication cost of the edge $e_{i,j}$

f_t	pdf of the predicted target location t
f_j	detection pdf of the sensor node j
$F(x)$	Fourier transform of x
$g_{i,j}$	channel gain for the edge $e_{i,j}$
G	network connectivity graph with V nodes and E edges
G_{V_T}	graph resulting from restricting the graph G to the subset V_T
h_{jm}	measurement of the m-th (scalar) parameter at node j
$\boldsymbol{h}_j^T$	j-th row of the matrix $\boldsymbol{H}$
$\boldsymbol{H}$	known measurement matrix, $\boldsymbol{H} \in \mathbb{R}^{n\times m}$
$\mathcal{I}(i)$	set of incoming flows to node i
$j \in \mathcal{N}(i)$	node j is one of the neighbor of node i
$\boldsymbol{J}$	Jacobian matrix
K	size of the selected subset
l	uniformly spaced number of quantization levels
L	lower bound of the optimization problem
L_{IDA}	objective value of the IDA algorithm
L_{FTRA}	optimal objective value of the FTRA algorithm
$L_{\text{FTR-AQ}}$	objective value of the FTR-AQ algorithm
$L_{\text{LO-AQ}}$	optimal objective value of the LO-AQ algorithm
n	total number of sensor nodes
$n+1$	identity of the sink node
n_s	nearest sensor to s
N_j	Gaussian noise with power spectral density
$\mathcal{N}(i,j)$	power spectral density with mean i and variance j
m	total number of parameters
$O(n)$	computational complexity of the oder of n
parent_j	identity of the parent of node j
p^*	optimal objective value of the optimization problem
p_j	identity of the parent of node j
p_{v_j}	pdf of the dither signal v_j
P_{FTRA}	total power incurred by the FTRA algorithm out of available $P_{\max}$
P_{gap}	power gap that is equal to $P_{\max}$ minus the total power incurred
$\overline{P}_{\text{gap}}$	average value of the power gaps for several scenarios considered in the simulation
$P_{\max}$	maximum allowed power budget
P_0	constant power budget

P_{tot}	total communication cost incurred by the algorithms
$\mathcal{Q}_j$	quantized message with B_j bits per sample
Q_j	quantization function at sensor node j
r	number of hops between a given node to the sink node
$\boldsymbol{s}$	the mean of the expected location values of the source target
S	sink node for final estimation
S_0	desired SNR at each sensor node
S_j	SNR value at node j
$\boldsymbol{S}_+^m$	set of symmetric positive semidefinite $m \times m$ matrices
S_{th}	threshold SNR value
t	expected location of the source target
$\mathcal{T}$	time available for each sensor node to perform the EF operation
T	nonspannig routing subtree
$\boldsymbol{T}$	transition matrix from time τ_{k-1} to τ_k
v_j	dither signal
V	set of the total sensor nodes
V_T	set of sensors associated to the tree T
w_j	quantizer input consisting of the local estimate and the dither signal, $w_j = \hat{\theta}_j + v_j$
$\mathcal{W}$	base-band bandwidth
(x_{1t}, x_{2t})	the coordinates of the location of the target t
y_j	local measurement at sensor node j
$\boldsymbol{y}$	vector observation, $\boldsymbol{y} \in \mathbb{R}^n$
z_j	additive Gaussian noise at sensor node j
$\boldsymbol{z}$	additive Gaussian noise and $\boldsymbol{z} \in \mathbb{R}^n$

Greek Symbols

α	path-loss exponent
β	signal decay exponent
∇x	gradient of x
$\nabla^2 x$	Hessian of x
δ	uniform quantization level set
Δ	step size of the quantizer
η	global scalarization parameter
φ_j	total number of flows coming into sensor node j
$\phi(.)$	log barrier function
γ_j	weighting factor to perform the suitable trade-off locally

	at sensor j
ι	unit imaginary number $\sqrt{-1}$
Λ	Lagrangian function
$\lambda_j^{(i)}$	utility function defined for node j, where i is the receiving node identity
$\lambda_0, \boldsymbol{\lambda_1}$	Lagrangian multipliers
μ	modulation scheme dependent parameter
ν	parameter that sets the accuracy of the approximation in ϕ
Ω_j	set of all sensors in a single path between sensor j and the sink node S
$\Sigma_{\hat{\theta}}$	distortion in estimation for one-dimentional scalar parameter
$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$	error covariance matrix associated to $\hat{\boldsymbol{\theta}}$
$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:j]$	aggregated MSE for all nodes between node 1 to node j
$\sigma_{z_j}^2$	variance in the sensor measurement at node j
$\sigma_{\mathcal{Q}_n}^2$	total MSE obtained at the sink node contributed by all n nodes
$\sigma_{q,j}^2$	quantization error when quantizing $\hat{\theta}_j$
τ_{act}	time consumed while activating sensors in IDA algorithm for a given power budget $P_{\max}$
τ_{bb}	time consumed while creating a backbone path
τ_{EF}	time consumed by the network to perform the EF operation
τ_{max}	time consumed while detecting maximum SNR sensor
τ_{sel}	time consumed while selecting a subset of sensors
θ	unknown deterministic scalar parameter
$\boldsymbol{\theta}$	unknown deterministic vector parameter
$\hat{\theta}$	estimator of θ
$\hat{\boldsymbol{\theta}}$	estimator of $\boldsymbol{\theta}$
$\hat{\theta}_j[j]$	local estimate at sensor node j: scalar case
$\hat{\boldsymbol{\theta}}[j]$	local estimate at sensor node j: vector case
$\hat{\boldsymbol{\theta}}[i:j]$	aggregated estimation between sensor i and j
ϑ	Lagrange dual function

Superscripts

$'$	first order derivatives
r	relaxed value of the variable
$*$	objective value related to the solution of the standard

	optimization problem
a	objective value related to the solution of the approximate solution
T	transpose of a matrix

Other Symbols

$\{b_j^{r*}\}_{=1}^{n}$	solution of the relaxed optimization problem
$\circ$	Hadamard (elementwise) product
(i, j)	directed edge from node i to node j, $i \rightarrow j$
$\mathcal{L}_1 \leq_P \mathcal{L}_2$	polynomial time reduction of problem $\mathcal{L}_1$ into problem $\mathcal{L}_2$
$\partial \boldsymbol{b}_{nt}^{r}$	Newton's search step
ℓ	defines an index
$\lvert x \rvert$	cardinality of x
Tr(x)	trace of x
$\lVert x \rVert_2$	norm-2 of x
$\lfloor x \rceil$	nearest integer value of x
$[x]^+$	defines $[x]^+ = x$ if $x \geq 0$ and $[x]^+ = \infty$ otherwise

Acronyms

AQ	Adaptive Quantization
BLUE	Best Linear Unbiased Estimator
cf	characteristic function
DES	Decentralized Estimation Scheme
DHP	Direct Hamiltonian Path
DQS	Dithered Quantizing System
EF	Estimate-and-Forward
EKF	Extended Kalman Filtering
FC	Fusion Center
FTR-AQ	Fixed-Tree Relaxation-based Adaptive Quantization
FTRA	Fixed-Tree Relaxation-based Algorithm
ID	Innovation Diffusion
IDA	Iterative Distributed Algorithm
iff	if and only if
i.i.d.	independent and identically distributed
KKT	Karush-Kuhn-Tucker
LO-AQ	Local Optimization-based Adaptive Quantization
LP	Linear Programming
LP-MF	Linear Programming based on Measure-and-Forward scheme

LSN	Local Sink Node
MAC	Medium Access Control
MF	Measure-and-Forward
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MST	Minimum Spanning Tree
MT	Measure-and-Transmit
NLP	Non-Linear Programming
NP-Hard	Non-deterministic Polynomial-time hard
pdf	probability density function
SNR	Signal to Noise Ratio
SPT	Shortest Path Tree
SPT-CC	SPT based only on Communication Cost
UDE-MT	Universal Decentralized Estimation based on Measure-and-Transmit
w.r.t.	with respect to
WSN	Wireless Sensor Network

Chapter 1

Introduction

> "*An equation means nothing to me unless it expresses a thought of God.*"
> — Srinivasa Aiyangar Ramanujan (1887-1920), India.

> "*La ciencia, a pesar de sus progresos increíbles, no puede ni podrá nunca explicarlo todo. Cada vez ganará nuevas zonas a lo que hoy parece inexplicable. Pero las rayas fronterizas del saber, por muy lejos que se eleven, tendrán siempre delante un infinito mundo de misterio.*"
> — Gregorio Marañón (1887-1960), Spain.

Recent advances in Wireless Sensor Networks (WSNs) have led to the emergence of small, inexpensive, limited sensing capability, limited computational and limited communication power sensors [1; 40; 83; 117; 129], which enable us to deploy a large-scale sensor network [36]. It allows to perform some intelligent tasks by deploying different types of sensors ubiquitously and pervasively in various geographical areas such as mobile networks, administrative buildings, habitat monitoring, forest fires, health care, traffic monitoring, battlefield surveillance, wildlife reserves and many more. Thus, WSNs can provide several services in very important applications [54; 55; 80] for the society.

Among the various existing constraints related to WSNs, such as power efficiency, resource-constrained processing and storage, the one that affects more importantly in the design of algorithm for these networks is usually the power efficiency since sensor nodes are usually dependent on a limited battery power. Since radio is the main power consumption in a sensor node, transmission and reception of the information should be limited as much as possible [7; 13; 19; 29; 38; 44; 55; 62; 73; 75; 83; 105]. Notice also that the wireless communication in a sensor node is directly proportional to the amount of data to be transmitted. These limitations and the specific design of sensors, demand investigating energy efficient algorithms and good data reduction approaches.

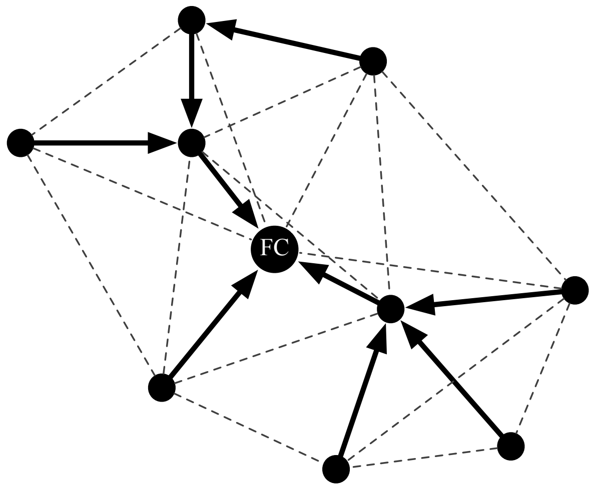


Figure 1.1: The model of a wireless sensor network with a sink node or Fusion Center (FC).

In a large number of applications, WSNs are required to cover a large geographical area, but in fact, many incidences take place locally and an estimation of some parameter related to localized phenomena is required, therefore there exists at least one Fusion Center (FC) (see Figure 1.1) to perform the final estimation. Notice that the presence of an FC in an estimation task does not mean that there exists a centralized structure in the network around this special node. In fact, all the algorithms could be fully distributed even if there is an FC. For instance, we can think of the FC as the querying node in the network. A user wants to perform an estimation of the parameter of interest, and for that purpose, access the network at one node that becomes the querying node at this time. This could possibly change over time, with different users querying at different nodes. Therefore, it is not possible to have a predetermined centralized structure in the network and our algorithms should be flexible enough to take account of various circumstances. For example, some applications can be: the energy captured by acoustic amplitude sensors where the sound source is localized in a certain spatial point [54; 55; 103], forest fire detection and verification [67], direction-of-arrival sensors for localization [32; 38; 54; 63; 73; 86; 107; 118; 129], the Fukushima inverse problem [72], which is based on sparse regularization [12; 116], or any other locally generated diffusive source. Intuitively, in these scenarios, sensor nodes that are located far away from the source will have significantly less informative measurements than the sensor nodes that are closer to it. Furthermore, the direct wireless transmission of the measurements (data) from all sensor nodes to an FC would certainly result in an inefficient use of the energy resources since the energy consumption in radio transmission grows as a power of the distance. An alternative is to allow multihop data transmission using appropriate routing strctures [2; 5; 6; 15; 25; 28; 44; 45; 46; 47; 61; 78; 82; 84; 90; 93; 101], which can significantly reduce the total transmission power, and thus enhance the network lifetime [15; 42; 44; 49; 61; 85; 111]. In addition, network lifetime can be further enhanced by activating only a subset of most informative sensor nodes [9; 48; 60; 62; 70; 73; 77; 84; 85; 87; 88; 102; 109; 113; 120; 124; 126] from which information is taken for further signal processing. Data fusion [2; 28; 38; 57; 65; 68; 71; 76; 85; 113; 120] can also be employed for in-network processing in order to save energy because it reduces the amount of data to be forwarded to other nodes. Finally, optimizing the data quantization using adaptive bit-rate allocation [42; 45; 54; 60; 68; 101; 120] can also be employed in order to further enhance the network life time since energy consumption also grows with the number of bits being transmitted.

In summary, the operations of multihop data routing, selection of active sensors, data fusion and bit-rate allocation, are very important to consider and to jointly optimize for the WSN in order to improve its lifetime. To the best of our knowledge, the joint optimization of all these operations has not yet been considered in the literature.

1.1 Motivation

In the recent literature, it has been extensively shown that the multihop data transmission is much more power efficient than direct data transmission, where each measurement is directly transmitted (Measure-and-

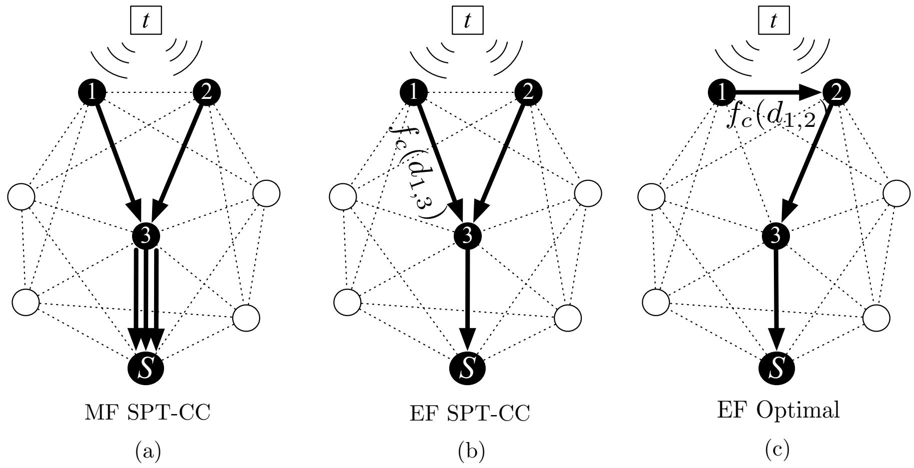


Figure 1.2: The Shortest Path Tree based on Communication Cost (SPT-CC) is not the optimal routing structure, in general, when an optimal trade-off has to be achieved between the total communication cost and estimation distortion, regardless of using the MF or EF scheme. Here, S and t denote the sink node and the source target, respectively. Discontinuous links show all potential available links. The function $f_c(.)$ represents the communication cost between every two sensor nodes.

Transmit (MT) scheme) to the querying (sink) node, which is responsible to perform the final parameter estimation [11; 56; 70; 87; 88; 109; 120]. Notice that we assume that the sink node is not allowed to use its own measurement. A multihop routing structure strategy allows to send the measurements to the sink node via two schemes: either by simply forward the measurements via intermediate sensor nodes in a chosen routing tree or fuse measurements altogether at each intermediate sensor node, and then forward only an aggregated estimate. Notice that the fusion of the measurements at each intermediate sensor node provides an improvement in estimation quality, at negligible extra computational cost. Notice that in practice the communication cost is much higher than the computational cost [40]. This leads to two schemes as illustrated in Figure 1.2(a) and (b):

1. *Measure-and-Forward (MF)*: in this scheme, sensor nodes simply

forward the measurements that are received from their child sensor nodes towards the sink node along with its parent node on a chosen multihop routing structure. Therefore, no aggregated estimation is calculated by the intermediate sensor nodes, and as a consequence, data flows sent through the links grow with the number of sensor node of the routing grows.

2. *Estimate-and-Forward (EF)*: in this scheme, a sequential estimation approach is considered. For a given multihop routing structure, each sensor node fuses all other measurements that are received from its child sensor nodes together with its own measurement, in order to obtain a local aggregated estimation, and then sends only one flow of fused data to its parent sensor node in the chosen routing structure, and the same operation is repeated until the sink node is reached. Thus, the sink node has to receive the final estimation.

The ideas behind the two schemes are illustrated in Figure 1.2(a) and Figure 1.2(b), respectively, where a localized source target is represented by t. The EF scheme has clearly several interesting advantages over the MF scheme. First of all, the EF scheme is more power efficient since an active sensor node in a route has only to forward the fused estimation (one piece of fused information to be transmitted), instead of forwarding its own measurement plus the measurements from its child sensor nodes, which are further away from the sink node on the routing structure. Moreover, we have the fact that the intermediate sensor nodes in the route have an estimation of the parameter, which gets better as the sensor node is closer to the sink node (more data fused as we approach to the sink node). One of the other major drawbacks of the MF scheme is that the sensor nodes near the sink node can be very strictly overburdened and died first, due to the large number of data flows that they have to forward, which creates a bottleneck effect.

Hence, given a WSN with a certain underlying network connectivity graph, a certain sink node and a localized source target (see Figure 1.2), we consider the problem, assuming the EF scheme, in order to jointly optimize over the sensor selection, multihop routing structure and the bit-rate allocation (quantization level) for sensor measurements. Thus, the total distortion in estimation (estimation error) at the sink node, subject to a total multihop communication cost, is minimized. In this work, we show that the Shortest Path Tree (SPT) [13; 17; 44; 45; 47; 77] is not the optimal routing structure when both the communication cost and estimation accuracy are taken into account. Figure 1.2 illustrates a simple example where the SPT based only on Communication Cost (SPT-CC) is not the optimal routing structure in general, when an optimal trade-off has to be achieved between the total communication cost and estimation distortion, regardless of the fact that one uses the MF or EF scheme. The MF scheme always requires more flows than the EF scheme, thus the MF scheme is clearly not power efficient. Moreover, as a simple example illustrated in Figure 1.2, assuming that the power budget is restricted so that no more than three sensors are selected and assuming that the EF scheme is used, then comparing the routing structures shown in Figure 1.2(b) and Figure 1.2(c), the routing structure in Figure 1.2(c) is optimal since $f_c(d_{1,3}) > f_c(d_{1,2})$, where $f_c(.)$ is a function representing the communication cost between every two sensor nodes. Furthermore, it can also be noted that in Figure 1.2(c), the bit-rate to sensor node 1 can be made higher than that of the same sensor node 1 of Figure 1.2(b), such that the cost $f_c(d_{1,2})$ becomes similar to $f_c(d_{1,3})$, resulting in a significant improvement of the quality of the final estimation at sink node S. Notice that according to the Shannon's information theory the bit-rate is proportional to the communication cost.

In our multihop distributed estimation, as we head towards the sink node, we need to provide higher bit-rates to quantize the fused estimation since the estimation accuracy improves as the more information is fused at intermediate sensor nodes. Therefore, bit-rate allocation depends on the number of hops from a sensor to the sink node in such a way that there is a need to provide higher bit-rates as heading towards the sink node in the multihop path (i.e., small hops from nodes to sink node). On the other hand, the location of the source target also influences the bit-rate assignment for each sensor measurement. For instance, if a sensor is close to the source target (high SNR), then even though a large number of hops required reaching the sink node, we need to assign a reasonably high bit-rate. In addition, bit-rate allocation also depends on the communication cost between each two sensors. Accordingly, there is a clear need to design an adaptive quantizer at each sensor node in order to provide an appropriate bit-rate, which depends on the compromise among the number of hops, the source target location and the communication cost between each two sensors.

We present an example as shown in Figure 1.3, illustrating the three operations: sensor selection, multihop routing structure and bit-rate allocation to each selected sensor measurement.

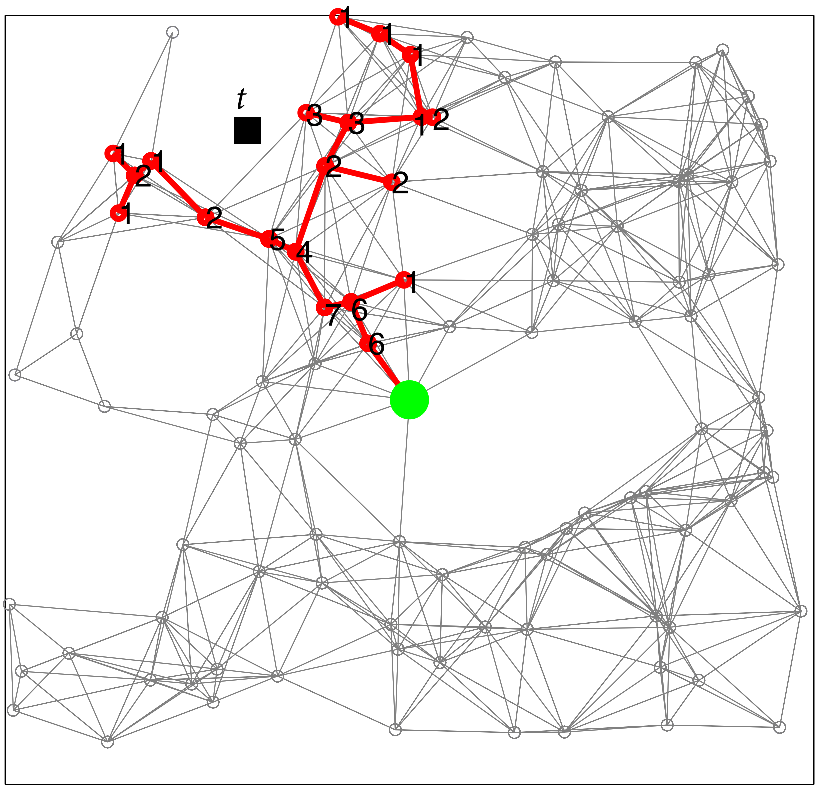


Figure 1.3: A typical solution with 100 sensors (small circles) where the following elements can be seen: a subset of selected sensors, an associated multihop routing structure (thick red edges) and bit-rate allocation for each selected sensor (marked numbers at selected sensors). The sink node is located at the center by big circle. We assumed a localized source target t, which is located around the top left corner (black square). Thin edges represent the underlying network connectivity graph and thick edges belong to the selected multihop routing structure. The numbers marked at the selected sensors represent the bit-rate assigned to quantize their measurements.

1.2 Main Contributions

The main contributions of this thesis can be enumerated as follows:

1. The problem of joint optimization of the sensor selection, multihop routing structure and adaptive bit-rate allocation (for sensor measurements) for distributed estimation under a given total power budget is formulated and analyzed, both in terms of algorithm design and complexity analysis, proving that this problem is NP-Hard when the EF scheme is used. We also provide a lower bound for the optimal solution of the original NP-Hard optimization problem.

2. First, we consider the problem to jointly optimize the sensor selection and multihop routing structure considering fine quantization (high bit-rates) is available to each sensor measurement. Then, we present a Fixed-Tree Relaxation-based Algorithm (FTRA) (see Section 4.1) that is based on a relaxation of our original optimization problem, and which decouples the choice of the routing structure from the sensor selection.

3. Then, we also design a novel and very efficient Iterative Distributed Algorithm (IDA) (see Section 4.2), which jointly optimizes locally and distributively the sensor selection and multihop routing structure. We also show experimentally that our IDA generates a solution that is close to the lower bound obtained to the original NP-Hard problem, thus approaching optimality.

4. Second, bit-rate allocation is also considered in the optimization problem along with the sensor selection and multihop routing structure to jointly optimize, in this case the problem becomes a nonlinear non-convex optimization problem. Note that in the first part, the objective function was linear, but the constraints were non-convex. Since the problem in the second part is a nonlinear non-convex optimization problem, very interestingly, we address this nonlinear non-convex optimization problem using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of linear (convex) subproblems that can be solved efficiently.

5. To this second optimization problem, we develop two algorithms,

namely, Fixed-Tree Relaxation-based Adaptive Quantization (FTR-AQ) algorithm (see Section 5.1), which is the extension of the FTRA algorithm while considering also adaptive quantization in the optimization problem, and Local Optimization-based Adaptive Quantization (LO-AQ) algorithm. Notice that the solution of the LO-AQ is obtained, by performing some additional operations (see Section 5.2) on the solution obtained from the FTR-AQ algorithm, which provides better estimation accuracy for the same given total power budget, at the expenses of additional computational complexity.

6. Finally, we compare our algorithms with the other relevant related works presented previously in the literature [47; 61; 84; 122], showing clearly superior performance in terms of the distortion in the estimation for the same given total power budget.

1.3 Organization of the Dissertation

The first two chapters are related to the general introduction and the state-of-the-art in distributed parameter estimation for ad-hoc WSNs, the sensor selection, routing structure and data quantization. The following chapters 3, 4, and 5 address the main contributions made by the corresponding published [96; 97; 98; 100; 101] and submitted [99] papers. Finally, we draw some conclusions and present the future work in chapter 6. The organization of this dissertation is as follows:

Chapter 2: this chapter deals with the state-of-the-art in distributed parameter estimation for ad-hoc WSNs, the sensor selection, routing structure and data quantization, where we summarize the required background knowledge to understand the problem presented in this thesis. It reviews the basics of distributed estimation in WSNs analyzing the various possible operations to be optimized, such as the sensor selection, routing structure and data quantization, and also all the relevant previous work presented in the literature.

Chapter 3: in this chapter, we formulate an optimization problem to jointly optimize the sensor selection and associated multihop routing structure so that the total distortion in estimation is minimized for a given total power budget. In this problem, fine quantization is assumed to be available, thus the bit-rate allocation optimization is ignored. In this case, the objective function is linear and the restrictions on resource optimization problem are not convex, which give rise to a complex (NP-Hard) problem, and which motivates us to use a relaxation method to solve this problem. This relaxation method transforms the original problem into a convex optimization problem and that can be solved easily by using standard convex optimization methods. For this problem, we also provide a lower bound for the optimal solution of our original NP-Hard optimization problem and show experimentally that our optimization algorithms generate a solution that is close to this lower bound.

The technical contributions related to this chapter have been published in one journal and the proceedings of three core conference papers:

- Shah, S.; Beferull-Lozano, B., "Joint sensor selection and multihop routing for distributed estimation in ad-hoc wireless sensor networks," *Signal Processing, IEEE Transaction on*, vol. 61, no. 24, pp. 6355-6370, December 15, 2013. [100]
- Shah, S.; Beferull-Lozano, B., "In-network iterative distributed estimation for power-constrained wireless sensor networks," *The 8th IEEE International Conference on Distributed Computing in Sensor Systems 2012*, DCOSS '12, pp. 239-246, Hangzhou, China, May 2012. (The Best Paper Award). [96]
- Shah, S.; Beferull-Lozano, B., "Power-aware joint sensor selection and routing for distributed estimation: a convex optimization approach," *The 8th IEEE International Conference on Distributed Computing in Sensor Systems 2012*, DCOSS '12, pp. 230-238, Hangzhou, China, May 2012. [98]
- Shah, S.; Beferull-Lozano, B., "In-network local distributed estimation for power-constrained wireless sensor networks," *IEEE 75th Vehicular Technology Conference*, VTC'12, pp. 1-5, Yokohama, Japan, May 2012. [97]

Then, we add the bit-rate allocation variables to the first problem, which becomes a nonlinear non-convex optimization problem. We address this nonlinear non-convex optimization problem using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of convex subproblems that can be solved efficiently. The technical contributions related to this formulation have been submitted in one journal and published in the proceeding of one core conference paper, which are:

- Shah, S.; Beferull-Lozano, B., "Energy-efficient multihop progressive estimation and adaptive quantization for ad-hoc wireless sensor networks," *Signal Processing, IEEE Transaction on*, 2014, (submitted). [99]
- Shah, S.; Beferull-Lozano, B., "Adaptive quantization for multihop progressive estimation in wireless sensor networks," *21st European Signal Processing Conference 2013*, EUSIPCO-2013, Marrakech, Morocco, September 2013. [101]

Chapter 4: this chapter deals with the algorithms that are designed for the problem formulated in Chapter 3, where we considered to jointly optimize the sensor selection and multihop routing structure using the EF scheme while assuming that fine quantization is available for each sensor measurement. Then, we provide two algorithms for this formulation, namely, Fixed-Tree Relaxation-based Algorithm (FTRA) and Iterative Distributed Algorithm (IDA). The FTRA is based on a relaxation of our original optimization problem (as stated above) and that decouples the choice of the sensor selection and routing structure. The algorithm IDA jointly optimizes both metrics, resulting in a superior performance, as compared to the FTRA algorithm. We analysis their complexity and show the results through simulation. The technical contributions related to this chapter have been published in [96; 97; 98; 100].

We also provide in this chapter, the basic assumptions, definitions and preliminary simulation results for the single moving target detection, estimation and tracking, which will be an immediate extension of this study.

Chapter 5: in this chapter, we develop two algorithms in order to jointly optimize the subset of sensors, the bit-rate allocation and associated multihop routing structure. Since this problem is a nonlinear non-convex optimization problem, very interestingly, as stated above, we address this problem using several relaxation steps and then solving the relaxed convex version over different variables, resulting in a sequence of two linear (convex) subproblems that can be solved efficiently. The algorithms are Fixed-Tree Relaxation-based Adaptive Quantization (FTR-AQ) algorithm and Local Optimization-based Adaptive Quantization (LO-AQ) algorithm. Notice that the solution of the LO-AQ is obtained, by performing some additional operations on the solution obtained from the FTR-AQ algorithm, thus it provides better estimation accuracy for the same given total power budget, at the expenses of additional computational complexity. Finally, we evaluate the performance and show the savings of our algorithms with respect to the other related similar works, which are cited in Chapter 2. The technical contributions related to this chapter have been published in [101] and submitted in [99].

Chapter 6: this chapter concludes the dissertation by summarizing its main contributions and by discussing some future work.

Chapter 2

State of the Art

Distributed parameter estimation is a fundamental problem arising from WSN applications in which the state of a dynamic system or a parameter related to a *localized phenomena* is to be estimated via a collection of scattered measurements received from the field. It plays an important role in many applications, such as communication, source localization, monitoring and tracking of a moving source, etc. This chapter describes the fundamental aspects of distributed parameter estimation over WSNs and the state of the art in this area.

More precisely, in this chapter, we introduce the background of the WSN applications for distributed estimation in relation to our work. First, we introduce in Section 2.1 and Section 2.2 the background of WSN applications and network architectures for data fusion, respectively. An introduction to distributed estimation and a comparison between centralized and decentralized estimation are presented in Section 2.3. Finally, a description of the related work and comparisons in terms of the different transmission structures and bit-rate allocations are presented in Section 2.4.

2.1 An Introduction to WSNs

A WSN is a distributed system of sensor nodes organized into a cooperative network [39], which are deployed in the region where the local phenomenon can take place. Usually, sensor nodes are battery powered and the processing capability involves using one or more micro-controllers, CPUs or DSP chips. There may have different types of memory (program, data and flash memories), an RF transceiver (usually with a single omni-directional antenna), a power source (e.g., batteries or solar cells) and can house several sensors and actuators. Sensor nodes communicate wirelessly and self-organize in terms of communication and processing after being deployed in an ad-hoc fashion. In recent years, great attention has been devoted to multi-sensor data integration (fusion) for numerous application areas (where location information is important) including, but not limited to, civilian and military applications.

2.1.1 WSNs in Civilian Applications

WSNs are useful in a variety of civilian applications worldwide such as building automation, residential appliance control, commercial equipment control, industrial machinery control, etc. For example, a WSN can be deployed to monitor the most important areas of a building by using several inter-connected sensor nodes to provide building automation (building management system). In this application, various physical quantities from these areas, are measured and reported by the sensor nodes, which enables better power-efficient uses of heating, lighting, air conditioning, ventilating, etc. The key features of typical civilian applications can be summarized as follows:

- *Known sensor positions:* the sensing locations can be determined based on the application and type of sensors prior to planning and deployment of the WSN. For instance, temperature sensors can be placed in each room so that the heating and cooling can be automatically adjusted for each room taking into account individual preferences; similarly motion control sensors can be deployed at the entrance of a room so that the light can be switched on or off automatically when people walk in or out, respectively.
- *Availability of heterogeneous devices:* a civilian WSN usually consists of a large number of small size devices, where each device performs different tasks, such as sensing different physical quantities, routing and relaying, in-network data processing and fusion, clustering and cross-node coordination, etc. On the other hand, manufacturers of varying expertise can provide products with different functionalities for various applications at different prices.

These products may also be different in terms of power supply, operational frequency band, storage capacity, computation ability, communication range, etc.

2.1.2 WSNs in Military Applications

WSNs in the military applications usually consists of sensors that measure different types of quantities, such as: electromagnetic energy, light, pressure, sound, presence of people or objects, or detect chemical, biological and explosive materials in solid, liquid and vapor form. For military applications, WSNs must be power-efficient when collecting the information about the environment and the various possible actors in the field. There are a wide variety of possible uses for WSNs in military applications that extend from information collection to enemy tracking. A typical example is battlefield surveillance, where the goal is to deploy sensor nodes in an area where monitoring is desired. In this scenario, the expected outcome from a WSN is to relay information back to a base station, where the unexpected movement of the enemy can be monitored in real-time, and an estimate of the trajectory can be generated. For example, the work done in the context of battlefield surveillance, as in [31; 53; 108; 125], outlines a hybrid sensor network based architecture for the tracking of moving targets with the aim of detecting troops of vehicular movements by using acoustic and magnetic sensors.

2.2 Network Architectures for Data Fusion

Data fusion is the task of combining information to estimate or predict the states of some process that is taking place in the environment. It can be defined as “a task dealing with the automatic detection, association, correlation, estimation, and combination of data and information from multiplier sources” [66]. Data fusion techniques use the observations and processed information of events from multiple sensor nodes as the inputs and integrate them altogether in order to achieve increased accuracies (such as distortion estimation and target location estimation accuracies) that could not be achieved using only an individual observation. Thus, in WSNs, data fusion can be used for improving data accuracy and saving every.

WSNs present many advantages over using only single sensor processing since WSNs are: dislocated over large regions, provide diverse characteristics (e.g., viewing angles) of the observed phenomenon, provide more robustness against node failures, can perform in-network inference of different types (detection, estimation, prediction, control, tracking, etc.). There are several possible network architectures [64] to process all the multiple sensor observations, the most important being: centralized with a sink node, hierarchical (with feedback or without feedback) with nodes of different levels of priority and distributed, which is the focus of this thesis.

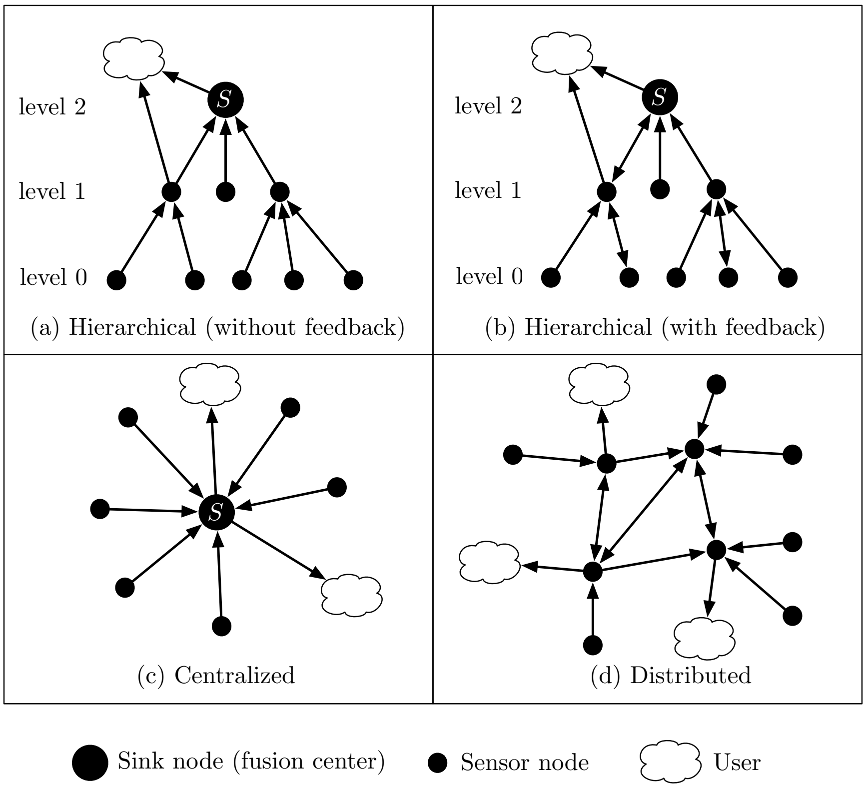


Figure 2.1: A simple illustration of the possible network architectures for data fusion. The role of a user, such as an entity or accessory to ask the information from certain sink nodes for further processing.

2.2.1 Centralized Architecture

In this network architecture, as shown in Figure 2.1(c), a central processor fuses all the information collected by all the active sensor nodes, where each information arrives at the sink node either by a Measure-and-Forward (MF) or a Measure-and-Transmit (MT) scheme.

- *Advantages:* this method is simple because erroneous information can be easily detected based on certain characteristics of the received information signals, for example, if a temperature sensor reading is received out of the expected range, it can be categorized as an erroneous information. The global fusion rule is usually simplified, sensor nodes are usually cheaper since they do not require computational capabilities but only transmission capabilities.

- *Disadvantages:* it is inflexible to sensor changes that is after deployment addition and deletion of extra sensor nodes can not be self determined by the central node (sink node), and the workload is concentrated at a single point sink node. The central node is a single point of failure, that is, if the sink node fails, the whole system becomes inoperative.

2.2.2 Hierarchical Architecture

In this architecture, sensor nodes are partitioned into hierarchical levels with fixed master/slave relationship among them; lower level sensor nodes assigned increasingly simple actions, selected by the higher level sensor nodes. The lower level sensor nodes are labeled starting from zero and the sink node being at the highest level. The information flows from the lower levels to the higher ones until the sink node is reached, as shown in Figure 2.1(a). In this case, the workload is balanced among sensor nodes. If the higher-level nodes are allowed to send the fused data back to the lower-level nodes, it is the architecture with feedback, as shown in Figure 2.1(b). Using the feedback architecture, the accuracy of a local estimate at a certain level node can be enhanced if the estimates from nodes at a higher level are fed back to lower level nodes and to be combined to the current estimate.

