Universidade Vigo

ESCOLA DE
ENXEÑARÍA DE TELECOMUNICACIÓN

Ph.D. programme in Signal Theory and Communications

Ph.D. Thesis

Submitted for the International Doctor Mention

FLAT FADING CHANNEL ESTIMATION BASED ON DIRTY PAPER CODING

Author: Gabriel Domínguez-Conde
Advisors: Pedro Comesaña-Alfaro
Fernando Pérez-González

2015

This work was partially funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) under projects TACTICA, COMPASS (TEC2013-47020-C2-1-R) and COMONSENS (TEC2015-69648-REDC), by the Galician Regional Government and ERDF under projects "Consolidation of Research Units" (GRC2013/009), REdTEIC (R2014/037) and AtlantTIC, and by the EU 7th Framework Programme under project NIFTy (HOME/2012/ISEC/AG/INT/4000003892).

Abstract

Channel estimation is a transversal problem in signal processing (for example, it is used in digital communications, image restoration, digital forensics, acoustics, etc.). Among channel estimation algorithms, pilot-based estimation techniques stand out as being among the most frequently used. These techniques devote part of the total available power, which is usually limited, to send pilot signals that are used later to estimate the channel. The frequent need to send pilot signals in order to be able to track the channel variations, which lowers the information rate, becomes as one of their major drawbacks.

Recently, the idea of concurrently sending a known training sequence with the information-bearing signal (also known as host) by means of arithmetically adding both sequences was proposed. These techniques are usually referred as superimposed training techniques. By implementing this idea, there is no drop in the information rate; however, part of the power available to send the information must be used by the added superimposed sequence thus causing a power loss in the information-bearing sequence. In addition, the original signal interferes with the pilot sequence of the superimposed training techniques, causing a decrease in the estimate performance, which is measured in terms of mean square error between the estimation and the actual channel gain. To tackle this issue, some solutions have been provided that use part of the power to partial cancel the host-interference. In this thesis, we have found a connection between superimposed training and digital watermarking. Indeed, this partially cancellation of the host of pilot sequences, known as Partially-Data-Dependent superimposed training (PDD) was independently proposed in digital watermarking, where is called Improved Spread Spectrum (ISS). We propose to obtain full cancellation of host-interference for estimation by applying the Dirty Paper Coding (DPC) paradigm that successfully was used in digital watermarking with several implementations (e.g., Scalar Costa Scheme, Distortion Compensated Dither-Modulation, etc.).

Specifically in this thesis, first we focus on the study of the flat fading channel estimation based on dirty paper coding for the case of real valued signals. Due to its interesting asymptotic properties, we design our estimation technique using Maximum-Likelihood Estimation (MLE). In order to do that, the

probability density function (pdf) of the random variables modeling the involved signals is required; in general, those pdfs are hard to handle mathematically and, as a consequence, so is the MLE cost function. Therefore, we have proposed a set of approximations of the pdf whose accuracy is validated in the cases for which they have been designed. In addition, a modification of the technique whenever the variances of the original signal and the channel noise are unknown is presented. In addition, this thesis proposes how to make full use of the Spread-Transform (ST) (an established concept of digital watermarking) to estimate the channel gain.

In addition, a theoretical study is introduced following an estimation theory perspective, which indicates that asymptotically our scheme is not only not harmed by the host but it helps for estimation, and an information theory perspective, whose results determine that the induced structure of the transmitted signal helps the estimation of the gain of the channel. Both analyses show an improvement on the estimation performance of our technique with respect to Spread-Spectrum (SS) and PDD.

The computational and time requirements needed to implement MLE, even using our pdf approximations, are not affordable in many applications. To tackle this, we introduce a set of MLE-based practical algorithms for estimation, designed with computational and temporal constraints. These algorithms take advantage of the statistical and deterministic properties of the problem. Several performance tests, measuring the accuracy of our algorithm, indicate that it outperforms other existing techniques whenever the structure of the sent signal becomes patent, and requires much shorter computational time than other existing DPC-based estimation techniques.

With the aim of gaining insight into the wide range of practical uses of our algorithms, this thesis presents a set of applications of the proposed technique. For example, we use our algorithms to make dirty paper coding watermarking robust to gain attacks. By using both synthetic signals and real images, the obtained results validate the efficacy of our techniques in dealing with such attacks. We also show, in a flat fading channel communications scenario, how to equalize the gain estimated with our algorithms. The results show that our techniques improve the performance with respect to equalizing techniques based either on the second moment estimation or on superimposed training. Finally, we also propose how to adapt our estimation algorithm to the case of complex signals and complex gains, whose performance indicates that the host also helps in the estimation.

A Laura

Agradecimentos

Chegados ao final da tese de doutoramento quero lembrar-me de toda a gente que me acompanhou neste processo que abarcou anos importantes da minha vida e que me deixarom, sem dúvida, pegada.

Primeiro, quero lembrar-me dos meus diretores de tese. Agradecer-lhe a Fernando e Pedro por me terem oferecido fazer a tese doutoramento e por terem-me dado o seu apoio durante a mesma. A tese resultou ser uma viagem altamente interessante, formativa e moi divertida. Estou altamente agradecido a Fernando por todas as oportunidades que me deu para o meu desenvolvimento profissional. Aprendim moito de Pedro, do que admiro, à parte do seu extraordinário nível técnico, a sua capacidade de trabalho e sacrifício.

Não queria tampouco esquecer-me dos outros professores do GPSC: os professores Carlos Mosquera, Roberto López e Nuria González. Sempre dispostos a ajudar e profundar na integração das duas almas do grupo. Outra das pessoas do grupo de quem tivem sempre a ajuda que pedim foi Carmen Touriño: obrigadíssimo.

Por outra parte, também queria agradecer-lhe à professora Deepa Kundur e ao seu grupo a acolhida que me derom em Texas e as interessantes conversas que mantivemos. Tampouco me queria esquecer dos meus companheiros que tão bem me integraram na comunidade chinesa e indonésia local, moi especialmente a Ivan Alexander e a Suhendra.

Foi um prazer trabalhar com David (alias o Cambadês, Clim, Chute, etc.). Nos meus anos de liceu, nunca conceberia fazer-me amigo dum Cambadês e especialmente se era de Castrelo. Sempre está surpreendendo, que se é nativo em francês, que um crack no origami,... Que dizer de Juan (alias JR, Troncosoft, etc.)! À parte de ser "una bestia pardísima" tecnicamente, saco-me o sombreiro por ser tão bom companheiro, sempre disposto a ajudar.

Dos membros do TSC5-Shore, não poderia esquecer-me de Miguel Masciopinto (alias M^2, M. Masciomaragota ou Mike Femto-polbos). Apontava já maneiras quando concordávamos na admiração a Caniggia e Maradona. Ainda que recentemente deixou o TSC5, queria lembrar-me de Iria (outra das integrantes do

comando PRNU), foi um prazer trabalhar com ela. Como firme crente no Pedrouzismo, não me podo esquecer do seu líder Alberto Pedrouzo e a sua perfeita mistura de silêncio/gargalhadas-no-momento-clave. De Simon [saimon], gostei das suas lições sobre fauna urbana assim como adoro que seja um sistema não ergódico: sempre fluindo.

Passei-no moi bem com Paulinha (uma reintegrata no armário): lembro-me o dia que fomos torcer por Portugal à Valença quando jogava contra Espanha. Desejar-lhe os maiores sucessos na sua etapa investigadora. Gonzalo (alias Vigas), apresentou-nos Clim num Mercadona (que romântico!!!), é moi bom tipo. Sempre admirarei a sua capacidade para os detalhes, incrível, um HPF em toda regra. Magui e as discussões sobre as diferenças entre pêssegos e melocotões... Lembrar-me também de Marta que às vezes não sei se toma a sério as parvadas que digo. E como não a Rocío: não se pode ser mais boinha. Gostei de trabalhar com Montse, falar cuma pessoa coa perspectiva de humanidades sempre é moi interessante. Graças a Leo e Esteban por abrir-me as portas da sua amizade e dar-me essas master classes de fotografia. Foi um prazer trabalhar com Mela, obrigado por ter-me descoberto inventos tão bons como o pinpong de escritório. Também estou agradecido a Sandra pola ajuda prestada nesta fase final da tese e desejo-lhe o melhor dos futuros.

Ainda lembro os inícios de Gordon Brown e Barack Obama num quartucho emprestado da escola... Que siga assim Luis (alias LPF) que lhe auguro o maior dos sucessos. A Dani dizer-lhe que temos que seguir vendo-nos para pescar anualmente, e que nunca tive um tripulante coma ele: temos que seguir! Foi um prazer compartilhar tempo Fran Campillo, desejo-lhe o melhor e obrigadíssimo por ter-me ajudado tanto. Não quero esquecer-me doutros clássicos TSC5 como Abu, Eli e outros tantos companheiros com os que passei tão bons momentos: obrigado a todos!

Gosto de ter amigos como Miguel Vilarinho (alias o Leonard Cohen de Teleco), desfruto do seu companherismo, da sua amizade, do seu pensamento crítico e das sugerências musicais. Como desfrutei com Maceda (alias Longo) das nossas conversas sobre música, literatura, filmes, séries e a beleza plástica do melhor Barça. A Álex (alias o Bulldozer, o enterrador de chefes, etc.) que siga assim, que não pare que vai lograr o seu objetivo de retirar-se aos 40. Obrigado a Antom pola sua amizade e a sua relativa recente conversão. Molaria manter esse costume de ver-nos e falar com calma no Culturgal cada ano. Que bom é visitar a Rober em Bueu e poder falar com ele sobre as últimas jogadas da política galega desfrutando do seu Spa com vistas à Ria. Juli foi das primeiras pessoas que conhecim que amava o cinema, lembrarei-me toda a vida dum dia que me fum para casa sobre as seis da manhã convencido por ele em ver Viridiana. Passei-no moi bem com Rayco, dobrando com os nossos possíveis projetos juntos de modulações, ou escrevendo AQUEL monólogo para o Clube da Comédia. Sempre é uma ledícia ver a Emiliano (alias o social-democrata) que sempre recebe a

um com o seu amplo sorriso e um abraço bem forte. Som moi afortunado de ter um amigo como Manu (alias XXXXX) que sempre o dá todo, incluindo a festinha surpresa que me dera em Madrid (Espanha) um dia que passei ali o meu aniversário. Não me podo esquecer tampouco de outros companheiros, som feliz de que sejam os meus amigos: Lois o brasileiro, o Ghaiteiro Nino, o dorneiro Arturo, de Tania, de Patri, de Paula, de Deborah (a outra meca de Teleco), de Kiko (e os seus vídeos), de MB, de Jose Baiona, de Jose Cedeira, de Aitor, de Pitu, de Montoto, e de outros moitos que seguro que me esqueço.

Deixando Vigo e passando a Ogrobe, quero agradecer-lhe a Marcos (alias Chabo, Chapu) que seja meu amigo. Desde que nos conhecemos, penso que desde os sete anos, sempre o vim como um reflexo de mim. Diria que moitas das cousas em que acredito não sei se som cousa minha ou por falar com ele. Espero que conservemos a nossa amizade para sempre. A Pelu, que siga assim, que entre pela porta da sua casa e nunca falte uma bebida. Gosto de Solla e Lote que poden passar anos (ou quase) sem ver-nos e quase sem falar, e que coincidir seja como se ainda nos viramos o dia anterior. Sempre flipo quando falo com Cabecho (alias Cobech, Miguelín, etc.) de música. Nunca sabes por onde vai tirar a conversa, pode acabar falando de Django Reinhardt ou de Peter Gabriel. Queria lembrar-me de outros moitos amigos mecos como: Iria (ainda que é picheleira), Marta Mariño, Melo, Webo, Nouro, Bibis, Nati, Marta, Figue, e outros moitos que seguro estou esquecendo.

Não me podo esquecer da minha família, começando pela minha nai Sefa, que sempre nos deu todo ao meus irmãos e a mim. Admiro a sua capacidade de loita e sacrifício. Tampouco de meu pai que tantas cousas me adiantou sobre a vida com frases em latim, como bom seminarista que foi, e que efetivamente se forom cumprindo. Da minha irmã Marta dizer que estou moi agradecido por ser tão boa irmã maior e ao seu marido também por dar-me dous sobrinhos tão cracks. Ao meu irmão pequeno Manu (alias Ziyas) que siga assim, vai chegar moi longe, é um valente. Queria lembrar-me dos finados dos meus avós Manolos, e das suas mulheres Mucha e Carme que me deixarom definitivamente este ano. A primeira foi-se indo devagar desde há mais de quinze e Carme sem avisar. Estarão sempre na minha memória. A minha tia Mari Carmen sempre está aí (como Alfonso mas este mais silencioso) e o seu filho Martín (alias Mac Barracuda Páramo) é das pessoas mais creativas que conheço. Queria reconhecer também o apoio incondicional dos meus padrinhos Pepe e Teresa, estou em dívida com eles.

Tenho muita sorte de ter uma namorada como Laura: fai-me moi feliz. O seu apoio em moitos momentos foi fundamental, sem ela todo teria sido moito mais difícil (por não dizer impossível). Como dixo Uxío Novoneyra "Non, a forza do noso amor non pode ser inútil!".

Ogrobe, Outubro de 2015.

List of my Publications

[1] Pedro Comesaña, Gabriel Domínguez-Conde, and Fernando Pérez-González. Fast DPC-based flat fading channel estimation. *IEEE Transactions on Signal Processing*, To be Submitted.

[2] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. Flat fading channel estimation based on dirty paper coding. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, 2014.

[3] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. A new look at ML step-size estimation for Scalar Costa Scheme data hiding. In *IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.

[4] Gabriel Domínguez-Conde and Deepa Kundur. Improving the visual performance of S/DISCUS. In *IEEE International Conference on Multimedia and Expo (ICME)*, Barcelona, Spain, July 2011.

[5] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. Performance analysis of Fridrich-Goljan self-embedding authentication method. In *IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, November 2009.

[6] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. Performance analysis of Fridrich-Goljan self-embedding authentication method. *IEEE Transactions on Information Forensics and Security*, 4(3):570–577, September 2009.

[7] Patent Title: Method and system for embedding information and authenticating a H.264 video using digital watermark.
Priority Date: 2 -September-2013.
Inventors: Luis Pérez-Freire, Gabriel Domínguez-Conde, David Vázquez-Padín, and Lukasz Z. Dzianach.
Holder Entity: Centum Research & Technology S.L.U.

Contents

Abbreviations and Acronyms

Add-SS Additive Spread Spectrum. xi, 12–15, 39, 53, 55, 57, 59, 61, 62, 64–66, 69, 74, 77, 79, 80, 109, 166, 169

AGC Automatic Gain Control. xi, 4, 161

AWGN Additive White Gaussian Noise. xi, 3, 12, 13, 141–144, 149, 159–161

BER Bit Error Rate. xi, 136–144

CLT Central Limit Theorem. xi, 26, 31, 36, 89, 93, 126, 134, 136, 140, 147

CRB Cramér-Rao Bound. xi, 9, 17, 18, 39–41, 43, 46–51, 83, 93, 97–106, 108, 150, 156, 168, 169, 171

DC-DM Distortion Compensated Dither-Modulation. i, xi, 18

DC-QIM Distortion Compensated Quantization Index Modulation. xi, 3, 13–15, 118, 165, 167

DCT Discrete Cosine Transform. xi, 135, 136, 138, 139

DNR Document-to-Noise Ratio. xi, 15, 46, 48, 49

DPC Dirty Paper Coding. i, ii, xi, 3, 8, 9, 59, 62, 71, 97, 131, 153, 155, 159–161, 165–168, 171, 172

DPCE Dirty Paper Coding Estimation. xi, 9, 15, 39, 41, 47, 50, 53, 58–60, 62, 64, 67–69, 75, 76, 79–81, 99, 100, 102, 103, 107–109, 112, 114, 116, 119, 131, 132, 140–144, 155, 160, 165

DPCEH Dirty Paper Coding Estimation for High-SNR Case. xi, 97–105, 107, 108, 110, 112–117, 141, 142

DPCEL Dirty Paper Coding Estimation for Low-SNR Case. xi, 97–106, 112–118

DPCEU Dirty Paper Coding Estimation for Unknown Variances. xi, 107–109, 111

DWR Document-to-Watermark Ratio. xi, 15, 22–25, 28, 29, 39, 47–51, 53, 54, 56–62, 65–72, 74–76, 79–81, 83, 97–118, 120, 125, 128, 137, 138, 143, 144, 151–156, 160

GPS Global Positioning System. xi, 6

HQR Host-to-Quantizer Ratio. xi, 15, 16, 19–28, 33–35, 37, 41–43, 45–48, 62–64, 100, 106, 109, 124, 125, 130, 133, 146, 149

HWNR Host-plus-Watermark-to-Noise Ratio. xi, 141–143

IDCT Inverse DCT. xi, 135

ISI Intersymbol Interference. xi, 4

ISS Improved Spread Spectrum. i, xi, 3, 12–15, 165, 167

KLD Kullback-Leibler Divergence. xi, 23–25, 168

ML Maximum-Likelihood. xi, 3, 8, 9, 17, 26–31, 39, 87, 89, 90, 93, 95, 109, 114, 131, 133, 135, 140, 146–148, 150, 159, 166–169

MLE Maximum-Likelihood Estimation. i, ii, xi, 8, 17, 26, 27, 29, 149

MMSE Minimum Mean Square Error. xi, 95

MPDD Modification of PDD. xi, 153–155

MSE Mean Square Error. xi, 4, 11, 97–109, 111, 114, 139, 153–157, 169

OFDM Orthogonal Frequency-Division Multiplexing. xi, 3, 165

PAM Pulse-Amplitude Modulation. xi, 9, 140, 141, 143

PDD Partially-Data-Dependent superimposed training. i, ii, xi, 3, 8, 13, 15, 30, 97–102, 104, 105, 109, 113, 114, 141–144, 153–156, 165, 166

pdf probability density function. ii, xi, 9, 14, 18–24, 26–29, 39–41, 44, 46–50, 52–54, 58, 63, 69–75, 77, 80, 83, 89, 93, 94, 100–102, 106–109, 114, 118, 119, 126, 129, 133, 137, 146–150, 155, 159, 160, 168, 169

PSNR Peak Signal to Noise Ratio. xi, 7, 139

QIM Quantization Index Modulation. xi, 13

SCR Self-Noise-to-Channel-Noise Ratio. xi, 16, 20–26, 28, 33, 36, 42, 43, 45–48, 63, 64, 100, 106, 109, 121, 133, 146, 149