- *Advantages:*
  - lower communication cost in general since data does not have to be sent directly to the sink node and can be transmitted using multihop path.
  - higher robustness since there is no single point of failure associated with the sink node since intermediate nodes can fuse information.
- *Disadvantages:*
  - Architecture: the optimization of the distribution of the sensor nodes among the different levels is not a simple task. Moreover, this optimization may need to be performed when the scenario changes.
  - Communication issues: once the distribution of levels among the nodes is decided the communication among the nodes should also be optimized, for example, connectivity links and bit-rate allocation should be optimized. Furthermore, depending on the scenario, sensor nodes should decide whether they use an MF or an Estimate-and-Forward (EF) scheme over the information.
  - Local computational needs: fusion rules for high performance results need to be defined locally at each sensor node, therefore each sensor node must be computationally efficient.

2.2.3 Decentralized Architecture

In this case, as shown in Figure 2.1(d), information fusion occurs locally at every sensor node on the basis of local observations and the information obtained from neighbor sensors, that is, the EF scheme can be applicable similarly to the hierarchical architecture. This does not require any central node for information fusion.

- *Advantages:*
  - it can be designed to be more scalable and tolerant to the addition or loss of sensor nodes.

  - protocols and distributed algorithms can be designed in order to provide higher robustness to failures and reactivity to changes in the environment.

- *Disadvantages:*

  - the higher complexity of the protocols and algorithms that are running at each sensor node.

2.3 Distributed Estimation

For decades, a significant contribution from the signal processing society [22; 23; 24; 43; 58; 69; 107; 130] has led to many signal estimation algorithms from multiple sources and sources with heterogeneous modalities. There are many important tasks to enhance the quality of signals in different scenarios, such as signal enhancement (noise reduction), source localization, process control and source coding. It is interesting to consider these tasks in distributed WSNs [54; 55; 58; 63; 65; 70; 85; 107], so that a WSN can perform such tasks in a self-organized manner. However, most of the work in this field considers centralized data fusion. Some key differences between centralized and decentralized estimation are discussed next.

2.3.1 Centralized vs. Decentralized Estimation

The issue of centralized versus decentralized processing of sensor measurements is well known in the literature. The interest in this issue has been motivated by the observation that a sensor node, an energy-constrained system, should be used smartly and to the fullest possible extent. In many applications of interest WSN are required to be deployed, in general, at large scale. At the same time, advances in sensor technology, lowering their cost and the advent of efficient low cost microprocessors has pushed the era of distributed data processing, and thus it has thrown new fuel into the debate between centralization and decentralization. Each of the two approaches has several important advantages and disadvantages [26]. Some of the advantages and disadvantages can be enumerated as follows:

- *Advantages (disadvantages) of centralized (distributed) processing:*

  - operations economy: all data processing algorithms are implemented only at the central unit.
  - hardware economy of scale: cheaper sensor nodes with no computational capabilities are used.
  - simplified unified control algorithm running at the sink node.
  - easy interfile communications: inter system communication become very smooth as all the data, ones received from all sensor nodes, is being stored centrally.

- *Advantages (disadvantages) of distributed (centralized) processing:*
  - communication failsoft capability: in case of a sensor failing, network is capable of routing data via other sensors.
  - central site failsoft capability: in case of the central node failure, information can still be stored from the other aggregating sensor nodes.
  - lower communication data rates and costs: data fusion at intermediate sensor node improves the estimation quality even if the data rates are low to have low communication cost.
  - architecture and configuration flexibility.
  - better capabilities to react and adapt faster to environment dynamics.
  - modular upgrade of the network.

Based on the above list and application scenarios considered in this thesis, our focus is on distributed algorithms.

2.4 Related Work and Comparisons

In this thesis, the main goal of a WSN is to collect the most accurate information of a physical phenomenon. This involves keeping all deployed sensor nodes active for an extended period of time or to activate a sufficient number of sensor nodes in the vicinity of sources of interest. Due to the restricted energy availability and in order to save power consumptions and to enhance the WSN lifetime, the number of active sensor nodes should be kept to a minimum. In addition, by allowing variable bit-rate for quantizing sensor measurements can also improve the WSN lifetime since the power consumption in a communication link depends also on the number bits transmitted. On the other hand, this motivates the selection of a subset of most informative sensor nodes. Then, information from these sensor nodes can be transmitted via intermediate sensor nodes to the sink node, which motivates also the optimization of the routing structure. The following section explain how the sensor selection, routing structure, and adaptive data rate (bit-rate) allocation mechanisms work, reviewing also the state of the art in this area.

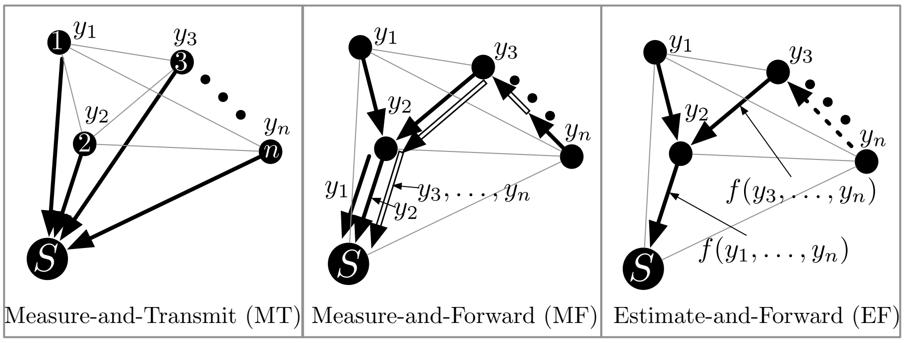


Figure 2.2: Network communication schemes, where $\{y_1, \ldots, y_n\}$ are the sensor measurements, $f(y_1, \ldots, y_n)$ is the fused data and S is the sink node.

2.4.1 Network Communication Schemes

The network communication schemes are generally classified into two types, either the information can be transmitted directly to the sink node or via a multihop path, that is, transferring measurements or preprocessed fused data multihop via the nearest neighbor to the sink node, as illustrated in Figure 2.2. In other words, direct transmission (MT scheme) or multihop transmissions (MF and EF schemes) to the sink node, as stated in Section 1.1. A significant amount of research work has been done in the context of WSNs for distributed estimation using direct wireless transmission to the sink node, thus ignoring (simplifying) the routing structure optimization. Direct wireless transmission is simple to implement, but for large scale deployments in general, it is not power efficient, which motivates the optimization of the routing structure as another key parameter to reduce the power consumption as well as to increase the overall network lifetime.

Several heuristic methods have been proposed to solve either the sensor selection or the data routing problems separately, for either centralized [6; 25; 78] or distributed (decentralized) [15; 16; 20; 21; 22; 27; 28; 47; 50; 59; 60; 61; 69; 70; 82; 84; 93; 102; 109; 110; 112; 119; 120; 121; 126; 128] estimation of deterministic parameters where the optimization of the sensor selection and routing are performed independently, while only in [61; 82] optimization of the sensor selection and data routing is jointly considered. Moreover, some of these methods use direct one-hop communication [11; 56; 70; 87; 88; 109; 120] while other use multihop communications [2; 5; 6; 15; 25; 28; 44; 45; 46; 47; 61; 78; 82; 84; 90; 93; 101] for the data transmission.

2.4.2 Sensor Selection

First, we analyze the problem of the sensor selection, where it is necessary to address how to dynamically determine, for a given objective, which nodes should sense (most informative ones) and what needs to be sensed, taking into account the communication cost incurred while transmitting these information directly to the sink node (MT scheme). Notice that when considering a localized event to monitor in a given field, sensor nodes far away from the event are not very informative, thus should be set to inactive sensing mode. By limiting the number of sensors to be active when performing some task, we can significantly reduce the total amount of power consumption of the network since less data will be transmitted to the sink node. In addition, we also reduce in general the collisions in the packet transmissions, which further reduce the power consumption; we assumed that the MAC protocol is available to each channel, which we will see in the next chapter. In other work, a certain level of estimation accuracy has to be attained by the selected subset of sensors.

The selection of a subset of sensors can be performed in principle either in a centralized or distributed fashion. The quality and usefulness of the data is heavily reliant upon the application and its use. For instance, in some applications (e.g., source localization and forest-fire detection, etc.), some individual sensor readings may not be very informative (e.g., low SNR signals) and better estimates can be obtained when aggregating (fusing) information from different sensors. Therefore, there is a clear need to define, for each scenario, performance criterion associated to the application, in addition to other elements such as sensing model and communication cost, as we will see in Chapter 3.

The problem of selecting a subset of sensor measurements from a set of available sensor nodes has been analyzed thoroughly in the literature [9; 48; 60; 62; 77; 85; 109; 124], but when the measurements are directly transmitted to the sink node, thus without involving any routing. In [48], a centralized solution is proposed based on performing a relaxation of an integer optimization problem using an efficient interior point method [10]. A distributed version of this interior point method is introduced in [9]. In [60; 62], a tradeoff between the number of active sensor nodes and the energy used (in data transmission directly to the sink node) by each active sensor is presented to minimize the Mean Square Error (MSE) estimation [52]. The problem of bandwidth constrained distributed estimation of deterministic parameters has been considered in [27; 69; 70; 87; 88; 102; 114; 120], where in [87; 88; 120], each sensor compresses its observation into a few bits, and then transmits the compressed information directly to the sink node for the parameter estimation. A problem of distributed estimation using quantized observations under bit-rate and power constraints is proposed in [42; 45; 54; 60; 102; 120; 126].

The problem of distributed estimation of a deterministic parameter under bandwidth and energy constrained WSN is also considered in [88; 120; 122], where information at each sensor is compressed and then transmitted directly to the sink node (MT scheme). In [122], a decentralized estimation algorithm is provided, where each sensor transmits its quantized measurement directly to the sink node where parameter estimation is obtained using the Best Linear Unbiased Estimator (BLUE) estimator [52]. In this work, bit-rate at each sensor is considered directly proportional to the received SNR. We refer this approach, as referred by the authors, a Universal Decentralized Estimation based on MT (UDE-MT).

Notice that none of these works consider the joint optimization of the sensor selection and routing structure.

2.4.3 Sensor Selection and Routing

In order to perform both sensor selection and routing operations, there exist basically two main methods:

1. *Separate sensor selection and routing*: where either the sensor selection operation can be performed given a fixed routing structure or vise versa, that is, first we select a subset of sensors and then find the optimal routing structure for this subset. Thus, in both the cases, decisions are taken in tandem and independently of each other. As we will show in Chapter 4, separate optimization of both operations may lead to suboptimal solutions.

2. *Joint sensor selection and routing*: in this case, the sensor selection and data routing structure are jointly optimized, that is, interplay between the sensor selection and associated routing structure is taken into account, which leads in general to better solutions as compared to separate optimization. In the problem of networked distributed estimation [61], this happens because the selection of a sensor node influences both the estimation quality and the routing possibilities.

Power efficient digital and analog communication schemes for progressive and consensus-based estimation algorithms in multihop WSNs are presented in [46; 47], where the total distortion in estimation is minimized using both centralized as well as distributed algorithms. However, these algorithms consider neither the routing optimization nor the sensor selection optimization of the two metrics (distortion in estimation and communication cost). The authors provide a detailed study on power efficient communication schemes. Although designing a detailed communication scheme is not the scope of this thesis, we consider the digital communication model presented in [46].

The results of this thesis are more related to the work of [61; 82; 84; 88; 120; 122] in the sense that the EF scheme is also used, except in [61; 122]. In [61], the MF scheme is used to jointly select a subset of sensors and their multihop routing structure and in [122], as explained earlier, the MT scheme is used. A problem involving joint sensor selection and routing using EF for distributed estimation of scalar parameter is introduced in [82], where a greedy distributed algorithm is proposed.

In this work, the trade-off between the distortion in estimation and communication cost is balanced by a scalar, which stores only two values: doubles the previous value or divide it by two. Thus, there is no optimal trade-off between two metrics.

In [61], the network lifetime maximization issue is considered for an estimation problem in energy-limited WSNs, where a Non-Linear Programing (NLP) problem is formulated. Later, the NLP is reformulated, using a convexification approach, into two Linear Programming (LP) problems that are used to optimize, using an MF scheme, the total number of bits at each sensor and the total communication cost of the multihop routing to the sink node; we name it as an LP-MF. Even though the use of an MF scheme simplifies the problem and reduces complexity as compared to using EF, the number of information flows generated in the former is substantially larger than in the later, and as a result, the total communication cost grows as the amount of measurement flows increases.

In [84], an Innovation Diffusion (ID) algorithm using the EF scheme is proposed, where the objective is to select a minimum subset of active sensors in order to save energy and prolong system lifetime. The main drawback of this algorithm is that it does not trade-off jointly the communication cost and estimation accuracy. Their multihop routing structure is determined by the next best sensor to be activated and that minimizes their objective function (distortion in estimation) without taking into account the communication cost.

In [47], a progressive distributed estimation using the EF scheme is proposed, where the objective is to estimate a deterministic parameter using the BLUE estimator, with the goal of minimizing the total energy. The main caveat of this approach is that there is no joint trade-off between both metrics. Their multihop routing structure is assumed to available, and it does not perform any selection of a subset of sensors.

2.4.4 Convex and Non-Convex Optimization

Many research problems in WSNs can be solved using mathematical techniques of minimization or maximization for a given objective subject to certain constraints. Convex optimization problems are the watershed between easy and hard optimization problems, while non-convex optimization problems involve several challenges and usually lead to NP-

Hard problems [17; 30; 57]. For example, in [122], the power allocation problem is formulated as a non-convex optimization problem where the goal is to minimize the total communication cost of the WSN under a given accuracy requirement for the parameter estimation. The variables to optimize are the bit-rate at each sensor node and transmission cost while transmitting a measurement directly from a sensor node to the sink node by taking into account both their local received SNRs and channel path losses. This problem involves integer variables and induces interplay between transmission cost and estimation, the problem become nonlinear non-convex optimization problem and can be shown to be NP-Hard problem. Even after relaxing the integer variables, problem remains non-convex. Thus, in general several relaxations or convexifications are required.

In [61], a distributed estimation problem is considered in WSNs under power constraints, where the goal is to maximize the network lifetime for a given distortion in estimation. In this work, the network lifetime maximization problem is also formulated as a non-convex optimization problem to optimize three variables of interest. The authors show that this non-convex problem can be decoupled into two linear problems by considering these variables separately, without loss of optimality, and reformulated as convex optimization problems, which provides a suboptimal solution to the original non-convex optimization problem.

In [48], the sensor selection problem in a distributed estimation framework is also formulated as a non-convex optimization problem. This problem is based on a linear and independent measurement model and their goal is to minimize the volume of the confidence ellipsoid, which can be expressed as a convex function, subject to a limitation on the cardinality of the selected subset of sensors (i.e, only K out of the n possible sensors can be selected). This is also a non-convex problem as the optimization is over binary variables to assign the status of active or inactive sensing. The authors propose to solve the problem via a single convex relaxation of the integer variables. However, the main drawback of this approach is that the nature of the network is not taken into account. Energy savings are only derived from the fact that we are keeping active only a limited number of sensors in the network, but do not consider the inherent transmission routing structure of the WSN. By imposing only a selection of a subset of sensors, the authors are equivalently assuming that the cost of transmitting a measurement from a sensor node to the fusion center is equal for all the sensors, which is not the case in a real multihop WSN. On the other hand, for scenarios where parameters related to localized phenomena have to be estimated, some sensor readings may be very informative, but could a high communication cost for sending its data to the sink node. In this case, it might be more power efficient to fuse a couple of measurements from closer nodes, which combined, can lead to a similar or even better approximation of the parameter we are interested in estimating.

2.4.5 Fixed-Rate and Adaptive-Rate Quantization

As mentioned before, for a typical WSN with limited resources (energy and bandwidth), it is important to limit the communication within the network. Therefore, it is desirable to limit the number of bits to be transmitted from local sensor nodes to the sink node via an efficient transmission scheme, while keeping the required level of performance. Moreover, in order to make the system power efficient, an adaptive rate quantization is to be preferred over the fixed-rate quantization, where in fixed-rate quantization [14; 35; 76; 94], all sensors quantize their observations using same number of bits. For example, in a region where sensor nodes receive measurements with very low sensing SNR, intuitively the number of bits that are used to quantize the measurement should be small since the contribution of these signals to the estimation is low. Thus, in order to take into account that these measurements are less important, it is convenient to consider adaptive bit-rate allocation.

As stated above in [122], which considers a decentralized estimation, where each sensor transmits its quantized measurement directly to the sink node for parameter estimation, the goal is to optimally choose the number of quantization levels and transmission powers for all sensor nodes, by taking into account both their local SNRs and path-losses associated to the links with the neighbors. The scheme presented, proposes that the sensor nodes connected to sink node with bad channels or having poor observation qualities, should decrease their quantization bit-rate or become inactive (no sensing) to save power. In this work, the MT scheme is used to transmit the information from selected nodes to the sink node, which is less power efficient than multihop routing structure. Furthermore, notice that the bit-rate allocation can be tied up with the choice of the routing structure, which can significantly provide an improved performance (as we will see by simulation results given in Chapter 5) compare to the bit-rate allocation method presented in [122].

The work in [61] jointly optimizes, using the MF scheme, the bit-rate allocation at each sensor, the total information bits generated by each sensor (received information bits from its children and its own information bits), and multihop routing structure. The main demerit of this approach is that it uses the MF scheme, that is, there is no fusion at intermediate sensor nodes. Thus, optimizing the total information bits at any intermediate sensor node will influence the quality of measurement (total bits) received from its child sensor nodes, which is not the case with the EF scheme.

In [120], a Decentralized Estimation Scheme (DES) is proposed where each node compresses its measurement into a certain number of bits, which is proportional to the logarithm of its local received SNR, and then uses an MT scheme. Then, each sensor node transmits its compressed measurement to the sink node for final estimation of an unknown parameter. The authors show that the proposed DES is universal in the sense that each sensor measurement compression scheme requires only the knowledge of local SNR of its measurement, rather than the noise probability distribution functions (pdf), while the final fusion step is also independent of the local noise pdfs. Furthermore, the authors show that the MSE of the proposed DES is within a small constant factor of what can be achieved with BLUE estimator.

In [60], the problem of joint sensor selection and rate allocation is also considered for a distributed estimation in WSNs. A simple scalar linear measurement model, where all the measurements are independent, is analyzed. The problem is formulated as a distortion minimization subject to a total bit-rate constraint in the network, where a uniform and probabilistic quantizer scheme, first introduced in [122], is used. In this quantization method, a local estimate is quantized as follows. If the local estimate lies between two quantization thresholds, for example, between δ_i and δ_{i+1}, then the local estimate is quantized to the threshold δ_i with probability $1 - p$ and it is quantized to the threshold δ_{i+1} with probability p. This scheme ensures that the quantization scheme works well regardless of the pdf of the local estimate, thus having a universality property. A distribution dependent quantization scheme would certainly increase the efficiency of the scheme at the expense of losing its universality. The sensor selection and rate allocation problem that is considered has no closed-form solution, authors provided a distributed algorithm based on an approximation with the centralized processing that acts in two phases. First, each node calculates its optimal quantization bit-rate based on its measurement SNR and only the sensors with the highest allocated bit-rates will transmit, while the others remain silent. Then, at the sink node distributed estimation algorithm performs the parameter estimation. The authors show that this algorithm has quasi-optimal performance, within a factor of 2.28 with respect to the optimal centralized quantization scheme with the same total bit-rate. However, no optimization is performed regarding the routing structure.

There are several other works in WSNs for bit-rate optimization that consider fixed-rate or adaptive-rate quantization methods. The adaptive-rate quantization is complex than fixed-rate quantization, but it provides flexibility to adapt the bit-rate depending on the quality of measurement, depending on the network topology and location of sink node that can improve the power efficiency of the network, as we show in the thesis.

Chapter 3

Optimization Problem Formulations

The purpose of this chapter is to provide the problem formulations of the works considered in this thesis. For a given WSN with a certain underlying network connectivity graph, a certain querying (sink) node and a localized source target, the problem of distributed parameter estimation arising from WSNs is considered, where the power constraints are imposed. We formulate an optimization problem, using the EF scheme, to minimize the total distortion in estimation (estimation error) subject to a total constraint on the communication cost (power budget). Thus, the three important operations of the WSN are jointly optimized, which are the *sensor selection*, *bit-rate allocation* across selected sensor nodes and associated multihop *routing structure* to send the aggregated information to the sink node.

First, we solve this problem by considering only the joint optimization of the sensor selection and routing structure, where fine quantization is assumed to be available, thus the bit-rate allocation optimization is ignored. We show that the objective function is linear and the restrictions on the resource optimization problem are not convex since it becomes an integer optimization problem. Then, we consider another problem where *bit-rate allocation* is also considered along with the sensor selection and routing structure to jointly optimize, which becomes a nonlinear non-convex optimization problem. Thus, the joint optimization problem becomes even harder to solve than the first optimization problem. Since the problem in the second part is a nonlinear non-convex optimization problem, very interestingly, we solve this problem by transforming it into a concatenation of two linear subproblems. We transform the nonlinear non-convex optimization problem by using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of linear (convex) subproblems. We prove that our problems belong to the class of NP-hard problems, by reducing from a simplified version of our problems to Directed Hamiltonian Problem (DHP). In Figure 1.3, we illustrate the three operations jointly carried out in this research work, which are sensor selection, multihop routing structure and bit-rate allocation to each selected sensor measurement.

We solve our relaxed and convexified optimization problems through standard convex optimization methods by interior-point methods, and we show that both problems can also be solved approximately but very efficiently.

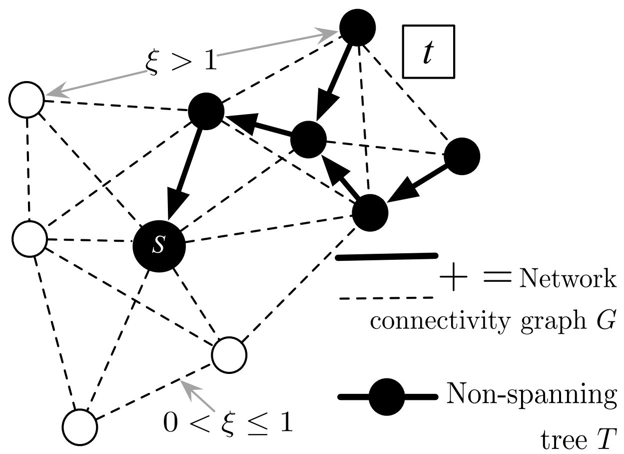


Figure 3.1: Symmetric network connectivity graph G. Node S denotes the sink node and t denotes the expected location of the source target.

3.1 Basic Assumptions and Definitions

We consider the problem of estimating some deterministic parameters generated by a spatially *localized phenomena*. There is a sink node whose goal is to obtain the best possible estimation under some total given power constraint, as it is the case in battery-powered WSNs. We consider stationary sensor nodes with one-to-one communication links deployed in a WSN, which is modeled as an undirected network connectivity graph $G = (V, E)$, where V is a set of sensors and E is a set of one-to-one orthogonal communication links. Generally speaking, our problem can be naturally formulated as an optimization problem [10] as follows:

$$\begin{aligned} \underset{\{T \subset G, B\}}{\text{minimize}} \quad & Distortion \\ \text{subject to} \quad & Comm\_cost \leq P_{\max} \end{aligned} \tag{3.1}$$

where $P_{\max}$ is the maximum power budget for the WSN in terms of communication cost. Our objective is the average estimation error (*Distortion*) obtained at the sink node, $T = (V_T, E_T) \subset G$ is in general a non-spanning routing tree (as represented in Figure 3.1 by the continuous edges, E_T), and $Comm\_cost$ is the total multihop communication cost associated to the tree T. Here, B is the bit-rate allocation optimizing variable. Therefore, our optimization problem minimizes, using the EF scheme, the total distortion in estimation subject to a total multihop communication cost. Thus, for a given total power budget $P_{\max}$ the best subset of sensors, V_T, (see bold nodes in Figure 3.1), the bit-rate allocation and the best associated multihop routing structure to send the aggregated information to the sink node, are achieved. Notice that if we activate more sensors from $V \in G$, the total distortion associated to the estimation can be usually reduced as the number of the measurements increases, although it will also depends on the quality of the measurements, bit-rate allocation and total given power budget. On the contrary, the total communication cost to deliver a fused measurement to the sink node increases, although it depends strongly on the chosen routing structure. Thus, there is a trade-off among the final distortion at the sink node, total communication cost and bit-rate allocation to each sensor measurement.

3.1.1 Network Model

In our network model, we assume that two sensors are neighbors if one sensor is in the transmission range of the other sensor and each sensor has a unique identity $j \in \{1, 2, ..., n\}$, where n is the number of deployed sensors. All n sensors are deployed in a square region in such a way that the distances between a sensor and any potential neighbor sensor can be expressed as ξd_{norm}, where ξ $(0 < \xi \leq 1)$ is uniformly distributed and d_{norm} is the normalized maximum distance reachable by each sensor, which is associated to a maximum power that a sensor node can use. We also assume an unoriented incidence matrix $\boldsymbol{A}$ of the underlying connectivity graph G, which is given by:

$$A_{i,j} = \begin{cases} 1 & \text{if} \quad d_{i,j} \leq d_{norm} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where $A_{i,j} = 1$ implies that the links (i, j) and (j, i) exist in the graph G and $A_{i,j} = 0$ there is no link exist between node i and node j, as shown in Figure 3.1. Here, $d_{i,j}$ is the distance between sensor i and sensor j.

We denote the $(n + 1)$-th sensor identity as the sink node S, where the final estimation is to be obtained. In this work, we assume that each sensor is equipped with an omnidirectional antenna and that the receiver has a Gaussian noise with power spectral density (p.s.d.) $N_j, j \in \{1, \ldots, n\}$ within the baseband $[-\frac{\mathcal{W}}{2}, \frac{\mathcal{W}}{2}]$. We assume time-invariant radio frequency channels between neighboring sensors with a certain bandwidth $\mathcal{W}$ and a duration of period $\mathcal{T}$, where $\mathcal{T}$ is the time available for each sensor to perform the EF operation. As a final approximation, we also assume that the various inter-sensor channels are orthogonal, thus we assume that there is an underlying MAC protocol that resolves the interference among these channels. We consider that each sensor can adjust its communication cost (transmission power) so that a desired SNR, $S_0 = \frac{\mu g f_c}{(N(2^B - 1))}$ (as defined precisely for instance in [46; 47]) at the receiver is achieved. In this sense, the communication cost from an active neighbor $j \in \mathcal{N}(i)$ to an active sensor i is given by:

$$f_c(d_{i,j}) = \frac{S_0 N_j (2^{B_j} - 1)}{g_{i,j} \mu} \tag{3.3}$$

with the channel gain $g_{i,j} = s_{i,j} d_{i,j}^{-\alpha}$, where α $(2 \leq \alpha \leq 6)$ [45; 84] is the path-loss exponent and $s_{i,j}$ is some constant chosen randomly from an exponential distribution with mean equal to one. Here, B_j is the bit-rate used to quantize samples at sensor j, and μ $(\mu > 0)$ is a parameter that depends on the particular modulation scheme.

3.1.2 Signal Model

We use a linear function in our signal model for sensor measurements because of its simplicity, and because it leads to the practical estimation approaches that give closed-form estimators. As we will see later in Section 3.2.3, even with a linear model, our optimization problem is already NP-hard, that is, a linear model still maintains essentially the complexity of our problem. In fact, assuming a linear model, the optimal estimator, as well as its performance, can be readily obtained. Let us consider the model:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{z} \tag{3.4}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is a vector observation, $\boldsymbol{\theta} \in \mathbb{R}^m$ is an unknown deterministic vector parameter to be estimated, whose observation is distorted by a known measurement matrix $\boldsymbol{H} \in \mathbb{R}^{n\times m}$ and corrupted by additive Gaussian noise $\boldsymbol{z} \in \mathbb{R}^n$, which is assumed to be independent and identically distributed (i.i.d.) with probability density function (pdf) $\mathcal{N}(\mathbf{0}, \boldsymbol{C})$, where $\boldsymbol{C} = \mathtt{diag}(\sigma^2_{z_1}, \dots, \sigma^2_{z_n})$ and $\sigma^2_{y_j} = \sigma^2_{z_j}, j = \{1, \dots, n\}$ are assumed to be known. We assume that each row vector $\boldsymbol{h}_j^T = [h_{j1}, \dots, h_{jm}]$ in matrix $\boldsymbol{H}$ follows the signal strength decay model [71; 92; 105], where $h_{jm} = d_{j,t}^{-\beta} a_{jm}$ describes the measurement of the m-th (scalar) parameter at sensor j, $d_{j,t}$ is the distance from a particular sensor node j to the source target t, and β is the signal decay exponent, which is assumed to be known (or estimated via training sequences [58; 105]). In order to account for the randomness in the observations, we consider that the vector element realizations $[a_{j1}, \dots, a_{jm}]$ are taken from a uniform and Gaussian distributions. In the evaluation of the algorithms, the location of the source target is taken at random from another Gaussian distribution

3.2 Optimization Problem under Fine Quantization Assumption

In our first setting, we implicitly assume the fixed fine quantization is available to each sensor measurement, where B_j in equation (3.3) is assumed to be very high, and then the communication cost becomes $f_c(d_{i,j}) \propto d^{\alpha}_{i,j}$. The optimal design of quantizer, which implies an optimization of the bit-rate allocation across sensor nodes, is discussed in Section 3.3. Thus, ignoring the bit-rate allocation constraint and considering in this section, only the joint optimization of the sensor selection and multihop routing structure.

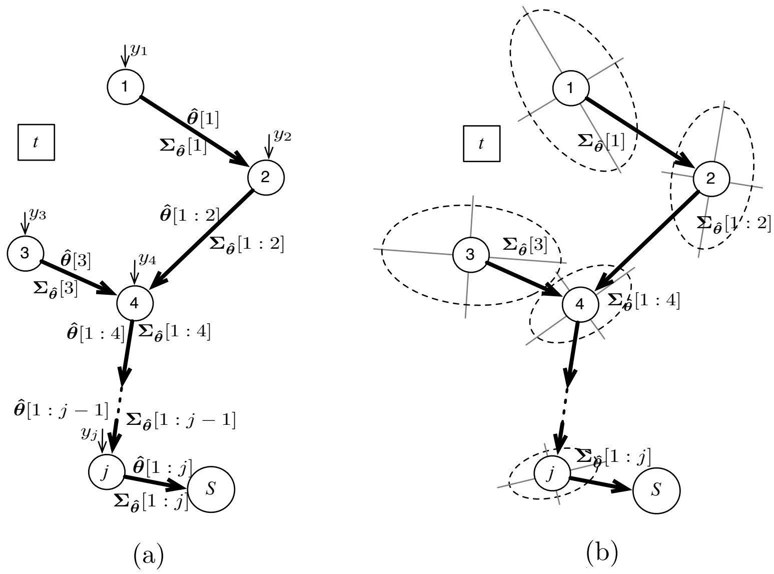


Figure 3.2: (a) A graphical representation of the distributed parameter estimation associated to a phenomenon located at source target t; (b) the distortion at each sensor, after performing the associated aggregation, is illustrated graphically with the associated confidence ellipsoids, which are defined by the corresponding error covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$. Notice that the volume of these ellipsoids decrease as more measurements are used and aggregated in the routing tree.

3.2.1 Distributed Parameter Estimation

The well known Best Linear Unbiased Estimator (BLUE) [45; 46; 52; 120; 128] is the optimal estimator for the linear problem (3.4), yielding the smallest possible MSE, thus it coincides in this case with the Minimum Mean Square Error (MMSE) estimator. The optimal estimator of $\boldsymbol{\theta}$ is readily given by:

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{y} \tag{3.5}$$

and the associated error covariance matrix is:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\boldsymbol{H}^T \boldsymbol{C}^{-1} \boldsymbol{H}\right)^{-1} \tag{3.6}$$

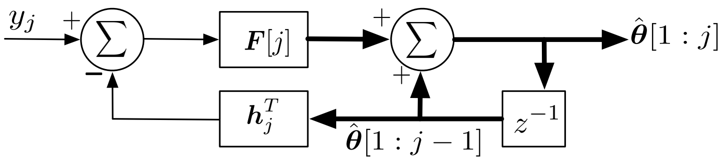


Figure 3.3: A standard sequential form of BLUE estimator as given in (3.7).

Moreover, the BLUE estimator has another advantage to be used in a WSN; it can be easily implemented in a sequential fashion when the measurement noises are independent. We denote the j-th row of the matrix $\boldsymbol{H}$ by $\boldsymbol{h}_j^T$, the steps for these updates are well known [52]:

- Estimator Update

$$\hat{\boldsymbol{\theta}}[1:j] = \hat{\boldsymbol{\theta}}[1:j-1] + \boldsymbol{F}[j]\left(y_j - \boldsymbol{h}_j^T \hat{\boldsymbol{\theta}}[1:j-1]\right) \tag{3.7}$$

  where

$$\boldsymbol{F}[j] = \frac{\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:j-1]\boldsymbol{h}_j}{\sigma_{y_j}^2 + \boldsymbol{h}_j^T \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:j-1]\boldsymbol{h}_j} \tag{3.8}$$

- MSE Update

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:j] = (\boldsymbol{I} - \boldsymbol{F}[j]\boldsymbol{h}_j^T)\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:j-1] \tag{3.9}$$

where $\hat{\boldsymbol{\theta}}[j]$ denotes the estimation at sensor j and $\hat{\boldsymbol{\theta}}[i{:}j]$, $\forall\ i,j \in T$, the aggregated estimation between sensor i and j (see Figure 3.2). In Figure 3.2(a), sensor 1 takes the measurement y_1 and the estimates $\boldsymbol{\theta}$, resulting in the estimation $\hat{\boldsymbol{\theta}}[1]$ and corresponding error covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1]$, which are both transmitted to sensor 2. Sensor 2 in turn uses its own measurement y_2 and the informations $\hat{\boldsymbol{\theta}}[1]$ and $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1]$ received from sensor 1 to compute a new estimation of $\boldsymbol{\theta}$ which is based on both measurements y_1 and y_2. However, notice that sensor 2 does not have access to y_1. The aggregated resulting estimation is denoted by $\hat{\boldsymbol{\theta}}[1{:}2]$ and the associated covariance matrix by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1{:}2]$. This information is now forwarded to sensor 4 and the process continues in the same way. Notice that there are no matrix inverses to be calculated since the denominator of $\boldsymbol{F}[j]$ in (3.8) is a scalar, thus it is also convenient in terms of implementation. Therefore, given a certain chosen routing structure, assuming that the EF scheme is used, each intermediate sensor takes the set of fused estimates and associated covariance matrices that are received from its child sensors and fuses them with its own measurement. Then, each sensor forwards the aggregated estimation and updated covariance matrix in the form of a single data packet (flow) to its parent sensor along the chosen routing tree.

3.2.2 Optimization Problem

Following our signal model in (3.4), let us consider the measurement from sensor j, that is, a linear combination of the vector parameters that need to be estimated, corrupted by independent additive Gaussian noise, which is given by:

$$y_j = \boldsymbol{h}_j^T \boldsymbol{\theta} + z_j, \quad j = 1, 2, ..., n \tag{3.10}$$

The final error covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$, based on the chosen active sensors, and which is obtained at the fusion center by using the BLUE estimator, can be expressed as:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\sum_{j=1}^{n} \frac{b_j \boldsymbol{h}_j \boldsymbol{h}_j^T}{\sigma_{y_j}^2} \right)^{-1} \tag{3.11}$$

where we assume a binary variable $b_j \in \{0, 1\}, j = 1, \ldots, n$, denoting the status of each sensor, namely, $b_j = 1$ denotes that the j-th sensor is active (i.e. chosen) and $b_j = 0$ denotes that the j-th sensor is inactive (i.e. not chosen). Without loss of generality, for the sake of simplicity, we assume that $\sigma_{y_j}^2 = 1 \; \forall j$.

From a theoretical point of view, in principle, in order to jointly minimize the distortion metric given by (3.11) subject to a total communication cost (power budget), we would need in principle to operate on the positive semidefinite cone[1] $\boldsymbol{S}_+^m$ [10], that is, the distortion ex-

[1] Matrix $\boldsymbol{S} \in \mathbb{R}^{m \times m}$ is positive semidefinite $\boldsymbol{S} \succeq 0$ if for all $x \in \mathbb{R}^m$, $x^T \boldsymbol{S} x \geq 0$.

pressed in terms of the confidence ellipsoid (as shown in Figure 3.2(b)) defined by the error covariance matrix $\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}$. This can be formulated as follows:

$$
\begin{aligned}
\underset{\{b_j,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}}\ (\text{w.r.t. } \mathbf{S}_+^m) \qquad & \mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1} \\
\text{subject to} \sum_{j=1,i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \geq 0 \\
& b_j \in \{0,1\},\ j = 1,\ldots,n
\end{aligned}
\tag{3.12}
$$

where $\boldsymbol{S}_+^m$ denotes the set of symmetric positive semidefinite $m \times m$ matrices and P_{gap} is the power gap that is equal to the maximum power allowed $P_{\max}$ minus the total power actually incurred P_{tot}. The second constraint ($b_j \leq b_i$) ensures that no sensor is selected if its parent on the tree is not selected. Third constraint ($b_{n+1} = 1$) together with the variable parent_j (parent of sensor j) enforces that the subset of selected sensors form a valid routing subtree $T \subset G$ rooted at the sink node, as long as sufficiently large power is available (large enough $P_{\max}$). In this problem, the sink node does not take the measurement of the event, that is, $\boldsymbol{h}_{n+1}^T = [0,\ldots,0]$.

We simplify our problem (3.12) further by applying some type of scalarization to the objective function so that we can cast our optimization problem as the minimization of a scalar objective function. Since minimizing the objective directly with respect to $\boldsymbol{S}_+^m$ in (3.12) is in practice a hard combinatorial problem for normal values of m and n [10; 79]. The three main scalarization techniques used in experiment design are: D-optimal design, E-optimal design and A-optimal design. Let's review them briefly:

- ***D*-optimal design**: in this scalarization, we seek to minimize the logarithm of the determinant of the MSE matrix. It is one of the most widely used options, where we try to minimize the volume of the confidence ellipsoid resulting from our estimator. In this case, the problem is posed as:

$$\begin{aligned}
\underset{\{b_j,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} (\text{w.r.t. } \mathbf{S}_+^m) \quad & \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \log\det\left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1} \\
\text{subject to} \sum_{j=1,i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \geq 0 \\
& b_j \in \{0,1\},\ j = 1,\dots,n
\end{aligned} \tag{3.13}$$

If we consider $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}$ as the objective function, which is related to the estimation quality, then since the logarithm is monotonic increasing function this minimization problem is equivalent to maximizing $\log\det\left(\sum_{j=1}^{n+1} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)$. As a consequence, our algorithms with this optimal design case, which we will see in the next chapter, are easier to implement than the other optimal design cases since there is no need to take the inverse of the error covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]$ associated to a local sensor j, which is not invertible since it is a rank($\boldsymbol{h}_j \boldsymbol{h}_j^T$)=1 matrix. This is the case that arises in iterative distributed algorithm presented in Chapter 4.