SCS Scalar Costa Scheme. i, xi, 3, 4, 9, 131, 132, 135, 166

SIT Superimposed Training. xi, 2, 3, 8, 39, 53, 55, 57, 59, 61, 62, 64–66, 69, 74, 77, 79, 80, 140, 169

SNR Signal-to-Noise Ratio. xi, 4, 16, 161

SS Spread-Spectrum. ii, xi, 8, 131

SSIM Structural Similarity. xi, 7

ST Spread-Transform. ii, xi, 119, 120, 140, 141, 159

ST-DM Spread-Transform Dither Modulation. xi, 30, 31, 107, 143, 144

TNHR Total-Noise-to-Host Ratio. xi, 16, 21–27, 33–35, 43, 46–48, 62, 63, 94, 106, 124, 125, 133

TNQR Total-Noise-to-Quantizer Ratio. xi, 16, 20, 22–26, 28, 30, 33, 37, 42–48, 63, 64, 77, 106, 108–110, 121, 133, 140, 143, 144, 146, 149

UAV Unmanned Aerial Vehicle. xi, 6

UMTS Universal Mobile Telecommunications System. xi, 2, 164

WNR Watermark-to-Noise Ratio. xi, 15, 22–25, 28, 29, 39, 47–49, 51, 62, 65, 66, 69, 79, 80, 97–118, 120, 125, 128, 136, 151–156, 160

Notation

X	A random variable
x	Realization of X
$\boldsymbol{X}$	A random vector
$\boldsymbol{x}$	Realization of $\boldsymbol{X}$
$\boldsymbol{x}^T$	Transpose of vector $\boldsymbol{x}$
x_i	ith element of vector $\boldsymbol{x}$
σ_X^2	Variance of X
$\mathrm{Var}\{X\}$	Variance of X
$\mathrm{E}\{X\}$	Expected value of X
$f_X(x)$	Probability density function of X
$p_X(x)$	Probability mass function of X
$\mathcal{Q}(\cdot)$	Quantizer
$\mathcal{Q}_\Delta(\cdot)$	Uniform scalar quantizer with step-size Δ
X^{ST}	X in the ST domain
$\lvert\mathcal{X}\rvert$	Cardinality of set $\mathcal{X}$
α	Distortion compensation parameter
α_{Costa}	Costa's α
$A \bmod B$	Modulo operation as $A - \mathcal{Q}_B(A)$
$\mathbb{Z}$	Set of integers
$\mathbb{N}$	Set of naturals

$\mathbb{C}$ . . . . . Set of complex numbers

$\mathbb{R}$ . . . . . Set of reals

$\|\boldsymbol{x}\|$ . . . . . Euclidean norm of $\boldsymbol{x}$

$|x|$ . . . . . Absolute value of x

$\measuredangle x$ . . . . . Phase of x

$\mathrm{Re}(x)$ . . . . . Real part of x

$\mathrm{Im}(x)$ . . . . . Imaginary part of x

$h(X)$ . . . . . Differential entropy of continuous random variable X

$I(X;Y)$ . . . . . Mutual information between X and Y

$< \boldsymbol{x}, \boldsymbol{y} >$ . . . . . Scalar product of $\boldsymbol{x}$ and $\boldsymbol{y}$

$\mathcal{N}(\mu, \sigma^2)$ . . . . . Gaussian distribution with mean μ and variance σ^2

$I_{L\times L}$ . . . . . Identity matrix of size $L \times L$

$\hat{x}$ . . . . . Estimate of x

Chapter 1

Introduction

Channel estimation is a transversal problem in signal processing. It is used in a number of applications, including digital communications (e.g., estimation of channel parameters, automatic gain control, signal-to-noise ratio estimation, etc.), image restoration (e.g., image deconvolution), digital forensics (e.g., estimation of the linear filter used for post-processing an image), and acoustics (e.g., estimation of the acoustic response of a room, echo cancellation, etc.).

One of the most prominent approaches of channel estimation is blind estimation. These techniques exploit certain underlying properties of the original signal (i.e., a transmitted signal in digital communications, an image in image restoration, etc.) and the channel to estimate the channel from the received signal. Those characteristics can be statistical, such as Higher Order Statistics [14], or deterministic, as in Constant Modulus Algorithms [56] or as in Deterministic Maximum Likelihood [32]. One of the main advantages of blind estimation is that these techniques do not need to modify the original signal to estimate the channel; however, blind estimation approaches suffer from slow convergence (i.e., a high number of samples of the received signal are required), and possible misconvergence [57]. Since it is not necessary to modify the original signal, this approach is usually selected in applications with this constraint (e.g., for petroleum exploration).

Arguably, pilot-based estimation is the other most widely-used channel estimation approach. In contrast to blind estimation, these techniques modify the original signal. Specifically, pilot-based systems use part of the total power budget to transmit a signal, referred to as pilot or training signal, that is known to the receiver, so it can be used to infer the channel response. In most cases, the pilot signal is transmitted in an orthogonal subspace to that of the information-bearing signal, most often through either time-domain or frequency-domain multiplexing.

Pilot-based approaches have a number of well-known drawbacks [59, 29, 30]: **1)** in fast-varying channels, the training signals must be sent frequently in order

to update the channel state information, thus wasting a significant amount of resources (in terms of bandwidth increase or loss in information rate[1]), **2)** the information-bearing signal has to be shut down, requiring the implementation of additional logic to synchronize the pilot sequence slots (in whatever domain they are allocated) both at the transmitter and the receiver, **3)** the estimate is based on particular locations of the pilot sequences (typically locations of time and/or frequency); therefore, interpolation is frequently required in order to obtain channel estimates at other times/frequencies.

Although they are less significant than the two channel estimation approaches described above, we would like to mention that there are estimation techniques called semi-blind estimation that use the statistics just as blind estimation does and known symbols as training algorithms do [15]. As their most important advantage, these techniques require shorter training sequences; however, they still need to use part of the time/frequency payload for sending training sequences.

1.1 Connections with Digital Watermarking

Although the basic idea was originally proposed in 1996 by Farhang-Boroujeny [22], recently the so-called Superimposed Training (SIT) has gained relevance as an alternative to the above approaches. In superimposed training a known pilot sequence (we will name it watermark due to the parallelism with data hiding) is added to the information-bearing signal (which, similarly, we will call host); to the best of our knowledge, the first reference to this relation between superimposed training and digital watermarking was mentioned in Mazzenga's work [40]. Essentially, these techniques use periodic sequences as watermarks to estimate the channel in order to take advantage of the induced cyclostationarity of the sent sequence. Since both signals are simply added (i.e., they are sent concurrently), explicit allocation of time/frequency slots for training purposes is not required, in contrast to traditional training methods [58, 41, 59]. However, assuming that the transmitter has some fixed power budget, the information-bearing signal will suffer from some power loss, and will be additionally distorted by the superimposed signal. Note that SIT is a precoding technique, which is not new in digital communications, since it has been extensively studied after being presented by Tomlinson-Harashima [53, 28] in order to take into account the side information regarding the channel state available at the transmitter.

Unfortunately, in superimposed training, the host and pilot sequences are not orthogonal; thus, the former will interfere with the pilot signal. This is a well-known problem in watermarking, where it is referred to as host interference, and it occurs in those schemes in which a watermark independent of

[1]Specifically, the training sequence in UMTS-TDD can be up to 20% of the payload.

the host is added to the latter (as in Additive Spread Spectrum schemes [13]). In both fields solutions have been proposed that devote some of the available power to partially cancel the host-interference in the direction of the added sequence. These schemes were independently developed by Malvar and Florêncio in 2003 [38] in the watermarking field, and by He and Tugnait in 2008 [29] for channel estimation (inspired by the work presented in 2005 for OFDM by Chen *et al.* [9][2]), and they were named Improved Spread Spectrum (ISS) and Partially-Data-Dependent superimposed training (PDD), respectively. Interestingly, this connection between PDD and ISS has not been reported before our work [16].

Both ISS and PDD only partially cancel the host interference, thus leaving room for improvement. In fact, full host-interference rejection has been achieved in data hiding by exploiting the Dirty Paper Coding (DPC) paradigm, initially proposed by Costa [10]. Adapting Costa's code construction, Chen and Wornell [8] proposed the use of Distortion Compensated Quantization Index Modulation (DC-QIM) which, thanks to its host-rejection feature, leads to substantial performance improvements with respect to ISS. The advantages of DPC techniques in watermarking have been widely recognized [10, 8, 18]. Specifically, DPC-based schemes can achieve the channel capacity for Additive White Gaussian Noise (AWGN) channels [21].

Since DPC is very sensitive to the gain attack (also known as linear valumetric attack), channel equalization has been studied in watermarking as one the possible solutions to this issue. In this case, the channel simply multiplies the watermarked signal by a constant real number, which can be cast as flat fading channel in traditional digital communications, resulting in very large probabilities of decoding error. Due to its relevance, several techniques have been proposed, based on channel equalization such as in Balado *et al.* [4] where a method based on uniform scalar quantizers and turbocodes was developed, which iteratively estimates the gain factor, compensates its effect, and decodes the embedded message. Shterev and Lagendijk [51] proposed an implementation based on exhaustive-search of the Maximum-Likelihood (ML) estimation of the gain factor; again, this value is used for equalizing the observations, and performing the decoding with the original codebook. However, the computational cost of [4, 51] is substantial, leaving room for improvement. This issue has been successfully addressed in our work [17], where we propose an ML approach but requiring far less computational resources than [51].

We should mention for the sake of completeness that other techniques tackle the DPC sensitivity to gain attacks in a way that can be called Robust Codebooks; in this case the typical Scalar Costa Scheme (SCS) codebooks [18] are

[2]The first work considering only full cancellation of host interference for SIT was proposed by Ghogho *et al.* in [26].

replaced by codebooks implicitly robust against the gain attack [47, 2, 42]. While [42] proposes the use of phase-based codebooks (as opposed to magnitude-based ones), in [2] the information is embedded by considering the maximum correlation between the host signal and a pseudo-randomly generated set of sequences, and in [47] a codebook that depends on the empirical statistics of the watermarked signal is used. Unfortunately, these techniques show several drawbacks as embedding distortion is difficult to control in phase quantization based techniques [42] and orthogonal dirty paper coding [2] (itself is also more computationally demanding than SCS), and the work in [47] requires a sample buffer to be filled before decoding can be performed in a robust way.

1.2 Applications of Channel Estimation

As previously indicated, channel estimation has applications in different technological fields. In this section, we discuss how channel estimation can be applied in some of them.

Digital communications has already been presented as one of the most active research areas for channel estimation. Primarily, channel estimation is used to help to mitigate the effects of Intersymbol Interference (ISI), which is a consequence of dispersive propagation channels. However, there exist several other well-known used applications in digital communications for channel estimation, including

- Automatic Gain Control (AGC) in Satellite Communications Channel: the estimation of the flat fading effect in satellite communications channel is not a trivial problem, even when that channel is time-invariant. Just as an example, in [25] De Gaudenzi and Luise propose a Non-Decision-Aided pseudo-Maximum Likelihood amplitude tracker, as part of a global all digital demodulator.

- Receiver adaptation: by considering the channel estimate, receiving filters could be designed in order to optimize the performance of the global system with respect to a target function. Different target parameters could be considered (e.g., minimization of the Mean Square Error (MSE), maximization of the Signal-to-Noise Ratio (SNR), minimization of the bit error rate, etc.).

- SNR estimation: a typical problem in communications is the estimation of the ratio between the received signal variance and the variance of the noise introduced by the channel. This is a non-trivial issue, mainly due to the possible scaling that the desired signal may have experienced.

In a different field, image restoration, one can be interested in recovering original information from a degraded image [35]. In order to study the alterations suffered by the image, they are usually modeled as a chain of operations applied to the image. These operations can include: blurring due to the optics, quantization, white balance adjustment, gamma correction, digital image post-processing, color filter array interpolation, etc. Some of those operations (e.g, blurring) can be modeled as bidimensional convolutions between the image and a filter that accounts for the alteration. If the filters can be estimated with enough accuracy, then their effect can be compensated by their inverse. In some applications, the image can be available at some point of the processing chain (e.g., the information obtained by the image sensor before performing the color array interpolation, JPEG compression, etc.) to aid filter estimation. However, in many cases, one cannot access to the original image, so one must perform a blind estimation (e.g., astronomical imaging, remote sensing, medical imaging, etc.). Besides image restoration, an estimate of the image filter is also interesting in image forensics, which studies the processing history of images in order to help determining their authenticity and reliability.

Acoustics also stands out as one of the areas that extensively uses channel estimation in many problems. For example,

- Room acoustic response estimation by using preprocessed audio: in some scenarios, the audio is prefiltered to compensate the acoustic response of the room to improve the listener's experience.

- Active noise control: in order to reduce the noise of an audio signal, which is usually modeled as an addition of an information-bearing signal and a noise signal, other additional noise signal is obtained (e.g., by using other sensor). This noise signal must be correlated with the noise of the audio signal, and it must be uncorrelated with the information-bearing signal. By estimating the acoustic channel through which the audio signal and the noise signal go, an active noise cancellation algorithm uses such estimate and the information that the noise signal has regarding the noise of the audio signal to cancel the latter at the listener (for example, by using a headphone/loudspeaker).

In petroleum exploration, channel estimation is used to determine the presence of oil reservoirs. In order to do that, the subsurface structure must be estimated; specifically, a general method presents the following steps: a source (in this case, controlled explosions on the surface) is emitted from the surface and its waves reflect between the union of different layers, part of the energy is reflected and recorded. In [37], the reflected waves are studied by modeling this effect as the convolution of an unknown source and a filter that accounts for the reflections.

1.3 A Brief Introduction to Digital Watermarking

During the last decade of the 20th century, the expansion of Internet, which allowed massive data sharing, and the digitization of the information with the consequently corresponding ability to copy digital contents (e.g., images, movies, music, etc.) cheaply and losslessly provoked that many digital contents with digital intellectual rights were distributed without respecting the rights of use. As a result, the digital content industry lost a relevant portion of its revenue[3].

In order to protect the intellectual property rights of digital contents, solutions based on encryption were proposed. In a basic scheme, a digital content is encrypted by the intellectual rights holder, afterwards the encrypted content is distributed to a legitimate user, decrypted and then consumed. However, one of the main drawbacks is that once the content is decrypted, it can be easily shared, the control of the content will be lost by the rights holder, and with it, the capability to profit from it.

Digital watermarking can be defined as the imperceptible and secure embedding of information into a signal. Almost 40 years after its first publication [23], this technology flourished driven by the idea that digital watermarking could help to maintain the business model of the commercial exploitation of digital content. In comparison to cryptography, digital watermarking in a regular digital right protection case is designed to insert data into the digital asset in such a way that the inserted data will be present as long as the digital content maintains its value (e.g., the alterations required to remove the watermark from a image must be large enough to deteriorate it to the point where it becomes worthless). Then, a special player can verify the presence of a valid watermark in the digital content before using it in order to enforce the compliance with its intellectual rights.

As digital watermarking was being developed for this major application (i.e., the intellectual property protection), many other uses arose. For example, for broadcast monitoring, a watermark can be embedded into an aired advertisement to be automatically tracked, in such a way that its statistics can be analyzed (e.g., number of times heard, its duration etc.). Another application of digital watermarking is to introduce metadata into digital contents. For example, Centum-RT developed an application for Unmanned Aerial Vehicles (UAVs) [7] that, in reconnaissance missions, can film terrain looking for a fire while at the same time embedding the Global Positioning System (GPS) coordinates or the time of the recording in order to make it possible to more easily determine the fire location and manage the footage.

[3]The conflict even provokes the pressure between governments. For instance, as Wikileaks has revealed, the US Government exhorted the Spanish Government to pass a stricter law of intellectual property [49].

Digital watermarking can also be used to discover the source of a leak. A watermark is embedded in order to be able to identify every copy of a digital content; if a copy is leaked, then this copy can be further analyzed to discover the source. For example in movie awards, copies of the films are distributed among the members of the jury, and if one of these copies ends up, for instance, in a peer-to-peer network, the license agreement breaker can be discovered. A famous case involved the actor Carmine Caridi who distributed movies in 2004. He was exposed and, as one the consequences, he was expelled from the Academy of Motion Picture Arts and Sciences [31]. This application can be also useful for the sensitive documents of corporations. A product called Shadow that uses watermarking for documents in traitor tracing and can expose the origin of the leaked document was recently presented by Gradiant [27].

Another example of application of digital watermarking is to increase the reliability of digital contents. For instance, to verify that footage taken by a videosurveillance system is authentic and has not been modified. Clearly such video has to be submitted as evidence in a court of law, and it has to be proven as completely genuine. As there are many powerful editing applications which allow an unskilled person to manipulate a video sequence in such a way as to alter it tracelessly, it is clear that a solution whereby a watermarking is embedded into the original is important. This can later be analyzed to detect manipulations or deletions. Academia has provided several watermarking-based solutions [35, 6, 19], as well as some commercial products, introduced by companies such as TRedess and its CWS Software [54].

In general, the selection of the digital watermarking algorithm to be used must depend on the requirements of the specific application. These needs can be described using the following basic characteristics of digital watermarking:

Blindness This property indicates if the original signal is required to extract the data (blind) or not (non-blind) from the received watermarked content. Typically, a blind digital watermarking algorithm is better than a non-blind one. Indeed, in many applications, due to practical constraints, only blind watermarking can be considered.

Perceptibility This feature measures the distortion which results from data embedding. Obviously, one is interested in producing an embedding distortion as low as possible. The particular selected metric depends on the required precision of the application and the limitations in its complexity. Due to the difficulties in modeling the human perception system, it is assumed that perceptual tests involving people are the most accurate but, at the same time, the most expensive. However, there are uses that do not require such precision, or where it cannot be afforded, and adopt simpler mathematical measures. For example, Peak Signal to Noise Ratio (PSNR) or more complex ones like Structural Similarity (SSIM) [61] can be used to measure the

difference between the original image and the watermarked image.

Security Indicates that ideally only the authorized users of the digital watermarking system can embed, modify, detect, or decode the embedded information. Attacks on security are defined in [48] as "those aimed at obtaining information from the secret parameters of the embedding and/or decoding functions".

Robustness Given an alteration of the watermarked signal, which can be intentional or unintentional, an algorithm can be designed to be fragile, semi-fragile, or robust, which is usually determined by the application. For example, in applications of protecting the copyright of digital contents, the owner is interested in using a robust digital watermarking technique.

Payload This indicates the amount of information that can be conveyed per time or use by the host signal. For instance, the number of bits that can be embedded into an image.