- ***E*-optimal design**: when using this scalarization we minimize the norm of the error covariance matrix (its maximum eigenvalue). Geometrically, this is interpreted as the minimization of the diameters of the confidence ellipsoid.

$$\begin{aligned}
\underset{\{b_j,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} (\text{w.r.t. } \mathbf{S}_+^m) \quad & \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left\| \left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1} \right\|_2 \\
\text{subject to} \sum_{j=1,i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \geq 0 \\
& b_j \in \{0,1\},\ j = 1,\dots,n
\end{aligned} \tag{3.14}$$

In this case, in our algorithms while executing the EF scheme distributively, at any local leaf sensor node (having no child nodes), the error covariance matrix $\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]$ associated to a leaf sensor node is rank-defficient (singular). Therefore, process of the algorithm with this scalarization is exactly same as the A-optimal case, which we will see below.

- ***A*-optimal design**: the scalarization technique we consider in this thesis is the so-called A-optimal formulation ([10], sec. 7.5) since this corresponds to minimizing the MSE. In this case, we minimize the trace of the error covariance matrix, the MSE since $\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}} = \mathbb{E}[\boldsymbol{e}\boldsymbol{e}^T]$ and Tr $(\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}) = \mathbb{E}[\text{Tr } \boldsymbol{e}\boldsymbol{e}^T] = \mathbb{E}||\boldsymbol{e}||_2^2 = \text{MSE}$. Thus, the scalarized version of problem (3.12), in this case, can be posed as:

$$\begin{aligned}
\underset{\{b_j,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} \; (\text{w.r.t. } \mathbf{S}_+^m) \quad & \mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}} = \text{Tr}\left[\left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
\text{subject to} \sum_{j=1, i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i, \; i = \text{parent}_j, \; A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \geq 0 \\
& b_j \in \{0,1\}, \; j = 1, \ldots, n
\end{aligned} \tag{3.15}$$

here we are considering the routing structure is given by the graph G, and where i is the parent of j and the edge $(i,j) \in G$. We let $\{b_j^*\}_{j=1}^n$ denote the optimal solution of this problem.

We can also construct an equivalent network flow [82] formulation for problem (3.15). Let us assume an oriented incidence matrix $\tilde{\boldsymbol{A}}$ for this case, which lies in $\mathbb{R}^{(n+1)\times M}$, where node $n+1$ is the sink node and $M \leq |E|$ is the total number of connected edges satisfying $d_{i,j} \leq d_{norm}$, and it is given by:

$$\tilde{A}_{i,j} = \begin{cases} 1 & \text{if} & i \text{ is the start node} \\ -1 & \text{if} & j \text{ is the start node} \\ 0 & \text{otherwise} & \end{cases} \tag{3.16}$$

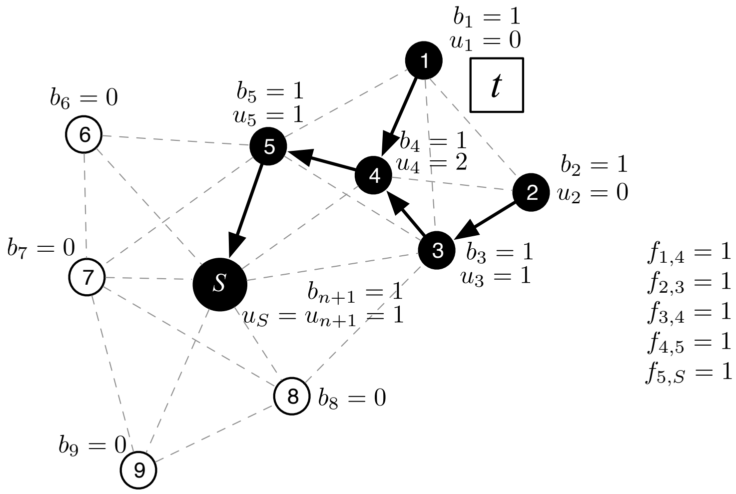


Figure 3.4: Example of a feasible flow following the notation presented in (3.17).

Moreover, let $f_{i,j}$ denote the flow going from node i to node j, $\mathcal{I}(i)$ the set of incoming flows to node i and b_i the variables that indicate the status of the sensor i. With this notation, problem (3.15) can be equivalently formulated as:

$$
\begin{aligned}
\underset{\{b_j, f_{i,j}\}}{\text{minimize}} \quad & \operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
\text{subject to} \quad & \tilde{\boldsymbol{A}} \boldsymbol{f} = \boldsymbol{s} - \boldsymbol{u} \\
& s_j = b_j \quad j = 1, \ldots, n \\
& s_{n+1} = 1 \\
& u_j = \sum_{i \in \mathcal{I}(j)} f_{i,j} \\
& u_{n+1} \geq 1 \\
& \sum_{(i,j) \in E} f_{i,j} f_c(d_{i,j}) \leq P_{\max} - P_{\text{gap}} \\
& f_{i,j} + f_{j,i} \leq 1 \quad \forall i \neq j \\
& f_{i,j}, b_j \in \{0,1\}, \; j = 1, \ldots, n \\
& P_{\text{gap}} \geq 0
\end{aligned}
\tag{3.17}
$$

where the first equality ensures the flow conservation in the network. The second and third constraints represent the status of activated sensors, which are used to generate flows in the network. The fourth equality constraint represents the total number of flows coming into a sensor node, which captures the idea of data fusion at each sensor node when using a sequential estimator (only one fused estimation is forwarded from each node). The fifth inequality ensures that the multihop routing structure is routed at the sink node. The sixth inequality constraint is related, as in (3.12), to the total power constraints. Finally, the seventh constraint prevents length-two cycles in a link. An example of a feasible solution is presented in the Figure 3.4. We have there a network consisting of nine sensor nodes with a sink node S, where nodes 1 to 5 are selected.

Notice that if we consider, for simplicity an independent scalar measurements, where we seek to estimate only one parameter, that is, now (3.10) becomes:

$$y_j = h_j\theta + z_j, \quad j = 1, ..., n \tag{3.18}$$

where in this case, the expression of the MSE is not a matrix anymore and we do not need to look for scalarizations. In this case, problem can be written as:

$$\begin{aligned}
\underset{\{b_j, \text{parent}_j, P_{\text{gap}}\}}{\text{minimize}} \quad & \Sigma_{\hat{\theta}} = \left(\sum_{j=1}^{n+1} b_j h_j^2\right)^{-1} \\
\text{subject to} \sum_{j=1, i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i, \ i = \text{parent}_j, \ A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \geq 0 \\
& b_j \in \{0, 1\}, \ j = 1, \ldots, n
\end{aligned} \tag{3.19}$$

The problem formulation and associated joint sensor selection and routing algorithms are presented in [96; 97].

3.2.3 Complexity Assessment: NP-Hardness

We prove the complexity of our problem using a reduction technique [30]. Basically, a problem $\mathcal{L}_1$ can be reduced to another problem $\mathcal{L}_2$: $\mathcal{L}_1 \leq_P \mathcal{L}_2$ if any instance of $\mathcal{L}_1$ can be converted to an instance of $\mathcal{L}_2$ (by a mapping that is computable in polynomial time), so that the solution to the instance of $\mathcal{L}_2$ allows to solve the instance of $\mathcal{L}_1$. Thus, if $\mathcal{L}_1$ reduces to another problem $\mathcal{L}_2$, then $\mathcal{L}_2$ is at least as hard to solve as $\mathcal{L}_1$.

Theorem 3.2.1 *The joint optimization problem (3.15) of the sensor selection and multihop routing structure for distributed estimation under a total power constraint, is NP-Hard.*

Proof The proof is based on performing a polynomial time reduction [30] from the Directed Hamiltonian Path (DHP) to our problem, that is, mapping every instance from the DHP problem to our problem. (see Appendix A for the detailed proof). ∎

3.2.4 Relaxation to the Integer Problem

Notice that problem (3.15) is an integer optimization problem, and it is NP-Hard, which motivates us to solve this problem approximately by first performing a relaxation over the variables $\{b_j\}_{j=1}^{n}$, and then solving the relaxed problem, and finally mapping back appropriately to the integer variables. Due to the relaxation, we can provide a feasible lower bound on the optimal value of problem (3.15) by simply solving the following relaxed problem.

$$\begin{aligned}
&\underset{\{b_j^r,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} && \operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
&\text{subject to} && \sum_{j=1, i=\text{parent}_j}^{n+1} b_j^r f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
&&& b_j^r \leq b_i^r,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
&&& b_{n+1}^r = 1 \\
&&& P_{\text{gap}} \geq 0 \\
&&& 0 \leq b_j^r \leq 1
\end{aligned} \tag{3.20}$$

where b_j^r is the relaxed version of variable b_j, leading in this case to a well defined convex (relaxed) problem since the objective is a convex function of b_j^r and all equalities and inequalities are linear on b_j^r.

Lemma 3.2.2 *The optimal solution to the relaxed problem (3.20), provides a lower bound on the optimal objective value p^* of problem (3.15). Then, the lower bound L is given by:*

$$L = Tr\left[\left(\sum_{j=1}^{n+1} b_j^{r*} \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \leq p^* \tag{3.21}$$

and the solution is given by the Minimum Spanning Tree[1] *(MST) of the directed graph $G_m(V, E_m)$ composed of edges: $e_{i,j} \in E_m$ with cost $c_{i,j} = b_i^{r*} f_c(d_{i,j})$ and $e_{j,i} \in E_m$ with cost $c_{j,i} = b_j^{r*} f_c(d_{i,j})$, for any pair of nodes $i, j \in V$. Here, E_m is the set of edges associated to the MST.*

Proof The solution $\{b_j^{r*}\}_{j=1}^{n+1}$ to the problem (3.20) provides a tree that is routed at the sink node, which satisfies the following properties: it is the MST of the directed graph $G_m = (V, E_m)$, where for every pair of nodes $\{i, j\}$, there are two edges defined, namely, $e_{i,j} \in E_m$ with cost $c_{i,j} = b_i^{r*} f_c(d_{i,j})$ and $e_{j,i} \in E_m$ with cost $c_{j,i} = b_j^{r*} f_c(d_{i,j})$. Then, for each pair of nodes $\{i, j\}$ there will be an edge corresponding to the smallest cost ($c_{i,j}$ or $c_{j,i}$) that will form part of the optimal solution to problem (3.20). Therefore, the MST of the graph G_m provides the optimal solution to the problem (3.20) and since problem (3.20) is a relaxed version of (3.15), it will provide a valid lower bound for the optimal objective value p^* of problem (3.15). ∎

3.2.5 Approximate Solution

We can easily verify our lower bound L given in (3.21) using the Newton method by solving relaxed problem (3.20) approximately but very efficiently using for instance *log barrier method* ([10], Sec. 11.2), that is,

[1] An edge-weighted graph is a graph where we assign communication costs (weights) to each edge. A minimum spanning tree of an edge-weighted graph is a spanning tree whose the sum of the total communication cost of its all edges is no larger than the communication cost of any other spanning tree.

problem (3.20) can also be posed as:

$$
\begin{aligned}
\underset{\{b_j^r, \text{parent}_j, P_{\text{gap}}\}}{\text{minimize}} \quad & \phi(\mathbf{b}^r) = \operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] - \\
& -\frac{1}{\nu}\sum_{j=1}^{n+1}\left(\log(b_j^r) + \log(1-b_j^r) + \log(b_i^r - b_j^r) + \log P_{\text{gap}}\right) \\
\text{subject to} \quad & \sum_{j=1}^{n+1} b_j^r f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_{n+1}^r = 1
\end{aligned}
\tag{3.22}
$$

where $\nu > 0$ is a parameter that sets the accuracy of the approximation. The function ϕ is convex and smooth, thus problem (3.22) can be efficiently solved by the Newton method. Let $\{b_j^{r*}(\nu)\}_{j=1}^{n+1}$ denote the solution of the approximate relaxed problem (3.22), which depends on the parameter ν.

A standard result in interior-point methods (Sec. 11.2.2, [10]) is that $\{b_j^{r*}(\nu)\}_{j=1}^{n+1}$ and the solution $\{b_j^{r*}\}_{j=1}^{n+1}$ for the relaxed problem (3.20) are related as follows:

$$
\begin{aligned}
\operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^{r*}(\nu) \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] - \operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^{r*} \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] & \leq \frac{2n}{\nu} \\
\Rightarrow \operatorname{Tr}\left[\left(\sum_{j=1}^{n} b_j^{r*}(\nu) \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] - L & \leq \frac{2n}{\nu}
\end{aligned}
\tag{3.23}
$$

and as $\nu \to \infty$, the solution $\{b_j^{r*}(\nu)\}_{j=1}^{n+1}$ approach towards the lower bound L. We can use this bound to choose ν so that the increase in the gap generated by the term $\frac{2n}{\nu}$ is small. The steps to calculate the values of $b_j^{r*}(\nu)$ by Newton's method are explained in Appendix B.

3.3 Optimization Problem under Adaptive Quantization Assumption

In this section, we consider the problem in which bit-rate allocation optimization is also considered along with the sensor selection and multihop routing optimization. Therefore, the optimization problem now becomes to jointly optimize the sensor selection, bit-rate allocation and multihop routing structure so that the total distortion in estimation can be minimized subject to a total given power budget. Notice that each intermediate sensor node performs a fusion and quantizes the fused information with a certain number of bits, and then quantized information is transmitted progressively (using a multihop routing structure) to the sink node that receives the final estimation. Notice also that the bit-rate variable B_j in communication cost model (3.3) is considered in order to take into account the bit-rate allocation.

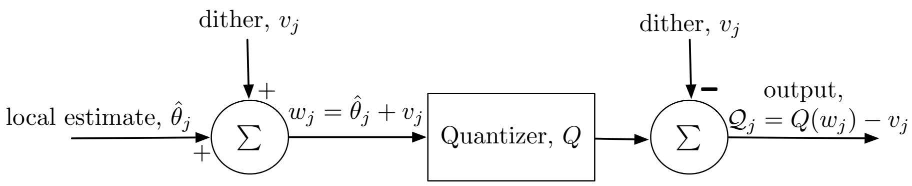


Figure 3.5: A dithered quantizer system [94].

Because of the strict energy limitation, each sensor measurement needs to quantize the data to be transmitted digitally, hence a local quantization [41; 54; 55; 61] is performed before the transmission. Suppose that we wish to obtain a quantized message $\mathcal{Q}_j$ with B_j bits per sample for a local estimate $\hat{\theta}_j$ at sensor j. For this, we consider a uniform quantization [14; 35; 104; 106; 115] with $l = 2^{B_j}$ uniformly spaced quantization levels, with thresholds given by a set $\delta = \{\delta_1, \ldots, \delta_l\}$ where $\delta_i = \delta_{i-1} + \Delta$ and $\Delta = \frac{2\mathcal{A}}{(2^{B_j}-1)}$, assuming that $\hat{\theta}_j$ is bounded within the range $[-\mathcal{A}, \mathcal{A}]$, and that we have the variance of the quantization error resulting from quantizing $\hat{\theta}_j$, is given by:

$$\sigma_{q,j}^2 = \frac{\Delta^2}{12} = \frac{4\mathcal{A}^2}{12(2^{B_j}-1)^2} \leq \frac{4}{3}\frac{\mathcal{A}^2}{2^{2B_j}} \tag{3.24}$$

where the last inequality holds under $B_j \geq 1$. Notice that this is an approximation only for high bit-rates [34].

It is important to note that the minimum loss of statistical data from the input occurs when the quantization error is made independent of the input signal [115] and that can be achieved by using dithering. We use a Dithered Quantizing Systems (DQS) defined as $Q_j(\hat{\theta}_j) = \mathcal{Q}_j(\hat{\theta}_j, B_j)$ [94; 104; 115], as shown in Figure 3.5. In this system, an additive random signal v_j, *called dither*, is being added to the local estimate $\hat{\theta}_j$ of sensor j, and then it is subtracted after the quantization operation. This dither signal is assumed to be a strict-sense stationary random process[1] and to be statistically independent of $\hat{\theta}_j$, this can be achieved in practice. Then, the quantizer input is given by $w_j = \hat{\theta}_j + v_j$, and thus in this system, w_j is not a deterministic function of $\hat{\theta}_j$, and neither is the total error $[Q(w_j) - v_j] - \hat{\theta}_j = Q(w_j) - w_j$ (see Figure 3.5). It is also important to note that the objective of the dithering is to control the statistics of the total error and its relationship to the system input since in undithered systems the error is clearly a deterministic function of the input.

It has been shown in [94] that the total error in a DQS system can be made uniformly distributed and statistically independently of the input for arbitrary input distributions (in this case, input is $\hat{\theta}_j$) *iff* the characteristic function[2] (cf) of the dither v_j, satisfies the following condition:

$$F\left(\frac{j}{\Delta}\right) = 0,\ j = \pm 1, \pm 2, \ldots \tag{3.25}$$

where the characteristic function is defined as:

$$F(u_j) = \int_{-\infty}^{\infty} f(v_j) e^{-2\pi i u_j v_j}\, dv_j \tag{3.26}$$

and where i is the unit imaginary number $\sqrt{-1}$ (*iota*) and dither pdf and cf are denoted as u_j and $F(u_j)$, respectively.

As shown in [104; 106], the total error in a DQS is statistically independent of the input signal *iff* (3.25) holds.

In this work, we are not considering more complex quantization methods such as vector quantization [28; 54; 55; 89; 127] since a close-form expression for the variance of a realistic vector quantization is generally difficult to obtain. Even for a single Gaussian random variable, in our case, the variance expression of the optimal quantization errors is not expressible.

[1]A strict-sense stationary random process is one whose n^{th} order distribution is time invariant.

[2]Characteristic function (cf) is a Fourier transform $F(.)$ of dithers pdf.

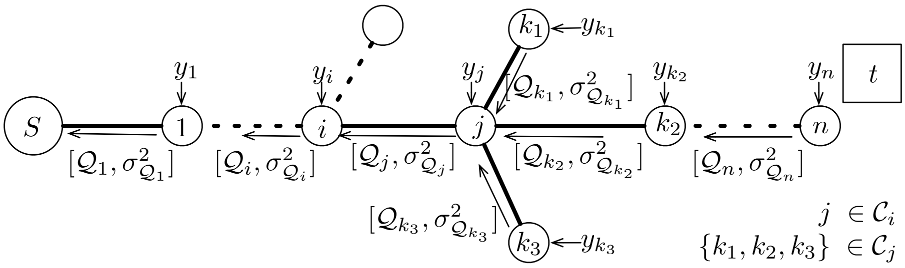


Figure 3.6: Illustration of a multihop progressive estimation scheme.

3.3.1 Progressive Parameter Estimation

Let us assume that the input to the j-th sensor (as shown in a simple example in Figure 3.6) consists of the sensor measurement y_j and information $[\mathcal{Q}_k, \sigma^2_{\mathcal{Q}_k}], k \in \mathcal{C}_j$ that are received from all its child nodes $\{k_1, k_2, k_3\} \in \mathcal{C}_j$, where $\mathcal{C}_j$ is the set of child nodes of sensor node j. We represent the quantized estimation that is received from sensor k by $\mathcal{Q}_k$ and its associated variance by $\sigma^2_{\mathcal{Q}_k}$. Then, the BLUE [47; 52] of θ based on y_j, $\{\mathcal{Q}_k, k \in \mathcal{C}_j\}$, and $\{\sigma^2_{\mathcal{Q}_k}, k \in \mathcal{C}_j\}$ at sensor j, is given by:

$$\hat{\theta}_j = \left(\frac{b_j h_j^2}{\sigma^2_{y_j}} + \sum_{k \in \mathcal{C}_j} \frac{1}{\sigma^2_{\mathcal{Q}_k}} \right)^{-1} \left(\frac{b_j h_j y_j}{\sigma^2_{y_j}} + \sum_{k \in \mathcal{C}_j} \frac{\mathcal{Q}_k}{\sigma^2_{\mathcal{Q}_k}} \right) \tag{3.27}$$

where the binary variable $b_j, j = 1, \ldots, n$ is for the sensor selection, that is, it holds the same meaning as before. Then, the variance of $\hat{\theta}_j$, which is the MSE of the estimator, is given by:

$$\sigma^2_{\hat{\theta}_j} = \left(\frac{b_j h_j^2}{\sigma^2_{y_j}} + \sum_{k \in \mathcal{C}_j} \frac{1}{\sigma^2_{\mathcal{Q}_k}} \right)^{-1} \tag{3.28}$$

Since the variance of the quantization error at sensor j, as in (3.24) is $\sigma^2_{q,j}$, then the variance of the quantized estimation $\mathcal{Q}_j$ at sensor j is given by $\sigma^2_{\mathcal{Q}_j} = b_j \sigma^2_{q,j} + \sigma^2_{\hat{\theta}_j}$, that is:

$$\sigma^2_{\mathcal{Q}_j} = b_j \sigma^2_{q,j} + \left(\frac{b_j h_j^2}{\sigma^2_{y_j}} + \sum_{k \in \mathcal{C}_j} \frac{1}{\sigma^2_{\mathcal{Q}_k}} \right)^{-1} \tag{3.29}$$

In order to provide insight into our problem, let us consider first a simple case where there is only one child k of sensor node j (that is, $|\mathcal{C}_j| = 1$) is available, where $|\mathcal{C}_j|$ denotes the cardinality of $\mathcal{C}_j$, then (3.29) is given by:

$$\begin{aligned} \sigma^2_{\mathcal{Q}_j} &= b_j\sigma^2_{q,j} + \left(\frac{b_j h_j^2}{\sigma^2_{y_j}} + \frac{1}{\sigma^2_{\mathcal{Q}_k}}\right)^{-1} \\ &= b_j\sigma^2_{q,j} + \frac{\sigma^2_{y_j}\sigma^2_{\mathcal{Q}_k}}{\sigma^2_{y_j} + \sigma^2_{\mathcal{Q}_k} b_j h_j^2} \end{aligned} \tag{3.30}$$

Since (3.29) is a nonlinear recursion of $\sigma^2_{\mathcal{Q}_k}$, even if we apply the same argument recursively in a 1-D network (that is, $|\mathcal{C}_j| = 1$), it is hard to find a closed-form expression for $\sigma^2_{\mathcal{Q}_n}$, which is the final overall MSE based on all n sensors. However, it can be easily seen that $\left(\frac{\sigma_{y_j}}{\sqrt{b_j}h_j} - \sigma_{\mathcal{Q}_k}\right)^2 \geq 0$, which gives $\frac{2\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j} \leq \left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)$. Then, (3.30) can be bounded as:

$$\sigma^2_{\mathcal{Q}_j} \leq b_j\sigma^2_{q,j} + \frac{\sigma^2_{y_j}}{4b_j h_j^2} + \frac{\sigma^2_{\mathcal{Q}_k}}{4} \tag{3.31}$$

And the equivalent form of (3.31) for $|\mathcal{C}_j| > 1$ of sensor j is:

$$\sigma^2_{\mathcal{Q}_j} \leq b_j\sigma^2_{q,j} + \frac{\sigma^2_{y_j}}{(1+|\mathcal{C}_j|)^2 b_j h_j^2} + \frac{1}{(1+|\mathcal{C}_j|)^2}\sum_{k\in\mathcal{C}_j}\sigma^2_{\mathcal{Q}_k} \tag{3.32}$$

It can be easily shown that expression (3.32) is satisfied when the following generalized inequality holds:

$$\begin{aligned} &(1+|\mathcal{C}_j|)^2\frac{\sigma^2_{y_j}}{b_j h_j^2}\prod_{k\in\mathcal{C}_j}\sigma^2_{\mathcal{Q}_k} \leq \\ &\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sum_{k\in\mathcal{C}_j}\sigma^2_{\mathcal{Q}_k}\right)\left(\frac{\sigma^2_{y_j}}{b_j h_j^2}\sum_{k\in\mathcal{C}_j}\prod_{s\in\mathcal{C}_j; s\neq k}\sigma^2_{\mathcal{Q}_s} + \prod_{k\in\mathcal{C}_j}\sigma^2_{\mathcal{Q}_k}\right) \end{aligned} \tag{3.33}$$

Expression (3.33) is the generalized form of $\frac{2\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j} \leq \left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)$ and it can be easily verified by taking $|\mathcal{C}_j| = 1$ or $|\mathcal{C}_j| = a > 1$ in (3.33), (see Appendix C for detailed proof).

The recursion for $\sigma^2_{\mathcal{Q}_j}$ leads to a generalized form of the total estimation error due to all n sensors. For this, first we write an equivalent form of (3.32) for the parent of sensor j ($i \leftarrow j \leftarrow k$), that is sensor i, as shown in Figure 3.6. In other words, since $j \in \mathcal{C}_i$, then (3.32) for sensor i becomes:

$$\sigma^2_{\mathcal{Q}_i} \leq b_i \sigma^2_{q,i} + \frac{\sigma^2_{y_i}}{(1+|\mathcal{C}_i|)^2 b_i h_i^2} + \frac{1}{(1+|\mathcal{C}_i|)^2} \sum_{j \in \mathcal{C}_i} \sigma^2_{\mathcal{Q}_j} \tag{3.34}$$

Using (3.32) and (3.34), we obtain:

$$\begin{aligned} \sigma^2_{\mathcal{Q}_i} \leq & \left(b_i \sigma^2_{q,i} + \frac{1}{(1+|\mathcal{C}_i|)^2} \sum_{j \in \mathcal{C}_i} b_j \sigma^2_{q,j} \right) + \\ & \left(\frac{\sigma^2_{y_i}}{(1+|\mathcal{C}_i|)^2 b_i h_i^2} + \frac{1}{(1+|\mathcal{C}_i|)^2} \sum_{j \in \mathcal{C}_i} \frac{\sigma^2_{y_j}}{(1+|\mathcal{C}_j|)^2 b_j h_j^2} \right) \\ & + \frac{1}{(1+|\mathcal{C}_i|)^2} \sum_{j \in \mathcal{C}_i} \frac{1}{(1+|\mathcal{C}_j|)^2} \sum_{k \in \mathcal{C}_j} \sigma^2_{\mathcal{Q}_k} \end{aligned} \tag{3.35}$$

If sensor k is the leaf node, that is, $|\mathcal{C}_k| = 0$, then (3.29) for sensor k becomes $\sigma^2_{\mathcal{Q}_k} = b_k \sigma^2_{q,k} + \frac{\sigma^2_{y_k}}{b_k h_k^2}$. In this case, (3.35) can be generalized as the sum of two terms:

$$\sigma^2_{\mathcal{Q}_n} \leq \sum_{j=1}^{n} \frac{b_j \sigma^2_{q,j}}{\mathcal{P}_j} + \sum_{j=1}^{n} \frac{\sigma^2_{y_j}}{\mathcal{P}_j (1+|\mathcal{C}_j|)^2 b_j h_j^2} \tag{3.36}$$

where $\sigma^2_{\mathcal{Q}_n}$ is the total MSE obtained at the sink node after the selection of all n sensors and $\mathcal{P}_j = \prod_{l \in \Omega_j} (1+|\mathcal{C}_l|)^2$, where Ω_j is the set of all sensors in a single path between sensor j and the sink node S.

3.3.2 Optimization Problem

In this section, we formulate a nonlinear non-convex optimization problem to jointly optimize the sensor selection, bit-rate allocation and multihop routing structure. So that for a given total power budget the total distortion in estimation is minimized. In this problem, to transmit B_j bits per sample reliably from sensor j to its parent i, we assume the minimum required communication cost must satisfy $f_c(d_{i,j}) =$

$\frac{S_0 N_j d_{i,j}^{\alpha}(2^{B_j}-1)}{\mu} < S_0 N_j d_{i,j}^{\alpha} 2^{B_j}$ from (3.3), based on the Shannon theory and a uniform quantization, where we assume that S_0 is constant for all sensors. Then, our optimization problem is formulated as follows:

$$\begin{aligned} \underset{\{i,j,B_j\}}{\text{minimize}} \quad & \sigma^2_{\mathcal{Q}_n} \\ \text{subject to} \quad & \sum_{j=1,(i,j)\in G}^{n} f_c(d_{i,j}) \leq P_{\max} \end{aligned} \tag{3.37}$$

and the equivalent form of (3.37) in terms of $\sigma^2_{\mathcal{Q}_n}$, $\sigma^2_{q,j} = \frac{4}{3}\frac{\mathcal{A}^2}{2^{2B_j}}$, and f_c is:

$$\begin{aligned} \underset{\{b_i,b_j,B_j\}}{\text{minimize}} \quad & \sum_{j=1}^{n}\left(\frac{4b_j\mathcal{A}^2}{3\mathcal{P}_j 2^{2B_j}} + \frac{\sigma^2_{y_j}}{\mathcal{P}_j(1+|\mathcal{C}_j|)^2 b_j h_j^2}\right) \\ \text{subject to} \quad & \sum_{j=1,(i,j)\in G}^{n} b_j N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max} - P_{\text{gap}} \\ & b_j \leq b_i, \text{ for } (i,j)\in G, \text{ and } A_{i,j}=1 \qquad (a) \\ & b_j \in \{0,1\}, j=1,\dots,n \qquad (b) \\ & P_{\text{gap}} \geq 0 \\ & B_j \geq 0, j=1,\dots,n \end{aligned} \tag{3.38}$$

where the binary variable b_j determines the sensor selection, constraints (a) and (b) impose the routing tree structure since the edge $(i,j)\in G$, which ensures a subtree $T\subset G$ from the selected sensors rooted at the sink node, and variable B_j is to assign the bit-rate to each sensor measurement. Here, P_{gap} holds the same meaning, that is, it is the power gap, which is equal to the maximum power allowed $P_{\max}$ minus the total power actually incurred. Notice that when $B_j = 0$, we allow sensor j to transmit its estimation using 1-bit per sample and when $B_j \geq 1$, then the total distortion estimation (3.36) converges to the best achievable performance $\sum_{j=1}^{n}\frac{\sigma^2_{y_j}}{\mathcal{P}_j(1+|\mathcal{C}_j|)^2 b_j h_j^2}$ exponentially as $B_j \to \infty$.

3.3.3 Complexity Assessment: NP-Hardness

Theorem 3.3.1 *Optimization problem (3.38) for distributed estimation subject to a total power constraint, is NP-Hard.*

Proof The proof follows the same procedure as the proof of the Theorem 3.2.1, (see also Appendix D for the detailed proof). ∎

Chapter 4

Joint Sensor Selection and Routing Algorithms: Fine Quantization

In this chapter, we consider the problem of jointly optimizing the sensor selection and routing structure assuming that fine quantization is available for all sensor measurements, so that the fused information (related to the parameter of interest) can be passed, using the multihop routing structure, to the sink node for the final estimation. We solve this problem by using two different approaches, namely, a Fixed-Tree Relaxation-Based Algorithm (FTRA) and a very efficient Iterative Distributed Algorithm (IDA) to optimize the sensor selection and routing structure. We also provide a lower bound for our optimization problem and show that our IDA provides a performance that is close to this bound, and it is substantially superior to the previous approaches presented in the literature. An important result from this work is the fact that because of the interplay between the communication cost and estimation gain when fusing measurements from different sensors, the traditional Shortest Path Tree (SPT) routing structure, widely used in practice, is no longer optimal. To be specific, our routing structure provides a better trade-off between the overall power efficiency and estimation accuracy. We also demonstrate the performance comparison with other related algorithms cited in Chapter 2.

We start, in this case, from the optimization problem (3.20), which has been derived in the previous chapter. If we rewrite the relaxed optimization problem (3.20) from the previous Chapter 3:

$$
\begin{aligned}
\underset{\{b_j^r,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} \quad & \text{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
\text{subject to} \quad & \sum_{j=1,i=\text{parent}_j}^{n+1} b_j^r f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j^r \leq b_i^r,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
& b_{n+1}^r = 1 \\
& P_{\text{gap}} \geq 0 \\
& 0 \leq b_j^r \leq 1
\end{aligned}
\tag{4.1}
$$

4.1 Fixed-Tree Relaxation-Based Algorithm

In this section, we consider a simple low complexity algorithm, denoted as Fixed-Tree Relaxation-Based Algorithm (FTRA). The main idea of this algorithm is to select a subset of sensors and a routing structure by decoupling the estimation process and routing structure controlling the communication cost. For this, first we generate a Shortest Path Tree based on Communication Cost (SPT-CC) rooted at the sink node, which provides the identity of the parent of each sensor, we store them as an edge set $\{(j, i)\}$ and define them as a directed edge $j \to i$, where i is the parent of j. Then, we re-write the relaxed version of the optimization problem (4.1) considering only the sensor selection so that the routing structure used for these sensors will be a subtree T of the SPT-CC that is also rooted at the sink node, which is given by:

$$
\begin{aligned}
\underset{\{b_j^r,T\subset\text{SPT-CC},P_{\text{gap}}\}}{\text{minimize}} \quad & \text{Tr}\left[\left(\sum_{j=1}^{n} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
\text{subject to} \quad & \sum_{j=1,i=\text{parent}_j}^{n+1} b_j^r f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j^r \leq b_i^r,\ i = \text{parent}_j \\
& P_{\text{gap}} \geq 0 \\
& 0 \leq b_j^r \leq 1
\end{aligned}
\tag{4.2}
$$

Notice that the relaxed problem (4.2) is a convex problem, but it is not equivalent to the original problem (3.15) since the solution $\{b_j^{r*}\}_{j=1}^n$ of this problem will not be binary in general. Notice also that in (4.2), we do not consider the variable b_{n+1}^r since SPT-CC is already routed at the sink node and since $\boldsymbol{h}_{n+1}^T = [0, \ldots, 0]$, calculating $\mathrm{Tr}\left[\left(\sum_{j=1}^n \widehat{b}_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right]$ is equivalent to $\mathrm{Tr}\left[\left(\sum_{j=1}^{n+1} \widehat{b}_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right]$. We use this solution to perform a suboptimal subset selection V_T by sorting the optimal values $\{b_j^{r*}\}_{j=1}^n$ in descending order and selecting the subset of K largest b_j^{r*}'s satisfying the power constraint. Then, denoting $\{\widehat{b}_j^r\}_{j=1}^n$ as the binary values such that $\widehat{b}_j^r = 1$ if $j \in V_T$ and $\widehat{b}_j^r = 0$ if $j \notin V_T$, we have that:

$$L_{\mathrm{FTRA}} = \mathrm{Tr}\left[\left(\sum_{j=1}^{n} \widehat{b}_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \geq p^* \tag{4.3}$$

where p^* is the optimal solution of problem (3.15). When making the sorting, because of the constraint $b_j^r \leq b_i^r$, this forces the routing structure to be a tree as long as $P_{\max}$ is large enough and the routing solution will be a subtree of the SPT-CC.

We can obtain the gap, $\delta_{\mathrm{FTRA}} = L_{\mathrm{FTRA}} - L$, using (3.21) and (4.3), in order to provide an assessment about how good suboptimal solution (4.3) is, since $L_{\mathrm{FTRA}} \geq p^* \geq L$.

Algorithm 1 Fixed-Tree Relaxation-Based Algorithm

Require: $P_{\max}$

1. Find SPT-CC using Bellman-Ford algorithm
2. Solve optimization problem (4.2) to find $\{b_j^{r*}\}_{j=1}^n$
3. Sort the optimal values $\{b_j^{r*}\}_{=1}^n$ in descending order
4. Select the subset of sensors corresponding to the largest b_j^{r*}'s while satisfying the power constraint
5. Choose the routing tree $T \subset$ SPT-CC that spans the subset of selected sensors
6. Calculate: L_{FTRA} using (4.3)

4.1.1 Main Disadvantages of the FTRA

The main disadvantages of the FTRA are the following:

1. this algorithm assumes that the SPT-CC is the optimal routing structure from which sensors are selected, thus ignoring the interplay between the communication cost and estimation error.
2. the sensor selection is optimized in a centralized manner, thus making it less convenient to be scalable in WSNs.

As we show in this work, the SPT-CC is not in general the optimal routing structure because of the interplay between the communication cost associated to a routing decision and estimation gain obtained when fusing measurements from different sensors. In other words, each routing decision affects both the communication cost and estimation gain, as already illustrated in the simple example of Figure 1.2.

4.2 Iterative Distributed Algorithm

In this section, we present a scalable Iterative Distributed Algorithm (IDA) that jointly performs the sensor selection and multihop routing allowing a trade-off between the metrics of the communication cost and distortion in estimation. In this case, there is not a pre-selected structured routing tree and our algorithm should iterate to jointly select both sensors and routes. Because of the limited look-ahead ability, our algorithm will provide an overall suboptimal solution to our original non-convex optimization problem (3.15), however, since it takes into account both metrics, as we will see below, this algorithm provides better results than FTRA algorithm, and in fact we show that it performs close to the optimal.

In this algorithm, we activate one sensor at each iteration based on a utility function $\lambda_j^{(i)}$ defined for each sensor j (where sensor j sends its measurement to sensor i), as described below. This utility function, in order to jointly minimize the communication cost and estimation error, leads to the selection of the best local neighbor sensor at each given sensor being processed. Next, we explain how $\lambda_j^{(i)}$ is obtained. Since sensor selection is based on the utility function $\lambda_j^{(i)}$, which we

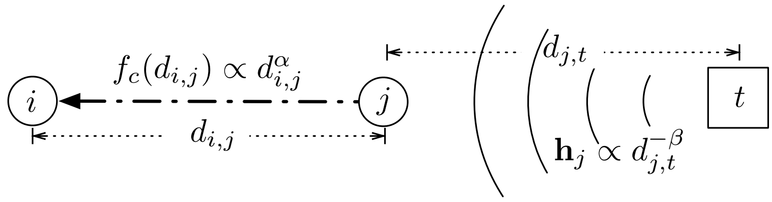


Figure 4.1: The communication cost and estimation gain model.

will show that it depends on distance dependent variables. Since our optimization problem involves minimizing the estimation error under a given constraint on the total power budget, which can be written, in this case, as follows:

$$\begin{aligned} \underset{\{d_{i,j},d_{j,t}\}}{\text{minimize}} \quad & \mathrm{Tr}(\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}) = \mathrm{Tr}\left[\left(\sum_{j=1}^{n} \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\ \text{subject to} \quad & \sum_{j=1}^{n} f_c(d_{i,j}) \leq P_{\max} \end{aligned} \tag{4.4}$$

where variable $d_{i,j}$ is the distance between sensor j to its parent sensor i and variable $d_{j,t}$ is the distance between sensor j to the source target t, as illustrated in Figure 4.1. In problem (4.4), we do not consider sensor selection variable b_j since in this case, in each iteration, we add one sensor and the associated communication cost until the power constraint is satisfied. Also, we do not consider the routing constraint $b_j \leq b_i$ since routing decision is taken locally based on the function $\lambda_j^{(i)}$ and a backbone, which we will see below.

The utility function $\lambda_j^{(i)}$ to be applied is a function that weights two metrics (communication cost and estimation gain) at a given local sensor j being processed, we consider an equivalent form of (4.4) only for this sensor, which is given by:

$$\begin{aligned} \underset{\{d_{i,j},d_{j,t}\}}{\text{minimize}} \quad & \mathrm{Tr}(\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]) = \mathrm{Tr}\left[\left(\boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\ \text{subject to} \quad & f_c(d_{i,j}) \leq P_0 \end{aligned} \tag{4.5}$$

where $P_0 = f_c(d_{norm})$, so that the inequality constraint in (4.5) is always satisfied.

To find an optimal trading-off function, we solve (4.5) analytically. In particular, we can write the Lagrangian function Λ for this problem as:

$$\Lambda(d_{i,j}, d_{j,t}, \lambda_j^{(i)}) = \mathrm{Tr}\big[\mathbf{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]\big) + \lambda_j^{(i)}\big(f_c(d_{i,j}) - P_0\big] \tag{4.6}$$

where $\lambda_j^{(i)}$ is a Lagrange multiplier that trades-off (weights) jointly both metrics. Clearly, the Lagrange multiplier $\lambda_j^{(i)}$ will depend on both variables $d_{i,j}$ (determining the communication cost associated to a link (j,i)) and $d_{j,t}$ (determining the estimation gain), so that it becomes a function of both metrics since both metrics dependent on these variables, that is, $f_c(d_{i,j}) \propto d_{i,j}^{\alpha}$ and $\boldsymbol{h}_j^T = [h_{j1}, \ldots, h_{jm}] \propto d_{j,t}^{-\beta}$ (see Figure 4.1), where $h_{jm} = d_{j,t}^{-\beta} a_{jm}$. Intuitively, for the sensors that are closer to the source target, we should provide more importance to the distortion metric $\boldsymbol{h}_j^T$ and less importance to the communication cost metric $f_c(d_{i,j})$, and the opposite for sensors that are farther from the source target. Taking the Lagrange multiplier $\lambda_j^{(i)} = \mathcal{F}(d_{i,j})\mathcal{G}(d_{j,t})$ as a function of $d_{i,j}$ and $d_{j,t}$, we show next that this function provides an optimal weight to both metrics at a given sensor j being processed. Solving the following two differential equations can derive this function.

The Lagrange dual function $\vartheta(\lambda_j^{(i)})$ is given by:

$$\vartheta(\lambda_j^{(i)}) = \inf_{\{d_{i,j}, d_{j,t}\}} \Lambda(d_{i,j}, d_{j,t}, \lambda_j^{(i)}) \tag{4.7}$$

Denoting $(.)'$ as the symbol of differentiation, then differentiating Λ w.r.t. $d_{i,j}$ yields:

$$\mathcal{F}'(d_{i,j})\mathcal{G}(d_{j,t})\big(f_c(d_{i,j}) - P_0\big) + \mathcal{F}(d_{i,j})\mathcal{G}(d_{j,t})f_c'(d_{i,j}) = 0$$

or, equivalently:

$$\mathcal{F}'(d_{i,j})f_c(d_{i,j}) + \mathcal{F}(d_{i,j})f_c'(d_{i,j}) = \mathcal{F}'(d_{i,j})P_0 \tag{4.8}$$

This can be written as:

$$\big(\mathcal{F}(d_{i,j})f_c(d_{i,j})\big)' = \mathcal{F}'(d_{i,j})P_0 \tag{4.9}$$

Integrating both sides w.r.t. $d_{i,j}$, we have:

$$\mathcal{F}(d_{i,j})f_c(d_{i,j}) = \mathcal{F}(d_{i,j})P_0 + c_1$$

$$\Rightarrow \mathcal{F}(d_{i,j}) = c_1\big(f_c(d_{i,j}) - P_0\big)^{-1} \tag{4.10}$$

where c_1 is a constant.

On the other hand, differentiating Λ w.r.t. $d_{j,t}$, we obtain:

$$\big(\mathrm{Tr}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j])\big)' + \mathcal{F}(d_{i,j})\mathcal{G}'(d_{j,t})\big(f_c(d_{i,j}) - P_0\big) = 0 \tag{4.11}$$

substituting the value of (4.10) in (4.11) provides:

$$\mathcal{G}'(d_{j,t}) = -\frac{1}{c_1}\big(\mathrm{Tr}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j])\big)' \tag{4.12}$$

Integrating both sides w.r.t. $d_{j,t}$ provides:

$$\mathcal{G}(d_{j,t}) = k_1 - \frac{1}{c_1}\mathrm{Tr}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]) \tag{4.13}$$

where k_1 is a constant.

Finally, $\lambda_j^{(i)}$ is given by using (4.10) and (4.13), that is:

$$\begin{aligned} \lambda_j^{(i)} &= \mathcal{F}(d_{i,j})\mathcal{G}(d_{j,t}) \\ \lambda_j^{(i)} &= \big(c - \mathrm{Tr}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j])\big)\big(f_c(d_{i,j}) - P_0\big)^{-1} \\ \Rightarrow \lambda_j^{(i)} &= \big(\mathrm{Tr}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]) - c\big)\big(P_0 - f_c(d_{i,j})\big)^{-1} \end{aligned} \tag{4.14}$$

where $c = k_1c_1$ and $P_0 \geq f_c(d_{i,j})$. Notice that $\lambda_j^{(i)}$ is obtained by minimizing the Lagrangian function Λ and is always positive due to the implicitly constraint in (4.5), therefore minimizing $\lambda_j^{(i)}$ in (4.6) minimizes the objective function in (4.5) under the given power constraint.

A small difficulty that arises in (4.14) is the fact that the covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]$ associated to an individual sensor j is rank-defficient (singular) since obviously rank$\left(\boldsymbol{h}_j\boldsymbol{h}_j^T\right)$=1. In order to avoid singularity of the rank one matrix $\boldsymbol{h}_j\boldsymbol{h}_j^T$ for (4.14), we use the matrix $\mathcal{B} = ||\boldsymbol{h}_j||_2^{-2}\left(\frac{\boldsymbol{h}_j}{||\boldsymbol{h}_j||_2}\right)\left(\frac{\boldsymbol{h}_j^T}{||\boldsymbol{h}_j||_2}\right)$, which is the closest (in Frobenius norm) positive semidefinite matrix to $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[j]$. Notice that $\mathrm{Tr}(\mathcal{B}) = ||\boldsymbol{h}_j||_2^{-2}$, thus (4.14) will be given by:

$$\begin{aligned} \lambda_j^{(i)} &= \big(||\boldsymbol{h}_j||_2^{-2} - c\big)\big(P_0 - f_c(d_{i,j})\big)^{-1} \\ \Rightarrow \lambda_j^{(i)} &= ||\boldsymbol{h}_j||_2^{-2}\big(P_0 - f_c(d_{i,j})\big)^{-1} \end{aligned} \tag{4.15}$$

where we have chosen constant $c = 0$ (this implies choice of $k_1 = 0$) since $||\boldsymbol{h}_j||_2^{-2} \geq 0$, in order to satisfy $\lambda_j^{(i)} \succ 0\ \forall j$.

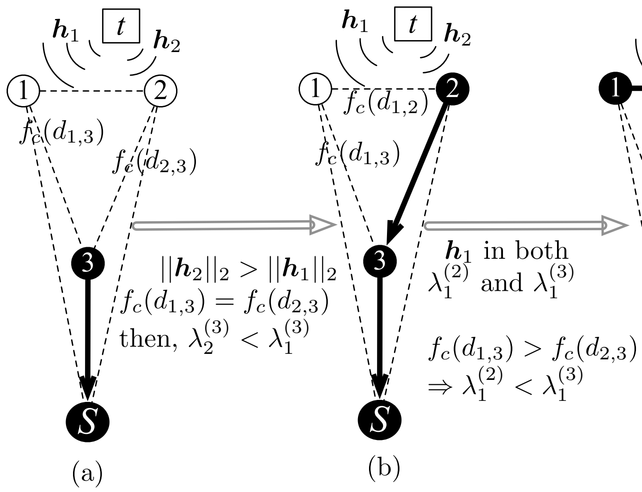


Figure 4.2: Sensor activation based on the utility function $\lambda_j^{(i)}$. In Figure (b), $\lambda_1^{(2)}$ and $\lambda_1^{(3)}$ are the utility functions for sensor 1, with respect to sensor 2 and sensor 3, respectively. Even though these tree structures are very simple, they represent the essential routing decisions that are taken locally at each node.

We illustrate, with a simple example, the activation of new sensors based on the utility function (4.15). In scenario 1, as illustrated in Figure 4.2(a), sensor 3 has two potential neighbor sensors, namely sensor 1 and sensor 2. Calculating $\lambda_1^{(3)}$ and $\lambda_2^{(3)}$ using (4.15), we can observe that $\lambda_2^{(3)} < \lambda_1^{(3)}$, thus in a first iteration, sensor 2 is activated (see Figure 4.2(b)). In scenario 2, as illustrated in Figure 4.2(b), there are two sensors $\{2, 3\}$ with a common neighbor sensor 1. In this case, the utility function $\lambda_1^{(2)}$ is smaller than the utility function $\lambda_1^{(3)}$, thus sensor 1 is activated with a routing link joining sensor 2 as in Figure 4.2(c). There will be some other scenarios such as several sensors calculating their 1-hop neighbors using the same utility function. Then the process will be to pick the best among all 1-hop neighbors, using for example message passing [81] or by gossiping [37] algorithms, based on the minimum utility value and activate the respective neighbor, then go for the next iteration.

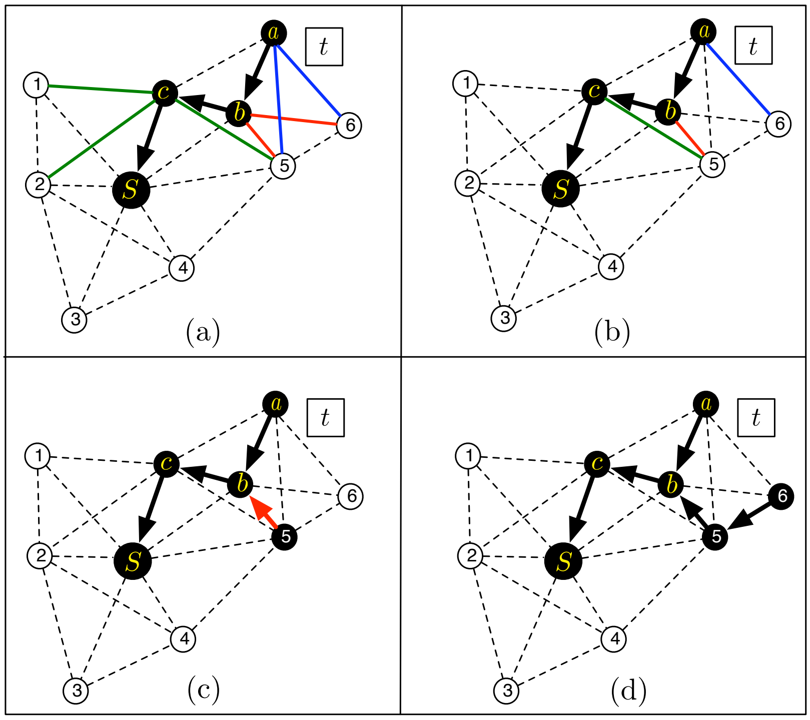


Figure 4.3: Illustration of the best neighbor sensor activation: in Figure 4.3(a), $\{a, b, c\} \in T \subset$ SPT-CC are the sensors forming a backbone, where $\{5, 6\} \in \mathcal{N}(a)$, $\{5, 6\} \in \mathcal{N}(b)$ and $\{1, 2, 5\} \in \mathcal{N}(c)$ are the 1-hop neighbor sets; in Figure 4.3(b), the best 1-hop neighbors $6 \in \mathcal{N}(a)$, $5 \in \mathcal{N}(b)$, and $5 \in \mathcal{N}(c)$ are selected based on metric $\lambda_j^{(i)}$; in Figure 4.3(c), the best neighbor sensor $5 \in \mathcal{N}(b)$ among all 1-hop neighbors $\{6, 5, 5\}$ is activated; and Figure 4.3(d) shows the selection of the next sensor 6 if the same previous steps are repeated. This process continues until power budget is utilized.

An important result we will observe is that, if we iterate to activate all sensors in the given field, it will provide the same routing structure as explained in Lemma 3.2.2 (for lower bound), which is the MST routed at the sink node. Therefore, selecting a subset of sensors and their associated routing structure using IDA algorithm will provide a better suboptimal solution to our main non-convex optimization problem (3.15).

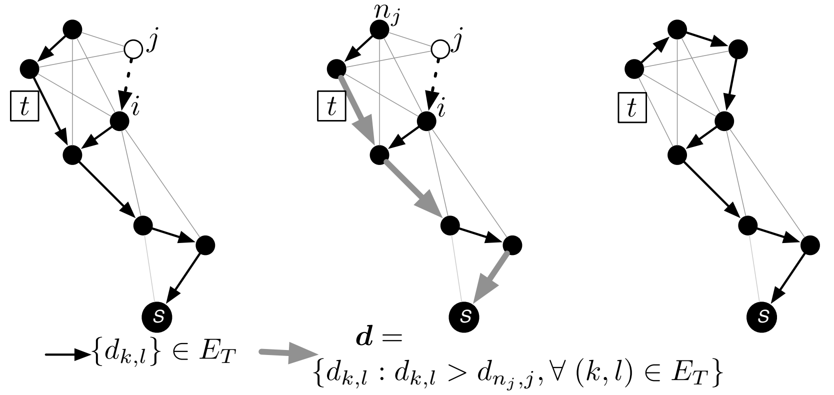


Figure 4.4: Successful backtracking operation performed by the algorithm.

The idea of the IDA algorithm is as follows: The whole process starts from the sensor that detects the phenomenon with the largest SNR, equivalently, the largest $||\boldsymbol{h}_j||$ (sensor a as shown in Figure 4.3(a)). Notice that the most power efficient way to send this measurement to the sink node is using the corresponding shortest path along the SPT-CC. This motivates us to select this single path as an initialization of our algorithm, forming a backbone (thick arrow path in Figure 4.3(a) from sensor a to the sink node S). Thus, we initially select all the intermediate sensors in this path and their associated routes from the SPT-CC (as illustrated in Figure 4.3(a) by the intermediate sensors b and c). Next, each of the currently selected sensors calculates locally the utility function (4.15) for all its 1-hop neighbors in the original connectivity graph (illustrated by thin continuous links in Figure 4.3(a)) and stores the identity of the neighbor that minimizes its associated utility function (such as neighbor 5 of sensor c, neighbor 5 of sensor b, and neighbor 6 of sensor a in Figure 4.3(b)). Finally, we activate the best neighbor that minimizes this utility function among all 1-hop neighbors (in this case, neighbor 5 of sensor b in Figure 4.3(c)). We iterate this process and activate one neighbor sensor in each iteration until the total power budget is utilized. The formal description of the whole process is provided in Algorithm 2.

Algorithm 2 Iterative Distributed Algorithm

Require: $P_{\max}$

$T = (V_T, E_T) \subset G$, constituting a multihop path (backbone) rooted at the sink node S from the sensor that is nearest to the source target t, where V_T and E_T are the subset of sensors and edges in the backbone, respectively;

j = next sensor to be activated and $j \in \mathcal{N}(i)$;

P_{IDA} = total communication cost of T;

n_j = sensor nearest to sensor j;

initialize: $w = 1$; $q = 1$;

T_{sink} = set of routing structures for V_T routed at the sink node;

- **while** $P_{\text{IDA}} <= P_{\max}$ **do**
  $(j, i) = \underset{\{i \in V_T, j \in \mathcal{N}(i) \backslash V_T\}}{\arg\min} \big[||\boldsymbol{h}_j||_2^{-2}\big(P_0 - wf_c(d_{i,j})\big)^{-1}\big]^+$
  $E_T = E_T \cup (j, i)$; $V_T = V_T \cup j$;
  $P_{\text{IDA}} = P_{\text{IDA}} + f_c(d_{i,j})$

  begin Backtracking
  Find $n_j \in V_T \backslash i$
  if $d_{n_j,j} \notin T$ **then**
  Find: $\boldsymbol{d} = \{d_{k,l} : d_{k,l} > d_{n_j,j}, \forall\, (k, l) \in E_T\}$;
  $\tilde{d}_{k,l} = \max\,\{d_{k,l} : (E_T \backslash d_{k,l} \cup d_{n_j,j}) \in T_{sink}, \forall\, d_{k,l} \in \boldsymbol{d}\}$;
  Update: $P_{\text{IDA}} = P_{\text{IDA}} - f_c(\tilde{d}_{k,l}) + f_c(d_{n_j,j})$;
  end if
  end Backtracking

  if $P_{\text{tot}} > P_{\max}$ **then**
  $E_T = E_T \backslash (j, i)$; $V_T = V_T \backslash j$;
  $P_{\text{IDA}} = P_{\text{IDA}} - f_c(d_{i,j})$;
  $P_{\text{gap}} = P_{\max} - P_{\text{IDA}}$;
  $w = \left[\frac{P_{\max}}{P_{\text{gap}}}\right]^q$; $q = q + 1$;
  else
  $q = 1$; $w = 1$;
  end if

- **end while**
  Calculate: $L_{\text{IDA}} = \text{Tr}\big[\big(\sum_{j \in V_T} \boldsymbol{h}_j \boldsymbol{h}_j^T\big)^{-1}\big]$

4.2.1 Backtracking Operation

Due to the greedy nature of our algorithm, while performing the selection of a new neighbor sensor, we also check whether an alternative route through the new selected sensor is more power efficient or not. This operation is called backtracking. We start form the current activated sensor j and then find the nearest sensor $n_j \in V_T \backslash i$ to sensor j based on the communication cost (see Figure 4.4), where i is the parent of j. Then, if link $d_{n_j,j} \in E_T$, then there is no other best alternative route, otherwise if $d_{n_j,j} \notin E_T$, then we consider the list $\boldsymbol{d}$ of all the links in E_T that have a larger communication cost than the communication cost of link $d_{n_j,j}$ and find the largest value from the list $\boldsymbol{d}$ such that while swapping it with $d_{n_j,j}$, ensures that T is still routed at the sink node. We update E_T by removing the largest link $\tilde{d}_{k,l}$, adding link $d_{n_j,j}$ and updating the total communication cost by $P_{\text{IDA}} = P_{\text{IDA}} - f_c(\tilde{d}_{k,l}) + f_c(d_{n_j,j})$, where $f_c(\tilde{d}_{k,l})$ is the communication cost of the removed link $\tilde{d}_{k,l}$. Finally, we update the identity of parents of each updated edges in T such that tree $T \in T_{sink}$, where T_{sink} is the set of routing structures for V_T routed at the sink node. Since, $f_c(\tilde{d}_{k,l}) > f_c(d_{n_j,j})$, this results in a reduction of the total communication cost with the same selected subset V_T.

In each iteration, after the backtracking operation, we also check if the P_{IDA} corresponding to the current subset of selected sensors is greater than $P_{\max}$. In this case, we need to swap the last activated sensor j with the next best sensor in terms of the weighted cost metric $f_c(d_{i,j})$ given by (4.15) while causing a total power budget $P_{\text{IDA}} < P_{\max}$. In order to perform this, we remove the last chosen sensor j as well as the associated communication cost $f_c(d_{i,j})$. This ensures that $P_{\text{IDA}} < P_{\max}$, generating a power gap $P_{\text{gap}} = P_{\max} - P_{\text{IDA}}$. Then, we re-iterate our algorithm to bring P_{gap} as close to zero as possible. For this, first we define $[x]^+ = x$ if $x \geq 0$ and $[x]^+ = \infty$ otherwise. Then, we update our utility function by placing a higher weight $w \propto \frac{P_{\max}}{P_{\text{gap}}}$ on the communication cost metric, so that if we have a very small gap P_{gap}, we can set an even higher weight on the communication cost in the utility function, $[||\mathbf{h}_j||_2^{-2}\big(P_0 - wf_c(d_{i,j})\big)^{-1}]^+$. With the new selected sensor, we restart the search for other sensors until the gap is filled as much as possible, as explained in Algorithm 2.

4.3 Implementation Issues and Complexity

At the beginning, we initialize a routing structure with the SPT-CC and the information related to the network connectivity graph G at the sink node. Then, the sink node broadcasts to each sensor separately a packet carrying the identities of 1-hop neighbors and the identity of the parent node. Thus, each sensor node is aware of identities of its 1-hop neighbors and the identity of its parent node[1]. Then, the WSN operates alternatively and periodically with two main phases, as shown in Figure 4.5. In the detection phase, all sensor nodes change their status periodically from sleeping to listening and from listening to sleeping mode until some of them (i.e., sensors near the source target) detect a physical quantity greater than a pre-defined threshold. In practice, these two sleeping and listening modes can be divided so that the network lifetime is controlled.

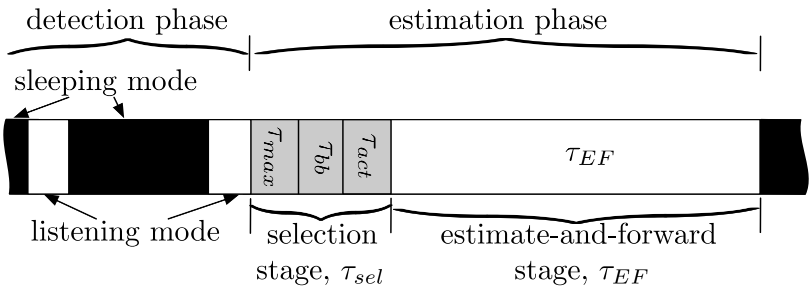


Figure 4.5: In our IDA, the estimation task is being carried out during the estimation phase, which is divided into two periodic segments. The first one is a selection stage of duration τ_{sel} where a subset of sensors V_T and an associated routing structure T are selected for a given total power budget. Then, the selected sensors perform the EF operation computing the total distortion in estimation distributively and incrementally during the estimate-and-forward stage of duration τ_{EF}. In general, $\tau_{EF} \gg \tau_{sel}$, thus the total communication cost is usually dominated by the estimate-and-forward stage.

Then, as soon as a sensor detects the presence of some source target, information about the occurring event is passed to the sink node via a

[1] The other option, which can be used locally at each node to obtain the identity of the parent node with the SPT-CC, requires the entire graph G.

single multihop path belonging to the SPT-CC. Then, the estimation phase starts, which is divided into two main periodic segments: first, a selection stage of duration τ_{sel}, where a subset of sensors V_T is selected together with the best associated routing structure $T = (V_T, E_T)$; and second, the estimate-and-forward stage τ_{EF}, where the EF scheme is applied distributively and incrementally on the subset V_T to estimate the parameter of interest while routing the fused estimation towards the sink node. At the end of estimate-and-forward stage, the WSN returns to the detection phase so that it can be ready to perform another estimation task. The duration τ_{EF} is usually in practice much longer than τ_{sel}. In most applications, a small fraction of the total power budget is consumed during the selection stage [84].

4.3.1 IDA Algorithm

In this algorithm, the selection stage τ_{sel} is further divided into three sub-stages: maximum SNR detection stage of duration τ_{max}; backbone path creation stage of duration τ_{bb}; and joint sensor selection (activation) and routing stage of duration τ_{act}, as shown in Figure 4.5.

- *Maximum SNR Detection Stage:* The sensor that senses with the highest SNR need to be identified; this can be obtained by either message passing [81] or by gossiping [37] algorithms. For example, if we define a threshold[1] SNR S_{th}, then only the sensors having an SNR above S_{th} (i.e., sensors within a certain distance from the source target) become active. Then, each active sensor starts messaging a value proportional to their own SNR to their nearest 1-hop active sensor. For example, consider that sensor j unicasts its SNR value S_j to its 1-hop nearest neighbor sensor i. Then, depending on the conditions $S_j > S_i$ or $S_j < S_i$, the sensor with higher SNR will be kept active and other will be set inactive. This process can continue until only one sensor is left active, which will be the one with the highest SNR.

- *Backbone Path Creation Stage:* Without loss of generality, let us assume that sensor i is the one with the highest SNR. Then the backbone path from i to the sink node is created within a certain

[1] We choose S_{th} in such a way that around 5% sensors remain active so that the communication overhead can be reduced.

time period τ_{bb}. In order to do this, we have to activate the parent of i, then the parent of the parent of i, and so on until the sink node is reached, where the identities of parents are taken from the SPT-CC. Furthermore, the initial backbone path provides an initial subset V_{bb} of sensors and a backbone path $T = (V_T, E_T) = (V_{bb}, E_{bb})$ with edges E_{bb}, and thus information needs to be forwarded back to all sensors in V_{bb}.

- *Joint Sensor Selection and Routing Stage:* With the knowledge of the initial subset of selected sensors V_T and a path $T = (V_T, E_T)$, we start activating incrementally other sensors until the given total power budget is utilized. In order to do this, Algorithm 2 is executed locally and synchronously at each selected sensor. First, as shown in Figure 4.3(b), each sensor calculates the utility function given by (4.15), namely $\lambda_j^{(i)}$, for all its current 1-hop unselected available neighbors and stores the identity of the neighbor that minimizes its associated utility value. Then, each selected sensor in V_T needs to use again the same message passing or gossiping method to find out the identity of the best 1-hop neighbor among all the locally selected 1-hop neighbors, as shown in Figure 4.3(c). For example, as shown in Figure 4.3(a), sensors $\{a, b, c\} \in V_T$ are in the backbone. Then, sensor a (leaf node) has to pass information to its parent sensor b, which includes a utility value $\lambda_6^{(a)}$ corresponding to its best 1-hop neighbor $6 \in \mathcal{N}(a)$ and the edge $(6, a)$. Then, sensor b has to compare the received utility value $\lambda_6^{(a)}$ to its own utility value $\lambda_5^{(b)}$ (corresponding to its best 1-hop neighbor $5 \in \mathcal{N}(b)$) and pass all the information containing the minimum utility value and corresponding edge of the chosen best 1-hop neighbor, to its parent sensor c. Notice that if sensor b is receiving more than one information set from different neighbors (which is known to sensor b), then b has to wait first until it receives all the information from the various neighbors, and then find the minimum utility value among all of them and the corresponding edge of the best 1-hop neighbor, to forward to its parent sensor c. Then, sensor c follows the same procedure to find and forward the information set associated to the minimum utility value. Once this is done at each sensor in V_T, the information of the edge of the best 1-hop neighbor has to be forwarded back to all other sensors of V_T so that the next step of the algorithm at each sensor can be processed. The next step is to include this best 1-hop sensor into

V_T, update the routing subtree T and update the total incurred communication cost P_{IDA}.

Then, the next step is the backtracking operation, as explained before Subsection 4.2.1, which is performed locally at each previously selected sensor, and in case there is an improvement, the sets V_T and E_T of the current subtree T and P_{IDA} have to be updated at each sensor. At each iteration, after the backtracking operation, we also check if $P_{\text{IDA}} > P_{\max}$, in which case, in order to utilize $P_{\max}$ effectively, we need to swap the last activated sensor with the next best sensor based on the weighted cost metric $f_c(d_{i,j})$ given by (4.15) causing the total power budget $P_{\text{IDA}} < P_{\max}$, as explained earlier in Algorithm 2. At the end of Algorithm 2, the rest of sensor nodes is set back to inactive state.

- *Estimate-and-Forward Stage:* In this stage, each selected sensor is aware of its children and its parent in the selected routing tree after the sensor selection and final routing structure has been generated. Thus, each sensor waits until it receives all the information from its children, and then applies the EF scheme to fuse all the information together with its own measurement. Then, it forwards the fused information to its parent node on the generated multi-hop routing structure. The estimation process starts with the leaf nodes, let us denote a leaf node as sensor j. Then, we send a rank-1 matrix $\boldsymbol{h}_j\boldsymbol{h}_j^T$ to its parent sensor. We follow these steps until a full-rank matrix is formed at an intermediate sensor. This is necessary in order to be able to compute the covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ properly. Since $\boldsymbol{h}_j \in \mathbb{R}^m$, each matrix $\boldsymbol{h}_j\boldsymbol{h}_j^T \in \mathbb{R}^{m\times m}$ is a rank-1 matrix, thus adding the first m rank-1 matrix measurements and assuming that the measurement vectors $\boldsymbol{h}_i^T = [h_{i1}, \ldots, h_{im}]$ are linearly independent, this will provide a full-rank $m \times m$ matrix, thus invertible. Since $m \ll n$ in our scenario, in practice, we need to perform initially this local summation for only a few sensors. The m-th chosen sensor calculates an accumulative matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m] = \left(\sum_{i=1}^{m} \boldsymbol{h}_i\boldsymbol{h}_i^T\right)^{-1}$, which is invertible. Notice that a leaf sensor always needs to send a rank-1 matrix to its parent, and then its parent uses (3.9) to fuse all the measurements from its child sensors together. For the $(m+1)$-th sensor fusing all its measurements, corresponding the associated error covariance matrix, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m+1]$, which is based on this aggregation, will be given by:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m+1] = \left(\boldsymbol{I} - \frac{\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m]\boldsymbol{h}_{m+1}\boldsymbol{h}_{m+1}^T}{\sigma_{m+1}^2 + \boldsymbol{h}_{m+1}^T\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m]\boldsymbol{h}_{m+1}} \right) \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m] \tag{4.16}$$

which is equivalent to calculating $\left(\sum_{i=1}^{m+1} \boldsymbol{h}_i \boldsymbol{h}_i^T \right)^{-1}$ in a single step. Notice that in (4.16), we do not need to take the inverse of any matrix, which is computationally convenient. The computational effort required to calculate $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m+1] \in \mathbb{R}^{m\times m}$ given $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}[1:m]$ is $O(m^3)$. This process can be done multiple times to perform several estimation tasks within this period.

4.3.2 FTRA Algorithm

In the case of the FTRA algorithm, as soon as the WSN enters into the estimation phase, at the beginning of the period τ_{sel}, all sensor nodes are required to send their measurements to the sink node using the SPT-CC. It is important to note that in the FTRA, the selection stage is not composed of three sub stages. Since in this initial step, the routing structure is the SPT-CC, the most power efficient way to send these measurements is using the MF scheme. Notice that the MF scheme is required only to collect the individual measurements from each sensor. Then, the FTRA algorithm is executed at the sink node to find the best subset of sensors V_T and associated routing structure $T = (V_T, E_T)$ (subtree of the SPT-CC) for a given total power budget $P_{\max}$. Later, the sink node broadcasts the indices associated to the selected sensors, so that the rest of the sensor nodes are set back to inactive state. In addition, the information $T = (V_T, E_T)$ is also broadcasted to each active node by the sink node. Once this operation is performed, each selected sensor node performs the EF operation for distributed estimation during the estimate-and-forward stage of duration τ_{EF}, similarly as described before for the IDA algorithm.

4.3.3 Complexity Analysis

- *IDA Algorithm:* The complexities in the different stages are the following: SPT-CC from Bellman-Ford algorithm takes $O(Mn)$ operations, where $M = |E|$ and $n = |V|$ are the number of edges and vertices, respectively. The maximum SNR detection stage takes $O(2(K_1 - 1))$ operations, where K_1 is the number of activated sensors, that is, sensors with the SNR higher than S_{th}. The backbone path creation stage takes $O(2K_2)$ operations, where K_2 is the number of sensors in the backbone path. And finally, the joint sensor selection and multihop routing generation stage involves: sensor selection, backtracking, and a certain number of iterations. Assuming that the number of iterations is K_3, and the average number of 1-hop neighbors is K_4, then finding the best 1-hop neighbor at any given sensor node takes $O(K_4 \log K_4)$, thus finding the best 1-hop neighbor takes $O(2(K_2 + K_3)K_4 \log K_4)$. The backtracking takes $O(K_3(K_2 + K_3)\log(K_2 + K_3))$ operations. Thus, the overall computational cost has order $O(n^2 \log n)$.

- *FTRA Algorithm:* The main computational complexity in the FTRA comprises the following parts: (a) SPT-CC using Bellman-Ford algorithm; (b) the MF operation; (c) solving optimization problem (4.2) at the sink node; and (d) sorting the relaxed values of $\{b_j^{r*}\}_{j=1}^{n}$ in descending order. Bellman-Ford runs with a complexity of order $O(Mn)$. The MF scheme requires $O(n + \tilde{n})$, where $\tilde{n} : n < \tilde{n} \ll n^2$ is the total number of hops required by the MF scheme, thus requires $O(n^2)$, while solving the optimization problem (4.2) (same complexity as in problem (3.22)) comprises computing the Newton step $\Delta \mathbf{b}_{nt}^r$ at each iteration. First, we calculate $\sum_{j=1}^{n} b_j^r \mathbf{h}_j \mathbf{h}_j^T$, which costs $O(nm^2)$, and compute its Cholesky factorization, which costs $O(m^3)$. Then, we need to compute the Hessian $\nabla^2 \phi$ of the function ϕ (3.22) and run an iteration of the Newton method, which costs $O(n^2 m)$. And then, we also compute the Cholesky factorization for the Hessian, which costs $O(n^3)$. Once we have computed the Cholesky factorization of $\nabla^2 \phi$, we can compute $\Delta \mathbf{b}_{nt}^r$ with cost $O(n^2)$. Thus, the overall computational cost has order $O(n^3)$.

Furthermore, estimate-and-forward stage takes $O(Km^3)$ operation if K sensors are selected. Thus, in this stage, the overall computational cost has order $O(nm^3)$.

4.4 Simulation Results

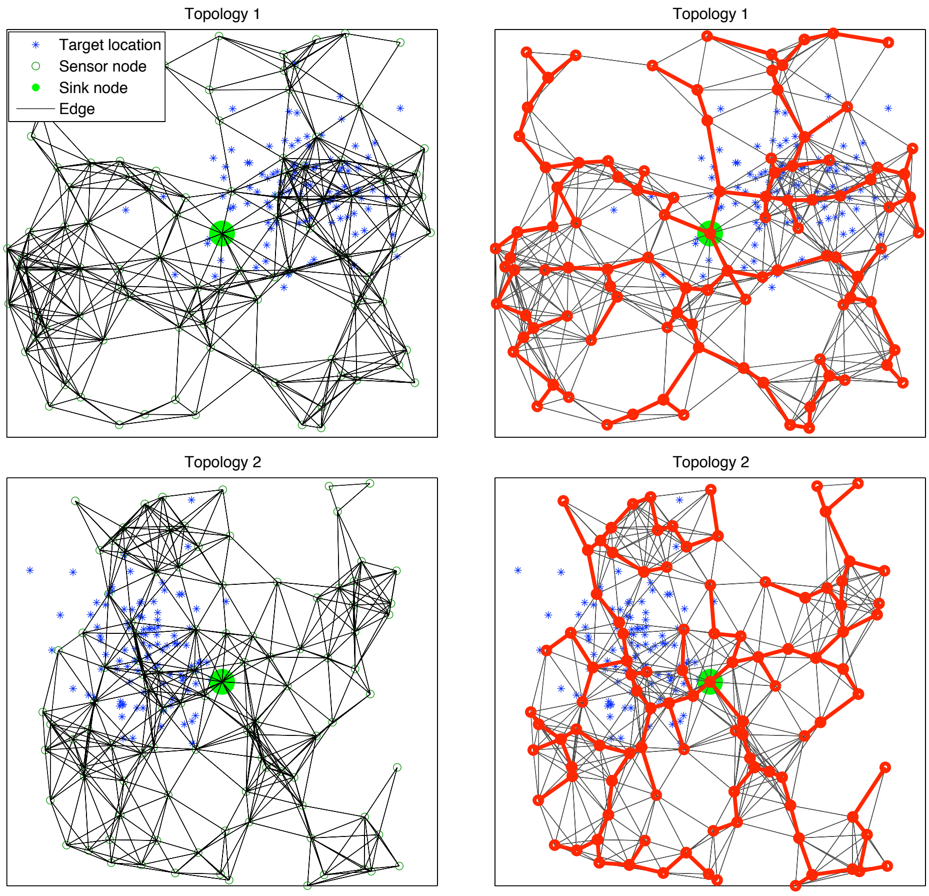


Figure 4.6: The topology examples used in the simulation, where thin edges belong to the network connectivity graph and thick edges (right hand side) represent the SPT-CC. The sink node is located at the center.

In this section, we show the performance comparison of the proposed algorithms through numerical simulations. We consider a WSN with $n = 100$ randomly deployed sensors in a square region and a vector of $m = 5$ parameters to be estimated. The distance between a sensor and any potential neighbor can be expressed as ξd_{norm}, where $\xi \in (0, 1]$ is uniformly distributed and d_{norm} is a normalizing factor. For the sake of simplicity, we assume $\alpha = 4$, $N_j = 1 \forall j$ for normalization and $\mu = 1$ for the communication cost model, thus $f_c(d_{i,j}) = \frac{S_0(2^B-1)}{g_{i,j}}$. We assume that each row vector $\boldsymbol{h}_j^T = [h_{j1}, \ldots, h_{jm}]$ in matrix $\boldsymbol{H}$ follows the usual signal strength decay model [58], where $h_{jm} = d_{j,t}^{-\beta} a_{jm}$, $d_{j,t}$ is the distance from a particular sensor node j to the source target t, and $\beta = 2$ is the signal decay exponent, which is assumed to be known (or estimated via training sequences [58; 105]). In order to account for the randomness in the observations, we consider two different distributions of vector elements $[a_{j1}, \ldots, a_{jm}]$, namely, independent uniform i.i.d. distribution $\mathcal{U}(0.25, 0.50)$ and a normal distribution $\mathcal{N}(0.375 \times \mathbf{1}, 5.8 \times 10^{-3}\boldsymbol{I})$ with the same mean and variance for the various components $a_{j1}, \ldots, a_{jm}$ in both cases. In order to demonstrate the robustness of the algorithms, we consider 100 expected locations (one per sensor, as shown in Figure 4.6) of the source target, taking these locations to be i.i.d. from a Gaussian distribution $\mathcal{N}(\boldsymbol{s}_{2\times1}, \boldsymbol{I}_{2\times2})$, where $\boldsymbol{s}$ is the mean of the expected location values. We test our algorithms using 100 different network topologies with the sink node located at the center of the region. For each topology, we generate different expected location values (see Figure 4.6 with topologies 1 and 2 out of 100 topologies used in the simulation) corresponding to the values of mean location $\boldsymbol{s}$. For each network topology, we have executed the algorithms for a range of maximum $P_{\max}$.