Method As indicated in the previous section, there are mainly two groups of digital watermarking techniques. The first to appear were the techniques that adopted Spread-Spectrum (SS), while the other class, the one that this work focuses on, is the Dirty Paper Coding.

1.4 Thesis Objectives and Outline

In this thesis, we propose the study of the flat fading channel estimation based on dirty paper coding, which will be addressed using Maximum-Likelihood Estimation (MLE). A set of practical ML-based algorithms will be proposed with strict complexity constraints (in contrast with [51] which employs exhaustive search). In addition, we want to analyze the performance of the technique in order to gain insights about its fundamental limits and to be able to determine if an analogous result to host interference cancellation obtained for digital watermarking could be achieved for estimation. The performance of the proposed algorithms must be tested under different conditions and compared against other estimation techniques (specifically, we will focus on estimators based on second order statistics as representative of blind estimators, and PDD, as an example of SIT). In addition to the main application framework, we will present further possible applications in a variety of technological fields to show the versatility of our approach.

In order to meet these objectives, the remaining chapters of this thesis are structured as follows:

- Chapter 2 briefly introduces watermarking techniques and channel estimation and presents the framework for this thesis. In addition, the hypotheses that will be used in this work are set and explained.

- In Chapter 3, the Maximum-Likelihood criterion is introduced to estimate the gain of the flat fading channel. There, the probability density function of the random variable that models the received signal is studied and approximated in several scenarios. Then, these probability density function (pdf) approximations are used to formulate an ML-based gain estimation algorithm. In addition, the lack of knowledge about the variance of the involved signals is studied, as well as a technique to control the effective conditions of the application cases.

- Chapter 4 presents a double theoretical analysis of the proposed technique following an estimation theory approach by studying the Cramér-Rao Bound (CRB), and also following an information theory perspective by analyzing the mutual information between the received sequence and a random variable modeling the gain of the flat fading channel.

- Chapter 5 presents a set of practical algorithms based on ML to obtain the gain of the flat fading channel with the restriction of requiring affordable computational resources. In addition, the accuracy of the technique is analyzed with respect to the key parameters of the system and compared with other channel estimation techniques.

- Chapter 6 extends the proposed estimation technique to other applications. Specifically, DPC estimation is used to make SCS more robust to gain attacks. In addition, Dirty Paper Coding Estimation (DPCE) is used in a digital communications framework for the case of using as host the symbols of Pulse-Amplitude Modulation (PAM) constellations. Two DPC-based techniques of complex gain estimation of complex flat fading channels are also presented.

- Finally, Chapter 7 summarizes the conclusions drawn from this thesis and describes several future research lines that we find interesting to address.

Chapter 2

Problem Formulation

In this chapter we formalize the gain estimation problem, and summarize some of the best known approaches to gain estimation and digital watermarking. Then, we introduce our new approach to gain estimation based on host quantization. Finally, we present the hypotheses that we will use throughout the remainder of this work.

2.1 Gain Estimation Problem Formulation

Following the discussion in the previous chapter, our design is expected to satisfy two constraints, specifically: 1) the transmission of the signal in the channel should not be interrupted, and 2) the estimation procedure must modify the transmitted signal only very slightly. As it was discussed in the Introduction, interruptions on the primary signal require additional logic, and may even be unfeasible. On the other hand, the original signal must keep its value (according to some objective distortion measure) when the estimation procedure is in place, i.e., after modifying the primary signal to assist the flat fading channel estimation (that is, gain estimation) performed at the receiver. Here we will use the MSE to quantify this distortion, but our results can be extended to any other topologically equivalent metric (e.g., all metrics induced by the p-norm in finite dimensional spaces). In terms of power budget, we aim at maximizing the ratio between the host power and the watermark power, and obtaining accurate estimates even when few observations are available.

Signal-aided gain estimation modifies the host signal $\boldsymbol{x}$, where the elements of this vector are continuous variables, to produce a transmitted signal $\boldsymbol{y}$ with some underlying structure that assists the estimation. We will denote $\boldsymbol{w} \triangleq \boldsymbol{y} - \boldsymbol{x}$; as indicated in the Introduction, this signal $\boldsymbol{w}$ is referred to as the *watermark* due to its similarity with the watermarking problem. We assume that the transmitted

signal goes through a flat fading channel (with gain t_0) which also introduces Additive White Gaussian Noise (AWGN) $\boldsymbol{n}$, so the received signal becomes $\boldsymbol{z} = t_0\boldsymbol{y} + \boldsymbol{n}$. A block diagram of this problem can be found in Fig. 2.1.

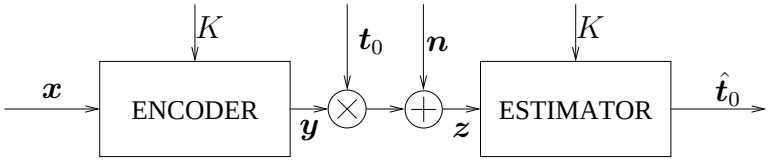


Figure 2.1: Block diagram of the flat fading channel (gain) estimation problem.

2.2 An Overview of Watermarking Techniques

Multibit digital watermarking aims at embedding a message m into a host signal $\boldsymbol{x}$, yielding the watermarked signal $\boldsymbol{y}$; in order to provide security against attacks, such embedding typically depends on a secret key K, which is shared by embedder and decoder. The watermark $\boldsymbol{w} \triangleq \boldsymbol{y} - \boldsymbol{x}$ is constrained to meet some cap on the embedding distortion σ_W^2, that is often derived from imperceptibility requirements.

In the simplest case, the signal at the decoder input $\boldsymbol{z}$ is modeled after some additive noise channel $\boldsymbol{n}$; therefore, $\boldsymbol{z} \triangleq \boldsymbol{y} + \boldsymbol{n}$. The decoder estimates m from $\boldsymbol{z}$ and the key K and provides an estimate $\hat{m}$. We consider the scenario where the watermarking system is typically designed to achieve the largest possible rate, maybe subject to additional complexity constraints. Fig. 2.2 illustrates the general block-diagram of the multibit digital watermarking problem.

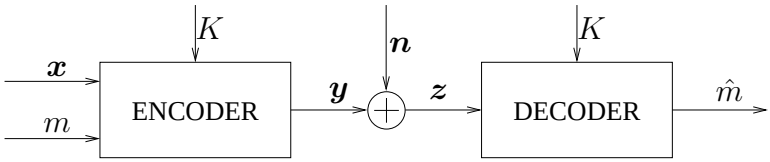


Figure 2.2: Block diagram of multibit digital watermarking.

Watermarking schemes differ in the way how they compute the watermark $\boldsymbol{w}$. One well-known instance is Additive Spread Spectrum (Add-SS) [13], where $\boldsymbol{w}$ is generated independently of the host $\boldsymbol{x}$; therefore, $\boldsymbol{x}$ interferes the decoding of m [18]. In order to mitigate the interference, ISS [38] subtracts part of the host projection onto the direction of the watermark; specifically, linear ISS computes the watermark as

$$\boldsymbol{w} = \left(\beta m - \lambda \frac{< \boldsymbol{x}, \boldsymbol{u} >}{\|\boldsymbol{u}\|^2}\right) \boldsymbol{u},$$

where β and λ are distortion controlling parameters, $< \cdot, \cdot >$ stands for the scalar product, $\boldsymbol{u}$ is a projecting vector which is generated from K independently of $\boldsymbol{x}$. In order to verify the distortion constraint on $\boldsymbol{w}$, for the binary case (i.e., $m \in \{0, 1\}$) then $\beta = \sqrt{\frac{L\sigma_W^2 - \lambda^2\sigma_X^2}{||\boldsymbol{u}||^2}}$. Pérez-Freire *et al.* proved in [46] that the maximum achievable rate of binary ISS is strictly larger than that obtained for Add-SS, although the channel capacity corresponding to a null host is not achieved; in other words, ISS only provides partial host rejection with respect to Add-SS.[1] This relationship between Add-SS and ISS is also relevant for the gain estimation problem, as the gain estimation counterpart of ISS (that is, PDD) also shows a performance improvement with respect to the gain estimation counterpart of superimposed pilots, due to this partial host rejection.

On the other hand, Quantization Index Modulation (QIM) watermarking [8] generates the watermarked signal by quantifying the host with a quantizer indexed by the embedded message, i.e., $\boldsymbol{y} = Q_m(\boldsymbol{x})$, so $\boldsymbol{w} = Q_m(\boldsymbol{x}) - \boldsymbol{x}$. Although QIM entirely removes the host interference, a modified version, named DC-QIM, is also proposed in [8] to achieve a tradeoff with the embedding distortion. When this scheme is applied, the watermarked signal is no longer at the quantizer centroid, but in the segment that connects it with the original host vector, the exact location depending on an optimization parameter explained below. Therefore, in DC-QIM a fraction of the quantization error, named *self-noise*, is added to the quantizer centroid, i.e.,

$$\boldsymbol{y} = Q_m(\boldsymbol{x}) + (1 - \alpha)[\boldsymbol{x} - Q_m(\boldsymbol{x})], \tag{2.1}$$

and consequently $\boldsymbol{w} = \alpha[Q_m(\boldsymbol{x}) - \boldsymbol{x}]$, where $\alpha \in (0, 1]$ is the so-called distortion compensation parameter. The larger α, the smaller the self-noise, i.e., the closer $\boldsymbol{y}$ will be to a quantization centroid, at the cost of a larger embedding distortion. Costa, based on channel capacity arguments [10], proposed to use $\alpha_{\text{Costa}} = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_N^2}$, which for our channel case should be modified to $\alpha_{\text{Costa}} = \frac{t_0^2\sigma_W^2}{t_0^2\sigma_W^2 + \sigma_N^2}$. We note that other designs for α, based on different target functions, can be found in the literature (e.g., [18]). We also note that QIM is simply a particular case of DC-QIM where $\alpha = 1$. The most extended implementation of DC-QIM in the watermarking literature corresponds to a message-dependent quantizer that is a shifted version of a prototype lattice quantizer, i.e., $Q_m(\boldsymbol{x}) = Q_\Lambda(\boldsymbol{x} - \boldsymbol{d} - \boldsymbol{d}_m) + \boldsymbol{d} + \boldsymbol{d}_m$, where $Q_\Lambda(\cdot)$ is the minimum-distance lattice quantizer, Λ denotes the prototype lattice, $\boldsymbol{d}$ (typically known as dither) is a pseudorandom vector dependent on K which is uniformly distributed on the fundamental Voronoi region of Λ, and $\boldsymbol{d}_m$ is message-dependent. Erez and Zamir [21] proved that there exists a sequence of lattices such that the maximum rate achievable for the Gaussian case when DC-QIM uses those lattices asymptotically (as L goes to infinity) approaches the capacity of an AWGN channel for no host interference (i.e., $\boldsymbol{x} =$

[1] Full host rejection would be achieved by making $\lambda = 1$, but this would entail an overly large distortion for most applications.

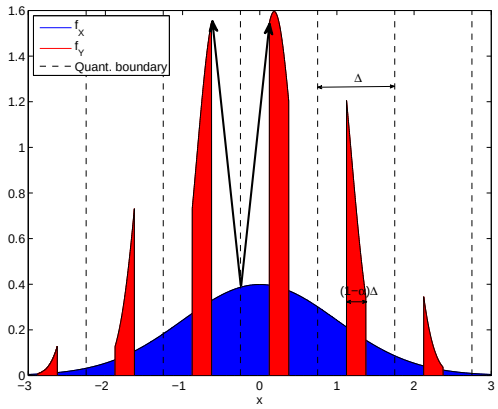


Figure 2.3: Comparison of f_X and f_Y, when $\sigma_X^2 = 1$, $\Delta = 1$, $d = 0.25$, and $\alpha = 0.75$.

0). This means that complete host rejection and optimal performance can be achieved with structured codes instead of the random ones used by Costa to prove his celebrated Dirty Paper Coding theorem [10]. In fact, practical schemes that approach Erez and Zamir's result have been proposed [20].

As we pointed out above, for the achievable rates of Erez and Zamir's lattice construction to approach the channel capacity it is necessary that the dimensionality L goes to infinity [21]. Even for large finite values of L the lattice quantization of the host sequence has a significant computational cost. For this reason, in most of practical applications $\Lambda = \Delta\mathbb{Z}^L$ (i.e., a uniform scalar quantizer) is used. Fig. 2.3 illustrates the effect of the watermark embedding on the pdf of the watermarked signal for the scalar case.

Summarizing, in terms of achievable rates, DC-QIM is superior than ISS, which in turn performs better than Add-SS; this ordering shows the worth of host-rejecting in watermarking, and highlights quantization-based techniques as the proper tools for achieving host rejection. As discussed in Chap. 1, a natural question is whether channel estimation can also benefit from these host-rejection methods. To show that this is so constitutes the main contribution of this thesis.

Finally, we would like to mention that some works in the watermarking literature [4, 51] studied the watermark decoding of quantization-based schemes when the received signal is a noisy scaled version of the transmitted one, i.e., $\boldsymbol{z} = t_0\boldsymbol{y} + \boldsymbol{n}$. In those cases, the primary target was not to provide an estimator of the scaling factor (as opposed to our previous work in [17]), but instead decode the embedded message. Watermarking quantization-based schemes are extremely sensitive to scaling, as it produces a codebook desynchronization problem. Our proposed method takes advantage of such weakness; it is known that random variables following a distribution sharply dependent on the parameter to be estimated often provide better estimates than those for which the dependence

is smoother [34].

2.3 Proposed Approach to the Gain Estimation Problem

In the previous chapter we have shown the parallelism in the development of watermarking and channel estimation techniques. As pointed out, superimposed pilots can be considered to be the channel estimation counterpart of Add-SS, and PDD the counterpart of ISS. Perhaps paradoxically, in the estimation literature there seems to be no correspondence to DC-QIM, that is, no quantization-based channel estimation method is currently known. Thus, the present PhD thesis comes to fill this existing gap, by proposing a DPCE method.

We thus propose to generate $\boldsymbol{y}$ by using a distortion compensated quantizer, i.e.,

$$\boldsymbol{y} = Q(\boldsymbol{x}) + (1-\alpha)[\boldsymbol{x} - Q(\boldsymbol{x})], \tag{2.2}$$

where $Q(\cdot)$ stands for a general quantizer, which will be typically based on the use of a lattice quantizer $Q_\Lambda(\cdot)$. We will focus on the case where a dithered uniform scalar quantizer with step-size Δ is used, so

$$\boldsymbol{y} = Q_\Delta(\boldsymbol{x} - \boldsymbol{d}) + \boldsymbol{d} + (1-\alpha)[\boldsymbol{x} - Q_\Delta(\boldsymbol{x} - \boldsymbol{d}) - \boldsymbol{d}], \tag{2.3}$$

where, Q_Δ is short for $Q_{\Delta\mathbb{Z}^L}$. Furthermore, from the uniform distribution of $\boldsymbol{D}$ on the fundamental Voronoi region of Λ, it follows that $\sigma_W^2 = \frac{\alpha^2\Delta^2}{12}$. A sufficient condition for the second design constraint in Sect. 2.1 to hold (i.e., that the estimation method can only slightly modify the transmitted signal) is that the Host-to-Quantizer Ratio (HQR), defined as HQR $\triangleq \frac{12\sigma_X^2}{\Delta^2}$ be much larger than 1 (HQR $\gg$ 1). Note that $\sigma_W^2 \leq \Delta^2/12$, and, consequently, the former is indeed a sufficient condition. We will also find it useful to define the Document-to-Watermark Ratio (DWR) (defined as $\frac{\sigma_X^2}{\sigma_W^2}$), Document-to-Noise Ratio (DNR) (defined as $\frac{\sigma_X^2}{\sigma_N^2}$), and the Watermark-to-Noise Ratio (WNR) (defined as $\frac{\sigma_W^2}{\sigma_N^2}$). Furthermore, following the watermarking terminology, the term $(1-\alpha)[\boldsymbol{x} - Q_\Delta(\boldsymbol{x}-\boldsymbol{d}) - \boldsymbol{d}]$ will be referred to as *self-noise*. The variance of the self-noise is $\frac{(1-\alpha)^2\Delta^2}{12}$. It is important to remark that the *only* difference between (2.1) and (2.2) is that now the quantizer is not indexed by the embedded message, as in our problem we are no longer interested in embedding a hidden message, but only in estimating the scaling factor t_0.[2]

[2]Notice, however, that our method can easily accommodate the embedding of such information, which may be interesting for some applications.

After presenting how the host is modified in order to estimate the gain of the channel, it is worth discussing the intrinsic limitations of our technique. For example, recalling the applications of channel estimation described in Sect. 1.2, it is reasonable to think that there are forensic applications in which one can modify an image in advance using our algorithm, in order to estimate a possible scaling attack afterwards. However, it is obvious that one cannot control the explosions on the surface with enough accuracy to be able to apply our scheme in petroleum exploration.

2.4 Work Hypotheses

We will consider two different practical scenarios, which will be described by the following quantities:

- **Self-Noise-to-Channel-Noise Ratio (SCR)**, which quantifies the relationship between the variance of the self-noise and the variance of the noise introduced by the channel when the scaling factor is t, both measured at the receiver, i.e., $\mathrm{SCR}(t) \triangleq \frac{(1-\alpha)^2 t^2 \Delta^2}{12\sigma_N^2}$.
- **Total-Noise-to-Quantizer Ratio (TNQR)**, which is the ratio between the total noise variance (self-noise after scaling, and channel noise) and the second moment of the scaled quantizer when the scaling factor is t, i.e., $\mathrm{TNQR}(t) \triangleq \frac{(1-\alpha)^2 t^2 \Delta^2/12 + \sigma_N^2}{t^2\Delta^2/12}$. Note that a realistic condition for correct centroid decoding (and consequently for good scaling estimation) is that $\mathrm{TNQR}(t_0) < 1$.
- **Total-Noise-to-Host Ratio (TNHR)**, which is the ratio between the total noise variance at the receiver (self-noise and channel noise) and the variance of the received host when the scaling factor is t, i.e., $\mathrm{TNHR}(t) \triangleq \frac{(1-\alpha)^2 t^2 \Delta^2/12 + \sigma_N^2}{t^2\sigma_X^2}$.

Although the three quantities we have just introduced are obviously functions of the scaling factor t (most of times we will evaluate them at the real scaling factor t_0), for the sake of notational simplicity, and whenever there is not risk of confusion, we will avoid to make explicit that dependence.