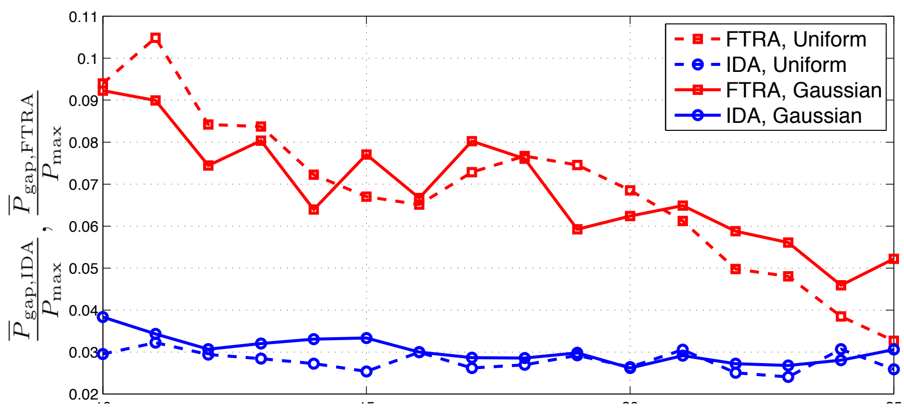
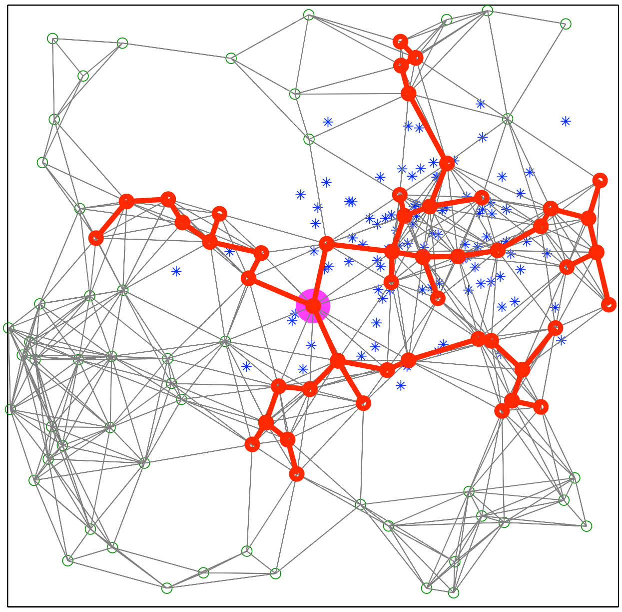


Figure 4.7: Power gap comparison for both algorithms: FTRA and IDA.



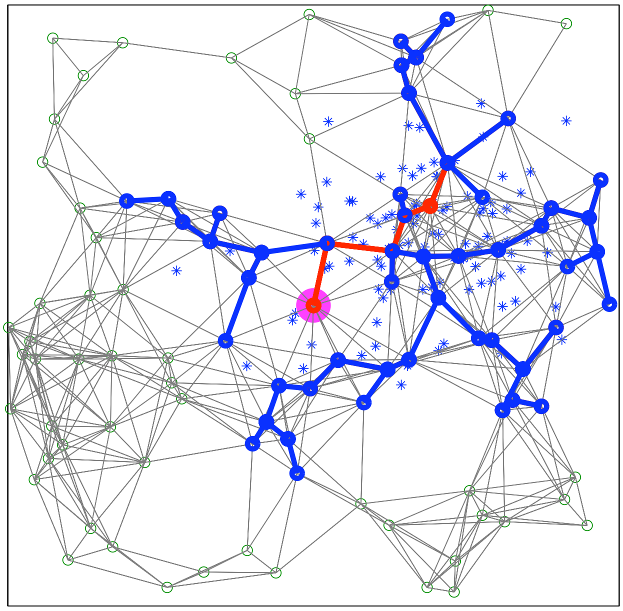


Figure 4.8: The subset of selected sensors and routing structure for FTRA (top) and for IDA (bottom) for a given power budget $P_{\max} = 20$. Path marked by dash red links in IDA is a backbone. Big circle at the center is the sink node.

In order to show the non-optimality of the FTRA, we calculate the average power gaps $\overline{P}_{\text{gap}}$ for both algorithms. We calculate the average gaps $\overline{P}_{\text{gap,FTRA}} = \overline{P_{\max} - P_{\text{FTRA}}}$ and $\overline{P}_{\text{gap,IDA}} = \overline{P_{\max} - P_{\text{IDA}}}$ using the estimated average of the gaps over all 100 topologies, that is, $\overline{P}_{\text{gap}} = \frac{1}{100}\sum_{i=1}^{100} P_{\text{gap},i} = \frac{1}{100}\sum_{i=1}^{100}(P_{\max} - P_{\text{gap},i})$. Figure 4.7 shows the percentage power gaps that results from the proposed algorithms. In many situations, we see an important gap between the maximum power allowed $P_{\max}$ and the total incurred power P_{FTRA} obtained when using the FTRA. It can also be seen that in the case of the IDA algorithm, the resulting gaps are small regardless of the different distributions used.

In Figure 4.8 (bottom), we observe that the IDA algorithm generates a different routing structure, due to the interplay between the estimation error and communication cost. It activates more informative sensors since some links that require higher communication cost are not present in it as opposed to the FTRA algorithm, where the routing structure is fixed and optimized independently as well as IDA efficiently utilizes the given power budget as shown in Figure 4.7.

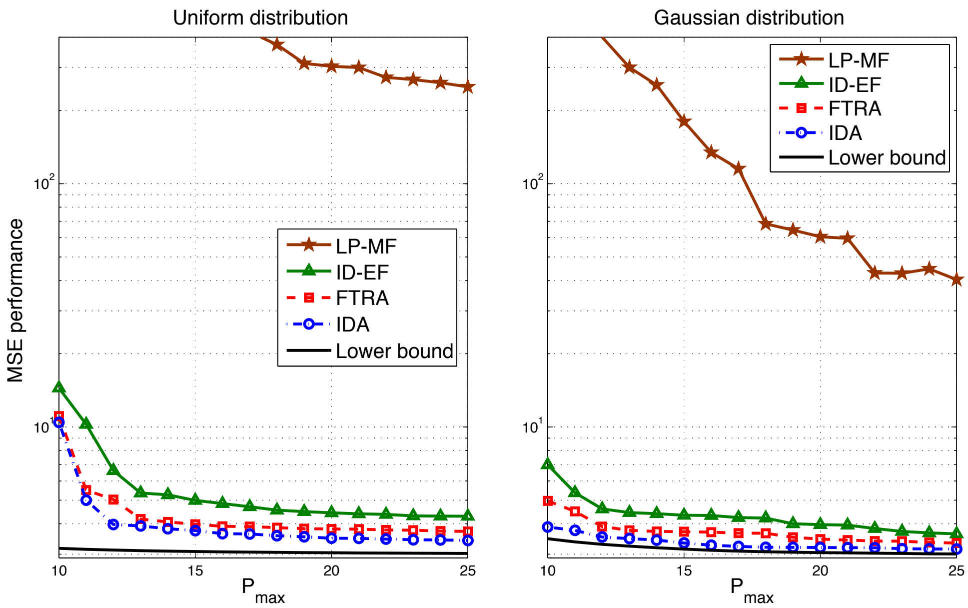


Figure 4.9: Comparison of IDA and other relevant algorithms, together with the lower bound L given by (3.21).

In order to show the MSE performance obtained when selecting a subset of sensors and a corresponding routing structure, we also evaluate the lower bound L given by (3.21). Let us denote L_{FTRA} the objective value corresponding to the solution obtained by the FTRA algorithm and L_{IDA} the objective value corresponding to the solution obtained by the IDA algorithm. Figure 4.9 shows L_{LP-MF}, L_{ID-EF}, L_{FRTA}, L_{IDA} and L, where L_{LP-MF} is the solution from the LP based algorithm given in [61], and L_{ID-EF} is the solution obtained from the ID based algorithm given in [84]. Notice that the solution L_{LP-MF} is based on the MF scheme but jointly optimizing both metrics, whereas the solution L_{ID-EF} is based on the EF scheme but independently optimizing both metrics without any trade-off. Figure 4.9 shows the performance in MSE for different values of the maximum power budgets $P_{\max}$ with the sink located at the center. Overall our IDA outperforms the FTRA algorithm as compared to the lower bound L, depending on the available power budget in both the distribution scenarios. Moreover, our proposed algorithms outperform the algorithms LP-MF and ID-EF, as illustrated in Figure 4.9.

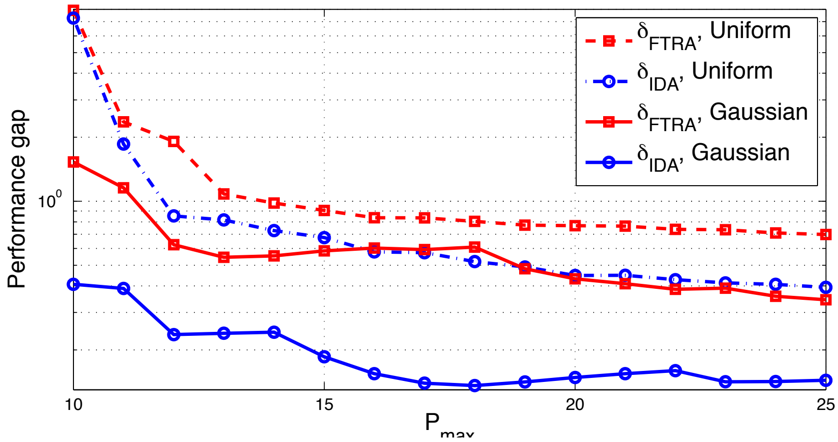
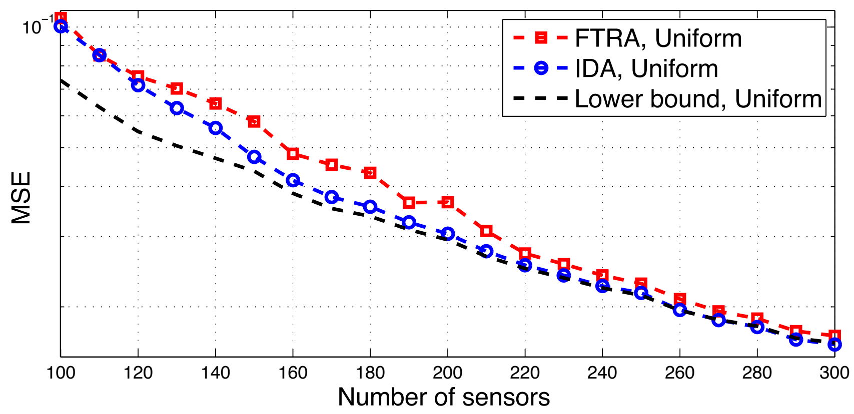


Figure 4.10: Gaps $\delta_{FTRA} = L_{FTRA} - L$ and $\delta_{IDA} = L_{IDA} - L$.

Figure 4.10 shows the gaps $\delta_{FTRA} = L_{FTRA} - L$ and $\delta_{IDA} = L_{IDA} - L$, where it can be seen that the IDA is close to the lower bound when the power budget is large enough. It can be also seen from these results that the SPT-CC routing structure is not optimal in general.



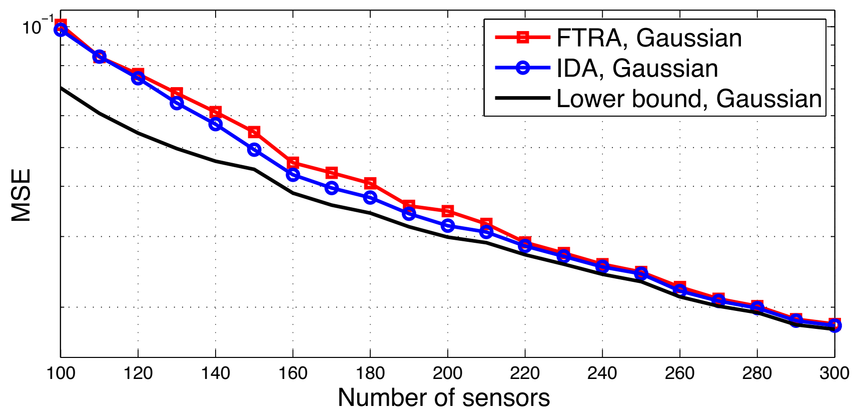


Figure 4.11: The MSE of the estimation obtained by the algorithms FTRA and IDA for different network sizes (from 100 sensors to 300 sensors) for the given fixed total power budget $P_{\max} = 15$.

In Figure 4.11, we illustrate how the distortion in the estimation behaves when considering denser networks, that is, networks with more sensors for the same area of deployment (thus, higher density) while keeping the total power budget fixed to $P_{\max} = 15$.

All the above results show that the performance of the algorithms are independent of the distributions used for $[a_{j1}, \ldots, a_{jm}]$. It further shows that the proposed algorithms are suitable for the different type of sensor measurements.

4.5 Tracking of Moving Target Sources

In this section, we present the problem formulation and a power efficient solution approach for the tracking of a moving target source, as an immediate extension of our work presented in this chapter. For the sake of simplicity, we assume that fine quantization is available to all sensor measurements and that a given total power budget is available for our optimization framework. As it is shown in this chapter, the approach of jointly selecting a subset of sensors and a multihop routing structure enhances the overall system lifetime. In the case of the moving target source, our goal is to find, at each time instance the optimal subset of sensors and an associated multihop routing structure so that the average estimation error is minimized for a given power budget.

4.5.1 Basic Assumptions and Definitions

We describe now the process of building a single-target tracking model and its problem formulation. First, we represent the source target motion dynamics by the discrete time Gauss-Markov [52] state model:

$$\boldsymbol{\theta}(k) = \boldsymbol{T}(k-1)\boldsymbol{\theta}(k-1) + \boldsymbol{z}_{\boldsymbol{\theta}}(k) \tag{4.17}$$

where $\boldsymbol{\theta}(k) = [x_{1t}(k),\ x_{2t}(k),\ x'_{1t}(k),\ x'_{2t}(k)]^T$ represents the state of the target at time $\tau_k = k\Delta\tau$ where $(x_{1t},\ x_{2t})$ is the coordinates of the location of the target t and x'_{1t}, x'_{2t} are the velocities of the target along the x and y axes, respectively. The matrix $\boldsymbol{T}$ is the transition matrix from time τ_{k-1} to τ_k. We assume a matrix $\boldsymbol{T}$ is given by:

$$\boldsymbol{T} = \begin{bmatrix} 1 & 0 & \Delta\tau & 0 \\ 0 & 1 & 0 & \Delta\tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.18}$$

which models a single mobility pattern given by:

$$\boldsymbol{\theta}(k) = \begin{bmatrix} x_{1t}(k-1) + \Delta\tau x'_{1t}(k-1) \\ x_{2t}(k-1) + \Delta\tau x'_{2t}(k-1) \\ x'_{1t}(k-1) \\ x'_{2t}(k-1) \end{bmatrix} + \boldsymbol{z}_{\boldsymbol{\theta}}(k) \tag{4.19}$$

where $\Delta\tau$ is the sampling period and $\boldsymbol{z}_{\boldsymbol{\theta}}(k)$ represents the Gaussian noise vector present in the system, and it is usually referred as the *driving* or *excitation noise*. We assume that the noise vector $\boldsymbol{z_\theta}(k)$ to be independent and white Gaussian with distribution $\mathcal{N}(0, \boldsymbol{C_\theta}(k)) = \sigma_\theta^2 \boldsymbol{I}$. Note that the transition matrix (4.18), for simplicity, is chosen to move in a straight line with constant velocity. Notice also that varying velocity can be attained by changing the third and fourth rows of matrix $\boldsymbol{T}$ by $[0\ 0\ 1 + \frac{1}{\Delta\tau}\ 0]$ and $[0\ 0\ 0\ 1 + \frac{1}{\Delta\tau}]$, respectively. Notice also that by changing $\boldsymbol{T}$, we can have different piece-wise straight trajectories or varying speeds.

On the other hand, the discrete time state equation for sensor measurements is given by:

$$\boldsymbol{y}(k) = \boldsymbol{h}(\boldsymbol{\theta}(k)) + \boldsymbol{z_y}(k) \tag{4.20}$$

where $\boldsymbol{y}(k) = [y_1(k),\ \ldots,\ y_K(k)]^T$ and K is the size of the selected subset of chosen (active) sensors, which depends on the allocated total power budget for each sampling period. Let us assume power budget is fixed for each sampling period as $P_{\max}$.

The individual j-th sensor measurement can be written from (4.20) as:

$$y_j(k) = \frac{g_{j,t}}{[(x_{1j}(k) - x_{1t}(k))^2 + (x_{2j}(k) - x_{2t}(k))^2]^{\beta/2}} + z_{y_j}(k) \tag{4.21}$$

where $h_j = \frac{g_{j,t}}{[(x_{1j}(k)-x_{1t}(k))^2+(x_{2j}(k)-x_{2t}(k))^2]^{\beta/2}}$ and (x_{1j}, x_{2j}) is the Cartesian coordinates of j-th sensor, which are known to sensor node j and $g_{j,t}$ is the acoustic energy generated (assuming omnidirectional) from the target t, as received at sensor node j. We assume, in this case, the noise vector $\boldsymbol{z_y}(k)$ independent compounds and white Gaussian with distribution $\mathcal{N}(0, \boldsymbol{C_y}(k)) = \sigma_{\boldsymbol{y}}^2 \boldsymbol{I}$.

4.5.2 Filtering for Target State Estimation

In order to estimate the target state, given a subset of active sensors, as usual we use an estimation filter. Notice that each individual sensor measurement (as given by (4.21)) is a nonlinear function of the parameter of interest with added Gaussian noise. Given this formulation, it is well known that Extended Kalman Filtering (EKF) [63; 91] is the best state estimation filtering technique. We will consider in this work the EKF filtering, given our problem setup, which makes EKF optimal. A Kalman filter that linearizes about the current mean and covariance is referred to as an EKF [52]. We can linearize the estimation around the current estimate using the partial derivatives (i.e., Taylor series) of the process and measurement functions to compute estimates even in the case of non-linear relationships. It can be seen that (4.21) is non-linear in the Cartesian distance. In the EKF estimation process, initially, we need to linearize the observation equation to determine its current prediction stages using a Taylor series expansion. Then, a standard Kalman filter can be applied on this linearized equation to predict and update the state of the system.

The EKF state estimation can be performed in two steps namely, predict and then update. First, current state estimate and current estimate error covariance, are computed using the knowledge of the state transition matrix $\boldsymbol{T}$. The prediction equations can be expressed as:

$$\widehat{\boldsymbol{\theta}}(k|k-1) = \boldsymbol{T}(k-1)\widehat{\boldsymbol{\theta}}(k-1|k-1) \tag{4.22}$$

$$\boldsymbol{\Sigma}(k|k-1) = \boldsymbol{T}(k-1)\boldsymbol{\Sigma}(k-1|k-1)\boldsymbol{T}(k-1)^T + \boldsymbol{C}_{\boldsymbol{\theta}}(k-1) \tag{4.23}$$

where $\widehat{\boldsymbol{\theta}}(k-1|k-1)$ and $\widehat{\boldsymbol{\theta}}(k|k-1)$ are the current and predicted estimated state vector, respectively. Similarly, $\boldsymbol{\Sigma}(k-1|k-1)$ and $\boldsymbol{\Sigma}(k|k-1)$ holds the same meaning for the current estimate error covariance matrix and predicted estimate error covariance matrix, respectively. Then, the predicted state estimate and predicted estimate error covariance are updated using (4.20), that is, the data from the active sensor nodes.

Proceeding with the derivation, we linearize $\boldsymbol{h}(\boldsymbol{\theta}(k))$ around the current estimate of $\boldsymbol{\theta}(k)$ using the well-known Jacobian matrix $\boldsymbol{J}(.)$ evaluated at the predicted state estimate, that is:

$$\boldsymbol{J}(k) = \left.\frac{\partial \boldsymbol{h}(\boldsymbol{\theta}(k))}{\partial \boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}(k|k-1)} \tag{4.24}$$

so that the linearized measurement equation (4.20) becomes:

$$\boldsymbol{y}(k) = \boldsymbol{J}(k)\boldsymbol{\theta}(k) + \boldsymbol{z}_{\boldsymbol{y}}(k) + \boldsymbol{h}(\widehat{\boldsymbol{\theta}}(k|k-1)) - \boldsymbol{J}(k)\widehat{\boldsymbol{\theta}}(k|k-1) \tag{4.25}$$

After the linearization of the measurements, the measurement residual $\tilde{\boldsymbol{y}}$, residual covariance matrix $\widetilde{\boldsymbol{C}_{\boldsymbol{y}}}$, and the Kalman gain $\boldsymbol{K}$ can be obtained by the following equations, respectively:

$$\tilde{\boldsymbol{y}}(k) = \boldsymbol{y}(k) - \boldsymbol{h}(\widehat{\boldsymbol{\theta}}(k|k-1)) \tag{4.26}$$

$$\widetilde{\boldsymbol{C}_{\boldsymbol{y}}}(k) = \boldsymbol{J}(k)\boldsymbol{\Sigma}(k|k-1)\boldsymbol{J}(k)^T + \boldsymbol{C}_{\boldsymbol{y}}(k) \tag{4.27}$$

$$\boldsymbol{K}(k) = \boldsymbol{\Sigma}(k|k-1)\boldsymbol{J}(k)^T\widetilde{\boldsymbol{C}_{\boldsymbol{y}}}(k)^{-1} \tag{4.28}$$

and the resulting estimation update equations will be given by:

$$\widehat{\boldsymbol{\theta}}(k|k) = \widehat{\boldsymbol{\theta}}(k|k-1) + \boldsymbol{K}(k)\tilde{\boldsymbol{y}}(k) \tag{4.29}$$

$$\boldsymbol{\Sigma}(k|k) = (\boldsymbol{I} - \boldsymbol{K}(k)\boldsymbol{J}(k))\boldsymbol{\Sigma}(k|k-1) \tag{4.30}$$

Notice that the EKF estimation performance is highly dependent on the initial state estimate and the system process model. Thus, the choice of the initial state estimate and system model is very important for the correct tracking of the target source.

4.5.3 Optimization Algorithm

To start the process of tracking and estimation of a moving target, first we jointly select, within a given sampling period, the subset of most informative sensors and an associated routing structure that needs to be routed at a leader node, where measurements from active (selected) sensors are to be sent for target state prediction and update. In order to perform this, initially, we need to choose the best leader node, which will also drive the next steps. We call this leader node a Local Sink Node (LSN). We assume that the initial target location is known, and that the choice of the first best LSN would be to select the node that receives the highest SNR. Once the LSN is known, the sensor selection and routing tree generation process starts from this node, as exactly given in Section 4.2 (taking the LSN as the root), by solving, in a distributed manner, the utility function (4.15) for 1-hop neighbors and the process continues until the given power budget is utilized. Thus, the most informative sensor nodes and an associated optimized routing structure routed at this LSN are found. Once this process is performed at the selected LSN, within the same sampling period, three operations need to be performed, which are: target state prediction, target state update and finding the next best LSN.

First, each selected sensor node is required to send its current measurement as in (4.21) via the selected multihop routing structure to the LSN. Then, the currently selected LSN is responsible to perform (4.22) and (4.23) to predict the estimated state vector as well as the estimated error covariance matrix. Once the prediction is over, LSN uses (4.26)-(4.30) to update the estimated state vector $\widehat{\boldsymbol{\theta}}(k|k)$ as well as the estimated error covariance matrix $\boldsymbol{\Sigma}(k|k)$. After performing these two operations, a single path (a backbone using the SPT-CC) from the LSN to the sink node is used to send the estimated target state to the sink node for the final target tracking and monitoring. Then, the third operation, that is, to select the next best LSN is performed, which is as follows.

A feasible approach to select the best LSN, using a prediction based selection scheme, is presented in [52], which we use in our work. This scheme selects a node, among the currently selected sensor nodes, with the highest predicted detection probability. We use the probability density function (pdf) of the predicted target location based on the predicted estimated state $\widehat{\boldsymbol{\theta}}(k|k-1)$ and the predicted estimated error covariance matrix $\boldsymbol{\Sigma}(k|k-1)$. Since the predicted target location has a Gaussian distribution, we can write the pdf of the predicted target location as:

$$f_t(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}(k|k-1)|^{1/2}} e^{\left[-\frac{1}{2}(\boldsymbol{x}-\widehat{\boldsymbol{\theta}}(k|k-1))^T \boldsymbol{\Sigma}(k|k-1)^{-1}(\boldsymbol{x}-\widehat{\boldsymbol{\theta}}(k|k-1))\right]} \tag{4.31}$$

where $\boldsymbol{x} = [x_1, x_2, x_1', x_2']^T$ and $|.|$ denotes the determinant.

Similarly, each sensor node has a detection pdf that determines the uncertainty region of the sensor node. We model (4.31), for sensor node j using a Gaussian distribution centered at sensor location (x_{1j}, x_{2j}), using a given fixed error covariance matrix $\boldsymbol{\Sigma}_0$, which is assumed to be known at each sensor node. Notice that depending on the type of sensors, $\boldsymbol{\Sigma}_0$ can change in general, the uncertainty region will be an ellipse. Therefore, for sensor node j, (4.31) can be equivalently written as:

$$f_j(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}_0|^{1/2}} e^{\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_j)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{x}_j)\right]} \tag{4.32}$$

where $\boldsymbol{x}_j = [x_{1j}, x_{2j}, 0, 0]^T$ and velocities in both coordinate directions for sensor nodes are taken to be zeros since sensor nodes are static.

Thus, the predicted detection probability for sensor node j can be calculated using (4.31) and (4.32) as:

$$F_j = \int\int f_t(\boldsymbol{x}) f_j(\boldsymbol{x})\, dx_1 dx_2 \tag{4.33}$$

Thus, before entering the next sampling period, all the selected sensor nodes are required to calculate (4.33) to obtain their detection probabilities and the highest detection probability sensor node is assigned as the next LSN[1]. Then, the previous LSN transmits all information of the target state to the newly selected LSN, so that all three operations are being performed during the next sampling period.

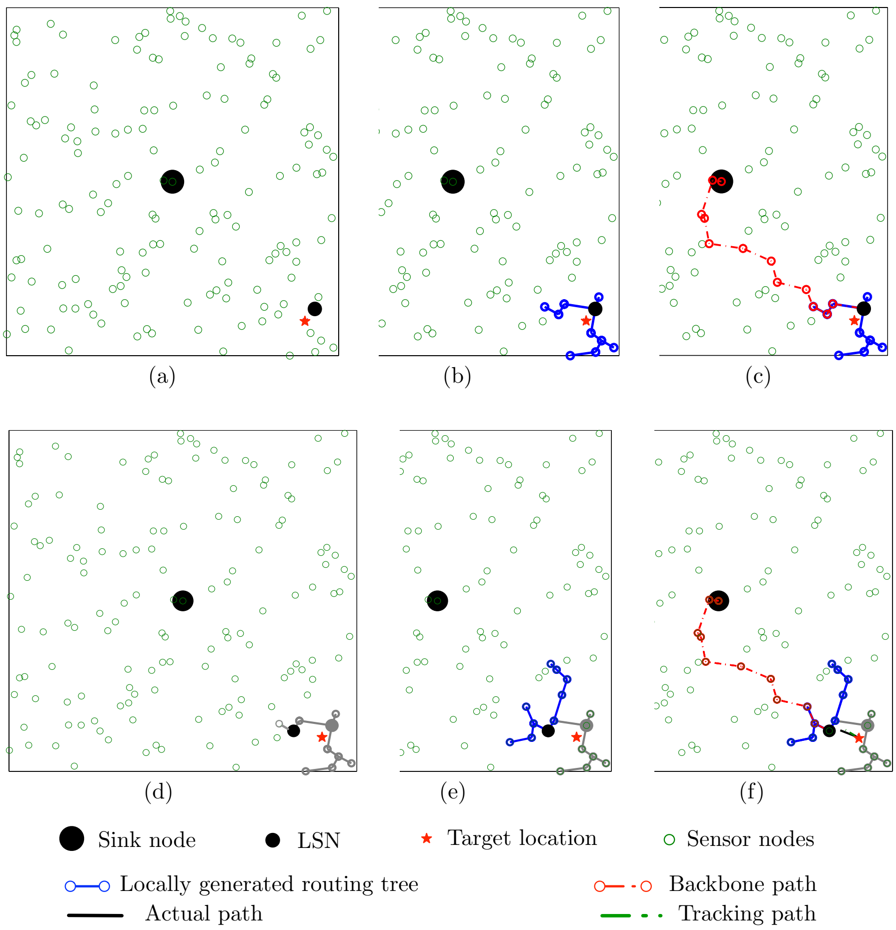


Figure 4.12: Algorithmic process of estimation and tracking of a target: (a) & (d) LSN selection, (b) & (e) routing tree generation and (c) & (f) sending target state to the sink node.

[1]This can be obtained by either message passing [81] or by gossiping [37] algorithms as explained in Section 4.3.1.

All three operations are illustrated sequentially in Figure 4.12, where in Figure 4.12(a), as a first step, an LSN based on the highest SNR is selected to initiate the process of target tracking and estimation. Then, the sensor selection and routing structure routed at this LSN is generated as illustrated in Figure 4.12(b), and then the measurements from the selected sensor nodes are received by the LSN via the selected routing structure to perform the prediction and update. After performing the target state prediction and update, the information (target state) is sent from this LSN to the sink node via a single SPT-CC path as illustrated in Figure 4.12(c). Finally, the third operation, that is, selecting the next best LSN based on the predicted detection probability (4.33) is illustrated in Figure 4.12(d) and all other sensor nodes become inactive. Then, all three operations are repeated at the newly selected LSN, as shown in Figure 4.12(e) and (f), and the process is repeated as long as the tracking of the target is required.

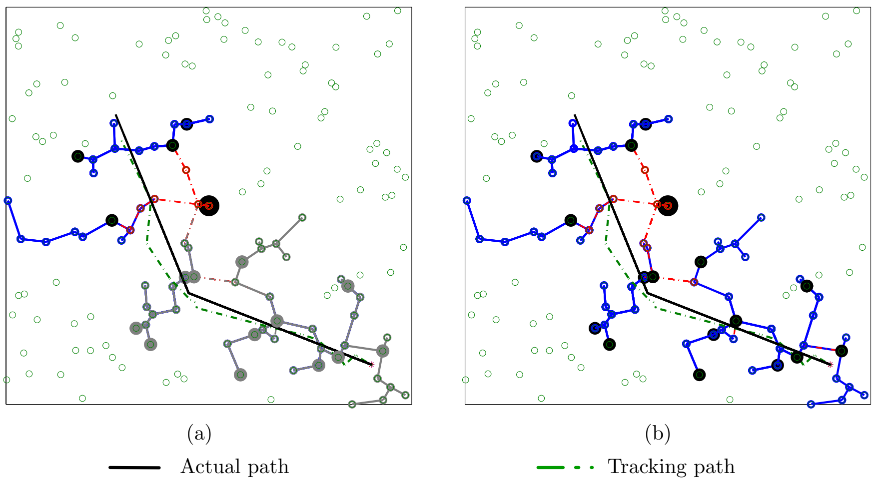


Figure 4.13: After several sampling periods, the actual target path and tracking path are illustrated by continuous and dashed lines, respectively.

Target tracking and actual paths are illustrated in Figure 4.13 by continuous and dashed lines, respectively. Some other intermediate stages of sensor selection and routing tree generation process are illustrated in Figure 4.13(a), with their associated single paths to the sink node. All sensors that are involved in this process, for tracking a target, given certain number of sampling periods, are illustrated in Figure 4.13(b).

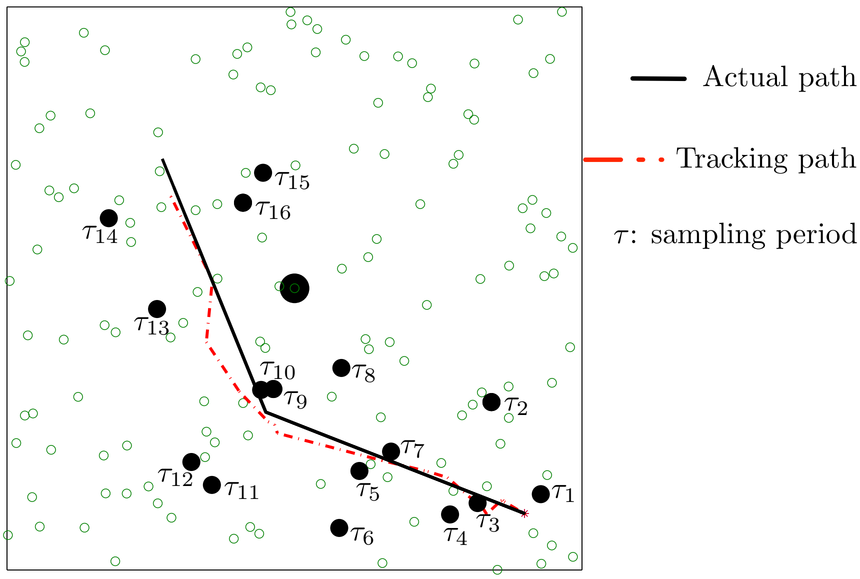


Figure 4.14: Illustration of LSNs with different sampling periods.

We also show the LSNs that are selected during different sampling periods as shown in Figure 4.14. It is important to note that sometimes the communication cost between two consecutive LSNs is very high. For example, transmitting information from sampling period τ_{12} to τ_{13} and from sampling period τ_{14} to τ_{15}, require high communication cost. Therefore, along with the predicted detection probability there must also be some other criteria to select the next LSN, that is, there could be a trade-off between predicted detection probability and the communication cost between two consecutive LSNs. However, this is out of the scope of this thesis and is left for the future work.

Chapter 5

Joint Sensor Selection and Routing Algorithms: Adaptive Quantization

In this chapter, we consider the problem of distributed parameter estimation in a WSN, where because of the power constraints each sensor transmits quantized information to its parent on a multihop path. The goal is to jointly optimize: (*i*) *sensor selection*, (*ii*) *routing structure*, and (*iii*) *number of bits per sample* for all selected sensor measurements, so that the total distortion in estimation is minimized for a given total power budget. To achieve this goal, we have already formulated in Chapter 3, a nonlinear non-convex optimization problem and shown (as in Appendix D) that it is an NP-Hard problem. Furthermore, we show that this nonlinear non-convex optimization problem can be addressed, by using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of two linear (convex) subproblems that can be solved efficiently, without loss of optimality. Then, we propose in this chapter a distributed algorithm using the Estimate-and-Forward (EF) scheme and a uniform dithered quantizer that trades-off first, using the first linear subproblem, the sensor selection and routing structure. Then, given the optimal solution of the first subproblem, we solve second subproblem, which provides an optimal bit-rate allocation to each selected sensor measurement.

If we recall the optimization problem (3.38) from Section 3.3.2:

$$\begin{aligned}
\underset{\{b_i,b_j,B_j\}}{\text{minimize}} \quad & \sum_{j=1}^{n}\left(\frac{4b_j\mathcal{A}^2}{3\mathcal{P}_j 2^{2B_j}}+\frac{\sigma_{y_j}^2}{\mathcal{P}_j(1+|\mathcal{C}_j|)^2 b_j h_j^2}\right)\\
\text{subject to} \quad & \sum_{j=1,(i,j)\in G}^{n} b_j N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max}-P_{\text{gap}}\\
& b_j \le b_i,\ \text{for } (i,j)\in G,\ \text{and } A_{i,j}=1 \qquad (a)\\
& b_j \in\{0,1\}, j=1,\ldots,n \qquad (b)\\
& P_{\text{gap}} \ge 0\\
& B_j \ge 0, j=1,\ldots,n
\end{aligned} \tag{5.1}$$

where the binary variable b_j determines the sensor selection, constraints (a) and (b) impose the routing tree structure since the edge $(i,j)\in G$, and variable B_j is to assign the bit-rate to each sensor measurement. This problem is clearly not convex since the Hessian of the objective is not positive semidefinite (see Appendix F for detailed proof).

5.1 Fixed-Tree Relaxation-Based Adaptive Quantization

A first approach to find an approximate solution to our problem consists of designing an algorithm that decouples the communication cost and estimation process. For this, first we generate a Shortest Path Tree for the whole connectivity graph based only on the Communication Cost (SPT-CC) with $B_j = 1$ bit per sample for all measurements, which is rooted at the sink node. Then, we store an edge-set $\{(j,i)\}$, where i is the parent of j, which is defined as a directed edge $j \to i$. Thus, the information for routing in (5.1) is given since the parent of each node is known by the SPT-CC. Then, we rewrite the optimization problem (5.1) to jointly solve the sensor selection (with variable b_j) and bit-rate allocation (with variable B_j) so that the routing structure used for the selected sensors will be a subtree of the SPT-CC and that has to be rooted at the sink node. We call this approach as Fixed-Tree Relaxation-based Adaptive Quantization (FTR-AQ) algorithm. Applying the EF scheme and relaxing problem (5.1), we have:

$$\begin{aligned}
\underset{\{b_j^r, B_j\}}{\text{minimize}} \quad & \sum_{j=1}^{n} \left(\frac{4 b_j^r \mathcal{A}^2}{3\mathcal{P}_j 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j (1+|\mathcal{C}_j|)^2 b_j^r h_j^2} \right) \\
\text{subject to} \quad & \sum_{j=1, j\in\mathcal{C}_i}^{n} b_j^r N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max} - P_{\text{gap}} \\
& b_j^r \leq b_i^r, j \in \mathcal{C}_i \text{ in SPT-CC} \\
& 0 \leq b_j^r \leq 1; P_{\text{gap}} \geq 0 \\
& B_j \geq 0
\end{aligned} \tag{5.2}$$

where b_j^r is the relaxed version of variable b_j, and $\mathcal{C}_i$ is the set of child nodes of sensor i, thus $j \in \mathcal{C}_i$ represents that sensor j is one of the child of sensor i, that is, sensor i is the parent of sensor j. Notice that the relaxed problem (5.2) is still not a convex optimization problem because the objective and equality constraint are the product of *affine* and *log-affine* functions. The proof of non-convexity of this problem can be seen in Appendix F. We propose a heuristic approach to solve this nonlinear non-convex problem [50; 61; 74; 93] in an efficient manner, by a concatenation of two linear subproblems, namely, *sensor selection* and *bit-rate allocation*, which are convex over the variables b_j^r and B_j, respectively. These subproblems are described as follows:

- *Sensor selection subproblem:* we solve problem (5.2) only for the variable b_j^r, where the variable B_j is assumed to be known, namely assuming fine quantization (high bit-rates), that is, $B_j \to \infty$. In this case, problem (5.2) becomes a convex optimization problem since then, the objective is convex and all other equality and inequalities are linear functions of b_j^r. This can be easily solved through standard convex optimization methods.