Then, with SNR used here to describe the set of hypotheses as the ratio between second moment of the scaled quantizer and the addition of the variances of the self-noise and the channel noise, the two considered scenarios are:

- **High-SNR**: characterized by $\mathrm{HQR} \gg 1$, $\mathrm{SCR}(t_0) \ll 1$, and $\mathrm{TNQR}(t_0) \ll 1$, meaning that the host is finely quantized, the self-noise is much smaller

than the noise introduced by the channel, and the addition of both noise components is much smaller than the quantization step-size. Be aware that the first and third inequalities imply $\mathrm{TNHR}(t_0) \ll 1$.

- **Low-SNR**: characterized by $\mathrm{HQR} \gg 1$, $\mathrm{SCR}(t_0) \ll 1$, $\mathrm{TNQR}(t_0) \gg 1$, and $\mathrm{TNHR}(t_0) \ll 1$, meaning, similarly to the previous case, that the host is finely quantized, and the self-noise is much smaller than the noise introduced by the channel. Furthermore, in this scenario the addition of both noise components is larger than the quantizer step-size, although it is much smaller than the host signal.

Chapter 3

Maximum-Likelihood Estimator

In order to obtain the estimate of the scaling factor t_0, we use the ML criterion which seeks the most likely value of t given a vector of observations $\boldsymbol{z}$ when there is no *a priori* knowledge about the distribution of the scaling factor, i.e., the ML estimator can be obtained as

$$\begin{aligned}\hat{t}_0(\boldsymbol{z}) &= \arg\max_t f_{\boldsymbol{Z}|T,\boldsymbol{K}}(\boldsymbol{z}|t,\boldsymbol{d}) \\ &= \arg\max_t \prod_{i=1}^{L} f_{Z|T,K}(z_i|t,d_i) \\ &= \arg\min_t -\sum_{i=1}^{L} \log\left(f_{Z|T,K}(z_i|t,d_i)\right) = \arg\min_t L(t,\boldsymbol{z}), \end{aligned} \tag{3.1}$$

where in the previous expression $f_{\boldsymbol{Z}|T,\boldsymbol{K}}(\boldsymbol{z}|t,\boldsymbol{d})$ and $f_{Z|T,K}(z_i|t,d_i)$ stand for the joint distribution of $\boldsymbol{Z}$ and the distribution of Z, respectively, given the scaling factor t and the dither sequence $\boldsymbol{d}$. The independence of the samples of the received watermarked sequence $\boldsymbol{z}$ allows us to simplify the expression by the second equality. In the third equality, the ML estimator is rewritten after using the monotonically increasing property of the logarithm, and $L(t,\boldsymbol{z})$ denotes the cost function to optimize in the fourth equality.

In many applications requiring parameter estimation, MLE is selected because it is a systematic approach and also due to its interesting properties [60]:

- The MLE is consistent, the solution of (3.1) converges in probability to t_0 as $L \to \infty$.
- It is asymptotically efficient, the value of the variance of $t_0 - \hat{t}_0(\boldsymbol{z})$ tends to the CRB as $L \to \infty$.
- The MLE asymptotically follows a Gaussian distribution with mean t_0 and variance the CRB.

The CRB is the lower bound of the variance of the unbiased estimators; this will be studied more deeply in Chapter 4.

3.1 Pdf of the Observations and Approximations

The pdf of $Z = t_0 Y + N$ given the scaling factor t_0 and the dither d is calculated, where Y models the Distortion Compensated Dither-Modulation (DC-DM) watermarked signal using uniform scalar quantizers and $N \sim \mathcal{N}(0, \sigma_N^2)$ denotes the noise of the channel. First, the pdf of Y given d can be expressed as

$$f_{Y|K}(y|d) = \sum_{i=-\infty}^{\infty} p_{I|K}(i|d) f_{S|K,I}(y|d,i), \tag{3.2}$$

where $p_{I|K}(i|d)$ denotes the probability mass function of the active centroid index i for known d, and $f_{S|K,I}(y|d,i)$ denotes the pdf of Y given d and i, i.e., the distribution of the self-noise of the ith centroid for known d. These two functions can be mathematically expressed as

$$p_{I|K}(i|d) = \int_{i\Delta-\frac{\Delta}{2}+d}^{i\Delta+\frac{\Delta}{2}+d} f_X(\tau)\, d\tau, \tag{3.3}$$

$$f_{S|K,I}(y|d,i) = \begin{cases} \frac{1}{p_{I|K}(i|d)(1-\alpha)} f_X\left(\frac{y-\alpha(i\Delta+d)}{(1-\alpha)}\right) & \text{if} \quad i\Delta - \frac{(1-\alpha)\Delta}{2} + d \le y \\ & \le i\Delta + \frac{(1-\alpha)\Delta}{2} + d \\ 0 & \text{otherwise,} \end{cases}$$

where in the previous expressions X denotes the random variable modeling the host, Δ denotes the step-size of the used scalar uniform quantizer $\mathcal{Q}_\Delta(\cdot)$, and α stands for the distortion-compensation parameter.

The pdf of $R = t_0 Y$ can be determined from the pdf of Y by applying the well-known scaling of a continuous random variable property [43] (i.e., if $A = cB$, then $f_A(a) = f_B(a/c)/c$) as

$$\begin{aligned} f_{R|T,K}(r|t_0,d) &= \frac{f_{Y|K}\left(\frac{r}{t_0}|d\right)}{t_0} \\ &= \frac{1}{t_0} \sum_{i=-\infty}^{\infty} p_{I|K}(i|d) f_{S|K,I}\left(\frac{r}{t_0}\Big|d,i\right). \end{aligned}$$

Finally, the resulting pdf of the addition of two independent random variables can be calculated as the convolution (operation denoted by $*$) of their respective

pdfs. Using this, the pdf of Z is expressed as

$$f_{Z|T,K}(z|t_0,d) = f_{R|T,K}(z|t_0,d) * f_N(z) \tag{3.4}$$

$$= \frac{1}{t_0}\sum_{i=-\infty}^{\infty} p_{I|K}(i|d) f_{S|K,I}\left(\frac{z}{t_0}\Big|d,i\right) * f_N(z). \tag{3.5}$$

After particularizing the pdf of Z given t_0 and d for zero-mean Gaussian distributed hosts, we have found that the obtained expression of the distribution of Z, which is not shown here, is difficult to handle. In order to deal with this, we propose the approximations of the pdf of these signals based on the hypotheses described in Sect. 2.4 of the previous chapter. Therefore, we divide the analysis of the pdf of Z for Gaussian distributed hosts into the two considered scenarios: the low-SNR case and the high-SNR case.

3.1.1 Low-SNR Case

In order to obtain the approximation for the pdf of Z for the low-SNR case, we first consider that under HQR $\gg 1$, the host will be asymptotically uniformly distributed inside each quantization bin, so the distribution of the self-noise will be the same for each considered centroid, implying that the used centroid and the self-noise are independent, allowing us to write

$$f_{Y|K}(y|d) \approx f_{U|K}(y|d) * f_S(y),$$

where U is the continuous random variable modeling the active centroid, and S models the self-noise that follows a uniformly distribution in $[-(1-\alpha)\Delta/2, (1-\alpha)\Delta/2]$; note that in this scenario this distribution does not depend on the index of the centroid.

The function $f_{U|K}(u|d)$ can be written as $f_{U|K}(u|d) = h_I(u)\sum_{k=-\infty}^{\infty}\delta(u - k\Delta - d)$, where $h_I(u)$ is defined as

$$h_I(u) = \int_{u-\Delta/2}^{u+\Delta/2} \frac{e^{-\frac{\tau^2}{2\sigma_X^2}}}{\sqrt{2\pi}\sigma_X} d\tau = f_X(u) * \text{rect}\left(\frac{u}{\Delta}\right), \tag{3.6}$$

where in the previous expressions, the integral is replaced by a convolution of the pdf of X, which follows a Gaussian distribution, and a rectangular function in the interval $[-\Delta/2, \Delta/2]$ and amplitude 1 (i.e., rect $\left(\frac{x}{\Delta}\right)$). By using the convolution, the Fourier transform of (3.6) can be seen to be

$$H_I(f) = e^{-2(\pi\sigma_X f)^2} \cdot \Delta\text{sinc}\left(\Delta f\right),$$

where $\text{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$.

Therefore, the Fourier transform of the approximation of $f_{U|K}(u|d)$ is

$$\begin{aligned} F_{U|K}(f|d) &= H_I(f) * \left(e^{-j2\pi fd} \frac{1}{\Delta} \sum_{i=-\infty}^{\infty} \delta\left(f - \frac{i}{\Delta}\right)\right) \\ &= \left(e^{-2(\pi\sigma_X f)^2} \text{sinc}\left(\Delta f\right)\right) * \left(e^{-j2\pi fd} \sum_{i=-\infty}^{\infty} \delta\left(f - \frac{i}{\Delta}\right)\right). \end{aligned}$$

The pdf of the $R = t_0 Y$ can be approximated by

$$f_{R|T,K}(r|t_0,d) = \frac{f_{U|K}(\frac{r}{t_0}|d)}{t_0} * \frac{1}{(1-\alpha)\Delta t_0} \text{rect}\left(\frac{r}{(1-\alpha)\Delta t_0}\right).$$

Therefore, the pdf of Z can be obtained as in (3.4), whose corresponding Fourier transform is

$$\begin{aligned} F_{Z|T,K}(f|t_0,d) &= F_{R|T,K}(f|t_0,d) \cdot F_N(f) \\ &\approx \left(\left(e^{-2(\pi\sigma_X t_0 f)^2} \text{sinc}\left(\Delta t_0 f\right)\right) * \left(e^{-j2\pi f dt_0} \sum_{i=-\infty}^{\infty} \delta(f - \frac{i}{t_0\Delta})\right)\right) \\ &\quad \times \text{sinc}\left((1-\alpha)\Delta t_0 f\right) e^{-2(\pi\sigma_N f)^2}. \end{aligned} \tag{3.7}$$

In order to simplify $F_{Z|T,K}(f|t_0,d)$ further, we will exploit the facts that $\sigma_X^2 \gg \Delta^2/12$, due to HQR $\gg 1$, and that $\sigma_N^2 \gg (1-\alpha)^2\Delta^2 t_0^2/12$, due to $\text{SCR}(t_0) \ll 1$, implying that in both cases the exponential is much sharper than the sinc it is multiplying to. Therefore, it will make sense to provide a local approximation of the sinc around its maximum; having that target in mind, we will use the second order Taylor series expansion of $\log(\text{sinc}(ax))$ around $x = 0$ (valid whenever $|x| \ll 1/a$), given by $\log\left(\text{sinc}\left(ax\right)\right) \approx -\frac{\pi^2 a^2 x^2}{6}$ to approximate the sinc function by $\text{sinc}\left(ax\right) \approx e^{\frac{-\pi^2 a^2 x^2}{6}}$. In addition to this approximation, by using $\text{TNQR}(t_0) \gg 1$, one can accurately just keep $i = -1, 0, 1$ in the sum in (3.7). These approximations allow us to write

$$\begin{aligned} F_{Z|T,K}(f|t_0,d) \approx &\left(\left(e^{-2(\pi t_0 f)^2\left(\sigma_X^2 + \frac{\Delta^2}{12}\right)}\right) * \left(e^{-j2\pi f dt_0} \sum_{i=-1}^{1} \delta(f - \frac{i}{t_0\Delta})\right)\right) \\ &\times e^{-2(\pi f)^2\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}, \end{aligned}$$

where the corresponding inverse transform is

$$\begin{aligned} f_{Z|T,K}^{\text{low-SNR}'}(z|t_0,d) \approx &\frac{e^{-\frac{z^2}{2\left(\sigma_N^2 + \left(\frac{(1+(1-\alpha)^2)\Delta^2}{12} + \sigma_X^2\right)t_0^2\right)}}}{\sqrt{2\pi\left(\sigma_N^2 + \left(\frac{(1+(1-\alpha)^2)\Delta^2}{12} + \sigma_X^2\right)t_0^2\right)}} \Bigg(1 \\ &+ 2e^{\frac{2\pi^2\left(-\left(\frac{\Delta^2}{12} + \sigma_X^2\right)\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)\right)}{\Delta^2\left(\sigma_N^2 + \left(\frac{(1+(1-\alpha)^2)\Delta^2}{12} + \sigma_X^2\right)t_0^2\right)}} \cos\left(2\pi\left(\frac{\left(\frac{\Delta^2}{12} + \sigma_X^2\right)t_0 z}{\Delta\left(\sigma_N^2 + \left(\frac{(1+(1-\alpha)^2)\Delta^2}{12} + \sigma_X^2\right)t_0^2\right)} - \frac{d}{\Delta}\right)\right)\Bigg), \end{aligned}$$

and since HQR $\gg 1$, $\Delta^2/12$ can be neglected in comparison with σ_X^2, yielding

$$f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d) \approx \frac{e^{-\frac{z^2}{2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\left(1+2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\right.$$
$$\left.\times\cos\left(\frac{2\pi\sigma_X^2 t_0 z}{\Delta\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}-\frac{2\pi d}{\Delta}\right)\right). \tag{3.8}$$

By applying $\text{TNHR}(t_0) \ll 1$ to (3.8) in order to neglect σ_N^2 in comparison with $t_0^2\sigma_X^2$, a more simplified approximation of the pdf of Z is obtained

$$f_{Z|T,K}^{\text{low-SNR,2}}(z|t_0,d) \approx \frac{e^{-\frac{z^2}{2t_0^2\sigma_X^2}}}{\sqrt{2\pi t_0^2\sigma_X^2}}\left(1+2e^{-\frac{2\pi^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}{\Delta^2 t_0^2}}\cos\left(\frac{2\pi z}{\Delta t_0}-\frac{2\pi d}{\Delta}\right)\right). \tag{3.9}$$

3.1.2 High-SNR Case

Since HQR $\gg 1$ holds, $p_{I|K}(i|d)$ used in (3.2) and defined in (3.3) can be accurately approximated by

$$p_{I|K}(i|d) \approx \frac{\Delta e^{-\frac{(i\Delta+d)^2}{2\sigma_X^2}}}{\sqrt{2\pi\sigma_X^2}}.$$

As stated above, using HQR $\gg 1$, one can assume that the self-noise S is uniformly distributed. Based on this and since S and N are independent, the pdf of $S+N$ can be obtained as the convolution of the pdf of a random variable uniformly distributed in $[t_0(i\Delta-(1-\alpha)\Delta/2+d), t_0(i\Delta+(1-\alpha)\Delta/2+d)]$, corresponding to the scaled self-noise, and the pdf of N; thus, $f_{Z|T,K}(z|t_0,d)$ can be approximated as

$$f_{Z|T,K}(z|t_0,d) \approx \sum_{i=-\infty}^{\infty}\frac{p_{I|K}(i|d)}{t_0(1-\alpha)\Delta}\int_{t_0(i\Delta-(1-\alpha)\Delta/2+d)}^{t_0(i\Delta+(1-\alpha)\Delta/2+d)}\frac{e^{-\frac{(z-\tau)^2}{2\sigma_N^2}}}{\sqrt{2\pi\sigma_N^2}}d\tau.$$

Using $\text{SCR}(t_0) \ll 1$, the result of the convolution can be approximated, using the same reasoning in the transform domain as in the previous section, by a Gaussian distribution with variance the addition of the variances of the scaled self-noise

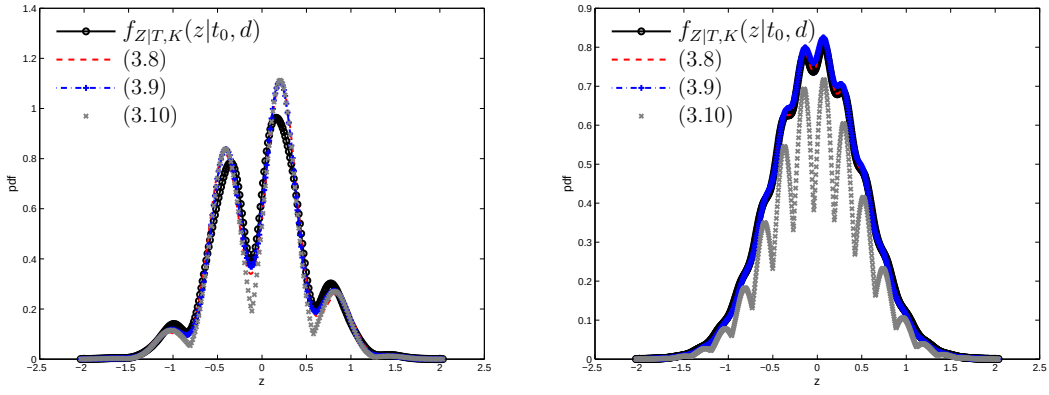
and the channel noise. Using this approximation, the pdf of Z given t_0, the dither d, and the centroid index i, can be expressed as

$$f_{Z|T,K,I}(z|t_0,d,i) \approx \frac{e^{-\frac{(z-t_0 i\Delta - t_0 d)^2}{2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}}.$$

Under $\mathrm{TNQR}(t_0) \ll 1$, one can accurately approximate the reconstruction point $i\Delta$ by $\mathcal{Q}_\Delta(z/t_0 - d)$; therefore, the approximation of the pdf of Z can be expressed as

$$\begin{aligned} f_{Z|T,K}^{\text{high-SNR}}(z|t_0,d) &\approx \sum_{i=-\infty}^{\infty} \frac{\Delta e^{-\frac{(i\Delta+d)^2}{2\sigma_X^2}}}{\sqrt{2\pi\sigma_X^2}} \frac{e^{-\frac{(z-t_0 i\Delta - t_0 d)^2}{2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}} \delta\left(i\Delta - \mathcal{Q}_\Delta\left(\frac{z}{t_0}-d\right)\right) \\ &\approx \frac{\Delta e^{-\frac{z^2}{2t_0^2\sigma_X^2}}}{\sqrt{2\pi\sigma_X^2}} \frac{e^{-\frac{((z-t_0 d) \mod (t_0\Delta))^2}{2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}}. \end{aligned} \tag{3.10}$$

3.1.3 Experimental Results



(a) $\alpha = 0.25$ and $d = \Delta/3$ (b) $\alpha = 0.75$ and $d = \Delta/5$

Figure 3.1: Pdf of Z and its approximations. DWR = 20 dB, WNR = 0 dB, and $t_0 = 0.5$. For (a): HQR = 6.25, $\mathrm{SCR}(t_0) = 2.25$, $\mathrm{TNQR}(t_0) \approx 8.13 \times 10^{-1}$, and $\mathrm{TNHR}(t_0) = 1.3 \times 10^{-1}$. For (b): HQR $\approx 5.63 \times 10$, $\mathrm{SCR}(t_0) \approx 2.78 \times 10^{-2}$, $\mathrm{TNQR}(t_0) \approx 2.31$, and $\mathrm{TNHR}(t_0) = 4.11 \times 10^{-2}$.

Figs. 3.1 and 3.2 show three examples of the true pdf of Z and our approximations.

The two examples of Fig. 3.1 were obtained for DWR = 20 dB, WNR = 0 dB, and $t_0 = 0.5$. In the left pane $\alpha = 0.25$, while on the right one $\alpha = 0.75$. It is worth pointing out that the proposed low-SNR approximations are tight in the presented scenarios, although only the four hypotheses used to obtain that approximation (i.e., HQR $\gg 1$, SCR$(t_0) \ll 1$, TNQR$(t_0) \gg 1$, and TNHR$(t_0) \ll 1$) can be simultaneously fulfilled for the right pane. Although the used hypotheses to obtain it are not verified in these two cases (i.e., HQR $\gg 1$, SCR$(t_0) \ll 1$, and TNQR$(t_0) \ll 1$), the high-SNR approximation performs rather well for $\alpha = 0.25$ case. In Fig. 3.2 a typical high-SNR scenario is proposed; namely, DWR = 20 dB, WNR = 3 dB, $\alpha = 0.75$, and $t_0 = 3$. The resemblance between the actual pdf of Z and its approximation $f^{\text{high-SNR}}_{Z|T,K}(z|t_0, d)$ is high, which is coherent with the verification of the corresponding hypotheses for that case. In addition, as one can expect, the low-SNR approximations show worse performance in this case, even taking negative values.

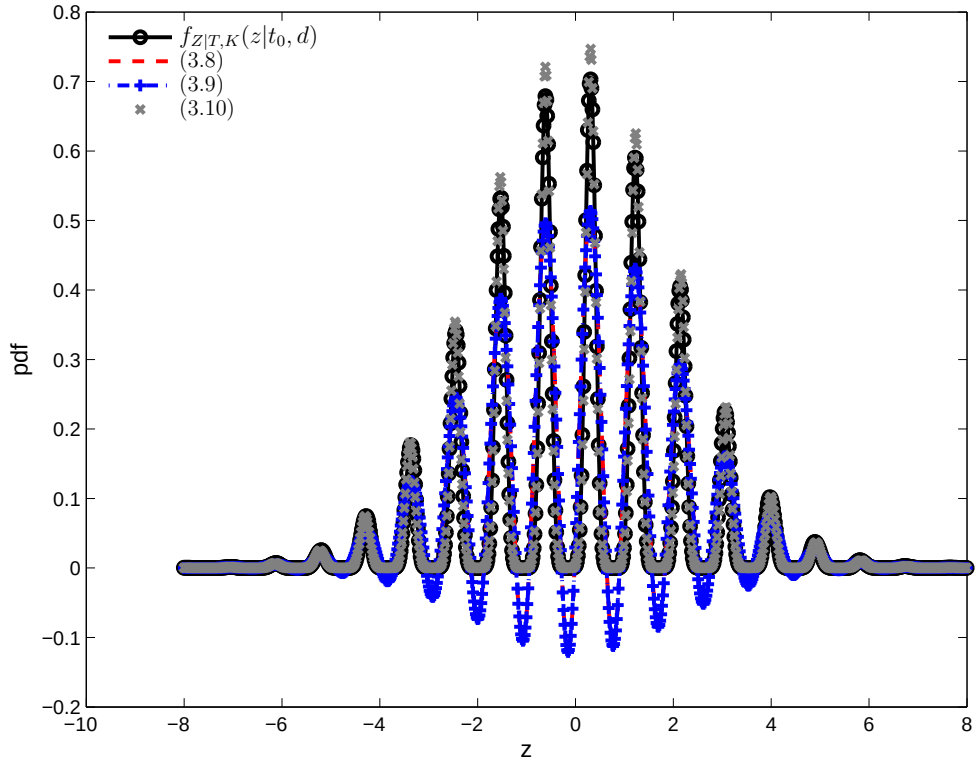
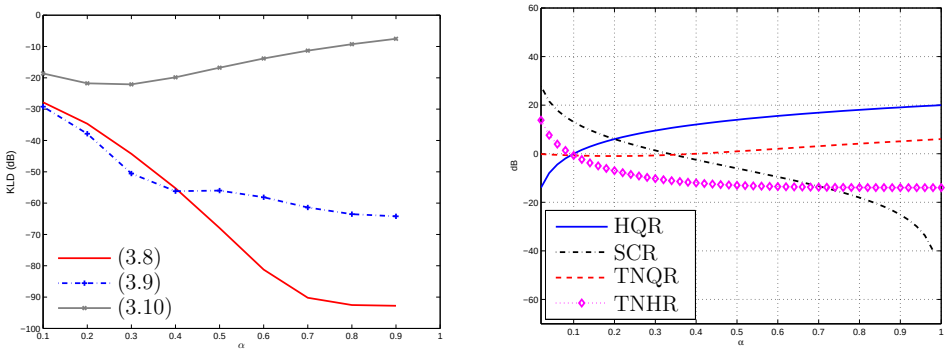


Figure 3.2: Pdf of Z and the its approximations shown in this manuscript. DWR = 20 dB, WNR = 3 dB, $\alpha = 0.75$, $t_0 = 2$ and $d = \Delta/5$. HQR $\approx 5.63 \times 10$, SCR$(t_0) \approx 8.87 \times 10^{-1}$, TNQR$(t_0) \approx 1.33 \times 10^{-1}$, and TNHR$(t_0) = 2.36 \times 10^{-3}$.

3.1.3.1 KLD of the pdf Approximations

In order to measure the adequacy of our approximations of the pdf of Z, the Kullback-Leibler Divergence (KLD) [11] is used as a measure of the distance be-



(a) KLD (b) Hypotheses

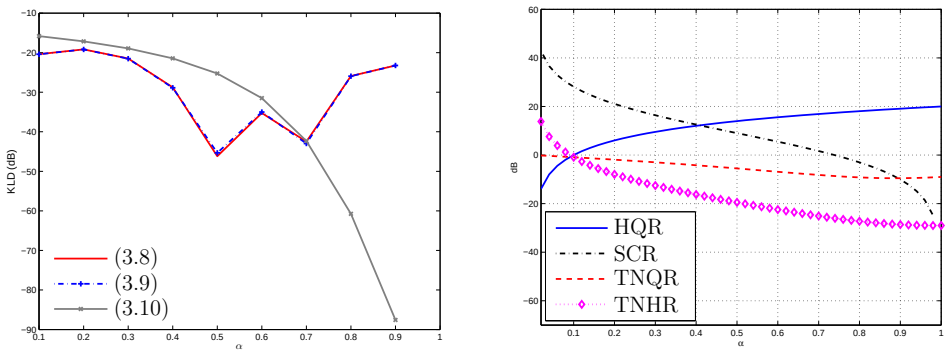
Figure 3.3: DWR = 20 dB, WNR = 0 dB, and $t_0 = 0.5$.

tween two distributions, in this case between the actual distribution and our approximations. The obtained results are analyzed taking into account the fulfillment of the hypotheses.

Figs. 3.3-3.5 show a set of pairs of figures: the KLD averaged with respect to the dither vs α for different values of DWR, WNR and t_0 on the left pane, and their corresponding curves of the HQR, SCR, TNQR, and TNHR in the right pane. The average KLD is obtained as

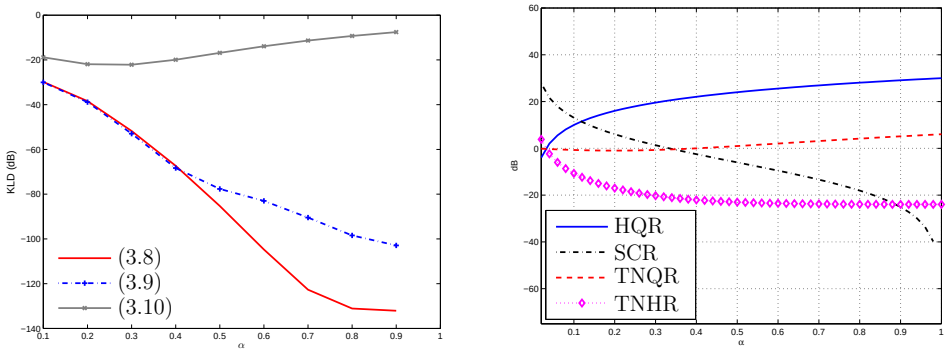
$$\text{KLD} \triangleq \frac{1}{\Delta}\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}}\int_{-\infty}^{\infty} f_{Z|T,K}(\tau|t_0,\upsilon)\log\left(\frac{f_{Z|T,K}(\tau|t_0,\upsilon)}{f_{Z|T,K}^{approx}(\tau|t_0,\upsilon)}\right)d\tau d\upsilon,$$

where in the previous expression *approx* denotes the approximation of the pdf of Z: for low-SNR (3.8), for low-SNR verifying $\text{TNHR}(t_0) \ll 1$ (3.9), or for the high-SNR cases (3.10). The aim of these average KLD curves is to quantify the average accuracy of the approximations $f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)$, $f_{Z|T,K}^{\text{low-SNR},2}(z|t_0,d)$, and $f_{Z|T,K}^{\text{high-SNR}}(z|t_0,d)$ for the considered distribution of D.



(a) KLD (b) Hypotheses

Figure 3.4: DWR = 20 dB, WNR = 3 dB, and $t_0 = 2$.



(a) KLD (b) Hypotheses

Figure 3.5: DWR = 30 dB, WNR = 0 dB, and $t_0 = 0.5$.

Before drawing conclusions from the average KLD figures, one must take into account that:

- the larger α and DWR, the more reasonable it is to assume that hypothesis $\mathrm{HQR} \gg 1$ holds.
- the larger α and the smaller t_0, the more reasonable it is to assume that $\mathrm{SCR}(t_0) \ll 1$ holds.
- for a given WNR, the larger the DWR, the more reasonable the $\mathrm{TNHR}(t_0) \ll 1$ is.

Figs. 3.3 and 3.5 show the average KLD as function of α for WNR = 0 dB, $t_0 = 0.5$, DWR = 20 and 30 dB, respectively. For $\alpha > 0.45$, where the hypotheses for the low-SNR case can be considered jointly fulfilled according to their HQR, $\mathrm{SCR}(t_0)$, $\mathrm{TNQR}(t_0)$, and $\mathrm{TNHR}(t_0)$ curves, one can see that the average KLD curve for $f_{Z|T,K}^{\text{low-SNR}}(z|t_0, d)$ is more accurate than the corresponding curve for $f_{Z|T,K}^{\text{low-SNR},2}(z|t_0, d)$. In contrast, the performance of the approximation for the high-SNR case $f_{Z|T,K}^{\text{high-SNR}}(z|t_0, d)$ is poor as does not fulfill its corresponding set of hypotheses.

In Fig. 3.4, the KLD curves are depicted for DWR = 20, WNR = 3 dB, and $t_0 = 2$. In this case, the low-SNR hypotheses are not verified, while for $\alpha \geq 0.7$ one can consider that the high-SNR hypotheses hold and, therefore, as one can expect, its KLD is significantly smaller than for the low-SNR cases.

3.2 Approximation of the Maximum Likelihood Function

The pdf of Z given a scaling factor t and a dither sample d_i is required by MLE. By using the generic expression (3.5) introduced at the beginning of Sect. 3.1, the expression of $L(t, \boldsymbol{z})$ becomes

$$L(t, \boldsymbol{z}) = -\sum_{i=1}^{L} \log \left(\frac{1}{t} \sum_{l=-\infty}^{\infty} p_{I|K}(l|d_i) f_{S|K,I} \left(\frac{z_i}{t} \middle| d_i, l \right) * f_N(z_i) \right);$$

which, in general and as stated above, it is difficult to handle. In order to deal with this issue, three approximations of the pdf of Z for Gaussian distributed hosts for low-SNR and high-SNR scenarios have been proposed in the previous section and they are used to obtain accurate and more tractable closed-form approximations of the cost function.

3.2.1 Low-SNR Case

Considering Gaussian distributed hosts for low-SNR scenarios, the approximation of $L(t, \boldsymbol{z})$ is calculated from the corresponding approximation of the pdf of Z (3.8), and also by using that $\log(1 + h) \approx h$ for $|h| \ll 1$, approximation based on its series expansion [1, p. 68], which will be valid if the ratio in the argument of the second exponential function of that pdf verifies

$$\frac{2\pi^2 \sigma_X^2 \left(\sigma_N^2 + \frac{(1-\alpha)^2 \Delta^2 t^2}{12} \right)}{\Delta^2 \left(\sigma_N^2 + \sigma_X^2 t^2 \right)} \gg 1;$$

thus, for values of t sufficiently close to t_0 the inequalities HQR $\gg 1$, SCR$(t) \ll 1$, and TNQR$(t) \gg 1$ are satisfied, and $L(t, \boldsymbol{z})$ can be approximated as

$$\begin{aligned} L(t, \boldsymbol{z}) \approx & \frac{\|\boldsymbol{z}\|^2}{2\left(\sigma_N^2 + \sigma_X^2 t^2\right)} + \frac{L}{2} \log\left(2\pi\left(\sigma_N^2 + \sigma_X^2 t^2\right)\right) \\ & - \sum_{i=1}^{L} 2e^{-\frac{2\pi^2 \sigma_X^2 \left(\sigma_N^2 + \frac{(1-\alpha)^2 \Delta^2 t^2}{12} \right)}{\Delta^2 \left(\sigma_N^2 + \sigma_X^2 t^2 \right)}} \cos\left(\frac{2\pi \sigma_X^2 t z_i}{\Delta\left(\sigma_N^2 + \sigma_X^2 t^2\right)} - \frac{2\pi d_i}{\Delta} \right). \end{aligned} \quad (3.11)$$

Using Central Limit Theorem (CLT), we study the expectation of the ML cost function in App. 3.A in order to gain insight into its properties for $L \to \infty$:

1. $\mathrm{E}\{L(t, Z)\}$ only has a minimum for $t/t_0 > \sigma_N/(\sigma_X t_0)$ (where $\sigma_N^2/(\sigma_X^2 t_0^2)$ tends to zero under TNHR$(t_0) \ll 1$) that is located close to t_0,

2. for any $\sigma_N^2/(\sigma_X^2 t_0^2)$, there is an arbitrary $\eta > 0$, such that $\mathrm{E}\{L(t,Z)\}$ monotonically decreases in $\eta < t/t_0 < 1$; therefore, since $\sigma_N^2/(\sigma_X^2 t_0^2) \to 0$ then η can be arbitrarily close to 0, $\mathrm{E}\{L(t,Z)\}$ monotonically decreases.

3. $\mathrm{E}\{L(t,Z)\}$ monotonically increases for $t > t_0$.

The cost function (3.11) can be further simplified by applying $\mathrm{TNHR}(t_0) \ll 1$, yielding

$$L(t,\boldsymbol{z}) \approx \frac{\|\boldsymbol{z}\|^2}{2\sigma_X^2 t^2} + \frac{L}{2}\log\left(2\pi\sigma_X^2 t^2\right) - \sum_{i=1}^{L} 2e^{-\frac{2\pi^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)}{\Delta^2 t^2}} \cos\left(\frac{2\pi z_i}{\Delta t} - \frac{2\pi d_i}{\Delta}\right), \tag{3.12}$$

which corresponds to the cost function using (3.9).

3.2.1.1 Gaussian Distributed Z Approximation

It is worth noting that if the dither is unknown (or it is not used for estimation) and $\mathrm{HQR} \gg 1$ holds, then Z can be approximated by a Gaussian distribution $Z \sim \mathcal{N}(0, t_0^2(\sigma_X^2+\sigma_W^2)+\sigma_N^2)$ (i.e., there is no structure in the pdf of Z). Therefore, from the expression of MLE at the beginning of this chapter, it is straightforward to obtain the ML estimator as

$$\hat{t_0}(\boldsymbol{z})_{\mathrm{var}} = \sqrt{\frac{\frac{\|\boldsymbol{z}\|^2}{L} - \sigma_N^2}{\sigma_X^2 + \sigma_W^2}}, \tag{3.13}$$

which will be referred to as the variance-based estimator of t_0 throughout this thesis, name based on its variance matching nature. Due to its simplicity, this estimator can be used to coarsely estimate t_0, even in cases where the conditions listed at the beginning of this section are not verified.

3.2.2 High-SNR Case

Using the approximation of the pdf of Z for zero-mean Gaussian distributed hosts in high-SNR scenarios (3.10) calculated in Sect. 3.1.2, the cost function can be approximated as

$$L(t,\boldsymbol{z}) \approx L\log\left(\frac{2\pi\sigma_X^2}{\Delta^2}\right) + \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2} + L\log\left(2\pi\left(\sigma_N^2 + \frac{(1-\alpha)^2 t^2\Delta^2}{12}\right)\right) + \frac{\|(\boldsymbol{z} - t\boldsymbol{d}) \mod (t\Delta)\|^2}{\sigma_N^2 + \frac{(1-\alpha)^2 t^2\Delta^2}{12}}, \tag{3.14}$$

where a factor of 1/2 multiplying the four components was removed. For the sake of notational simplicity, we denote the previous expression as $L(t, \boldsymbol{z})$ in the sense that the result of the optimization using this expression and the original approximation cost function is the same. In addition, the term $L \log\left((2\pi\sigma_X^2)/\Delta^2\right)$ is neglected due to its independence of t, obtaining

$$L(t, \boldsymbol{z}) \approx \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2} + L \log\left(2\pi\left(\sigma_N^2 + \frac{(1-\alpha)^2 t^2 \Delta^2}{12}\right)\right) + \frac{\|(\boldsymbol{z} - t\boldsymbol{d}) \mod (t\Delta)\|^2}{\sigma_N^2 + \frac{(1-\alpha)^2 t^2 \Delta^2}{12}}, \tag{3.15}$$

the previous comment regarding the use of $L(t, \boldsymbol{z})$ is also applied here.