- *Bit-rate allocation subproblem:* once we obtain the solution $\{b_j^{r*}\}_{j=1}^n$ to the sensor selection subproblem, we select the best subset of sensors out of the total n sensors, using the sorting procedure as described in Section 4.1. Then, we solve problem (5.2) only for the variable B_j and subset of sensors obtained from the sensor selection subproblem, which is again a convex optimization problem. Thus, it is solvable through the standard convex optimization methods.

The detailed process for solving these subproblems is as follows:

We start by assigning $B_j \to \infty, j = 1, \ldots, n$ in (5.2), resulting in sensor selection subproblem, which is given by:

$$
\begin{aligned}
\underset{\{b_j^r\}}{\text{minimize}} \quad & \left(\sum_{j=1}^{n} \frac{b_j^r h_j^2}{\sigma_{y_j}^2} \right)^{-1} \\
\text{subject to} \quad & \sum_{j=1, j \in \mathcal{C}_i}^{n} b_j^r N_j d_{i,j}^{\alpha} = P_0 - P_{0,\text{gap}} \\
& b_j^r \leq b_i^r, j \in \mathcal{C}_i \\
& 0 \leq b_j^r \leq 1 \\
& P_{0,\text{gap}} \geq 0
\end{aligned}
\tag{5.3}
$$

where $P_0 = \frac{P_{\max}}{2^{B_j}}$ and it is assumed to be large enough as to satisfy the equality constraint and makes (5.3) feasible and $P_{0,\text{gap}} = \frac{P_{\text{gap}}}{2^{B_j}}$ is normalized power gap in this case. The detailed transformation of problem (5.2) into problem (5.3) is given in Appendix D.

Algorithm 3 Procedural steps to solve *sensor selection subproblem*

Require: $P_{\max}$, $B_j \to \infty, j = 1, \ldots, n$

1. Find SPT-CC using Bellman-Ford algorithm
2. Solve problem (5.3) to find $\{b_j^{r*}\}_{j=1}^n$
3. Sort the optimal values $\{b_j^{r*}\}_{j=1}^n$ in descending order
4. Select a subset V_T^* of K sensors corresponding to the largest b_j^{r*}'s
5. Obtained $\mathcal{P}_j^* = \prod_{l \in \Omega_j} \left(1 + |\mathcal{C}_l^*|\right)^2$ for the updated number of children $|\mathcal{C}_l^*|$ of each of the selected sensors
6. Choose the routing tree $T^* \subset$ SPT-CC that span the subset V_T^* of selected sensors

We obtain the optimal solution $\{b_j^{r*}\}_{j=1}^n$ by solving the convex optimization problem (5.3) over the variables b_j^r's. Then, the suboptimal subset selection can be obtained by sorting the optimal values $\{b_j^{r*}\}_{j=1}^n$ in descending order and selecting a subset V_T^* of K largest b_j^{r*}'s. The best size of the subset of sensors, K, depends on the total allowed power budget $P_{\max}$. It can be chosen in such as way that the total distortion in estimation is minimized for the given power budget, where bit-rate allocation strongly influences the choice of the best K. For example, small K can be chosen while providing higher bit-rates and large K otherwise for the same given total power budget; the total distortion in estimation also influences the choice of the best K. Notice also that because of the subset selection, the number of children $|\mathcal{C}_l|$ for each sensor in V_T^* needs to be updated, thus $\mathcal{P}_j = \prod_{l\in\Omega_j}\left(1+|\mathcal{C}_l|\right)^2$ becomes $\mathcal{P}_j^* = \prod_{l\in\Omega_j}\left(1+|\mathcal{C}_l^*|\right)^2$ with the updated value $|\mathcal{C}_l^*|$. Then, denoting $\{\widehat{b}_j^r\}_{j=1}^n$ a set of binary values such that $\widehat{b}_j^r = 1$ if $j \in V_T^*$ and $\widehat{b}_j^r = 0$ if $j \notin V_T^*$, we have a subset of sensors, where for each selected sensor, its bit-rate allocation B_j needs to be optimized. To do this, we need to solve bit-rate allocation subproblem only for the subset V_T^* while satisfying the power constraint $P_{\max}$. Thus, bit-rate allocation subproblem for V_T^*, given $\widehat{b}_j^r = 1, j \in V_T^*$, can be written as:

$$
\begin{aligned}
\underset{\{B_j\}}{\text{minimize}} \quad & \sum_{j\in V_T^*}\left(\frac{4\mathcal{A}^2}{3\mathcal{P}_j^* 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j^*(1+|\mathcal{C}_j^*|)^2 h_j^2}\right) \\
\text{subject to} \quad & \sum_{j\in V_T^*, j\in\mathcal{C}_i^*} N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max} - P_{\text{gap}} \\
& B_j \geq 0 \\
& P_{\text{gap}} \geq 0
\end{aligned}
\tag{5.4}
$$

This problem is also convex since the objective and equality are convex over the variable B_j and all other inequalities are linear. The solution of this problem provides a suboptimal solution for B_j^* that needs to be rounded up to the nearest integer, $B_j^* = \lfloor B_j^* \rceil$. Furthermore, it is to be noted that because of the relaxation and constraint $b_j^r \leq b_i^r$ in (5.3), sorting the variables $\{b_j^{r*}\}_{j=1}^n$ forms correctly a subtree T^* of SPT-CC, which is the routing structure, and it is routed at the sink node.

Algorithm 4 Procedural steps to solve *bit-rate allocation subproblem*

Require: $P_{\max}$, V_T^*, $\mathcal{C}_j^*$, $\mathcal{P}_j^*$, $j \in V_T^*$

1. Solve problem (5.4), only for sensors of the subset V_T^*, to find $\{B_j^*\}_{j\in V_T^*}$
2. Round B_j^* to the nearest integer, that is, $B_j^* = \lfloor B_j^* \rceil$

5.1.1 Approximate Solution

Although subproblems (5.3) and (5.4) can be solved through standard convex optimization methods, for example, by interior-point methods (Sec. 11.2.2, [10]), we show below that both can also be solved approximately but very efficiently. First, we solve sensor selection subproblem (5.3) with the Newton's method using, for instance, *log barrier method* ([10], Sec. 11.2), that is, subproblem (5.3) can also be posed as:

$$\begin{aligned} \underset{\{b_j^r\}}{\text{minimize}} \quad & \phi(\boldsymbol{b}^r) = \left(\sum_{j=1}^{n} \frac{b_j^r h_j^2}{\sigma_{y_j}^2} \right)^{-1} - \\ -\frac{1}{\eta} \sum_{j=1}^{n} & \left(\log(b_j^r) + \log(1-b_j^r) + \log(b_i^r - b_j^r) + \log P_{0,\text{gap}} \right) \qquad (5.5) \\ \text{subject to} \quad & \sum_{j=1, j \in \mathcal{C}_i}^{n} b_j^r N_j d_{i,j}^{\alpha} = P_0 - P_{0,\text{gap}} \end{aligned}$$

where $\eta > 0$ is a positive parameter that controls the quality of approximation, and because of the implicit constraints $0 \leq b_j^r \leq 1$ in the problem (5.5), the function ϕ is convex and smooth, thus it can be solved efficiently by the Newton method. Let $\{b_j^{r*}(\eta)\}_{j=1}^n$ denote the solution of (5.5), which depends on the parameter η.

In particular, it can be seen as in interior-point methods that $\{b_j^{r*}(\eta)\}_{j=1}^n$ is no more than $\frac{2n}{\eta}$ suboptimal for the the solution $\{b_j^{r*}\}_{j=1}^n$ of the subproblem (5.3), that is:

$$\left(\sum_{j=1}^{n} \frac{b_j^{r*}(\eta) h_j^2}{\sigma_{y_j}^2} \right)^{-1} \leq \left(\sum_{j=1}^{n} \frac{b_j^{r*} h_j^2}{\sigma_{y_j}^2} \right)^{-1} + \frac{2n}{\eta} \qquad (5.6)$$

This confirms the intuitive idea that $\{b_j^{r*}(\eta)\}_{j=1}^n$ converges to the solution $\{b_j^{r*}\}_{j=1}^n$ as $\eta \to \infty$. A description of the Newton method to calculate $\{b_j^{r*}(\eta)\}_{j=1}^n$ is given in the Appendix E.

Once we obtain the solution $\{b_j^{r*}(\eta)\}_{j=1}^n$ of the problem (5.5), we use the same sorting procedure, as described above in Algorithm 3, to select a subset V_T^a of K sensors, where we represent the approximate solution with the superscript ^{a}. Then, we update $\mathcal{P}_j$ and $\mathcal{C}_j$ corresponding to V_T^a, to get $\mathcal{P}_j^a$ and $\mathcal{C}_j^a$, respectively. In this case, a subtree for the subset V_T^a is given by $T^a \subset G$. Then, in order to solve subproblem (5.4) analytically, we write the Lagrangian function Λ:

$$\begin{aligned}\Lambda(\boldsymbol{B}, \lambda_0, \boldsymbol{\lambda}) = \sum_{j \in V_T^a} \left(\frac{4\mathcal{A}^2}{3\mathcal{P}_j^a 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j^a (1 + |\mathcal{C}_j^a|)^2 h_j^2} \right) + \\ + \lambda_0 \left(\sum_{j \in V_T^a, j \in \mathcal{C}_i^a}^{N} N_j d_{i,j}^{\alpha} 2^{B_j} - P_{\max} + P_{\text{gap}} \right) - \sum_{j \in V_T^a} \lambda_j B_j\end{aligned} \tag{5.7}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{|V_T^a|}$ and $\boldsymbol{B} \in \mathbb{R}^{|V_T^a|}$.

The Lagrangian function (5.7) gives the following Karush-Kuhn-Tucker (KKT) conditions [10]:

$$\frac{-8\log(2)\mathcal{A}^2}{3\mathcal{P}_j^a 2^{2B_j}} + \log(2)\lambda_0 N_j d_{i,j}^{\alpha} 2^{B_j} - \lambda_j = 0, j \in V_T^a \tag{5.8}$$

$$\begin{aligned}\lambda_0 \left(\sum_{j \in V_T^a} N_j d_{i,j}^{\alpha} 2^{B_j} - P_{\max} + P_{\text{gap}} \right) = 0 \\ \sum_{j \in V_T^a}^{N} N_j d_{i,j}^{\alpha} 2^{B_j} - P_{\max} + P_{\text{gap}} \leq 0 \\ \lambda_0 \geq 0\end{aligned} \tag{5.9}$$

$$\begin{aligned}\lambda_j B_j = 0, j \in V_T^a \\ \lambda_j \geq 0, j \in V_T^a \\ B_j \geq 0, j \in V_T^a\end{aligned} \tag{5.10}$$

It is easy to verify from the KKT condition (5.10) that $\lambda_j = 0$ for all $B_j > 0$, $j \in V_T^a$. Then, given $\lambda_j = 0$, we can also verify from (5.8) that $\lambda_0 > 0$ for all $j \in V_T^a$, which further from (5.9) implies $\sum_{j \in V_T^a} N_j d_{i,j}^{\alpha} 2^{B_j} - P_{\max} + P_{\text{gap}} = 0$. Then, to find the optimal solution for $B_j, j \in V_T^a$ given the above conditions, require to solve the above KKT conditions, which gives from (5.8) the following:

$$B_j^a = \frac{1}{3} \log_2 \left(\frac{4\mathcal{A}^2}{3\lambda_0 N_j d_{i,j}^{\alpha} \mathcal{P}_j^a} \right), j \in V_T^a \tag{5.11}$$

In practice, each B_j^a needs to be rounded up to the nearest integer value, that is, $B_j^a = \lfloor B_j^a \rceil$. Notice that even after rounding if any $B_j^a < 1$, we set its value to 1. The value of λ_0 is obtained by solving the KKT conditions (5.8) and (5.9), which is given by:

$$\lambda_0 = \frac{8\mathcal{A}^2}{3(P_{\max} - P_{\text{gap}})^3} \sum_{j \in V_T^a}^{N} \frac{N_j^2 d_{i,j}^{2\alpha}}{\mathcal{P}_j^a} \tag{5.12}$$

Then, given these solution, we can obtain the objective value of the relaxed problem (5.2). In this solution, for simplicity, P_{gap} can be chosen equal to zero since our aim is to minimize this gap as close to zero as possible.

5.2 Local Optimization-based Adaptive Quantization

It can be noticed that the operations described in the previous section to find a possible subset V_T of K selected sensors, their assigned bit-rate $B_j, j \in V_T$, and an associated routing structure T, can be further improved by performing an edge-swap method on the subset V_T of sub-tree (routing structure) T and optimizing B_j for the swapped edges. As described in [98], the edge-swap method is used to swap a higher communication cost edge with another appropriate smaller communication cost edge, allowing routing structure T to be routed at the sink node. Then, due to the currently swapped edge having smaller communication cost, we can provide even higher bit-rate for this edge, which significantly improves the total distortion in estimation. The process of edge-swapping and bit-rate optimization need to be carried out for all the possible (swappable) edges in V_T. We call the algorithm described in this section as Local Optimization-based Adaptive Quantization (LO-AQ) algorithm.

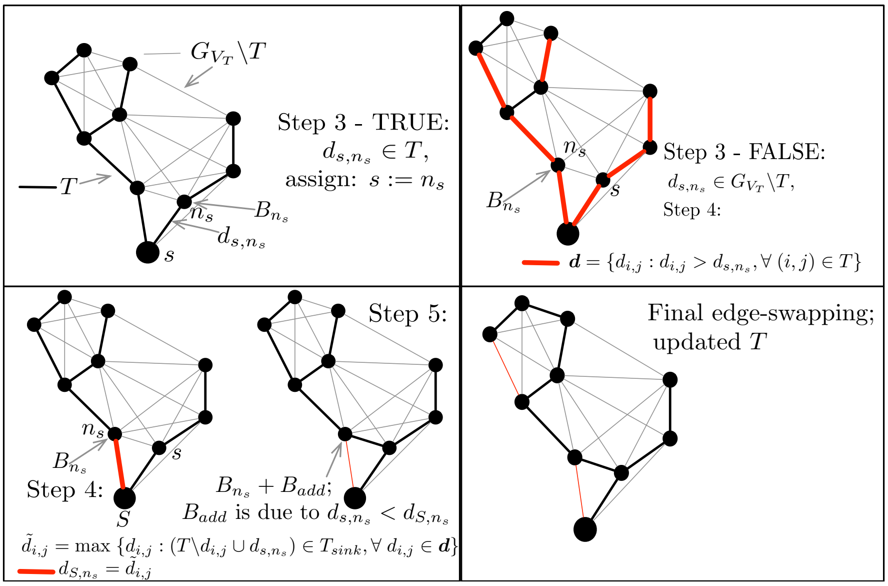


Figure 5.1: Graphical representation of the edge-swap method [3; 4; 98]; Step 3, Step 4 and Step 5 of Algorithm 5 are shown.

In order to perform these operations, first we define the restricted subgraph G_{V_T} as follows:

Definition 1 *Given a graph $G = (V, E)$, the restriction G_{V_T} of G to a subset $V_T \in V$ of K nodes, is a graph $G_{V_T} = (V_T, E_T)$ where for every pair of nodes $i, j \in V_T$, if $(i, j) \in E_T$, then $(i, j) \in E$.*

Then, given the subgraph G_{V_T}, we perform swaps among the edges in T and the edges in $G_{V_T} \backslash T$ in such a way that the new resulting tree after an update (swap) in the edges, remains a non-spanning tree of a graph G_{V_T} and that must be routed at the sink node. Notice also that after each possible swapping, we update the bit-rate for the corresponding edge. The steps of the procedure are given in Algorithm 5 and then are illustrated in Figure 5.1.

Algorithm 5 LO-AQ Algorithm

Require: V_T, T, B_j

Initialization:
$s = n + 1$, identity of the sink node;
index $\ell = 1$;
n_s= nearest sensor to s;
$K = |V_T|$, number of sensor in V_T;
T_{sink} = set of routing structures for V_T routed at the sink node;

- **Step 1:** find n_s
- **Step 2:** if $\ell = K$, Stop; otherwise continue
- **Step 3:** if $d_{s,n_s} \in T$, go to Step 6; otherwise continue
- **Step 4:** find a set of edges,
  $\boldsymbol{d} = \{d_{i,j} : d_{i,j} > d_{s,n_s}, \forall\, (i,j) \in T\}$, and then find,
  $\tilde{d}_{i,j} = \max\,\{d_{i,j} : (T \backslash d_{i,j} \cup d_{s,n_s}) \in T_{sink}, \forall\, d_{i,j} \in \boldsymbol{d}\}$;
  Remove the edge $\tilde{d}_{i,j}$, $\quad T := T \backslash \tilde{d}_{i,j}$;
  Update child nodes, $\quad |\mathcal{C}_i| := |\mathcal{C}_i| - 1$; $j \in \mathcal{C}_i$
  Add edge d_{s,n_s}, $\quad T := T \cup d_{s,n_s}$;
  Update child nodes, $\quad |\mathcal{C}_s| := |\mathcal{C}_s| + 1$; $n_s \in \mathcal{C}_s$
  Set $B_{n_s} = B_j$ and calculate the power gap:
  $$P_{\text{gap}} = N_j \tilde{d}^{\alpha}_{i,j} 2^{B_j} - N_{n_s} d^{\alpha}_{s,n_s} 2^{B_{n_s}};$$
- **Step 5:** update B_{n_s} for the edge d_{s,n_s} such that $P_{\text{gap}} \to 0$
- **while** $P_{\text{gap}} > 0$ **do**
  $B_{n_s} := B_{n_s} + 1$
  $P_{\text{gap}} = N_j \tilde{d}^{\alpha}_{i,j} 2^{B_j} - N_{n_s} d^{\alpha}_{s,n_s} 2^{B_{n_s}}$;
- **end while**
  $B_{n_s} := B_{n_s} - 1$;
- **Step 6:** assign $s := n_s$; $\ell := \ell + 1$ go to Step 1

A detailed description of these steps is the following: first, we assume that the operations from the previous algorithm generates the following results: V_T, $B_j, j \in V_T$, and $T = (V_T, E_T)$ of K sensors. Then, we assign $s = n + 1$ (identity of the sink node) and initiate the index $\ell = 1$, and then we find the nearest sensor n_s to s based on the communication cost using the solution $B_j, j \in V_T$. If $d_{s,n_s} \notin T$. Then, we consider a list $\boldsymbol{d}$ of all the edges in T that have the larger communication cost than the communication cost of the edge d_{s,n_s}, and find the largest value $\tilde{d}_{i,j}$ from this list so that while swapping it with d_{s,n_s}, ensures that the new tree $T \in T_{sink}$[1] is still routed at the sink node. We update T and number of children $|\mathcal{C}|$ by removing the largest edge $\tilde{d}_{i,j}$ and including the edge d_{s,n_s}, and then we calculate the corresponding power gap $P_{\text{gap}} = N_j \tilde{d}_{i,j}^{\alpha} 2^{B_j} - N_{n_s} d_{s,n_s}^{\alpha} 2^{B_{n_s}}$ with $B_{n_s} = B_j$. Then, the power gap $P_{\text{gap}} > 0$ (since measurement flows from n_s to s and $d_{s,n_s} < \tilde{d}_{i,j}$) can be minimized by optimizing B_{n_s} locally (Step 5 of Algorithm 5). Increasing B_{n_s} from its current value tends to reduce P_{gap} toward zero, thus reducing the quantization error. If at any moment, the update in B_{n_s} produces $P_{\text{gap}} < 0$, we deduct one from the current value of B_{n_s}, so that the constraint $P_{\text{gap}} \geq 0$ is satisfied. After optimizing B_{n_s}, we assign $s := n_s$, increase the index ℓ by one, and return to Step 1 to repeat the process until we scan all the edges in T, that is, $\ell = K$. On the other hand, if edge $d_{s,n_s} \in T$ (Step 3 of Algorithm 5), then we do not perform any swap and assign $s := n_s$ (Step 6 of Algorithm 5), increase ℓ by one, and then we repeat the process. Notice that each edge in T is scanned only once.

5.3 Implementation Issues and Complexity

Notice that in our algorithms, a subset of active sensors, an associated multihop routing structure and the bit-rate allocation to all active sensor measurements are jointly optimized at the sink node. Then, the selected (activated) sensors perform the EF operation computing the total distortion in estimation distributively and incrementally while routing the fused estimation towards the sink node. In order to do this, first we initialize a routing structure using the SPT-CC and information related to the network connectivity graph G at the sink node. Then, the sink node broadcasts to each sensor separately the identity of their respective parent nodes, so that each sensor node is able to send its information to the sink node via its parent node.

[1] For a given graph G_{V_T} there will be at least one spanning tree routed at the sink node, and the list of all possible spanning trees routed at the sink node is the T_{sink}.

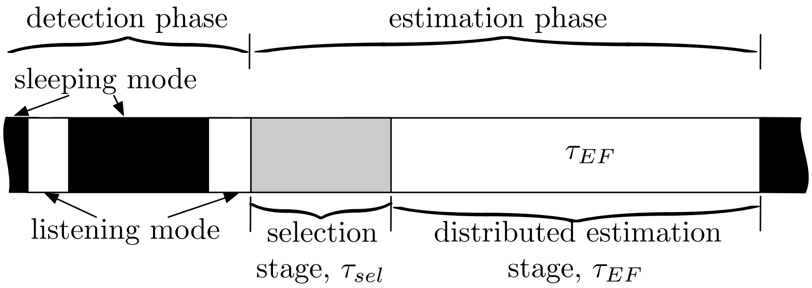


Figure 5.2: In our algorithm, the estimation task is being carried out during the estimation phase, which is divided into two periodic segments. The first one is a selection stage of duration τ_{sel}, where for a given total power budget a subset of sensors V_T their bit-rate allocation $B_j, j \in V_T$ and an associated routing structure T, are to be selected. Then, the selected sensors perform the EF operation computing the total distortion in estimation distributively and incrementally during the distributed estimation stage of duration τ_{EF}. Generally, in practice, $\tau_{EF} \gg \tau_{sel}$, thus the total communication cost is usually dominated by the distributed estimation stage.

The WSN operates alternatively and periodically with two main phases, as shown in Figure 5.2. In the detection phase, all sensor nodes change their status periodically from sleeping to listening and from listening to sleeping mode until some of them (i.e., sensors near the source target) detect a physical quantity greater than a pre-defined threshold. In practice, these two sleeping and listening modes can be divided so that the network lifetime is controlled, but this is outside of the scope of this work. Then, as soon as a sensor detects the presence of some source target, information about the occurring event is passed to the sink node via a single multihop path belonging to the SPT-CC (i.e., via parent nodes, see Figure 3.6). Then, the estimation phase starts, which is divided into two main periodic segments: first, a selection stage of duration τ_{sel}, where a subset of sensors V_T is selected together with their optimal bit-rate allocation and the best associated routing structure $T = (V_T, E_T)$; and second, the distributed estimation stage τ_{EF}, where the EF scheme is applied distributively and incrementally on the subset V_T to estimate the parameter of interest while routing the fused estimation towards the sink node. At the end of the distributed estimation stage, the WSN returns to the detection phase so that it can be ready to perform another estimation task. The duration τ_{EF} is usually in practice much longer than τ_{sel}. In most of the applications, a small fraction of the total power budget is consumed during the selection stage [84]. Next, we explain the differences in implementation for algorithms FTR-AQ and LO-AQ.

5.3.1 FTR-AQ Algorithm

In the case of FTR-AQ algorithm, as soon as the WSN enters into the estimation phase, at the beginning of the selection stage, all sensor nodes are required to send their measurements to the sink node via their parent nodes, which are obtained using the SPT-CC. Since in this initial step, the routing structure is the SPT-CC, the most power efficient way to send these measurements is using the MF scheme. Notice that the MF scheme is required only to collect the individual measurements from each sensor. Then, the FTR-AQ algorithm is executed at the sink node to find the best subset of sensors V_T^*, their optimal bit-rate allocation and the associated routing structure $T^* = (V_T^*, E_T^*)$ (subtree of the SPT-CC) for a given total power budget $P_{\max}$. Later, the sink node broadcasts the indices associated to the selected sensors, so that the rest of the sensor (non-selected) nodes are set back to inactive state. In addition, the information $T^* = (V_T^*, E_T^*)$ and $B_j^*, j \in V_T^*$ are also broadcasted by the sink node to each active sensor node. Once this operation is performed, each selected sensor node performs the EF operation during the distributed estimation stage of duration τ_{EF}.

5.3.2 LO-AQ Algorithm

As stated above, in this algorithm, the inputs to this algorithm are generated by performing the operations of the FTR-AQ algorithm, which generates, for example the solutions[1]: V_T, B_j, and T. Then, the LO-AQ algorithm, as described in Algorithm 5, is executed locally at the sink node, in order to improve the multihop routing structure and bit-rate allocation for the same given total power budget $P_{\max}$. Next, the updated information of V_T, B_j, and T are transmitted to each active

[1]Notice that these are the same solutions as V_T^*, B_j^*, and T^* obtained from the FTR-AQ. In order to have consistency with the description of the LO-AQ algorithm, we denote here as: V_T, B_j, and T.

sensor node so that in the next stage, the distributed estimation stage is performed to obtain the total distortion estimation.

5.3.3 Distributed Estimation Stage

In this stage, each selected sensor is aware of its allocated bit-rate and its children and parent nodes from the selected routing tree, after the sensor selection, bit-rate allocation and final routing structure have been generated by the algorithms. Thus, each sensor waits until it receives all the information from its children, and then applies the EF scheme to fuse all the information (as shown in Figure 3.6) together with its own measurement. Then, it forwards the fused information to its parent node on the generated multihop routing structure. It is important to note that the estimation process starts with the leaf nodes. For a leaf sensor node k, we send the information $[\mathcal{Q}_k, \sigma^2_{\mathcal{Q}_k}]$ to its parent sensor j. Later, sensor j waits to receive all the information from all its child sensors $\mathcal{C}_j$ (see sensors $\{k_1, k_2, k_3\}$ shown in Figure 3.6). Then, sensor j uses (3.29) to fuse all the information received from all its children $\mathcal{C}_j$ together with its own measurement and quantizes the resulting fused information using bit-rate B_j, and then forwards the information $[\mathcal{Q}_j, \sigma^2_{\mathcal{Q}_j}]$ to its parent node i, as shown in Figure 3.6. This process continues until the sink node is reached, which will store the final distortion estimation due to the measurements of all active sensor nodes. This process can be done multiple times to perform several estimation tasks within this period of duration τ_{EF}.

5.3.4 Complexity Analysis

- *FTR-AQ Algorithm:* The main computational complexity in FTR-AQ algorithm consists of: (a) SPT-CC using Bellman-Ford algorithm; (b) the MF operation and (c) solving convex subproblems (5.3) and (5.4) at the sink node. Bellman-Ford runs with a complexity of order $O(Mn)$, where $M = |E|$ and $n = |V|$ are the number of edges and vertices of graph G, respectively. The MF scheme requires $O(n + \tilde{n})$, where $\tilde{n} : n < \tilde{n} \ll n^2$ is the total number of hops required by the MF scheme, thus it requires $O(n^2)$ operations, while solving the optimization subproblems (5.3) and (5.4), using interior-point methods, each one requires $O(n^3)$ operations. Interior-point method typically requires a few tens of iterations and each iteration can be carried out with a complexity of $O(n^3)$ operations. Thus, the overall computational cost has order $O(n^3)$.

- *LO-AQ Algorithm:* The computation effort required to perform the edge-swap method on K sensors takes $O(K^2 \log K)$ operations and improving the total distortion involves finding their best B_j's that takes $O(K \log K)$ operations. If the steps taken in our LO-AQ are no faster than n^3 (i.e., the additional operations in LO-AQ are no faster than the operations carried out in the FTR-AQ), then the total computational effort of this algorithm will be $O(n^3)$, the same as solving FTR-AQ. This is always true even though we consider $K = n$ since $n^2 \log n + n \log n < n^3$.

5.4 Simulation Results

This section presents extensive simulations that illustrate the effectiveness of the proposed adaptive quantization algorithms FTR-AQ and LO-AQ. The network we consider is shown in Figure 1.3, where $n = 100$ sensors are deployed. The network is modeled so that the distance between a sensor and its parent sensor is given by ξd_{norm}, where ξ is uniformly distributed within the range $(0, 1]$. We assume the path loss exponent $\alpha = 4$, which is a typical value for urban macrocell environments or multi-level office buildings [33] and $N_j = 1\ \forall j$ in our transmission model $f_c(d_{i,j}) < S_0 N_j d_{i,j}^{\alpha} 2^{B_j}$. Without loss of generality, we assume that $\sigma_{z_j}^2 = \sigma_{y_j}^2 = 1\ \forall j$ and $\mathcal{A} = 1$. The measurement gain in our example is assumed to be $h_j = 1/d_{j,t}^{\beta}$, where $d_{j,t}$ is the distance from sensor j to the source target t and $\beta = 2$ is the signal decay exponent, which is assumed to be known (or estimated via training sequences [58; 105]). We assume that the exact location of the source target is not known, and hence we consider 100 expected locations (one per sensor) of the source target (as shown in Figure 5.3). We take these locations to be i.i.d. from a Gaussian distribution $\mathcal{N}(\boldsymbol{s}_{2\times1}, \boldsymbol{I}_{2\times2})$, where $\boldsymbol{s}$ is the mean of the expected location values. We test our algorithms using 100 different network topologies with the sink node located at the center of the region. For each topology, different expected location values corresponding to the values of the mean location $\boldsymbol{s}$ are generated (see Figure 5.3 for topologies 1 and 2 out of 100 topologies used in the simulation).

For each network topology, we have executed the algorithms for a range of maximum power budgets.

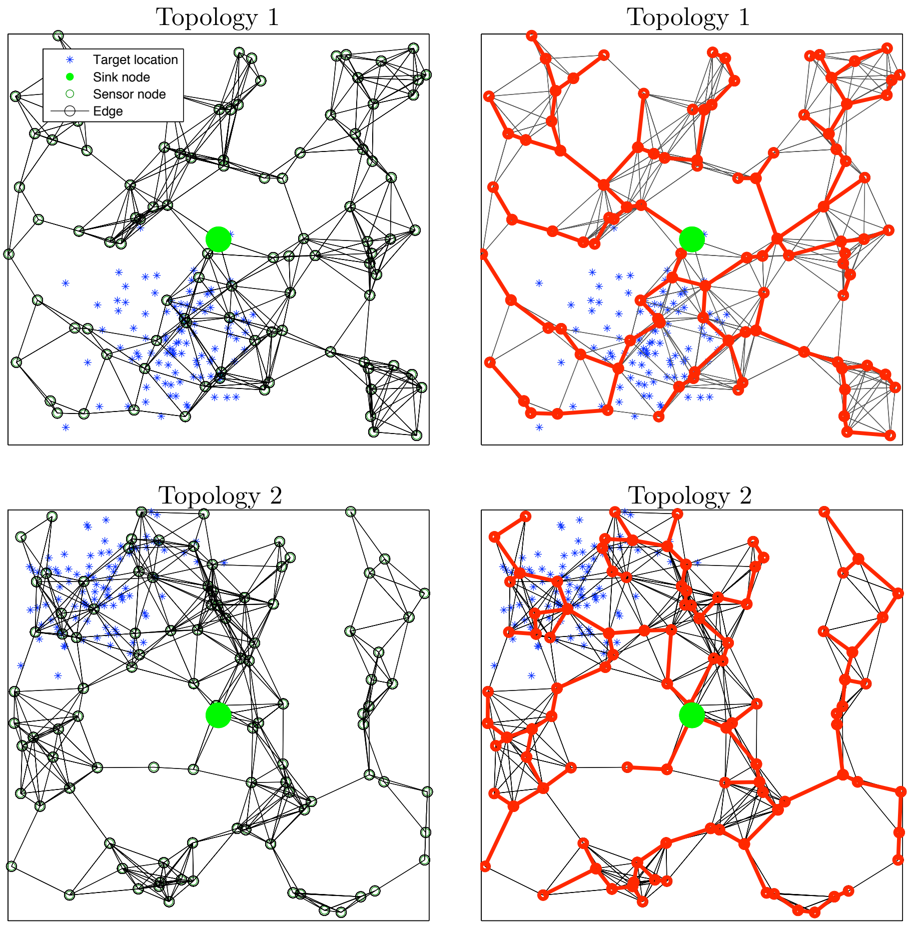


Figure 5.3: Topology examples used in the simulation, where thin edges belong to the network connectivity graph and thick edges (right hand side) represent the SPT-CC. The sink node is located at the center of the region.

Next, we show the results obtained by averaging the performance over 100 network topologies.

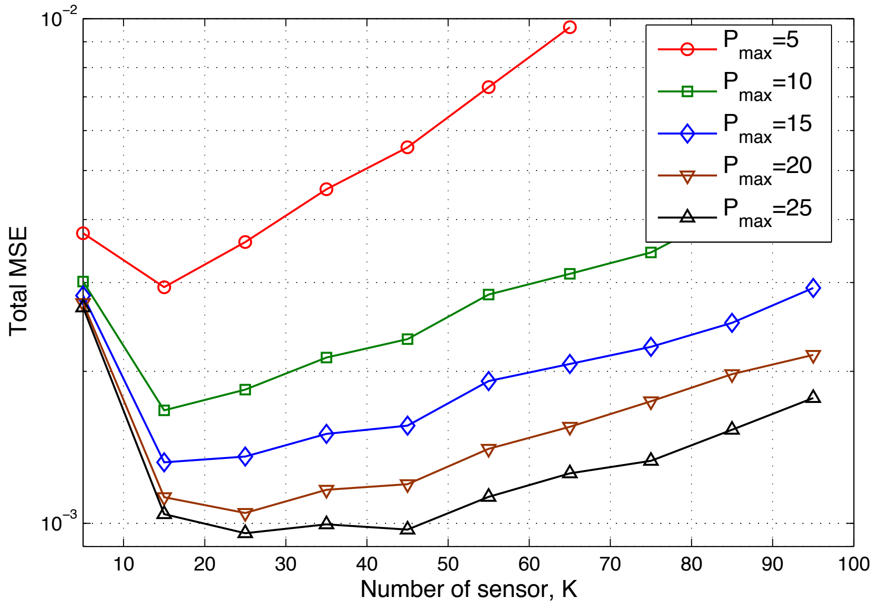


Figure 5.4: The best choice for the size K of the subset of sensors, for different total power budgets, ranging from $P_{\max} = 5$ to $P_{\max} = 25$.

Notice that the total MSE, as given in (3.36), is the sum of the two terms, one proportional to the quantization error and other proportional to the MSE of the estimator. Therefore, an optimal trade-off between these two terms, for a given total power budget $P_{\max}$, can be achieved by selecting a proper size K of the subset of sensors, or visa versa. For example, for a given $P_{\max}$ and a small K, the quantization error term can be made small if $P_{\max}$ is high enough (due to small size K, higher bit-rates can be allocated to each measurement), but the MSE of the estimator term can not be made small (less number of measurements to fuse). On the other hand, a large K can provide a small MSE of the estimator, but the quantization error can be large when low bit-rates are allocated, which is controlled by $P_{\max}$. Therefore, it is necessary to find the best K for a given total power budget $P_{\max}$, so that both terms can be jointly minimized.

In our simulation setting, on average, the best choice of K for a given range of the total power budgets $P_{\max} = 5$ to 25 is between $K = 15$ to 25 out of $n = 100$ deployed sensors, as shown in Figure 5.4. It can be also observed that as we increase $P_{\max}$, the best choice of K also increases. This is simply because we have more power available to select more informative sensors and to allocate higher bit-rates. Thus, in order to compare the MSE performances for different values of K, in Figure 5.5, we show a comparison between fixed and adaptive quantization.

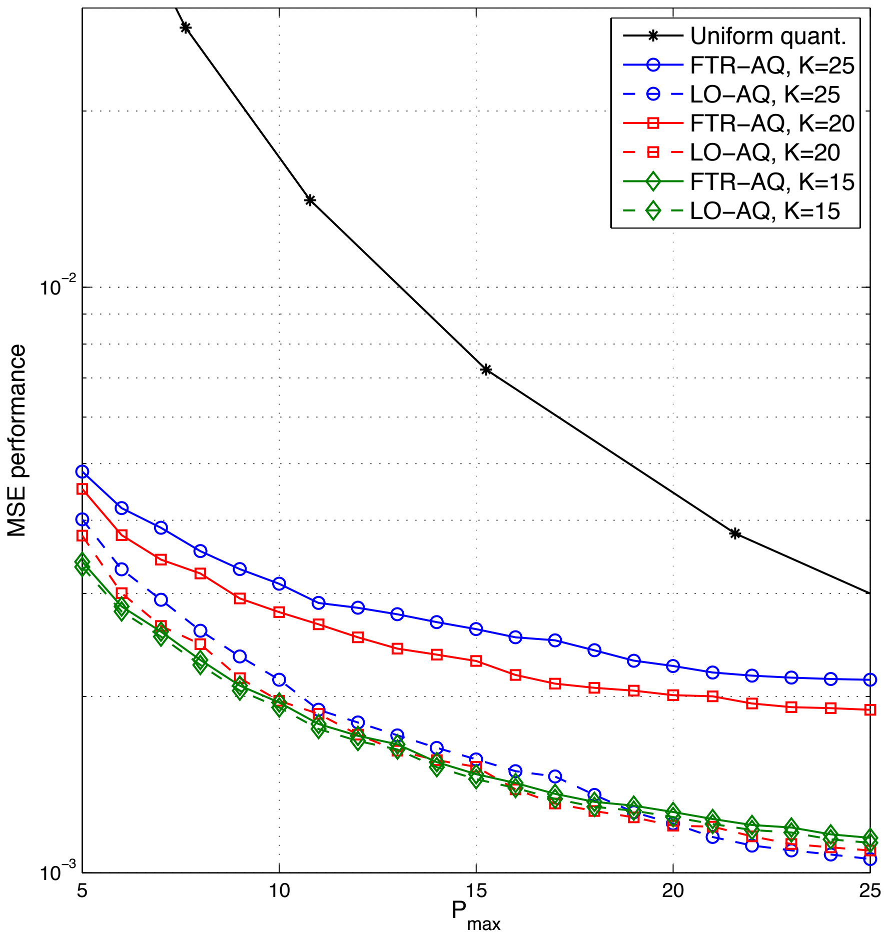


Figure 5.5: MSE performance of our proposed approaches, when activating different subsets of sensors of different sizes ($K = 15, 20, 25$), for a given range of total power budgets ($P_{\max} = 5$ to 25) and comparing their performances with a fixed uniform quantization algorithm.

It can be seen that our adaptive quantization approaches perform superior to the fixed uniform quantization. It can also be seen that as we use higher values of $P_{\max}$, the LO-AQ performance increases compared to FTR-AQ, which is caused by the updates in B_j due to the swapped edges; B_j can be updated to higher values when higher values of $P_{\max}$ are allowed. It is clearly seen that the performance of LO-AQ, when using a subset of $K = 15$ sensors with a high value of $P_{\max}$, degrades as compared to using $K = 20$ and $K = 25$ sensors. This is because higher values of $P_{\max}$ allow allocating higher bit-rates, and thus it helps to minimize the total quantization error, however at the same time, the estimation error is limited due to the small number of sensors $K = 15$.