Similarly to the low-SNR case, we are interested in studying the asymptotic properties of the ML cost function (e.g., the location of the global minima/maxima, etc.). If we assume for a certain channel gain t sufficiently close to t_0 the inequalities $\mathrm{HQR} \gg 1$, $\mathrm{SCR}(t) \ll 1$, and $\mathrm{TNQR}(t) \ll 1$ are satisfied, then the expectation of the cost function, calculated in App. 3.B, can be approximated as

$$\mathrm{E}\left\{L(t, \boldsymbol{z})\right\} \approx L + L \log\left(2\pi\sigma_N^2\right) + L\frac{(t_0 - t)^2\sigma_X^2 + \sigma_N^2}{\sigma_N^2}.$$

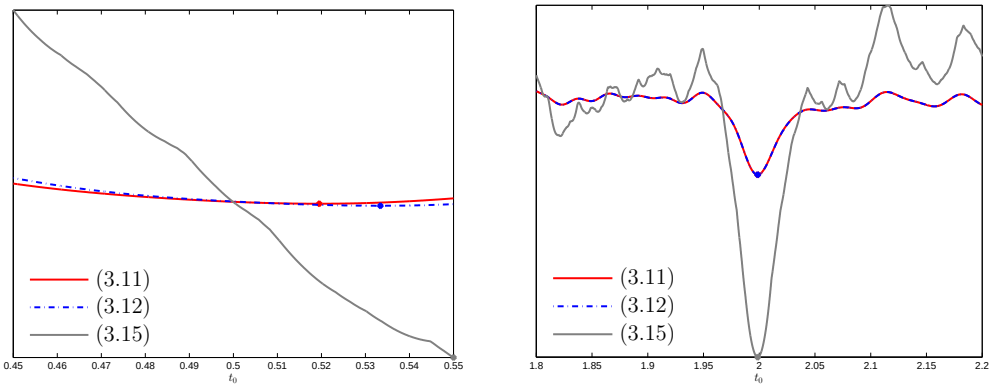
It is straightforward to verify that the third term only has a minimum with respect to t for $t > 0$ at t_0.

3.2.3 Discussion and Examples

By comparing the approximations of $L(t, \boldsymbol{z})$ given in (3.11), (3.12), and (3.15), one can notice that these three expressions share a common structure: the first and second terms are related to the envelope of the pdf of Z, and the third term takes into account the structure of the pdf of Z induced by the embedding.

As an illustrative example, Fig. 3.6 shows the representations of the proposed ML cost function approximations for WNR = 0 dB, $\alpha = 0.75$, and $L = 10^2$. The left-most case corresponds to DWR = 20 dB and $t_0 = 0.5$, that is selected as an example of a low-SNR scenario where (3.11) and (3.12) clearly show better performance than (3.15). In Fig. 3.6.b with DWR = 40 dB and $t_0 = 2$, which can be considered a high-SNR scenario, the high-SNR approximation to the ML cost function shows good accuracy; in addition, it is worth noting that the low-SNR approximations to the ML cost functions also have the minimum close to t_0, i.e., they are accurate.

Although ML estimation shows interesting properties, as those enumerated above, also it presents some drawbacks. For example, as the target functions can have several local maxima or minima, well-known optimization algorithms cannot be systematically applied; thus, the design of *ad-hoc* optimization algorithms is



(a) DWR = 20 dB and $t_0 = 0.5$. (b) DWR = 40 dB and $t_0 = 2$.

Figure 3.6: Representation of the three proposed approximations of $L(t, \boldsymbol{z})$ for WNR = 0 dB, $\alpha = 0.75$, and $L = 10^2$.

required in many cases. In addition, it is also widely assumed that MLE requires more computational resources than other estimation methods [57] (e.g., methods based on moments).

3.3 Unknown Host and Channel Noise Variances

So far we have worked on an ML-based approach, which requires the exact distribution of the host signal and the channel noise to be known at the estimator. Nevertheless, in practical scenarios these requirements must be typically relaxed, since although the kind of distribution of those signals can be reasonably assumed to be known (e.g., the channel noise is customarily modeled as Gaussian), prior knowledge of their variances may be not available. Consequently, in such case the estimator can be designed by taking into account that the estimation problem is affected by some unwanted/nuisance parameters (i.e., the host and noise variances). Following typical estimation procedures, the ML target function is optimized with respect to those unwanted parameters, and then with respect to the parameter of interest (in the current case, t) [55]. For the sake of simplicity, we will focus the analysis in this section on the high-SNR target function.

Mathematically, in a first step we optimize the high-SNR scenario pdf of $\boldsymbol{Z}$ given σ_X^2, σ_N^2, and t with respect to σ_X^2 and σ_N^2; therefore, we use the complete expression of that pdf, derived in the previous section (i.e., formula (3.14)).

Specifically, the considered ML target function is

$$L\log\left(\frac{2\pi\sigma_X^2}{\Delta^2}\right)+\frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2}+L\log\left(2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2t^2\Delta^2}{12}\right)\right)+\frac{\|(\boldsymbol{z}-t\boldsymbol{d})\mathrm{mod}(t\Delta)\|^2}{\sigma_N^2+\frac{(1-\alpha)^2t^2\Delta^2}{12}}. \tag{3.16}$$

The values of σ_X^2 and σ_N^2 optimizing that target function are $(\sigma_X^2)^* = \frac{\|\boldsymbol{z}\|^2}{Lt^2}$, and $(\sigma_N^2)^* = \frac{\|(\boldsymbol{z}-t\boldsymbol{d})\mathrm{mod}(t\Delta)\|^2}{L} - \frac{(1-\alpha)^2t^2\Delta^2}{12}$, respectively. By replacing those values in (3.16), one obtains a new target function, which no longer depends on σ_X^2, nor σ_N^2,

$$L\log\left(\frac{2\pi e\|\boldsymbol{z}\|^2}{\Delta^2Lt^2}\right)+L\log\left(\frac{2\pi e\|(\boldsymbol{z}-t\boldsymbol{d})\mathrm{mod}(t\Delta)\|^2}{L}\right);$$

it is straightforward to see that the minimization of that function with respect to t is equivalent to the minimization of

$$\frac{\|(\boldsymbol{z}-t\boldsymbol{d})\mathrm{mod}(t\Delta)\|^2}{t^2},$$

i.e., the estimator will look for that scaling factor that minimizes the Euclidean distance (normalized by the scaling factor) between observations and quantizer centroids. In other words, whenever *a priori* knowledge of the statistics of the host signal and the noise is not available, the estimation will only rely on the signal structure introduced at the transmitter, i.e., the estimator will look for that scaling factor that matches the patterns of the observations and the quantizer centroids.

For the sake of comparison, it is worth mentioning that although the PDD-based ML estimator depends on σ_X^2 and σ_N^2, when one performs the optimization of the ML target function with respect to the unwanted parameters, then the estimation algorithm described in [29] is achieved.

3.4 Spread-Transform Dirty Paper Coding Estimation

As it was discussed in the previous section, the proposed estimator when *a priori* information on the host and noise variances are not available, exploits the structure introduced by the quantizer at the embedding. However, in those cases where the TNQR(t_0) is very large, that structure will be lost. This is a well-known problem in watermarking, where one of the most extended solutions is the so-called Spread-Transform Dither Modulation (ST-DM) [8]. This technique is based on projecting the L dimensions of the considered signals onto L_{ST} ($L_{\mathrm{ST}} \leq L$) dimensions. By doing so, the ratio Δ^2/σ_N^2 is increased by a factor L/L_{ST}, obtaining a smaller TNQR(t_0), and consequently enhancing the signal structure.

Namely, the watermark is embedded in a projected version of $\boldsymbol{x}$, $\boldsymbol{x}^{\mathrm{ST}} = V^T\boldsymbol{x}$, where V is an $L \times L_{\mathrm{ST}}$ orthonormal matrix (this matrix can be easily generated, for example by applying the Gram-Schmidt orthogonalization algorithm [36] to a randomly generated matrix). Then, the watermarked signal in the projected domain is computed (similarly to (2.2)) as $\boldsymbol{y}^{\mathrm{ST}} = Q(\boldsymbol{x}^{\mathrm{ST}}) + (1-\alpha)[\boldsymbol{x}^{\mathrm{ST}} - Q(\boldsymbol{x}^{\mathrm{ST}})]$, and $\boldsymbol{y} = \boldsymbol{x} + V(\boldsymbol{y}^{\mathrm{ST}} - \boldsymbol{x}^{\mathrm{ST}})$. Note that the watermark distortion per dimension of $\boldsymbol{x}$ is now $\sigma_W^2 = \frac{L_{\mathrm{ST}}\alpha^2\Delta^2}{12L}$, so for a fixed σ_W^2 the quantization step Δ is increased with respect to the case $L = L_{\mathrm{ST}}$.

Other advantages of the proposed ST-DM-based estimation are: 1) even if the distributions of the host and the noise are not Gaussian, due to the CLT they will asymptotically converge to Gaussian for large values of L and the most extensively used designs of V (e.g., matrices generated by $\{-1, +1\}$ uniform random generators), 2) due to the small cost of projecting, and the reduction of the dimensionality of the observations fed to the ML estimator, the computational cost is reduced.

Appendix

3.A Analysis of $\mathrm{E}\{L(t,Z)\}$ for Low-SNR Scenarios

The objective function $L(t,\boldsymbol{z})$ for low-SNR, defined in Sect. 3.2.1, can be approximated as

$$L(t,\boldsymbol{z}) = \frac{\|\boldsymbol{z}\|^2}{2\left(\sigma_N^2+\sigma_X^2 t^2\right)} + \frac{L}{2}\log\left(2\pi\left(\sigma_N^2+\sigma_X^2 t^2\right)\right) - \sum_{i=1}^{L} 2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2 t^2\right)}} \cos\left(\frac{2\pi\sigma_X^2 t z_i}{\Delta\left(\sigma_N^2+\sigma_X^2 t^2\right)} - \frac{2\pi d_i}{\Delta}\right);$$

this approximation makes it possible to obtain of a closed form of its expectation, which is written as

$$\begin{aligned}
&\mathrm{E}\left\{L(t,Z)\right\} \approx \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2}\left(\int_{-\infty}^{\infty} f_{Z|T,K}^{\text{low-SNR}}(\tau|t_0,\nu)L(t,\tau)d\tau\right)d\nu \\
&= \frac{\sigma_N^2+\sigma_X^2 t_0^2}{2\left(\sigma_N^2+\sigma_X^2 t^2\right)} + \frac{1}{2}\log\left[2\pi\left(\sigma_N^2+\sigma_X^2 t^2\right)\right] \\
&\quad -2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2 t^2}{\sigma_N^2+\sigma_X^2 t^2}+\frac{4\sigma_X^2 t t_0}{\sigma_N^2+\sigma_X^2 t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2 t t_0\right)^2}{\left(\sigma_N^2+\sigma_X^2 t^2\right)^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2 t_0^2}{\sigma_N^2+\sigma_X^2 t_0^2}\right)}{\Delta^2}} \\
&= \mathrm{E}\{L_0(t,Z)\} + \mathrm{E}\{L_1(t,Z)\},
\end{aligned} \tag{3.17}$$

where in the previous expression $\mathrm{E}\{L_0(t,Z)\} = (\sigma_N^2+\sigma_X^2 t_0^2)/(2\left(\sigma_N^2+\sigma_X^2 t^2\right)) + 1/2\log\left[2\pi\left(\sigma_N^2+\sigma_X^2 t^2\right)\right]$ and $\mathrm{E}\{L_1(t,Z)\}$ denotes the remaining term.

In order to analyze the maxima/minima of $\mathrm{E}\left\{L(t,Z)\right\}$, we separately analyze $\mathrm{E}\{L_0(t,Z)\}$ and $\mathrm{E}\{L_1(t,Z)\}$. For both functions, we prove asymptotically as $\mathrm{HQR} \gg 1$, $\mathrm{SCR}(t_0) \ll 1$, $\mathrm{TNQR}(t_0) \gg 1$, and $\mathrm{TNHR}(t_0) \ll 1$ hold that they have a minimum with respect to t close to t_0, they increase for $t > t_0$, and they decrease for $t < t_0$. Therefore, it can be assured that $\mathrm{E}\left\{L(t,Z)\right\}$ only has a minimum asymptotically close to t_0.

On one hand, the first derivative with respect to t of $\mathrm{E}\{L_0(t,Z)\}$ is calculated, obtaining

$$\frac{\sigma_X^4 t(t^2-t_0^2)}{\left(\sigma_N^2+\sigma_X^2 t^2\right)^2},$$

where the previous expression only has a positive root at t_0 that corresponds to a minimum, as one can verify that this expression is positive for $t>t_0$ and negative for $t<t_0$.

On the other hand, in order to analyze $\mathrm{E}\{L_1(t,Z)\}$, $-\log(-(\cdot))$ is applied to that expression and the term $-\log(2)$ dismissed to yield

$$-\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t^2}{\sigma_N^2+\sigma_X^2t^2}+\frac{4\sigma_X^2tt_0}{\sigma_N^2+\sigma_X^2t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2tt_0\right)^2}{\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t_0^2}{\sigma_N^2+\sigma_X^2t_0^2}\right)}{\Delta^2},$$

so its first derivative with respect to t is

$$\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2\sigma_N^2t\left(\sigma_N^2+\sigma_X^2t^2\right)-12\left(\sigma_N^4\sigma_X^2t_0+\sigma_X^6t^3t_0(-t+t_0)+\sigma_N^2\sigma_X^4t\left(2t^2-t_0^2\right)\right)\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)^3}.$$

Therefore, the local maxima/minima of the previous expression with respect to t will be located at the solutions of

$$\begin{aligned}g(t)\triangleq\ &(1-\alpha)^2\Delta^2\sigma_N^2t\left(\sigma_N^2+\sigma_X^2t^2\right)\\&-12\left(\sigma_N^4\sigma_X^2t_0+\sigma_X^6t^3t_0(-t+t_0)+\sigma_N^2\sigma_X^4t\left(2t^2-t_0^2\right)\right)=0.\end{aligned}\quad(3.18)$$

Since the polynomial of the previous expression is fourth degree, by applying the fundamental theorem of algebra [62], one can guarantee that there are four complex solutions. In order to identify the solutions of (3.18), the following properties can be used

- $g(0)=-12\pi^2\sigma_N^4\sigma_X^2t_0$, and

  $$g(2\sigma_N^2/(\sigma_X^2t_0))=\frac{2\left(-48\sigma_N^6\sigma_X^4t_0^2+6\sigma_N^4\sigma_X^6t_0^4+(1-\alpha)^2\Delta^2\sigma_N^6\left(4\sigma_N^2+\sigma_X^2t_0^2\right)\right)}{\sigma_X^4t_0^3},$$

  that asymptotically takes positive values as $-48\sigma_N^6\sigma_X^4t_0^2$ can be neglected compared to $6\sigma_N^4\sigma_X^6t_0^4$ by applying $\mathrm{TNHR}(t_0)\ll 1$. There is at least a sign change of $g(t)$ in $[0,2\sigma_N^2/(\sigma_X^2t_0)]$; therefore, there is at least one root within this interval. By applying the $\mathrm{TNHR}(t_0)\ll 1$, one can guarantee that there is a minimum of $\mathrm{E}\{L_1(t,Z)\}$ asymptotically close to $t/t_0=0$, as $2\sigma_N^2/(\sigma_X^2t_0^2)$ tends to 0.

- As stated in the previous paragraph, $g(2\sigma_N^2/(\sigma_X^2t_0))$ takes positive values; and, on the other hand, $g(\sigma_N/\sigma_X)=2\sigma_N^5\left((1-\alpha)^2\Delta^2-12\sigma_X^2\right)/\sigma_X$ is negative as $(1-\alpha)^2\Delta^2$ can be neglected compared to $12\sigma_X^2$ by invoking $\mathrm{HQR}\gg 1$. So there is at least a sign change of $g(t)$ in $[2\sigma_N^2/(\sigma_X^2t_0),\sigma_N/\sigma_X]$; therefore,

there is at least a maximum of the third element of the target function within this interval. By applying $\mathrm{TNHR}(t_0) \ll 1$, a maximum is asymptotically close to $t/t_0 \to 0$. An analogous analysis can be carried out for negative values of t taking into account that

$$g(-2\sigma_N^2/(\sigma_X^2 t_0)) = -\frac{2}{\sigma_X^4 t_0^3}\left((1-\alpha)^2\Delta^2\sigma_N^6\left(4\sigma_N^2+\sigma_X^2 t_0^2\right)\right.$$
$$\left.-6\left(32\sigma_N^8\sigma_X^2+8\sigma_N^6\sigma_X^4 t_0^2-3\sigma_N^4\sigma_X^6 t_0^4\right)\right),$$

is, using $\mathrm{TNHR}(t_0) \ll 1$, negative and $g(-\sigma_N/\sigma_X) = -2(1-\alpha)^2\Delta^2\sigma_N^5/\sigma_X + 24\sigma_N^5\sigma_X$ is positive as the HQR $\gg 1$ holds.

- Finally, $g(t_0) = \sigma_N^2\left((1-\alpha)^2\Delta^2 - 12\sigma_X^2\right)t_0\left(\sigma_N^2+\sigma_X^2 t_0^2\right)$ is negative under HQR $\gg 1$, while,

$$g(t_0(1+2\sigma_N^2/(\sigma_X^2 t_0^2))) = \sigma_N^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)$$
$$\times\frac{\left(12\sigma_X^6 t_0^4+(1-\alpha)^2\Delta^2\left(8\sigma_N^4+6\sigma_N^2\sigma_X^2 t_0^2+\sigma_X^4 t_0^4\right)\right)}{\left(\sigma_X^4 t_0^3\right)},$$

is positive; therefore there is at least a root of (3.18) in $[t_0, t_0(1+2\sigma_N^2/(\sigma_X^2 t_0^2))]$ (note that with this, the four roots of (3.18) are identified). By applying $\mathrm{TNHR}(t_0) \ll 1$, one can guarantee that there is a minimum asymptotically close to t_0.

Therefore, one can guarantee that, under the four hypotheses of the low-SNR case presented in Chap. 2, there is a minimum of the third term of (3.17) asymptotically at $t/t_0 = 0$, two maxima asymptotically close to $t/t_0 = 0$ and a minimum asymptotically close to t_0. Specifically, the corresponding disjoint intervals of the roots of the third term of (3.17), arranged in increasing order of t, are: a maximum in $[-\sigma_N/\sigma_X, -2\sigma_N^2/(\sigma_X^2 t_0)]$, a minimum in $[0, 2\sigma_N^2/(\sigma_X^2 t_0)]$, a maximum in $[2\sigma_N^2/(\sigma_X^2 t_0), \sigma_N/\sigma_X]$, and a minimum $[t_0, t_0(1+2\sigma_N^2/(\sigma_X^2 t_0^2))]$. A graphical example of these four intervals containing the roots of $g(t)$ is shown in Fig. 3.7 for DWR = 20 dB, WNR = 0 dB, $\alpha = 0.5$, and $t_0 = 0.7$.

Since $\mathrm{E}\{L_0(t,Z)\}$ has only one minimum located at t_0 and $\mathrm{E}\{L_1(t,Z)\}$ has only a minimum in $[\sigma_N/\sigma_X, \infty)$, one can state that $\mathrm{E}\{L(t,Z)\}$ has only a minimum for $t > \sigma_N/\sigma_X$, which asymptotically corresponds to t_0. Moreover, one can state that, asymptotically when the four hypotheses hold, this expectation monotonically decreases for $\sigma_N/\sigma_X < t < t_0$ and monotonically increases for $t > t_0$ within this interval.

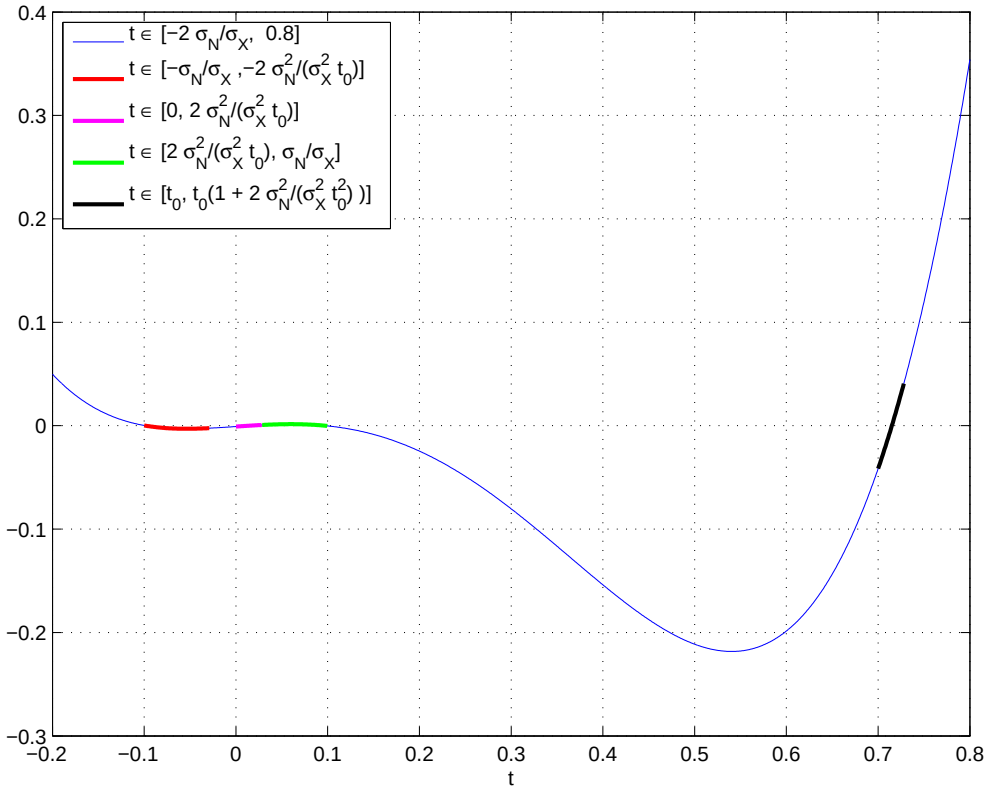


Figure 3.7: Example of $g(t)$. The corresponding intervals where the four roots are located are shown. DWR = 20 dB, WNR = 0 dB, $\alpha = 0.5$, and $t_0 = 0.7$.

3.B Expectation of the Cost Function for the high-SNR Case

The expectation of the cost function for (3.15) can be approximated by

$$\mathrm{E}\{L(t,\boldsymbol{z})\} \approx L\frac{(\sigma_X^2+\alpha^2\Delta^2/12)t_0^2+\sigma_N^2}{\sigma_X^2 t^2} + L\log\left(2\pi\left(\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12\right)\right) + L\frac{(t_0-t)^2\sigma_X^2+(t-\alpha t_0)^2\Delta^2/12+\sigma_N^2}{\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12},$$

where the assumption $L \to \infty$ is used to apply the CLT in the first and third terms of the expectation of the cost function to substitute them by power relations, i.e., the first term by $L((\sigma_X^2+\sigma_W^2)t_0^2+\sigma_N^2)$ and the third term by using the approximation (3.19) explained in App. 3.C. This expression can be further simplified obtaining

$$\mathrm{E}\{L(t,\boldsymbol{z})\} \approx L + L\log\left(2\pi\sigma_N^2\right) + L\frac{(t_0-t)^2\sigma_X^2+\sigma_N^2}{\sigma_N^2},$$

where the closeness of t and t_0 is used in addition to $\mathrm{SCR}(t) \ll 1$ to neglect $(1-\alpha)^2t^2\Delta^2/12$ and $(t-\alpha t_0)^2\Delta^2/12$ compared to σ_N^2 in the second and third

terms. In addition HQR $\gg 1$ is invoked to neglect $\alpha^2\Delta^2/12$ compared to σ_X^2 in the first term, and simultaneously use HQR $\gg 1$ and TNQR$(t_0) \ll 1$ to drop σ_N^2 compared to $\sigma_X^2 t_0^2$ also in the first term.

3.C Modulo-Lattice Reduction of the Received Signal

In this appendix we analyze the quantization of the received signal by a scaled version of the lattice used at embedding; indeed, we will consider the case where the scaling factor of the lattice, which we will denote by t, might be different from the gain of the flat fading channel the transmitted signal has gone through, i.e., t_0. Based on basic properties of the modulo-lattice reduction, it is easy to check that $(\boldsymbol{z} - t\boldsymbol{d}) \bmod (t\Lambda)$ is equal to

$$
\begin{aligned}
& \Big[t_0 \Big((1-\alpha)\boldsymbol{x} + \alpha\Big[\boldsymbol{x} - (\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big]\Big) + \boldsymbol{n} - t\boldsymbol{d}\Big] \bmod (t\Lambda) \\
= \; & \Big[t_0 \Big(\boldsymbol{x} - \alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big]\Big) + \boldsymbol{n} - t\boldsymbol{d}\Big] \bmod (t\Lambda) \\
= \; & \Bigg[(t_0 - t)\boldsymbol{x} - (t_0 - t)\alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big] + \boldsymbol{n} + t\Big(\boldsymbol{x} - \boldsymbol{d} - \alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big]\Big)\Bigg] \bmod (t\Lambda) \\
= \; & \Bigg[(t_0 - t)\boldsymbol{x} - (t_0 - t)\alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big] + \boldsymbol{n} \\
& + t\Big((\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda - \alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big]\Big)\Bigg] \bmod (t\Lambda) \\
= \; & \Bigg[(t_0 - t)\boldsymbol{x} - (t_0 - t)\alpha\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big] + \boldsymbol{n} + t(1-\alpha)\Big((\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big)\Bigg] \bmod (t\Lambda) \\
= \; & \Big[(t_0 - t)\boldsymbol{x} + (t - \alpha t_0)\Big[(\boldsymbol{x}-\boldsymbol{d}) \bmod \Lambda\Big] + \boldsymbol{n}\Big] \bmod (t\Lambda). \qquad (3.19)
\end{aligned}
$$

This sequence of equalities highlights the strong dependence of the modulo-lattice reduction of the received signal with respect to the difference between the considered scaling factor t and the real flat fading channel scaling factor t_0, as $(t_0 - t)$ multiplies to $\boldsymbol{x}$, whose variance is much larger that the second moment of the considered lattice. Therefore, if the distribution of the host signal is smooth (e.g., Gaussian) even for small values of $|t_0 - t|$ the distribution of the quantization error will quickly converge to the uniform distribution in the fundamental Voronoi region of Λ. Indeed, this behavior is the ultimate reason for the decoding issues analyzed in [4] and [51]; however, in this thesis this rapid evolution of the quantization error distribution with respect to $t_0 - t$ is an asset that enables it improve the estimator performance.

Chapter 4

Theoretical Analysis of DPCE

In this chapter, a theoretical analysis is carried out in order to obtain the fundamental limits of DPCE and also to understand how its asymptotic performance depends on the scheme parameters (e.g., DWR, WNR, α, etc.).

In Sect. 4.1, the theoretical analysis is conducted following an estimation-theoretic approach. Specifically, as in many estimation works (e.g., [52], [24], etc.), the CRB is studied, since it determines a lower bound on the variance of the error of any unbiased estimator of t_0. Approximations of the CRB are introduced for the proposed approximate pdfs of Z, i.e., for the low-SNR and high-SNR cases. It is worth pointing out that, as indicated above, the ML estimator is asymptotically efficient when L tends to infinity; therefore, by comparing the performance of the ML estimator with the CRB, one can evaluate the efficiency of our proposed estimators (this performance comparison is presented exhaustively in Chapter 5).

In Sect. 4.2, we analyze the scaling estimation problem from an information-theoretic perspective. Specifically, we compare the performance of the proposed scheme with that achieved by Add-SS/SIT in terms of the mutual information between between Z and T given the secret key K focusing our analysis, for the sake of simplicity, on the single letter (i.e., $L = 1$) case. Since mutual information measures the information that Z contains about T, the rationale supporting the use of this metric is that the larger the mutual information, the more information available to estimate the gain. Although in the rest of the work t is deterministic, throughout that section T is supposed to follow a Rayleigh distribution with mode σ_T (note that this distribution is widely used for the multiplicative part of flat-fading channels [39]).

4.1 Estimation Theory: Cramér-Rao Bound

Given an unbiased t_0 estimator $\hat{t}_0(\boldsymbol{z})$, $\mathrm{Var}\left\{\hat{t}_0(\boldsymbol{z})\right\} \geq 1/I(t_0)$ [34]. $I(t_0)$ is known as the Fisher information and its inverse is called the Cramér-Rao Bound (CRB). The CRB can be calculated as

$$\begin{aligned}\mathrm{CRB} = \frac{1}{I(t_0)} &= \left(-\mathrm{E}\left\{\frac{\partial^2}{\partial t_0^2}\log f_{\boldsymbol{Z}|T,K}(\boldsymbol{z}|t_0,\boldsymbol{d})\right\}\right)^{-1} \\ &= \left(-L\mathrm{E}\left\{\frac{\partial^2}{\partial t_0^2}\log f_{Z|T,K}(z|t_0,d)\right\}\right)^{-1}; \end{aligned} \tag{4.1}$$

the third equality of the previous expression results of taking into account that the elements of $\boldsymbol{z}$ and $\boldsymbol{d}$ are independently distributed (the elements of $\boldsymbol{d}$ are also identically distributed).

As indicated above, the actual pdf of Z is difficult to handle mathematically and, thus, we have proposed three approximations of its pdf (two for low-SNR scenarios and one for high-SNR scenarios). For the same reasons, we calculate the CRB using these pdf approximations.

4.1.1 Low-SNR Case

The Fisher information is calculated using the approximation of the pdf of Z $f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)$ (3.8) obtained in Sect. 3.1.1 as

$$I(t_0) \approx -L\mathrm{E}_D\left\{\mathrm{E}_{Z|T,K}\left\{\frac{\partial^2}{\partial t_0^2}\log f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)\right\}\right\};$$

in the previous expression $\mathrm{E}_D\{\cdot\}$ denotes the expectation with respect to the distribution of the dither sequence, and $\mathrm{E}_{Z|T,K}\{\cdot\}$ stands for the expectation with respect to Z for a given gain t_0 and dither d. From (3.8)

$$\begin{aligned}\log\left(f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)\right) &= \log\left(\frac{e^{-\frac{z^2}{2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\right) \\ &+ \log\left(1+2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2 t_0^2}{12}\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\cos\left(\frac{2\pi\sigma_X^2 t_0 z}{\Delta(\sigma_N^2+\sigma_X^2 t_0^2)}-\frac{2\pi d}{\Delta}\right)\right) \\ &= \log\left(f_{Z|T,K}^{\text{low-SNR},a}(z|t_0,d)\right)+\log\left(f_{Z|T,K}^{\text{low-SNR},b}(z|t_0,d)\right).\end{aligned}$$

The previous expression is divided into two components: one corresponding to the logarithm of the scaled host and noise distribution $\log\left(f_{Z|T,K}^{\text{low-SNR},a}(z|t_0,d)\right)$ and

the other component corresponding to the logarithm of the effect of the DPCE technique $\log\left(f_{Z|T,K}^{\text{low-SNR},b}(z|t_0,d)\right)$. For the sake of simplicity, the calculation of the CRB based on that approximated pdf is also split in these two halves.

4.1.1.1 $\log\left(f_{Z|T,K}^{\textbf{low-SNR},a}(z|t_0,d)\right)$

To calculate the part of the Fisher information corresponding to $\log\left(f_{Z|T,K}^{\text{low-SNR},a}(z|t_0,d)\right)$, we take its second derivative with respect to the scaling factor t_0 is required, so

$$\begin{aligned}\frac{\partial^2 \log\left(f_{Z|T,K}^{\text{low-SNR},a}(z|t_0,d)\right)}{\partial t_0^2} &= \frac{\partial^2 \log\left(\frac{e^{-\frac{z^2}{2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\right)}{\partial t_0^2} \\ &= \frac{\sigma_X^2\left(-\sigma_N^4+\sigma_X^4 t_0^4+\left(\sigma_N^2-3\sigma_X^2 t_0^2\right)z^2\right)}{\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^3}. \end{aligned} \tag{4.2}$$

In order to obtain the expectation of the previous expression, we calculate $\mathrm{E}\left\{Z^2|T=t_0,K=d\right\}$ using $f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)$ (3.8) as an approximation of the pdf of Z, yielding

$$\begin{aligned}&\int_{-\infty}^{\infty}\tau^2 f_{Z|T,K}^{\text{low-SNR}}(\tau|t_0,d)d\tau = \sigma_N^2+\sigma_X^2 t_0^2 \\ &+\frac{2e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+12\sigma_X^2\right)t_0^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}\left(-4\pi^2\sigma_X^4 t_0^2+\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)\right)\cos\left(\frac{2\pi d}{\Delta}\right)}{\Delta^2} \\ &\approx \sigma_N^2+\sigma_X^2 t_0^2 \\ &+\frac{2e^{-\frac{2\pi^2\sigma_X^2}{\Delta^2}}\left(-4\pi^2\sigma_X^4 t_0^2+\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)\right)\cos\left(\frac{2\pi d}{\Delta}\right)}{\Delta^2};\end{aligned}$$

where, in the previous expression HQR $\gg 1$ was used to simplify the argument of the exponential function. By using this result in (4.2), one approximates the expectation of (4.2),

$$\begin{aligned}&\int_{-\infty}^{\infty}\frac{\partial^2\log\left(f_{Z|T,K}^{\text{low-SNR},a}(\tau|t_0,d)\right)}{\partial t_0^2}f_{Z|T,K}^{\text{low-SNR}}(\tau|t_0,d)d\tau \approx \frac{2\sigma_X^2}{\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^3}\left(-\sigma_X^2 t_0^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)\right. \\ &\left.+\frac{e^{-\frac{2\pi^2\sigma_X^2}{\Delta^2}}\left(\sigma_N^2-3\sigma_X^2 t_0^2\right)\left(-4\pi^2\sigma_X^4 t_0^2+\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)\right)\cos\left(\frac{2\pi d}{\Delta}\right)}{\Delta^2}\right),\end{aligned}$$

and the expectation with respect to D of the previous expression can be straightforwardly calculated, obtaining

$$\int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta} \left(\int_{-\infty}^{\infty} \frac{\partial^2 \log\left(f_{Z|T,K}^{\text{low-SNR},a}(\tau|t_0,\kappa) \right)}{\partial t_0^2} f_{Z|T,K}^{\text{low-SNR}}(\tau|t_0,\kappa) d\tau \right) d\kappa$$
$$\approx -\frac{2\sigma_X^4 t_0^2}{(\sigma_N^2 + \sigma_X^2 t_0^2)^2}. \tag{4.3}$$

4.1.1.2 $\log\left(f_{Z|T,K}^{\textbf{low-SNR},b}(z|t_0,d) \right)$

The verification of the first three hypotheses of the low-SNR case (i.e., HQR $\gg 1$, SCR$(t_0) \ll 1$, TNQR$(t_0) \gg 1$) entails a large negative value of the argument of the exponential of $\log\left(f_{Z|T,K}^{\text{low-SNR},b}(z|t_0,d) \right)$; thus, as one can approximate $\log(1+h) \approx h$ for $|h| \ll 1$, $\log\left(f_{Z|T,K}^{\text{low-SNR},b}(z|t_0,d) \right)$ can be approximated by

$$\log\left(f_{Z|T,K}^{\text{low-SNR},b}(z|t_0,d) \right) \approx 2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+(1-\alpha)^2\Delta^2 t_0^2/12\right)}{\Delta^2\left(\sigma_N^2+\left(\sigma_X^2\right)t_0^2\right)}} \cos\left(\frac{2\pi\sigma_X^2 t_0 z}{\Delta\left(\sigma_N^2+\sigma_X^2 t_0^2\right)} - \frac{2\pi d}{\Delta} \right). \tag{4.4}$$

The expectation with respect to D and Z of the second derivative of (4.4) with respect to t_0 is calculated in App. 4.B, and can be written as

$$\int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta} \left(\int_{-\infty}^{\infty} \frac{\partial^2 \log\left(f_{Z|T,K}^{\text{low-SNR},b}(\tau|t_0,\kappa) \right)}{\partial t_0^2} f_{Z|T,K}^{\text{low-SNR}}(\tau|t_0,\kappa) d\tau \right) d\kappa \tag{4.5}$$
$$\approx \frac{e^{-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2 t_0^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}} 2\pi^2\sigma_X^2}{9\Delta^4\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^4} \Bigg((1-\alpha)^2\Delta^4\sigma_N^2 \Big[\left(\pi^2(1-\alpha)^2+6\right)\sigma_N^2\sigma_X^2 t_0^2 - 3\sigma_N^4$$
$$+9\sigma_X^4 t_0^4 \Big] - 12\Delta^2\sigma_X^4 t_0^2 \left(\left(2\pi^2(1-\alpha)^2+3\right)\sigma_N^4 + 6\sigma_N^2\sigma_X^2 t_0^2 + 3\sigma_X^4 t_0^4 \right) + 144\pi^2\sigma_N^4\sigma_X^6 t_0^2 \Bigg).$$

4.1.1.3 Complete CRB Approximation

The expectation with respect to Z and D of the second derivative with respect to t_0 of the logarithm of $f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)$ can be obtained by adding (4.3) to (4.5). The corresponding Fisher information approximation can be simplified by taking

into account that $\mathrm{HQR} \gg 1$, $\mathrm{SCR}(t_0) \ll 1$, and $\mathrm{TNQR}(t_0) \gg 1$, obtaining

$$\begin{aligned} I_{\text{low-SNR}}(t_0) &\approx -L\mathrm{E}\left\{\frac{\partial^2}{\partial t_0^2}\log f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)\right\} \\ &\approx L\frac{2\sigma_X^4 t_0^2}{\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^2} + L\frac{2e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2 t_0^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}}}{3\Delta^4\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^4} \\ &\times\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^4\sigma_N^6 - 48\pi^2\sigma_N^4\sigma_X^6 t_0^2 + 12\Delta^2\left(2\sigma_N^2\sigma_X^6 t_0^4+\sigma_X^8 t_0^6\right)\right) \\ &\triangleq I_{\text{low-SNR}}^a(t_0) + I_{\text{low-SNR}}^b(t_0). \end{aligned} \tag{4.6}$$

In the previous expression, $I_{\text{low-SNR}}(t_0)$ is divided into two parts. One part

$$I_{\text{low-SNR}}^a(t_0) \triangleq L\frac{2\sigma_X^4 t_0^2}{\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^2},$$

corresponds to the to the Fisher information of the case when the involved signals are approximately Gaussian distributed, (i.e., the inverse of the CRB (4.40) presented in App. 4.A). The other part $I_{\text{low-SNR}}^b(t_0)$ corresponds to the Fisher information due to the structure of the signal produced by the embedding algorithm.