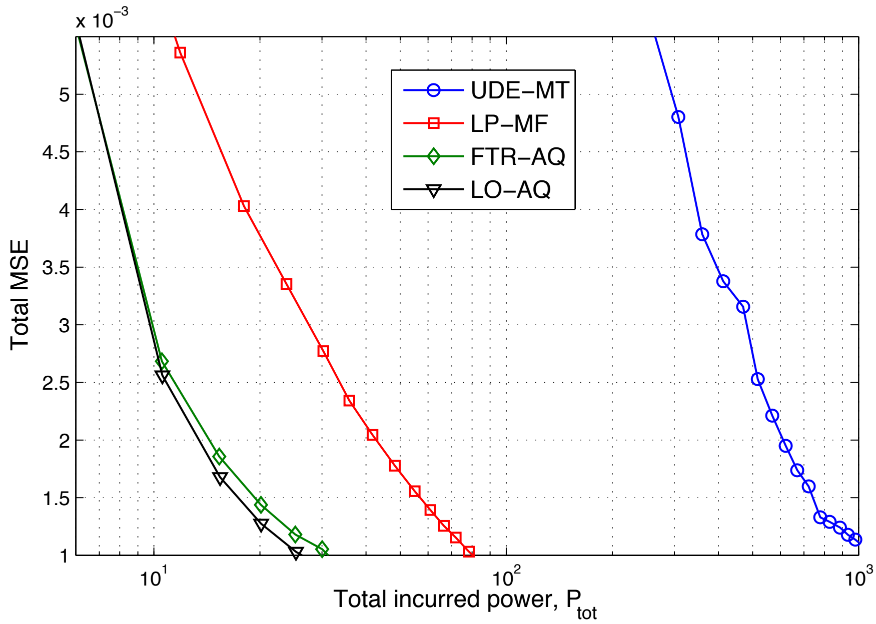


Figure 5.6: Performance comparison in terms of MSE among the proposed and other previously proposed algorithms (UDE-MT [122] and LP-MF [61]) for different total power budgets.

Next, we compare our algorithm to two heuristic algorithms (as described in the related work, Section 2.4), namely, algorithms UDE-MT [122] and LP-MF [61]. Algorithm UDE-MT does not consider the routing optimization while LP-MF does and uses MF scheme. The total MSE versus the total incurred power is shown in Figure 5.6 for UDE-MT, LP-MF, FTR-AQ and LO-AQ algorithms, where the maximum allowed total power range is available from $P_{\max} = 5$ to 1000. We can observe that our approaches outperform clearly the both algorithms UDE-MT and LP-MF, which provides the significance of the use of the EF scheme.

Now, we present an example illustrating the solutions: $\{b_j^*\}_{j=1}^n$, $\{B_j^*\}_{j \in V_T^*}$, and the routing structure for one of the topologies (out of 100 topologies) used for the simulation.

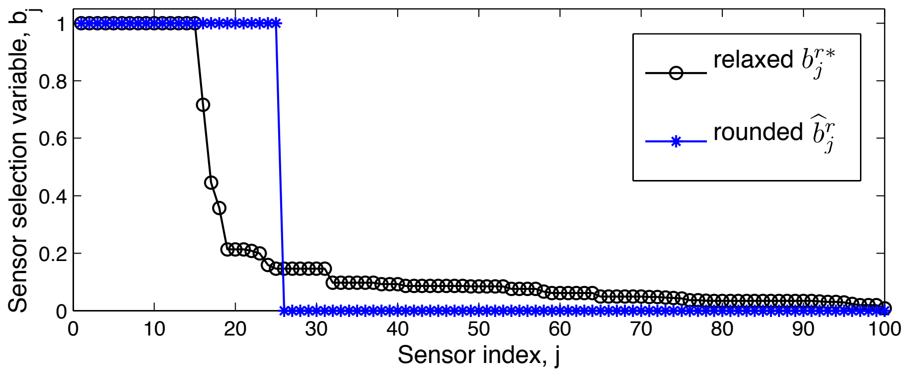


Figure 5.7: Solution of sensor selection subproblem $\{b_j^{r*}\}_{j=1}^n$ and its rounded values $\{\widehat{b}_j\}_{j=1}^n$ in descending order for the topology 1 shown in Figure 5.3.

The solution (relaxed and rounded values) of sensor selection subproblem, that is, $\{b_j^*\}_{j=1}^n \in \{0,1\}$ and $\{\widehat{b}_j\}_{j=1}^n \in [0,1]$ are shown in Figure 5.7, where these values are represented in descending order, from which, for example, we select a subset V_T^* of $K = 25$ sensors. The solution, values of bit-rate allocation problem, that is, $B_j^*, j \in V_T^*$, is shown in Figure 5.8. Notice that the solution of sensor selection (after rounding) also provides a routing structure T^* constituting the subset V_T^*, the LO-AQ algorithm improves this routing structure as T^a and the bit-rate allocation $B_j^a, j \in V_T^*$. The routing structures and bit allocations obtained for FTR-AQ and LO-AQ are shown in Figure 5.8. The routing structure associated to FTR-AQ is represented by thick green edges and the routing structure associated to LO-AQ is represented by thin blue edges. A bit-rate allocation for a node is represented as $[B_j^*, B_j^a]$ where

B_j^* is the bit-rate obtained from FTR-AQ and B_j^a the one obtained from LO-AQ. The underlying network connectivity graph is represented by thiner gray edges.

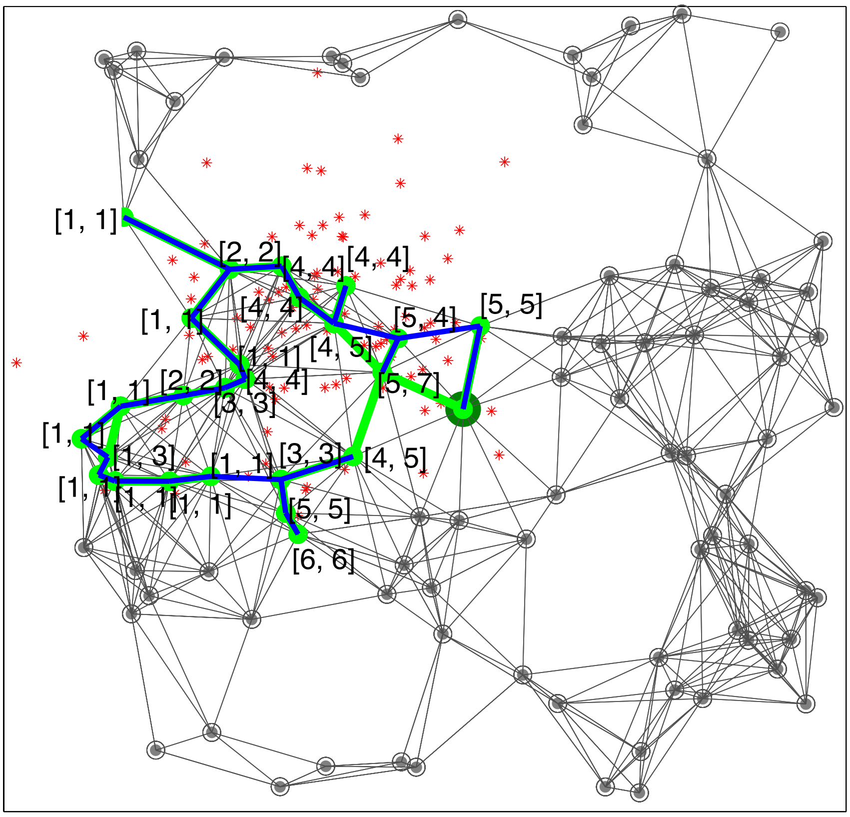


Figure 5.8: Subset of selected sensors and routing structure for FTR-AQ (thick green edges) and for LO-AQ (thin blue edges), for a given power budget $P_{\max} = 15$, where $MSE_{\text{FTR-AQ}} = 2.71 \times 10^{-3}$ and $MSE_{\text{LO-AQ}} = 8.96 \times 10^{-4}$. Thinner edges belong to the underlying network connectivity graph and big circle at the center is the sink node. Numbers in brackets represent the values of bit-rates for both algorithms, that is, $[B_j^*, B_j^a]$.

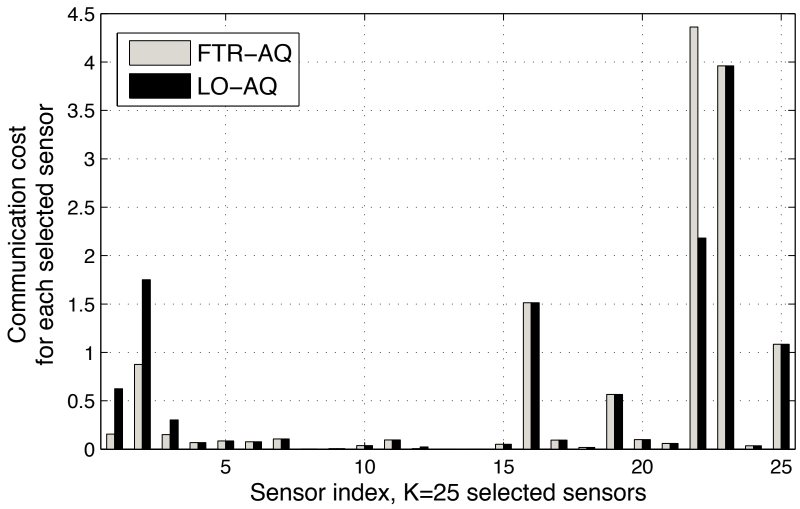


Figure 5.9: Amount of communication cost consumed by each individual sensor from the selected subset shown in Figure 5.8.

Then, we present in Figure 5.9 the amount of communication cost consumed by each individual sensor from the selected subset (see Figure 5.8). The total communication costs for FTR-AQ and LO-AQ are 13.82 and 13.75 for a given $P_{\max} = 14$, respectively.

We also present a simulation result assuming all sensors are selected from the solution of the sensor selection subproblem. In Figure 5.10, we show an example of bit-rate allocation obtained from our proposed algorithms. Notice that the quantization error term at the sink node is given by $\frac{\sigma^2_{q,j}}{\mathcal{P}_j} = \frac{\sigma^2_{q,j}}{\prod_{l \in \Omega_j} \left(1+|\mathcal{C}_l|\right)^2} = \frac{\sigma^2_{q,j}}{\left(1+|\mathcal{C}_l|\right)^{2(r-1)}}$ (see (3.36)), where r is the number of hops between sensor j and the sink node S, hence $r-1$ is the size of set Ω_j. It can be noticed that sensors close to the sink node (small r, that is, small normalizing distance) need higher bit-rates in order to maintain the term $\frac{\sigma^2_{q,j}}{\mathcal{P}_j}$ low. On the other hand, if r is large, low bit-rates are sufficient to maintain the term $\frac{\sigma^2_{q,j}}{\mathcal{P}_j}$ small. There may be few sensors that are far away from the sink node (large normalizing distance) and need higher bit-rate. This is because their parents may be very close to them as well as to the source target.

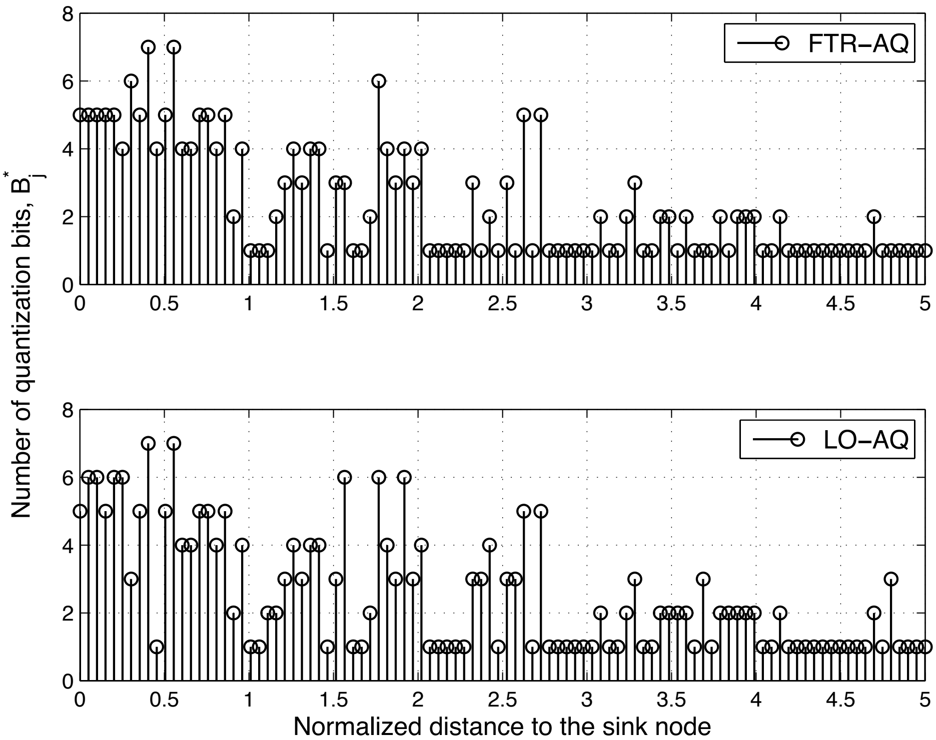


Figure 5.10: Number of quantization bits B_j allocated to each sensor versus the normalized Euclidean distance from the sensor to the sink node while activating all sensors for a fixed power budget $P_{\max} = 35$.

Chapter 6

Conclusions and Future Work

6.1 Summary of Contributions

In this dissertation work, given a WSN with a certain underlying network connectivity graph, a certain querying (sink) node and a localized source target, we have considered the problem of distributed parameter estimation, where power constraints are imposed. Then, in order to achieve a certain estimation task (e.g., forest-fire detection, localization based on direction-of-arrival sensors or estimation of any other localized phenomena), we have considered the problem, using an EF scheme (see Section 1.1), of jointly optimizing the subset of sensors, bit-rate allocation and associated multihop routing structure to send the aggregated information to the sink node. Our goal is to minimize the total distortion in the estimation subject to a given certain total power budget.

In the literature, it has been widely shown that multihop wireless data transmission in WSNs is usually much more power efficient than direct data transmission in which each measurement is directly transmitted to the sink node. Multihop data transmission allows to send measurements to the sink node via two schemes either by simply forwarding the measurements via intermediate sensor nodes in the routing tree (an MF scheme, see Section 1.1) or fusing them along the routing path at each intermediate sensor node, thus forwarding only the aggregated estimate, and thus using an EF scheme. Notice that fusing the measurements at each intermediate sensor node provides an improvement in overall estimation quality with a very small extra computational cost. Thus, in our work, we have considered Pareto optimization between two important metrics namely, the distortion in estimation and total multihop communication cost. In addition to the routing structure and sensor selection, we have also considered the quantization process at the nodes so that each node sends its quantized information to its parent sensor on the chosen multihop path. Notice that the information at each sensor is fused with other received quantized information (if available) from its child sensors, and then the resulting fused information is sent via a multihop path to the sink node for the final estimation. This dissertation provides the following contributions:

- First, we have considered an optimization problem in order to jointly optimize the sensor selection and multihop routing structure assuming that a fine quantization is available for each sensor measurement (as shown in Section 3.2.2). This problem is a non-convex problem (integer optimization problem), and it has been shown to be an NP-Hard problem. Furthermore, in order to solve it efficiently, it has been simplified into a convex problem by means of relaxation of the integer constraints (see Section 3.2.4), and then mapping back the solution from the later problem to the original one.

- We have solved our optimization problem by two different approaches, namely, Fixed-Tree Relaxation-based Algorithm (FT-RA) (see Section 4.1) and Iterative Distributed Algorithm (IDA) (see Section 4.2). The FTRA is based on a relaxation of our original optimization problem (as stated above) and that decouples the choice of the sensor selection and routing structure. On the other hand, the algorithm IDA jointly optimizes both metrics, resulting in a better performance compared to the FTRA algorithm. We have also provided a lower bound (see Lemma 3.2.2) for the optimal solution of our original NP-Hard optimization problem and have shown experimentally that our IDA generates a solution that is close to this lower bound, thus approaching optimality.

- Second, we have considered *bit-rate allocation* as an additional variable in the first problem, which becomes a nonlinear non-convex optimization problem. The resulting joint optimization problem is also NP-Hard and becomes even harder to solve than the first optimization problem.

- For this other optimization problem, we have developed two approaches to jointly optimize the subset of sensors, the bit-rate allocation across the nodes and the associated multihop routing structure. Since this problem is a nonlinear non-convex optimization problem, very interestingly, we address this nonlinear non-convex optimization problem using several relaxation steps and then solving the relaxed convex version over different variables in tandem, resulting in a sequence of linear (convex) subproblems that can be solved efficiently. We call our two approaches Fixed-Tree Relaxation-based Adaptive Quantization (FTR-AQ) algorithm (see Section 5.1) and Local Optimization-based Adaptive Quantization (LO-AQ) algorithm (see Section 5.2). As we have shown the LO-AQ algorithm provides better estimation accuracy for the same given total power budget, at the expenses of additional computational complexity.

6.2 Conclusions

In this PhD thesis, we consider a deterministic parameter estimation scenario where a WSN covers a large geographical area and it monitors a *localized phenomenon*. An important result of our work is to show clearly that the optimal routing structure is not an SPT-CC in general, due to the interplay between the communication cost and gain in estimation when fusing measurements from different sensors. Depending on the location of sensors, the information provided by a sensor about the phenomenon will be more or less important. On the other hand, the cost of communicating each measurement to the sink node through a multihop route may also vary widely from sensor to sensor. There is a need to design efficient joint sensor selection and multihop routing algorithms for battery-powered WSNs where the choice of a subset of sensors and a routing subtree affects both communication cost and estimation accuracy. Furthermore, adaptive quantization at each node also plays an important role as the sensors that are far away from the source target require less quantization levels since they have very low sensing SNR (as illustrated in Figure 5.10).

Most of the solutions that have been proposed in the recent literature, intend to reduce the problem to independently the sensor selection or to the source coding (quantization) optimization, ignoring the joint optimization of the routing structure with the sensor selection and quantization optimization. Optimizing the routing structure is an important variable in the problem, since in general, transmitting an information that is far away from the sink node requires more energy than one that is close to it.

6.3 Future Research Lines

Next, we provide a description list of the possible future research lines related to the problems addressed in this thesis. We have already developed some initial algorithms in this context, but there is still a substantial amount of open problems and extensions that can still be done afterwards. The list can be enumerated as below:

1. Use of a random parameter instead of a deterministic parameter in our signal model (3.4). This would allow us to introduce correlated measurements (even if the additive noise is independent across nodes). In this case, the joint optimization of the sensor selection, bit-rate allocation and multihop routing will result in solution that differ from the ones consider here since the value of an additional measurements would certainly depend on the already selected sensor measurements. A local whitening strategy [52] can be employed so that the correlated measurements can be transformed into uncorrelated measurements, and thus the BLUE estimator can be used to estimate the unknown parameter.

2. Considering more realistic link models in the optimization, instead of only path-loss model. In this case, for instance, an estimate-amplify-and-forward technique could be used [123].

3. In our work, a localized static source target is considered. However, considering the tracking of a moving source target and for this purpose optimizing, for each sampling period, the sensor subset selection and an associated multihop routing structure is a promising line of research, and an immediate extension of this work. In this

thesis, we have only consider the problem formulation and have obtained some initial results, as given in Section 4.5.

4. As we have already been noticed in Section 4.5.3 that sometimes the communication cost between two consecutive LSNs is very high. Therefore, along with the predicted detection probability there must also be some other criteria to select the next LSN, that is, there could be a trade-off between predicted detection probability and the communication cost between two consecutive LSNs.

5. Our work is also focused in a scenario where sensors are static, using dynamic mobile sensors can also be considered where a subset of sensors with certain assigned bit-rates can be updated every certain time interval, as well as its associated multihop routing structure.

6. It has been observed from our simulations that the sensor node that is closest to the sink node will generally die first (run out of battery) because of the amount of information packets that it will have to forward. This problem can be alleviated by considering multiple sink nodes and/or by considering optimizing an additional variable that keeps record of the remaining battery power level at each sensor node. In this way, at a certain instance the most powered nodes will be more used than the least powered nodes, which will significantly improve the network lifetime. On the other hand, adding this additional variable, the optimization problems will become even more complex and hard to solve, making it more challenging.

Appendix A

Proof of NP-Hardness: Fine Quantization

Proof In general, the joint sensor selection and routing problem (3.15):

$$
\begin{aligned}
\underset{\{b_j,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} (\text{w.r.t. } \mathbf{S}_+^m) \quad & \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \operatorname{Tr}\left[\left(\sum_{j=1}^{n+1} b_j \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] \\
\text{subject to} \sum_{j=1, i=\text{parent}_j}^{n+1} & b_j f_c(d_{i,j}) = P_{\max} - P_{\text{gap}} \\
& b_j \le b_i,\ i = \text{parent}_j,\ A_{i,j} = 1 \\
& b_{n+1} = 1 \\
& P_{\text{gap}} \ge 0 \\
& b_j \in \{0,1\},\ j = 1, \ldots, n
\end{aligned}
\tag{A.1}
$$

belongs to a class of integer optimization problems [18; 79] and it is typically NP-Hard [8; 17; 30; 57; 101]. For the sake of simplicity, we consider $m = 1$, that is, a one-dimensional scalar parameter, thus the distortion is given by $\Sigma_{\hat{\theta}} = (\sum_{j=1}^{n} b_j h_j^2)^{-1}$, as it can also be seen in an equivalent problem (3.19). Notice also that the sink node is not responsible to take the measurement, that is, $h_{n+1} = 0$. Moreover, minimizing $(\sum_{j=1}^{n} b_j h_j^2)^{-1}$ is equivalent to:

$$\begin{aligned}
\min \log \left(\sum_{j=1}^{n} b_j h_j^2 \right)^{-1} \quad &= \min \left(-\log \sum_{j=1}^{n} b_j h_j^2 \right) \\
&\Leftrightarrow \max \log \sum_{j=1}^{n} b_j h_j^2 \\
&\Leftrightarrow \max \sum_{j=1}^{n} b_j h_j^2 \\
&\Leftrightarrow \min \left(-\sum_{j=1}^{n} b_j h_j^2 \right) \qquad \text{(A.2)}
\end{aligned}$$

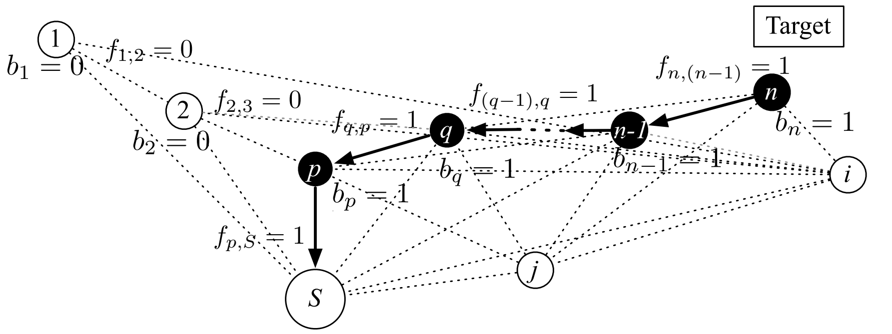


Figure A.1: Graphical representation of the graph G with n sensors.

Again, for simplicity, we assume that our problem is to find a single-path, in our graph of n sensors, from sensor n (nearest to the source target t) to the sink node S (as shown in Figure A.1). Then, we perform the reduction from another well-known NP-Hard problem to this problem, which is actually a simplification of our problem (3.15). Notice that a communication cost can be expressed in terms of the set of flows $f_{i,j} \in \{0, 1\}$ connecting each pair of sensors i and j, using the typical notation as in the network flow problem (3.17). Considering these facts, one can use an equivalent overall weighted objective by combining the objective and the constraint function to form Pareto optimization, thus in this case problem (3.17) can be written as:

$$
\begin{aligned}
\underset{\{b_j, f_{i,j}\}}{\text{minimize}} \quad & -\sum_{j=1}^{n} b_j h_j^2 + \eta \sum_{(i,j)\in E} f_{i,j} f_c(d_{i,j}) \\
\text{subject to} \quad & \tilde{\boldsymbol{A}}\boldsymbol{f} = \boldsymbol{\psi} = \boldsymbol{s} - \boldsymbol{u} \\
& \psi_n = 1 \\
& \psi_S = -1 \\
& \psi_j = 0 \ \ \forall j \neq n, S \\
& u_{n+1} \geq 1 \\
& \boldsymbol{f} \in \{0, 1\} \\
& b_j = \begin{cases} 1 & \text{if} \quad f_{i,j} \geq 0 \text{ or } f_{i,j} \geq 0 \forall i \neq j \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{A.3}
$$

where ψ represents the total flows at any given node, and therefore for a single path between node n and S (as in Figure A.1) all other nodes other than n and S having one incoming and one outgoing flow, which results $\psi_j = 0 \, \forall j \neq n, S$. The global scalarization parameter $\eta > 0$ controls the importance of both metrics in (A.3). Notice that, for each power budget constraint $\sum_{(i,j)\in E} f_{i,j} f_c(d_{i,j})$ the value of η changes as a function of $P_{\max}$, η increases as $P_{\max}$ decreases and visa versa. Our goal is to find an optimal subset of sensors that includes closest sensor n to the target, which is equivalent to a path between the sensor n and the sink node S in our graph. The first term in the objective function, for example, for a subset of K sensors can be re-written as (see Figure A.1):

$$
\begin{aligned}
\sum_{j=p}^{n} b_j h_j^2 &= b_p h_p^2 + b_q h_q^2 + \cdots + b_{n-1} h_{n-1}^2 + b_n h_n^2 \\
\sum_{j=p}^{n} b_j h_j^2 &= f_{q,p} h_p^2 + f_{(q-1),q} h_q^2 + \cdots + f_{n,(n-1)} h_{n-1}^2 + h_n^2 \\
\sum_{j=p}^{n} b_j h_j^2 &= \sum_{(i,j)\in E_K} f_{i,j} h_j^2 + h_n^2
\end{aligned}
\tag{A.4}
$$

then, the objective in (A.3) for a subset of K sensors becomes:

$$
\underset{\{f_{i,j}\}}{\text{minimize}} \quad \sum_{(i,j)\in E_K} f_{i,j} (\eta f_c(d_{i,j}) - h_j^2) - h_n^2 \tag{A.5}
$$

where the number of edges in a subset $E_K \subset E$ is K.

The objective (A.5) is independent of b_j (no sensor selection), and it is equivalent to a Direct Hamiltonian Path (DHP) problem from the sensor node n to the sink node S in a directed graph $G = (V, E)$ where the edges have a cost equal to $\tilde{f}_c(d_{i,j}) = \eta f_c(d_{i,j}) - h_j^2$. Notice that, when we weight less the communication cost term (i.e. a low η), we can create negative cycles in the equivalent graph and the standard DHP problem becomes unbounded (infinite flow on negative weights). Interestingly, this problem is much harder than the usual DHP problem when there are no negative cycles, where for instance Karp's algorithm [51] can be applied (polynomial time). In fact, it can be shown that the general DHP with negative weights is NP-Hard [95], as we show next.

Let us consider activating another sensor (for example, sensor i as in Figure A.1), then we prove the NP-Hardness of our problem by showing that the problem of generating a single path from i to S, namely, the (i, S) DHP, is NP-Hard. In order to do so, we show that the DHP problem, which is known to be NP-complete [30], reduces to the (i, S) DHP. We consider an instance for the DHP problem, such as the one shown in Figure A.1 and add to it an extra vertex i, which is connected to every vertex $v \in V$, adding the corresponding edge, that is, we add an extra edge (i, v). This will generate another directed graph $G_i = (V_i, E_i)$. If there is a Hamiltonian path in G_i, there will be an edge for i to the one of the vertex of G, which will ensure that the new Hamiltonian path starts at i and ends at S covering each vertex (in the selected subset) of G. Therefore, it ensures that there is a Hamiltonian path (i, S) in G_i. Conversely, if there is a Hamiltonian path (i, S) in G_i, the vertex excluding i will also form a Hamiltonian path in G. Since this transformation (adding extra vertex and edge) can be done in polynomial time, we have that $DHP \leq_P (i, S)$ DHP. Moreover, since (i, S) $DHP \in NP$, we can conclude that (i, S) DHP is NP-complete and its associated optimization problem is NP-Hard. ∎

Appendix B

Verification using Newton's Method

We describe Newton's method (see [10], Sec. 10.2 for more details) for solving our problem (3.22):

$$
\begin{aligned}
&\underset{\{b_j^r,\text{parent}_j,P_{\text{gap}}\}}{\text{minimize}} \qquad \phi(\mathbf{b}^r) = \mathrm{Tr}\left[\left(\sum_{j=1}^{n+1} b_j^r \boldsymbol{h}_j \boldsymbol{h}_j^T\right)^{-1}\right] - \\
&-\frac{1}{\nu}\sum_{j=1}^{n+1}\left(\log(b_j^r) + \log(1-b_j^r) + \log(b_i^r - b_j^r) + \log P_{\text{gap}}\right) \qquad \text{(B.1)}\\
&\quad \text{subject to} \quad \sum_{j=1}^{n+1} b_j^r f_c(d_{i,j}) = P_{\max} - P_{\text{gap}}\\
&\qquad\qquad\qquad\qquad b_{n+1}^r = 1
\end{aligned}
$$

We take $\texttt{diag}(\boldsymbol{b}^r)\boldsymbol{f_c} = (P_{\max}/n)\mathbf{1}$ as an initial (feasible) point (where $\boldsymbol{f_c}^T = [f_c(d_{1,p_1}),\ldots,f_c(d_{n,p_n})]$ and p_j is the parent of j) and the definition of Newton's search step $\partial\boldsymbol{b}_{nt}^r$ is modified at each step to take the equality constraints into account, leading to this expression:

$$
\partial\boldsymbol{b}_{nt}^r = -(\nabla^2\phi)^{-1}\nabla\phi + \left(\frac{\boldsymbol{f_c}^T(\nabla^2\phi)^{-1}\nabla\phi}{\boldsymbol{f_c}^T(\nabla^2\phi)^{-1}\boldsymbol{f_c}}\right)(\nabla^2\phi)^{-1}\boldsymbol{f_c} \qquad \text{(B.2)}
$$

where $\nabla\phi$ and $\nabla^2\phi$ are the gradient and Hessian of function ϕ, respectively. Then, we use a backtracking line search to choose a step size

$\delta \in (0, 1]$, and update $\boldsymbol{b}^r$ as:

$$\boldsymbol{b}^r[j+1] = \boldsymbol{b}^r[j] + \delta \partial \boldsymbol{b}^r_{nt} \tag{B.3}$$

We stop when the Newton decrement $(-\nabla \phi(\boldsymbol{b}^r)^T \partial \boldsymbol{b}^r_{nt})^{1/2} \leq \varepsilon$, for $\varepsilon > 0$ sufficiently small. The total number of steps required is typically ten or fewer. For completeness, expressions for the derivatives of ϕ in our problem, in terms of its gradient $\nabla\phi$ and the Hessian $\nabla^2\phi$ can be written as:

$$(\nabla\phi)_j = -\boldsymbol{h}_j^T \boldsymbol{M}^2 \boldsymbol{h}_j - \frac{1}{\nu}\left(\frac{1}{b^r_j} - \frac{1}{1-b^r_j} - \frac{1}{b^r_i - b^r_j}\right) \tag{B.4}$$

where $\boldsymbol{M} = (\sum_{j=1}^n b^r_j \boldsymbol{h}_j \boldsymbol{h}_j)^{-1}$.

The Hessian $\nabla^2\phi$ is given by:

$$\begin{aligned}\nabla^2\phi = 2(\boldsymbol{H}\boldsymbol{M}^{3/2}\boldsymbol{H}^T) \circ (\boldsymbol{H}\boldsymbol{M}^{3/2}\boldsymbol{H}^T) + \frac{1}{\nu}\,\texttt{diag}\Bigg(&\frac{1}{(b^r_1)^2} + \frac{1}{(1-b^r_1)^2} + \\ &+ \frac{1}{(b^r_{p_1} - b^r_1)^2} + \cdots + \frac{1}{(b^r_n)^2} + \frac{1}{(1-b^r_n)^2} + \frac{1}{(b^r_{p_n} - b^r_n)^2}\Bigg)\end{aligned} \tag{B.5}$$

where $\circ$ denotes the Hadamard (elementwise) product.

Appendix C

An Upper Bound for Nonlinear Recursion of $\sigma^2_{\mathcal{Q}_k}$

Recall that (3.33) is given by:

$$
\begin{aligned}
&(1+|\mathcal{C}_j|)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2} \prod_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k} \leq \\
&\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sum_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k}\right)\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} \sum_{k\in\mathcal{C}_j} \prod_{s\in\mathcal{C}_j; s\neq k} \sigma^2_{\mathcal{Q}_s} + \prod_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k}\right)
\end{aligned}
\tag{C.1}
$$

Without loss of generality, we assume that $\sigma^2_{\mathcal{Q}_{k_1}} \leq \sigma^2_{\mathcal{Q}_{k_2}} \leq \cdots \leq \sigma^2_{\mathcal{Q}_{k_a}}$ where $\{k_1, \ldots, k_a\} \in \mathcal{C}_j$ are the children noes of sensor node j, and then using (C.1), we define a function $f(.)$ as:

$$
f(|\mathcal{C}_j|) = \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sum_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k}\right)\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} \sum_{k\in\mathcal{C}_j} \prod_{s\in\mathcal{C}_j; s\neq k} \sigma^2_{\mathcal{Q}_s} + \prod_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k}\right)}{(1+|\mathcal{C}_j|)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2} \prod_{k\in\mathcal{C}_j} \sigma^2_{\mathcal{Q}_k}}
\tag{C.2}
$$

now, for all $|\mathcal{C}_j| \geq 1$ where $j = 1, \ldots, n$, we need to prove $f(|\mathcal{C}_j|) \geq 1$, which satisfies (C.1).

Case 1: $|\mathcal{C}_j| = 1$, that is, 1-D network where only one child node is available to each sensor node. In this case, (C.2) becomes:

$$f(1) = \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)}{4\frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}} = \left[\frac{\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}}{2\frac{\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j}}\right]^2 \tag{C.3}$$

$$f(1) = \left[\frac{\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}}{2\frac{\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j}} - 1 + 1\right]^2 = \left[\frac{\left(\frac{\sigma_{y_j}}{\sqrt{b_j}h_j} - \sigma_{\mathcal{Q}_k}\right)^2}{2\frac{\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j}} + 1\right]^2 \tag{C.4}$$

$$f(1) \geq 1$$

notice that the numerator term $\left(\frac{\sigma_{y_j}}{\sqrt{b_j}h_j} - \sigma_{\mathcal{Q}_k}\right)^2$ in (C.4) is positive, thus $f(1) \geq 1$ is satisfied for $|\mathcal{C}_j| = 1$. Furthermore, $\left(\frac{\sigma_{y_j}}{\sqrt{b_j}h_j} - \sigma_{\mathcal{Q}_k}\right)^2 \geq 0$ gives $\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k} \geq 2\frac{\sigma_{y_j}\sigma_{\mathcal{Q}_k}}{\sqrt{b_j}h_j}$, which is used to generate (3.31).

Case 2: $|\mathcal{C}_j| = a > 1$, that is, 2-D network, in this case each sensor node can have more than one child. Assume that there are a children for $\mathcal{C}_j$ denoted as $\{k_1, \ldots, k_a\} \in \mathcal{C}_j$. Then (C.2), in this case, can be written as:

$$f(a) = \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_{k_1}} + \cdots + \sigma^2_{\mathcal{Q}_{k_a}}\right)\left(\begin{array}{c} \frac{\sigma^2_{y_j}}{b_j h_j^2}\left(\sigma^2_{\mathcal{Q}_{k_2}}\sigma^2_{\mathcal{Q}_{k_3}}...\sigma^2_{\mathcal{Q}_{k_a}} + ... \right. \\ \left. + \sigma^2_{\mathcal{Q}_{k_1}}\sigma^2_{\mathcal{Q}_{k_2}}...\sigma^2_{\mathcal{Q}_{k_{a-1}}}\right) + \sigma^2_{\mathcal{Q}_{k_1}}\sigma^2_{\mathcal{Q}_{k_2}}...\sigma^2_{\mathcal{Q}_{k_a}} \end{array}\right)}{(1+a)^2\frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_{k_1}}\sigma^2_{\mathcal{Q}_{k_2}}...\sigma^2_{\mathcal{Q}_{k_a}}} \tag{C.5}$$

For the sake of simplicity, without loss of generality, we assume that $\sigma^2_{\mathcal{Q}_k} = \sigma^2_{\mathcal{Q}_{k_1}} = \cdots = \sigma^2_{\mathcal{Q}_{k_a}}$, then in this case (C.5) becomes:

$$\begin{aligned} f(a) &= \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + a\sigma^2_{\mathcal{Q}_k}\right)\left(\frac{a\sigma^2_{y_j}}{b_j h_j^2}\sigma^{2(a-1)}_{\mathcal{Q}_k} + \sigma^{2a}_{\mathcal{Q}_k}\right)}{(1+a)^2\frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^{2a}_{\mathcal{Q}_k}} \\ &= \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + a\sigma^2_{\mathcal{Q}_k}\right)\left(\frac{a\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)}{(1+a)^2\frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}} \end{aligned} \tag{C.6}$$

$$f(a) = \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} + a\sigma^2_{\mathcal{Q}_k}\right)\left(\frac{a\sigma^2_{y_j}}{b_j h_j^2} + \sigma^2_{\mathcal{Q}_k}\right)}{(1+a)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}} - 1 + 1 \tag{C.7}$$

$$f(a) = \frac{\frac{a\sigma^4_{y_j}}{b_j^2 h_j^4} + a\sigma^4_{\mathcal{Q}_k} + (1+a^2)\frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k} - (1+a)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}}{(1+a)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}} + 1 \tag{C.8}$$

$$f(a) = \frac{\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} - \sigma^2_{\mathcal{Q}_k}\right)^2}{(1+a)^2 \frac{\sigma^2_{y_j}}{b_j h_j^2}\sigma^2_{\mathcal{Q}_k}} + 1 \geq 1 \tag{C.9}$$

again the numerator term $\left(\frac{\sigma^2_{y_j}}{b_j h_j^2} - \sigma^2_{\mathcal{Q}_k}\right)^2$ in (C.9) is positive. Thus, ensures that $f(a) \geq 1$ for $|\mathcal{C}_j| = a > 1$.

Notice that even if all $\sigma^2_{Q_k}$ in (C.5) are different, the numerator term in (C.6) will still be positive. Therefore, $f(a) \geq 1$ for $|\mathcal{C}_j| = a > 1$ always holds.

Appendix D

Proof of NP-Hardness: Adaptive Quantization

Proof Rewriting the reduction technique from Subsection 3.2.3, which will help to follow the following procedure. It defines that a problem $\mathcal{L}_1$ is reducible to another problem $\mathcal{L}_2$ ($\mathcal{L}_1 \leq_p \mathcal{L}_2$), if a process for solving problem $\mathcal{L}_2$ efficiently, if it is solvable, could also be used as a subprocess to solve problem $\mathcal{L}_1$ efficiently. When this is true, then $\mathcal{L}_2$ is as hard to solve than $\mathcal{L}_1$.

Let us assume that $\mathcal{L}_1$ is our nonlinear non-convex optimization problem (3.38):

$$\begin{aligned}
\underset{\{b_i, b_j, B_j\}}{\text{minimize}} \quad & \sum_{j=1}^{n} \left(\frac{4 b_j \mathcal{A}^2}{3 \mathcal{P}_j 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j (1 + |\mathcal{C}_j|)^2 b_j h_j^2} \right) \\
\text{subject to} \quad & \sum_{j=1, j \in \mathcal{C}_i}^{n} b_j N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i, j \in \mathcal{C}_i, A_{i,j} = 1 \\
& b_j \in \{0, 1\}, j = 1, \ldots, n \\
& P_{\text{gap}} \geq 0 \\
& B_j \geq 0, j = 1, \ldots, n
\end{aligned} \tag{D.1}$$

and it can be reducible to another problem $\mathcal{L}_2$ in such a way that the variable B_j is given, or assume a fine quantization, thus we can assume that $B_j \to \infty$ in (D.1). Notice that for each $B_j \to \infty$, the first term in the objective of (D.1) is tending to zero that is $\sigma^2_{q,j} \rightarrow 0$, and therefore from (3.29), MSE of the estimator at sensor j is given by $\sigma^2_{\hat{\theta}_j} = \left(\frac{b_j h_j^2}{\sigma^2_{y_j}}\right)^{-1}$. Thus, the total MSE is given by $\left(\sum_{j=1}^{n} \frac{b_j h_j^2}{\sigma^2_{y_j}}\right)^{-1}$, which is now assumed as the problem $\mathcal{L}_2$; a reduced version of $\mathcal{L}_1$, and it is given by:

$$
\begin{aligned}
\underset{\{b_j, b_k\}}{\text{minimize}} \quad & \left(\sum_{j=1}^{n} \frac{b_j h_j^2}{\sigma^2_{y_j}}\right)^{-1} \\
\text{subject to} \quad & \sum_{j=1,(i,j)\in G}^{n} b_j N_j d^{\alpha}_{j,k} = P_0 - P_{0,\text{gap}} \\
& b_j \leq b_k, \ \text{for } (i,j) \in G, \ \text{and } A_{i,j} = 1 \\
& b_j \in \{0,1\} \\
& P_{\text{gap}} \geq 0
\end{aligned}
\tag{D.2}
$$

where $P_{\max} - P_{\text{gap}}$ is replaced by $P_0 - P_{0,\text{gap}} = \frac{P_{\max}}{2^{B_j}} - P_{0,\text{gap}}$ and P_0 is assumed sufficient enough as to satisfy the equality constraint in (D.2). Then, the optimization problem (D.2) for distributed estimation subject to a total power constraint, is NP-Hard. Since this problem is a combinatorial optimization problem, in which the variable $b_j \in \{0,1\}$.

We have already been shown that the problem (D.1) is reducible to another problem (D.2) and the process of solving (D.2) efficiently, will be used as a subprocess to solve the problem (D.1) efficiently, which is the case described in Section (5.1). Therefore, the problem (D.2) is at least as hard to solve than (D.1). Thus, it follows the reduction technique [30] and already defined in Appendix A, and we have already proven that the problem (D.2) is an NP-Hard (see in Appendix A). Thus, the nonlinear non-convex optimization problem (D.1) is an NP-Hard problem.

Appendix E

Solution of Problem (5.5)

We describe Newton's method for computing our problem (5.5):

$$\underset{\{b_j^r\}}{\text{minimize}} \quad \phi(\boldsymbol{b}^r) = \left(\sum_{j=1}^{n} \frac{b_j^r h_j^2}{\sigma_{y_j}^2}\right)^{-1} -$$

$$-\frac{1}{\eta}\sum_{j=1}^{n}\left(\log(b_j^r)+\log(1-b_j^r)+\log(b_i^r-b_j^r)+\log P_{0,\text{gap}}\right) \tag{E.1}$$

$$\text{subject to} \sum_{j=1, j\in\mathcal{C}_i}^{n} b_j^r N_j d_{i,j}^{\alpha} = P_0 - P_{0,\text{gap}}$$

more details can be seen, for example, in[96; 98] and ([10], Sec. 10.2). We start by initializing the variable b_j^r as $\texttt{diag}(\boldsymbol{b}^r)\boldsymbol{f_c} = \left(\frac{P_{\max}}{n}\right)\mathbf{1}$, where $\boldsymbol{f_c}^T = [f_c(d_{1,p_1}), \ldots, f_c(d_{n,p_n})])$. Then, we define Newton's search step by $\partial\boldsymbol{b}_{nt}^r$, and it is modified at each step to take the equality constraint into account, which is given by:

$$\partial\boldsymbol{b}_{nt}^r = -(\nabla^2\phi)^{-1}\nabla\phi + \left(\frac{\boldsymbol{f_c}^T(\nabla^2\phi)^{-1}\nabla\phi}{\boldsymbol{f_c}^T(\nabla^2\phi)^{-1}\boldsymbol{f_c}}\right)(\nabla^2\phi)^{-1}\boldsymbol{f_c} \tag{E.2}$$

where $\nabla\phi$ and $\nabla^2\phi$ are the gradient and Hessian of function ϕ, respectively. We then choose a step size $\delta \in (0, 1]$ based on a backtracking line search, and update $\boldsymbol{b}^r$ as:

$$\boldsymbol{b}^r[j+1] = \boldsymbol{b}^r[j] + \delta\partial\boldsymbol{b}_{nt}^r \tag{E.3}$$

We stop when the Newton decrement $(-\nabla\phi(\boldsymbol{b}^r)^T \partial \boldsymbol{b}^r_{nt})^{1/2} \leq \varepsilon$, for $\varepsilon > 0$ sufficiently small. The total number of steps required is typically ten or fewer. For completeness, Expressions for the derivatives of ϕ in our problem, in terms of its gradient $\nabla\phi$ and the Hessian $\nabla^2\phi$ can be written as:

$$(\nabla\phi)_j = -\left(\frac{(b^r_j)^2 h^2_j}{\sigma^2_{y_j}}\right)^{-1} - \frac{1}{\eta}\left(\frac{1}{b^r_j} - \frac{1}{1-b^r_j} - \frac{1}{b^r_k - b^r_j}\right) \tag{E.4}$$

The Hessian $\nabla^2\phi$ is given by:

$$\begin{aligned}\nabla^2\phi = 2\left(\sum_{j=1}^{n} \frac{(b^r_j)^3 h^2_j}{\sigma^2_{y_j}}\right)^{-1} + \\ + \frac{1}{\eta}\,\mathtt{diag}\left(\frac{1}{(b^r_1)^2} + \frac{1}{(1-b^r_1)^2} + \frac{1}{(b^r_{\mathcal{C}_1} - b^r_1)^2} + \right. \\ \left. \cdots + \frac{1}{(b^r_n)^2} + \frac{1}{(1-b^r_n)^2} + \frac{1}{(b^r_{\mathcal{C}_n} - b^r_n)^2}\right)\end{aligned} \tag{E.5}$$

Appendix F

Proof of Non-Convexity of Problem (5.1)

Recall that (5.1) is given by:

$$\begin{aligned}
\underset{\{b_i, b_j, B_j\}}{\text{minimize}} \quad & \sum_{j=1}^{n} \left(\frac{4b_j \mathcal{A}^2}{3\mathcal{P}_j 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j (1+|\mathcal{C}_j|)^2 b_j h_j^2} \right) \\
\text{subject to} \quad & \sum_{j=1,(i,j)\in G}^{n} b_j N_j d_{i,j}^{\alpha} 2^{B_j} = P_{\max} - P_{\text{gap}} \\
& b_j \leq b_i, \text{ for } (i,j) \in G, \text{ and } A_{i,j} = 1 \qquad (a) \\
& b_j \in \{0,1\}, j = 1, \ldots, n \qquad (b) \\
& P_{\text{gap}} \geq 0 \\
& B_j \geq 0, j = 1, \ldots, n
\end{aligned} \tag{F.1}$$

Taking Hessian of the objective function, we can prove the convexity of this problem, which is as follows:

$$\begin{aligned}
&\nabla^2 \left(\sum_{j=1}^{n} \left(\frac{4b_j \mathcal{A}^2}{3\mathcal{P}_j 2^{2B_j}} + \frac{\sigma_{y_j}^2}{\mathcal{P}_j (1+|\mathcal{C}_j|)^2 b_j h_j^2} \right) \right) = \\
&= \begin{bmatrix} \sum_{j=1}^{n} \left(\frac{2\sigma_{y_j}^2}{\mathcal{P}_j (1+|\mathcal{C}_j|)^2 b_j^3 h_j^2} \right) & -\sum_{j=1}^{n} \left(\frac{8\mathcal{A}^2 \log(2)}{3\mathcal{P}_j 2^{2B_j}} \right) \\ -\sum_{j=1}^{n} \left(\frac{8\mathcal{A}^2 \log(2)}{3\mathcal{P}_j 2^{2B_j}} \right) & \sum_{j=1}^{n} \left(\frac{16 b_j \mathcal{A}^2 \log^2(2)}{3\mathcal{P}_j 2^{2B_j}} \right) \end{bmatrix} \nsucceq 0
\end{aligned} \tag{F.2}$$

F. PROOF OF NON-CONVEXITY OF PROBLEM (5.1)

It can be seen that the Hessian of the objective function is not positive semidefinite, thus our problem 5.1 is a non-convex optimization problem.

Bibliography

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, **38**(4):393–422, 2002. 1

[2] Jamal N. Al-Karaki, Raza Ul-Mustafa, and Ahmed E. Kamal. Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms. *Computer Networks*, **53**(7):945–960, May 2009. 1, 2.4.1

[3] Edoardo Amaldi, Leo Liberti, Nelson Maculan, and Francesco Maffioli. Efficient edge-swapping heuristics for finding minimum fundamental cycle bases. In CelsoC. Ribeiro and SimoneL. Martins, editors, *Experimental and Efficient Algorithms*, **3059** of *Lecture Notes in Computer Science*, pages 14–29. Springer Berlin Heidelberg, 2004. (document), 5.1

[4] Edoardo Amaldi, Leo Liberti, Francesco Maffioli, and Nelson Maculan. Edge-swapping algorithms for the minimum fundamental cycle basis problem. *Mathematical Methods of Operations Research*, **69**:205–233, 2009. 10.1007/s00186-008-0255-4. (document), 5.1

[5] A. Anandkumar, A. Swami, J.E. Yukich, and Lang Tong. Energy scaling laws for distributed inference in random fusion networks. *Selected Areas in Communications, IEEE Journal on*, **27**(7):1203 –1217, September 2009. 1, 2.4.1

[6] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita. A line in the sand: A wireless

sensor network for target detection, classification, and tracking. *Computer Networks (Elsevier)*, **46**:605–634, 2004. 1, 2.4.1

[7] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak. Joint source-channel communication for distributed estimation in sensor networks. *Information Theory, IEEE Transactions on*, **53**(10):3629 –3653, oct. 2007. 1

[8] F. Bian, D. Kempe, and R. Govindan. Utility-based sensor selection. *Information Processing in Sensor Networks, The Fifth International Conference on*, IPSN 2006. pages 11–18, 2006. A

[9] Danny Bickson and Danny Dolev. Distributed sensor selection via gaussian belief propagation. *CoRR/abs /0907.0931*, 2009. 1, 2.4.2

[10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. (document), 2.4.2, 3.1, 3.2.2, 3.2.2, 3.2.2, 3.2.5, 3.2.5, 5.1.1, 5.1.1, B, E

[11] Kerem Bulbul and Ozgur Ercetin. Maximum precision-lifetime curve for joint sensor selection and data routing in sensor networks. *European Transactions on Telecommunications*, **18**(7):815–823, 2007. 1.1, 2.4.1

[12] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, **25**(2):21–30, 2008. 1

[13] Jae-Hwan Chang and L. Tassiulas. Maximum lifetime routing in wireless sensor networks. *Networking, IEEE/ACM Transactions on*, **12**(4):609–619, aug. 2004. 1, 1.1

[14] S. Chaudhari and V. Koivunen. Effect of quantization and channel errors on collaborative spectrum sensing. In *Signals, Systems and Computers, Conference Record of the Forty-Third Asilomar Conference on*, pages 528–533, November 2009. 2.4.5, 3.3

[15] Hongbin Chen. Performance-energy tradeoffs for decentralized estimation in a multihop sensor network. *Sensors Journal, IEEE*, **10**(8):1304–310, August 2010. 1, 2.4.1

[16] Junyong Chen and Weizhou Su. Optimal bit allocation for distributed estimation in wireless sensor networks. In *Control and Automation, IEEE International Conference on*, ICCA 2009. pages 92–97, December 2009. 2.4.1

[17] Razvan Cristescu, Baltasar Beferull-Lozano, Martin Vetterli, and Roger Wattenhofer. Network correlated data gathering with explicit communication: NP-completeness and algorithms. *IEEE/ACM Transactions on Networking*, **14**(1):41–54, February 2006. 1.1, 2.4.4, A

[18] Shuguang Cui, A. Goldsmith, and A. Bahai. Joint modulation and multiple access optimization under energy constraints. In *Global Telecommunications Conference, IEEE*, GLOBECOM '04. **1**, pages 151–155, December 2004. A

[19] Shuguang Cui, Jin-Jun Xiao, A.J. Goldsmith, Zhi-Quan Luo, and H.V. Poor. Energy-efficient joint estimation in sensor networks: analog vs. digital. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 2005. Proceedings (ICASSP '05). **4**, pages iv/745–iv/748, March 2005. 1

[20] A.K. Das and M. Mesbahi. Distributed linear parameter estimation over wireless sensor networks. *Aerospace and Electronic Systems, IEEE Transactions on*, **45**(4):1293–1306, October 2009. 2.4.1

[21] Véronique Delouille, Ramesh Neelamani, and Richard Baraniuk. Robust distributed estimation in sensor networks using the embedded polygons algorithm. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, IPSN '04, pages 405–413, New York, NY, USA, 2004. ACM. 2.4.1

[22] A. Dogandzic and B. Zhang. Distributed estimation and detection for sensor networks using hidden Markov random field models. *Signal Processing, IEEE Transactions on*, **54**(8):3200–3215, August 2006. 2.3, 2.4.1

[23] Y.C. Eldar. Minimax MSE estimation of deterministic parameters with noise covariance uncertainties. *Signal Processing, IEEE Transactions on*, **54**(1):138–145, January 2006. 2.3

[24] Y.C. Eldar, A. Ben-Tal, and A. Nemirovski. Linear minimax regret estimation of deterministic parameters with bounded data uncertainties. *Signal Processing, IEEE Transactions on*, **52**(8):2177–2188, August 2004. 2.3

[25] Ozgur Ercetin, Ozgur Gurbuz, Kerem Bulbul, and Aylin Aksu. Joint sensor selection and data routing in sensor networks. In *Proceedings of the 4th IFIP-TC6 international conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communication Systems*, NETWORKING'05, pages 828–839, Berlin, Heidelberg, 2005. Springer-Verlag. 1, 2.4.1

[26] Rockarc J. F. and J.S. Leventer. Centralization versus decentralization of information systems: An annotated bibliography. *CISR Report 22*, April 1976. 2.3.1

[27] Jun Fang and Hongbin Li. Distributed adaptive quantization for wireless sensor networks: from delta modulation to maximum likelihood. *Signal Processing, IEEE Transactions on*, **56**(10):5246–5257, October 2008. 2.4.1, 2.4.2

[28] Jun Fang and Hongbin Li. Hyperplane-based vector quantization for distributed estimation in wireless sensor networks. *Information Theory, IEEE Transactions on*, **55**(12):5682–5699, December 2009. 1, 2.4.1, 3.3

[29] J.A. Fuemmeler and V.V. Veeravalli. Energy efficient multi-object tracking in sensor networks. *Signal Processing, IEEE Transactions on*, **58**(7):3742–3750, July 2010. 1

[30] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990. 2.4.4, 3.2.3, 3.2.3, A, A, D

[31] D.S. Ghataoura, J.E. Mitchell, and G.E. Matich. Networking and application interface technology for wireless sensor network surveillance and monitoring. *Communications Magazine, IEEE*, **49**(10):90–97, 2011. 2.1.2

[32] Hana Godrich, Alexander M. Haimovich, and Rick S. Blum. Target localization accuracy gain in MIMO radar-based

systems. *Information Theory, IEEE Transactions on*, **56**(6):2783–2803, June 2010. 1

[33] ANDREA GOLDSMITH. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005. 5.4

[34] V.K. GOYAL. High-rate transform coding: how high is high, and does it matter? In *Information Theory, 2000. Proceedings. IEEE International Symposium on*, pages 207–, 2000. 3.3

[35] ROBERT M. GRAY AND DAVID L. NEUHOFF. Quantization. *Information Theory, IEEE Transactions on*, **44**(6):2325–29, 1998. 2.4.5, 3.3

[36] TING HE, S. BEN-DAVID, AND L. TONG. Nonparametric change detection and estimation in large-scale sensor networks. *Signal Processing, IEEE Transactions on*, **54**(4):1204–1217, April 2006. 1

[37] SANDRA M. HEDETNIEMI, STEPHEN T. HEDETNIEMI, AND ARTHUR L. LIESTMAN. A survey of gossiping and broadcasting in communication networks. *Networks*, **18**(4):319–349, 1988. 4.2, 4.3.1, 1

[38] W.R. HEINZELMAN, A. CHANDRAKASAN, AND H. BALAKRISHNAN. Energy-efficient communication protocol for wireless microsensor networks. In *System Sciences, Proceedings of the 33rd Annual Hawaii International Conference on*, **2**, pages 1–10, January 2000. 1, 1

[39] JASON HILL, ROBERT SZEWCZYK, ALEC WOO, SETH HOLLAR, DAVID CULLER, AND KRISTOFER PISTER. System architecture directions for networked sensors. *SIGARCH Computer Architecture News*, **28**(5):93–104, November 2000. 2.1

[40] JASON LESTER HILL AND JASON LESTER HILL. System architecture for wireless sensor networks. Technical report, University of California, Berkeley, 2003. 1, 1.1

[41] TING-WEI HOU, HOUNG-KUO KU, AND YUAN-TSUNG CHEN. An improved vector quantizer design method: the codebook reorganization algorithm. In *JCIS'06*. Atlantis Press, 2006. 3.3

[42] Y. Thomas Hou, Yi Shi, and Hanif D. Sherali. Rate allocation in wireless sensor networks with network lifetime requirement. In *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '04, pages 67–77, New York, NY, USA, 2004. ACM. 1, 2.4.2

[43] Y. Huang and J. Zhang. Lower bounds on the variance of deterministic signal parameter estimators using bayesian inference. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Proceedings (ICASSP '03). 2003 **6**, pages VI/745–VI/748, April 2003. 2.3

[44] Yi Huang and Yingbo Hua. Multi-hop progressive decentralized estimation of deterministic vector in wireless sensor networks. In *Signals, Systems and Computers, Conference Record of the Forty-First Asilomar Conference on*, ACSSC 2007. pages 2145–2149, November 2007. 1, 1, 1.1, 2.4.1

[45] Yi Huang and Yingbo Hua. Multihop progressive decentralized estimation in wireless sensor networks. *Signal Processing Letters, IEEE*, **14**(12):1004–1007, December 2007. 1, 1.1, 2.4.1, 2.4.2, 3.1.1, 3.2.1

[46] Yi Huang and Yingbo Hua. Energy planning for progressive estimation in multihop sensor networks. *Signal Processing, IEEE Transactions on*, **57**(10):4052–4065, October 2009. 1, 2.4.1, 2.4.3, 3.1.1, 3.2.1

[47] Yi Huang and Yingbo Hua. On energy for progressive and consensus estimation in multihop sensor networks. *Signal Processing, IEEE Transactions on*, **59**(8):3863–3875, August 2011. 6, 1, 1.1, 6, 2.4.1, 2.4.3, 3.1.1, 3.3.1

[48] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *Signal Processing, IEEE Transactions on*, **57**(2):451–462, February 2009. 1, 2.4.2, 2.4.4

[49] D.B. Jourdan and O.L. de Weck. Layout optimization for a wireless sensor network using a multi-objective genetic algorithm. In *Vehicular Technology Conference, IEEE 59th*, VTC 2004-Spring. 2004 **5**, pages 2466–2470, 2004. 1

[50] S. Kar, J.M.F. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *Information Theory, IEEE Transactions on*, **58**(6):3575–3605, June 2012. 2.4.1, 5.1

[51] R. M. Karp. Reducibility Among Combinatorial Problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. A

[52] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. 2.4.2, 3.2.1, 3.2.1, 3.3.1, 4.5.1, 4.5.2, 4.5.3, 1

[53] Wei-Chieh Ke, Bing-Hong Liu, and Ming-Jer Tsai. Efficient algorithm for constructing minimum size wireless sensor networks to fully cover critical square grids. *Wireless Communications, IEEE Transactions on*, **10**(4):1154–1164, 2011. 2.1.2

[54] Y.H. Kim and A. Ortega. Quantizer design and distributed encoding algorithm for source localization in sensor networks. In *Information Processing in Sensor Networks, Fourth International Symposium on*, IPSN 2005. pages 231–238, April 2005. 1, 1, 2.3, 2.4.2, 3.3, 3.3

[55] Yoon Hak Kim and A. Ortega. Quantizer design for energy-based source localization in sensor networks. *Signal Processing, IEEE Transactions on*, **59**(11):5577–5588, November 2011. 1, 1, 2.3, 3.3, 3.3

[56] Alexey Krasnopeev, Jin jun Xiao, and Zhi quan Luo. Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises. *EURASIP Journal on Wireless Communications and Networking*, pages 473–482, 2005. 1.1, 2.4.1

[57] L. Krishnamachari, D. Estrin, and S. Wicker. The impact of data aggregation in wireless sensor networks. In *Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference on*, pages 575–578, 2002. 1, 2.4.4, A

[58] Dan Li, K.D. Wong, Yu Hen Hu, and A.M. Sayeed. Detection, classification, and tracking of targets. *Signal Processing Magazine, IEEE*, **19**(2):17–29, March 2002. 2.3, 3.1.2, 4.4, 5.4

[59] Hongbin Li and Jun Fang. Distributed adaptive quantization and estimation for wireless sensor networks. *Signal Processing Letters, IEEE*, **14**(10):669 –672, October 2007. 2.4.1

[60] Junlin Li and G. AlRegib. Rate-constrained distributed estimation in wireless sensor networks. In *Computer Communications and Networks, Proceedings.15th International Conference on*, ICCCN 2006. pages 317–322, October 2006. 1, 2.4.1, 2.4.2, 2.4.5

[61] Junlin Li and G. AlRegib. Maximizing network lifetime for estimation in multi-hop wireless sensor networks. In *Computer Communications and Networks, Proceedings of 17th International Conference on*, ICCCN '08. pages 1–6, August 2008. 6, (document), 1, 6, 2.4.1, 2, 2.4.3, 2.4.4, 2.4.5, 3.3, 4.4, 5.1, 5.6, 5.4

[62] Junlin Li and G. AlRegib. Distributed estimation in energy-constrained wireless sensor networks. *Signal Processing, IEEE Transactions on*, **57**(10):3746–3758, October 2009. 1, 1, 2.4.2

[63] Teng Li, A. Ekpenyong, and Yih-Fang Huang. Source localization and tracking using distributed asynchronous sensors. *Signal Processing, IEEE Transactions on*, **54**(10):3991–4003, October 2006. 1, 2.3, 4.5.2

[64] Chee-Yee Chong, Kadar I., Alford M.G., Vannicola V., Thomopoulos S., Liggins, M.E. Distributed fusion architectures and algorithms for target tracking. *Proceedings of the IEEE*, **85**(1):95–107, January 1997. 2.2

[65] S. Lindsey and C.S. Raghavendra. PEGASIS: Power-efficient gathering in sensor information systems. In *Aerospace Conference Proceedings, IEEE*, **3**, pages 3–1125 – 3–1130, 2002. 1, 2.3

[66] James Llinas and DavidL Hall. *Multisensor Data Fusion*, pages 1–14. CRC Press, 2013/11/26 2008. 2.2

[67] Jaime Lloret, Miguel Garcia, Diana Bri, and Sandra Sendra. A wireless sensor network deployment for rural and forest fire detection and verification. *Sensors*, **9**(11):8722–8747, 2009. 1

[68] Huiyu Luo, Yu-Ching Tong, and G. Pottie. A two-stage DPCM scheme for wireless sensor networks. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Proceedings (ICASSP '05). **3**, pages iii/661–iii/664, March 2005. 1

[69] Zhi-Quan Luo. An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network. *Selected Areas in Communications, IEEE Journal on*, **23**(4):735–744, April 2005. 2.3, 2.4.1, 2.4.2

[70] Zhi-Quan Luo. Universal decentralized estimation in a bandwidth constrained sensor network. *Information Theory, IEEE Transactions on*, **51**(6):2210–2219, June 2005. 1, 1.1, 2.3, 2.4.1, 2.4.2

[71] Guoqiang Mao, Brian D. O. Anderson, and Bariş Fidan. Path loss exponent estimation for wireless sensor network localization. *Computer Networks*, **51**(10):2467–2483, July 2007. 1, 3.1.2

[72] Marta Martinez-Camara, Ivan Dokmanic, Juri Ranieri, Robin Scheibler, Martin Vetterli, and Andreas Stohl. The fukushima inverse problem. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, ICASSP 2013. pages 4330–4334, 2013. 1

[73] E. Masazade, Ruixin Niu, P.K. Varshney, and M. Keskinoz. Energy aware iterative source localization for wireless sensor networks. *Signal Processing, IEEE Transactions on*, **58**(9):4824–4835, September 2010. 1, 1

[74] M.P. Michaelides and C.G. Panayiotou. Plume source position estimation using sensor networks. In *Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation*, pages 731–736, June 2005. 5.1

[75] U. Monaco, F. Cuomo, T. Melodia, F. Ricciato, and M. Borghini. Understanding optimal data gathering in the energy and latency domains of a wireless sensor network. *Computer Networks*, **50**(18):3564–3584, December 2006. 1

[76] E.J. Msechu and G.B. Giannakis. Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *Signal Processing, IEEE Transactions on*, **60**(1):400–414, January 2012. 1, 2.4.5

[77] Frank Mufalli, Rajan Batta, and Rakesh Nagi. Simultaneous sensor selection and routing of unmanned aerial vehicles

for complex mission plans. *Computers & Operations Research*, **39**(11):2787–2799, November 2012. 1, 1.1, 2.4.2

[78] An Nguyen, N. Milosavljevic, Qing Fang, Jie Gao, and L.J. Guibas. Landmark selection and greedy landmark-descent routing for sensor networks. In *Computer Communications, 26th IEEE International Conference on*, INFOCOM 2007. pages 661–669, May 2007. 1, 2.4.1

[79] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1982. 3.2.2, A

[80] M. Perillo and W. Heinzelman. DAPR: a protocol for wireless sensor networks utilizing an application-based routing cost. In *Wireless Communications and Networking Conference, IEEE*, WCNC 2004. **3**, pages 1540–1545, March 2004. 1

[81] K.H. Plarre and P.R. Kumar. Extended message passing algorithm for inference in loopy gaussian graphical models. *Ad Hoc Networks*, **2**(2):153–169, 2004. 4.2, 4.3.1, 1

[82] E. Portu-Repollés and B. Beferull-Lozano. Joint sensor selection and routing for distributed estimation in wireless sensor networks. *EUSIPCO, Glasgow, Scotland*, August 2009. 1, 2.4.1, 2.4.3, 3.2.2

[83] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, **43**(5):51–58, May 2000. 1

[84] Zhi Quan, W.J. Kaiser, and A.H. Sayed. Innovations diffusion: A spatial sampling scheme for distributed estimation and detection. *Signal Processing, IEEE Transactions on*, **57**(2):738–751, February 2009. 6, 1, 6, 2.4.1, 2.4.3, 3.1.1, 4.3, 4.4, 5.3

[85] Zhi Quan and A.H. Sayed. Innovations-based sampling over spatially-correlated sensors. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, ICASSP 2007. **3**, pages III/509–III/512, April 2007. 1, 2.3, 2.4.2

[86] Vaidyanathan Ramadurai and Mihail L. Sichitiu. Localization in wireless sensor networks: A probabilistic approach, 2003. 1

[87] A. Ribeiro and G. B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks, part II. *Signal Processing, IEEE Transactions o*, **54**:2784–2796, 2006. 1, 1.1, 2.4.1, 2.4.2

[88] A. Ribeiro and G.B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case. *Signal Processing, IEEE Transactions on*, **54**(3):1131–1143, March 2006. 1, 1.1, 2.4.1, 2.4.2, 2.4.3

[89] G. Roe. Quantizing for minimum distortion (corresp.). *Information Theory, IEEE Transactions on*, **10**(4):384–385, October 1964. 3.3

[90] I. Rubin and Xiaolong Huang. Capacity aware optimal activation of sensor nodes under reproduction distortion measures. In *Military Communications Conference, IEEE*, MILCOM 2006. pages 1–8, October 2006. 1, 2.4.1

[91] N.F. Sandell and R. Olfati-Saber. Distributed data association for multi-target tracking in sensor networks. In *Decision and Control, 47th IEEE Conference on*, CDC 2008. pages 1085 –1090, dec. 2008. 4.5.2

[92] Rahul P. Sawant, Qilian Liang, Dan O. Popa, and Frank L. Lewis. Experimental path loss models for wireless sensor networks. In *Military Communications Conference, IEEE*, MILCOM 2007. pages 1–7, October 2007. 3.1.2

[93] I.D. Schizas, A. Ribeiro, and G.B. Biannakis. Consensus-based distributed parameter estimation in ad hoc wireless sensor networks with noisy links. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, ICASSP 2007. **2**, pages II/849–II/852, April 2007. 1, 2.4.1, 5.1

[94] L. Schuchman. Dither signals and their effect on quantization noise. *Communication Technology, IEEE Transactions on*, **12**(4):162–165, 1964. (document), 2.4.5, 3.5, 3.3

[95] R. Sedgewick. *Algorithms in Java*. Algorithms in Java. Addison-Wesley, 2004. A

[96] S. Shah and B. Beferull-Lozano. In-network iterative distributed estimation for power-constrained wireless sensor networks. In *Distributed Computing in Sensor Systems (DCOSS), 2012 IEEE 8th International Conference on*, pages 239–246, 2012. 1.3, 3.2.2, E

[97] S. Shah and B. Beferull-Lozano. In-network local distributed estimation for power-constrained wireless sensor networks. In *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pages 1–5, 2012. 1.3, 3.2.2

[98] S. Shah and B. Beferull-Lozano. Power-aware joint sensor selection and routing for distributed estimation: A convex optimization approach. In *Distributed Computing in Sensor Systems (DCOSS), 2012 IEEE 8th International Conference on*, pages 230–238, 2012. (document), 1.3, 5.2, 5.1, E

[99] S. Shah and B. Beferull-Lozano. Energy-efficient multihop progressive estimation and adaptive quantization for ad-hoc wireless sensor networks. *Signal Processing, IEEE Transactions on*, (Submitted), 2014. 1.3

[100] S. Shah and B. Beferull-Lozano. Joint sensor selection and multihop routing for distributed estimation in ad-hoc wireless sensor networks. *Signal Processing, IEEE Transactions on,*, **61**(24):6355–6370, December 2013. 1.3

[101] S Shah and Baltasar Beferull-Lozano. Adaptive quantization for multihop progressive estimation in wireless sensor networks. In *21st European Signal Processing Conference 2013 (EUSIPCO 2013)*, Marrakech, Morocco, September 2013. 1, 1.3, 2.4.1, A

[102] S.F.A. Shah, A. Ribeiro, and G.B. Giannakis. Bandwidth-constrained MAP estimation for wireless sensor networks. In *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, pages 215–219, November 2005. 1, 2.4.1, 2.4.2

[103] Xiaohong Sheng and Yu-Hen Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *Signal Processing, IEEE Transactions on*, **53**(1):44–53, January 2005. 1

[104] D. Sherwood. Some theorems on quantization and an example using dither. In *Circuits, Systems and Computers, 1985. Nineteeth Asilomar Conference on*, pages 207–212, 1985. 3.3, 3.3, 3.3

[105] S. Srinivasa and M. Haenggi. Path loss exponent estimation in large wireless networks. In *Information Theory and Applications Workshop, 2009*, pages 124–129, February 2009. 1, 3.1.2, 4.4, 5.4

[106] A.B. Sripad and D. Snyder. A necessary and sufficient condition for quantization errors to be uniform and white. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, **25**(5):442–448, 1977. 3.3, 3.3

[107] A. Sundaresan and P.K. Varshney. Location estimation of a random signal source based on correlated sensor observations. *Signal Processing, IEEE Transactions on*, **59**(2):787–799, February 2011. 1, 2.3

[108] Salil Kanhere, Branko Ristic, Travis Bessell, Mark Rutten, Tatiana Bokareva, Wen Hu and Sanjay Jha. Wireless sensor networks for battlefield surveillance. *In Proceedings of the Land Warfare Conference*, 2006. 2.1.2

[109] G. Thatte and U. Mitra. Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology. *Signal Processing, IEEE Transactions on*, **56**(7):2649–2661, July 2008. 1, 1.1, 2.4.1, 2.4.2

[110] P. Venkitasubramaniam, Lang Tong, and A. Swami. Quantization for maximin ARE in distributed estimation. *Signal Processing, IEEE Transactions on*, **55**(7):3596–3605, July 2007. 2.4.1

[111] C.T. Vu, R.A. Beyah, and Yingshu Li. Composite event detection in wireless sensor networks. In *Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE Internationa*, pages 264–271, April 2007. 1

[112] J.M. Walsh and P.A. Regalia. Expectation propagation for distributed estimation in sensor networks. In *Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007. IEEE 8th Workshop on*, pages 1–5, June 2007. 2.4.1

[113] Wei Wang, Martin J. Wainwright, and Kannan Ramchandran. Information-theoretic limits on sparse signal recovery: dense versus sparse measurement matrices. *Information Theory, IEEE Transactions on*, **56**(6):2967–2979, June 2010. 1

[114] Ye Wang, P. Ishwar, and V. Saligrama. One-bit distributed sensing and coding for field estimation in sensor networks. *Signal Processing, IEEE Transactions on*, **56**(9):4433–4445, September 2008. 2.4.2

[115] R.A. Wannamaker, S. Lipshitz, J. Vanderkooy, and J.N. Wright. A theory of nonsubtractive dither. *Signal Processing, IEEE Transactions on*, **48**(2):499–516, 2000. 3.3, 3.3

[116] Bernd Wernsperger and Clemens Schlosser. Noble gas monitoring within the international monitoring system of the comprehensive nuclear test-ban treaty. *Radiation Physics and Chemistry*, **71**(3-4):775 – 779, 2004. 1

[117] Kamin Whitehouse and David Culler. Calibration as parameter estimation in sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, WSNA '02, pages 59–67, New York, NY, USA, 2002. ACM. 1

[118] Kamin Whitehouse, Chris Karlof, Alec Woo, Fred Jiang, and David Culler. The effects of ranging noise on multihop localization: an empirical study. In *Proceedings of the 4th international symposium on Information processing in sensor networks*, IPSN '05, Piscataway, NJ, USA, 2005. IEEE Press. 1

[119] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese. Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *Automatic Control, IEEE Transactions on*, **27**(4):799–813, August 1982. 2.4.1

[120] J.-J. Xiao and Z.-Q. Luo. Decentralized estimation in an inhomogeneous sensing environment. *Information Theory, IEEE Transactions on*, **51**(10):3564–3575, October 2005. 1, 1.1, 2.4.1, 2.4.2, 2.4.3, 2.4.5, 3.2.1

[121] Jin-Jun Xiao, Shuguang Cui, Zhi-Quan Luo, and A.J. Goldsmith. Joint estimation in sensor networks under energy

constraints. In *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, pages 264–271, October 2004. 2.4.1

[122] Jin-Jun Xiao, Shuguang Cui, Zhi-Quan Luo, and A.J. Goldsmith. Power scheduling of universal decentralized estimation in sensor networks. *Signal Processing, IEEE Transactions on*, **54**(2):413–422, February 2006. 6, (document), 6, 2.4.2, 2.4.3, 2.4.4, 2.4.5, 5.6, 5.4

[123] Jin-Jun Xiao, Shuguang Cui, Zhi-Quan Luo, and A.J. Goldsmith. Linear coherent decentralized estimation. *Signal Processing, IEEE Transactions on*, **56**(2):757–770, 2008. 2

[124] Ou Yang and Wendi B. Heinzelman. Sensor selection cost function to increase network lifetime with QoS support. In *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '08, pages 370–374, New York, NY, USA, 2008. ACM. 1, 2.4.2

[125] Yourim Yoon and Yong-Hyuk Kim. An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks. *Cybernetics, IEEE Transactions on*, **43**(5):1473–1483, 2013. 2.1.2

[126] Chao Yu and G. Sharma. Distributed estimation and coding: A sequential framework based on a side-informed decomposition. *Signal Processing, IEEE Transactions on*, **59**(2):759–773, February 2011. 1, 2.4.1, 2.4.2

[127] K. Zeger, A. Bist, and T. Linder. Universal source coding with codebook transmission. *Communications, IEEE Transactions on*, **42**(234):336–346, feb/mar/apr 1994. 3.3

[128] K. Zhang and X.R. Li. Optimal sensor data quantization for best linear unbiased estimation fusion. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, **3**, pages 2656–2661, December 2004. 2.4.1, 3.2.1

[129] F. Zhao and L. Guibas. *Wireless Sensor Networks: An Information Processing Approach*. The Morgan Kaufmann Series in Networking. Elsevier Science, 2004. 1, 1

[130] Feng Zhao, Jaewon Shin, and J. Reich. Information-driven dynamic sensor collaboration. *Signal Processing Magazine, IEEE*, **19**(2):61–72, March 2002. 2.3