4.1.1.3.1 CRB for $f_{Z|T,K}^{\text{low-SNR},2}(z|t_0,d)$

The CRB corresponding to $f_{Z|T,K}^{\text{low-SNR},2}(z|t_0,d)$ (3.9) is calculated based on the $f_{Z|T,K}^{\text{low-SNR}}(z|t_0,d)$ CRB approximation obtained in the previous section and additionally considering $\mathrm{TNHR}(t_0) \ll 1$(i.e., by neglecting σ_N^2 compared to $\sigma_X^2 t_0^2$), obtaining

$$\begin{aligned} I_{\text{low-SNR},2}(t_0) &\approx -L\mathrm{E}\left\{\frac{\partial^2}{\partial t_0^2}\log f_{Z|T,K}^{\text{low-SNR},2}(z|t_0,d)\right\} \\ &\approx L\left(\frac{2}{t_0^2} + \frac{8e^{-\frac{1}{3}\pi^2\left((1-\alpha)^2+\frac{12\sigma_N^2}{\Delta^2 t_0^2}\right)}\pi^2\left(-4\pi^2\sigma_N^4+\Delta^2\sigma_X^2 t_0^4\right)}{\Delta^4 t_0^6}\right). \end{aligned} \tag{4.7}$$

As in (4.6), the first element of the previous expression is the corresponding Fisher information of the variance-based estimator of t_0, while the second term corresponds to the Fisher information part of the estimation due to the signal embedding.

4.1.2 High-SNR Case

For high-SNR scenarios, the minus logarithm of (4.1) is calculated using the pdf approximation of Z (3.10) considering, in the same way as (3.19) was developed in App 3.C, that $(z-td) \mod (t\Delta) = [(t_0-t)x + (t-\alpha t_0)\,[(x-d) \mod \Delta] + n] \mod (t\Delta)$, as

$$\begin{aligned}
&-\log\left(f^{\text{high-SNR}}_{Z|T,K}(z|t_0,d)\right)\\
&=\frac{1}{2}\left[\frac{\left([(t_0-t)x+(t-\alpha t_0)\,[(x-d) \mod \Delta]+n] \mod (t\Delta)\right)^2}{\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12}\right.\\
&\left.+\log\left(2\pi\left(\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12\right)\right)+\frac{((x+w)t_0+n)^2}{\sigma_X^2t^2}+\log\left(2\pi\sigma_X^2\right)-2\log(\Delta)\right].
\end{aligned} \tag{4.8}$$

In order to obtain the Fisher information, the second derivative of the previous expression is calculated with respect to t and subsequently evaluated at t_0. Specifically, jointly assuming that t and t_0 are close and thus $\text{TNQR}(t) \ll 1$, the modulo-$t\Delta$ reduction of the first term in (4.8) can be neglected, and thus that expression can be written as

$$\begin{aligned}
&-\log\left(f^{\text{high-SNR}}_{Z|T,K}(z|t_0,d)\right)\approx\frac{1}{2}\left[\frac{((t_0-t)x+(t-\alpha t_0)\,[(x-d) \mod \Delta]+n)^2}{\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12}\right.\\
&\left.+\log\left(2\pi\left(\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12\right)\right)+\frac{((x+w)t_0+n)^2}{\sigma_X^2t^2}+\log\left(2\pi\sigma_X^2\right)-2\log(\Delta)\right].
\end{aligned}$$

The first derivative of the previous expression with respect to t is

$$\begin{aligned}
&-\frac{\partial\log\left(f^{\text{high-SNR}}_{Z|T,K}(z|t_0,d)\right)}{\partial t}\\
&\approx\frac{1}{2}\Big(-\frac{2\left((t_0-t)x+(t-\alpha t_0)w+n\right)\left(\sigma_N^2(x-w)+(1-\alpha)^2\Delta^2/12t(n+t_0(x-\alpha w))\right)}{\left(\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12\right)^2}\\
&+\frac{2(1-\alpha)^2\Delta^2/12t}{\sigma_N^2+(1-\alpha)^2t^2\Delta^2/12}-\frac{2\left((x+w)t_0+n\right)^2}{\sigma_X^2t^3}\Big),
\end{aligned}$$

and the second derivative with respect to t is

$$\begin{aligned}
&-\frac{\partial^2\log\left(f^{\text{high-SNR}}_{Z|T,K}(z|t_0,d)\right)}{\partial t^2}\\
&\approx\frac{1}{2}\Bigg(\Bigg(\sigma_N^4(w-x)^2+(1-\alpha)^4\left(\frac{\Delta^2}{12}\right)^2t^2\left((x-\alpha w)t_0+n\right)\cdot\left((3t_0-2t)x-(3t_0\alpha-2t)w+3n\right)\\
&-(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2\Big[3t^2(w-x)^2+n^2-6tt_0(w-x)(\alpha w-x)+t_0^2(x-\alpha w)^2\\
&+2n(3t(w-x)+t_0(x-\alpha w))\Big]\Bigg)\cdot 2\cdot\left(\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t^2\right)^{-3}\\
&+\frac{2(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2-2(1-\alpha)^4\left(\frac{\Delta^2}{12}\right)^2t^2}{(\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t^2)^2}+\frac{6((x+w)t_0+n)^2}{\sigma_X^2t^4}\Bigg).
\end{aligned}$$

Calculating the expectation of the previous expression with respect to X, W, and N, one obtains

$$\mathrm{E}\left\{-\frac{\partial^2 \log\left(f_{Z|T,K}^{\text{high-SNR}}(z|t_0,d)\right)}{\partial t^2}\right\}\approx\Bigg(\sigma_N^4\left(\sigma_X^2+\alpha^2\tfrac{\Delta^2}{12}\right)+(1-\alpha)^4\left(\tfrac{\Delta^2}{12}\right)^2t^2\left(t_0(3t_0-2t)\sigma_X^2+t_0(3\alpha t_0-2t)\alpha^3\tfrac{\Delta^2}{12}+3\sigma_N^2\right)$$

$$-(1-\alpha)^2\tfrac{\Delta^2}{12}\sigma_N^2\left[3t^2\left(\sigma_X^2+\alpha^2\tfrac{\Delta^2}{12}\right)+\sigma_N^2-6tt_0\left(\sigma_X^2+\alpha^3\tfrac{\Delta^2}{12}\right)+t_0^2\left(\sigma_X^2+\alpha^4\tfrac{\Delta^2}{12}\right)\right]\Bigg)$$

$$\times\left(\sigma_N^2+(1-\alpha)^2\tfrac{\Delta^2}{12}t^2\right)^{-3}+\frac{(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2-(1-\alpha)^4\left(\frac{\Delta^2}{12}\right)^2t^2}{\left(\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t^2\right)^2}+\frac{3\left(\left(\sigma_X^2+\alpha^2\frac{\Delta^2}{12}\right)t_0^2+\sigma_N^2\right)}{\sigma_X^2t^4};$$

by evaluating this expression at $t=t_0$ and multiplying the result by L, the Fisher information approximation is obtained

$$\begin{aligned}
I'_{\text{high-SNR}}(t_0) &\approx L\Bigg(\left(\sigma_N^4\left(\sigma_X^2+\alpha^2\frac{\Delta^2}{12}\right)+(1-\alpha)^4\left(\frac{\Delta^2}{12}\right)^2t_0^2\left(t_0^2\sigma_X^2+t_0^2(3\alpha-2)\alpha^3\frac{\Delta^2}{12}+3\sigma_N^2\right)\right.\\
&-(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2\left[3t_0^2\left(\sigma_X^2+\alpha^2\frac{\Delta^2}{12}\right)+\sigma_N^2-6t_0^2\left(\sigma_X^2+\alpha^3\frac{\Delta^2}{12}\right)+t_0^2\left(\sigma_X^2+\alpha^4\frac{\Delta^2}{12}\right)\right]\Big)\\
&\times\left(\sigma_N^2+(1-\alpha)^2t_0^2\frac{\Delta^2}{12}\right)^{-3}+\frac{(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2-(1-\alpha)^4\left(\frac{\Delta^2}{12}\right)^2t_0^2}{\left(\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t_0^2\right)^2}+\frac{3\left(\left(\sigma_X^2+\alpha^2\frac{\Delta^2}{12}\right)t_0^2+\sigma_N^2\right)}{\sigma_X^2t_0^4}\Bigg)\\
&= L\left(I'^{,a}_{\text{high-SNR}}(t_0)+I'^{,b}_{\text{high-SNR}}(t_0)+I'^{,c}_{\text{high-SNR}}(t_0)\right).
\end{aligned}\tag{4.9}$$

This formula is studied in order to get a larger insight. The term $I'^{,a}_{\text{high-SNR}}(t_0)$ is simplified by taking into account that $\text{HQR}\gg 1$ and $\text{TNQR}(t_0)\ll 1$ as

$$\begin{aligned}
I'^{,a}_{\text{high-SNR}}(t_0) &\approx \frac{L\left(\sigma_X^2\left(\sigma_N^4+(1-\alpha)^4(\frac{\Delta^2}{12})^2t_0^4+2(1-\alpha)^2\frac{\Delta^2}{12}\sigma_N^2t_0^2\right)\right)}{\left(\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t_0^2\right)^3}\\
&= \frac{L\sigma_X^2}{\sigma_N^2+(1-\alpha)^2\frac{\Delta^2}{12}t_0^2},
\end{aligned}\tag{4.10}$$

going to infinity under the verification of the three hypotheses of the high-SNR case (i.e., $\text{HQR}\gg 1$, $\text{SCR}(t_0)\ll 1$, $\text{TNQR}(t_0)\ll 1$). The term $I'^{,b}_{\text{high-SNR}}(t_0)$ in (4.9)

$$I'^{,b}_{\text{high-SNR}}(t_0)\triangleq L\frac{(1-\alpha)^2\Delta^2/12\sigma_N^2-(1-\alpha)^4(\Delta^2/12)^2t_0^2}{(\sigma_N^2+(1-\alpha)^2\Delta^2/12t_0^2)^2},$$

it can be lower-bounded as

$$\begin{aligned}
L\frac{(1-\alpha)^2\Delta^2/12\sigma_N^2-(1-\alpha)^4(\Delta^2/12)^2t_0^2}{(\sigma_N^2+(1-\alpha)^2\Delta^2/12t_0^2)^2} &\geq L\frac{-(1-\alpha)^4(\Delta^2/12)^2t_0^2}{\left(\sigma_N^2+(1-\alpha)^2\Delta^2/12t_0^2\right)^2}\\
&\geq L\frac{-(1-\alpha)^4(\Delta^2/12)^2t_0^2}{\left((1-\alpha)^2\Delta^2/12t_0^2\right)^2}=\frac{-L}{t_0^2};
\end{aligned}$$

similarly, it can be upper-bounded as

$$L\frac{(1-\alpha)^2\Delta^2/12\sigma_N^2-(1-\alpha)^4(\Delta^2/12)^2t_0^2}{(\sigma_N^2+(1-\alpha)^2\Delta^2/12t_0^2)^2} \leq L\frac{(1-\alpha)^2\Delta^2/12\sigma_N^2}{(\sigma_N^2+(1-\alpha)^2\Delta^2/12t_0^2)^2} \leq L\frac{(1-\alpha)^2\Delta^2/12\sigma_N^2}{2(1-\alpha)^2\Delta^2/12\sigma_N^2t_0^2}=\frac{L}{2t_0^2}.$$

Finally, based on the verification of HQR $\gg 1$ and TNQR$(t_0) \ll 1$, the remaining term of (4.9) can be approximated as

$$I'^{,c}_{\text{high-SNR}}(t_0) \triangleq L\frac{3\left(\left(\sigma_X^2+\alpha^2\frac{\Delta^2}{12}\right)t_0^2+\sigma_N^2\right)}{\sigma_X^2t_0^4} \approx \frac{3L}{t_0^2}.$$

Therefore, under the hypotheses of this scenario and since $t_0 > 0$, the Fisher information will asymptotically converge to (4.10) as HQR $\gg 1$, SCR$(t_0) \ll 1$, and TNQR$(t_0) \ll 1$ due to the fact that the $I'^{,b}_{\text{high-SNR}}(t_0)$ and $I'^{,c}_{\text{high-SNR}}(t_0)$ can be neglected comparatively,

$$I_{\text{high-SNR}}(t_0) \approx \frac{L\sigma_X^2}{\sigma_N^2+\frac{(1-\alpha)^2}{\alpha^2}\sigma_W^2t_0^2}. \tag{4.11}$$

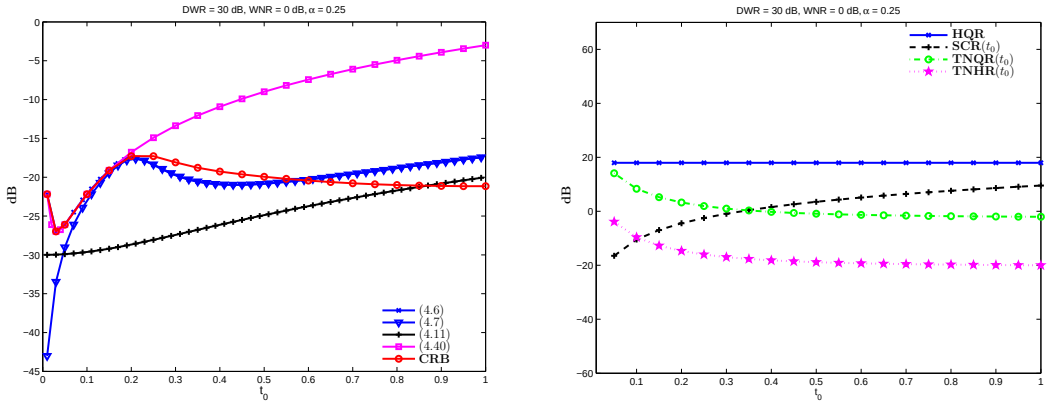
For a given σ_X^2 and σ_N^2, the previous expression is maximized with respect to α by setting $\alpha = 1$, i.e., removing the interfering effect of the watermark in the estimation. In this case, the obtained approximation of the Fisher information is LDNR, which corresponds to the Fisher information approximation of the case when $\boldsymbol{y}$ is known by the estimator. However, it is worth noting that $I_{\text{high-SNR}}(t_0) \approx L$DNR is also obtained as long as SCR$(t_0) \ll 1$, regardless the value of α.

4.1.3 Numerical Results

In the left pane of Figs. 4.1 and 4.2, the CRB vs. t_0 curves using the CRB approximations for low-SNR and high-SNR cases introduced in this thesis are shown; while in the right pane, the values taken by HQR, SCR(t_0), TNQR(t_0), and TNHR(t_0) are represented. In addition, the CRBs obtained numerically and using the variance-based estimator are also depicted.

By the analysis of the evolution of the CRB curves with respect to t_0 in these figures, three different intervals of t_0 can be differentiated in relationship with the pdf of Z:

- For very low values of t_0, the channel noise prevails over the scaled host, i.e., $\sigma_N^2 > t_0^2\sigma_X^2$; therefore, the channel noise severely interferes in the estimation. Since t_0 reduces the ratio $\sigma_N^2/(t_0^2\sigma_X^2)$ and, thus, such interference, the CRB decreases with t_0.



(a) CRB (b) Hypotheses

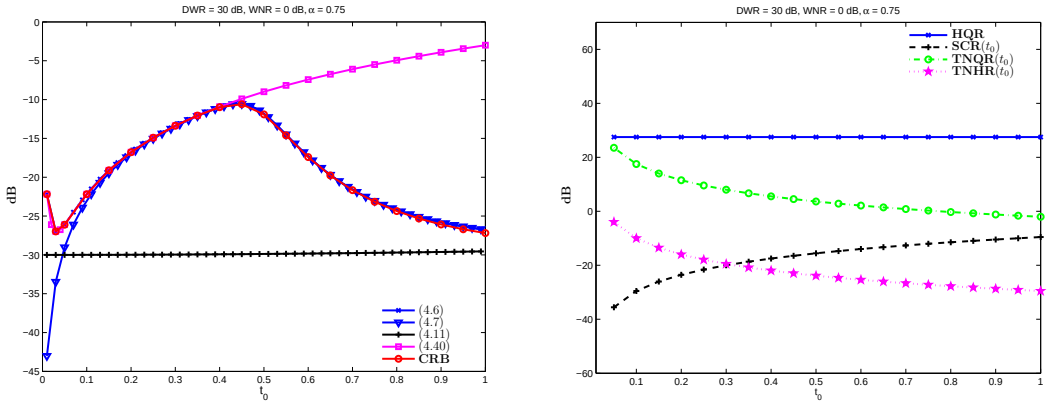
Figure 4.1: In (a), CRB vs. t_0 curves with our CRB approximations for low-SNR cases (the inverse of (4.6) and for $\text{TNHR}(t_0) \ll 1$ (4.7)) and for high-SNR cases (the inverse of (4.11)) are shown. In addition, the curves for the t_0 variance-based estimator (4.40) calculated in App. 4.A, and the CRB numerically calculated (CRB) are also depicted. DWR = 30 dB, WNR = 0 dB, and $\alpha = 0.25$. In (b), the corresponding values of the hypotheses are shown.

- For larger values of t_0, the scaled host dominates the estimation; however, there is no structure in the pdf of Z because the power of the channel noise is much larger than the power of the scaled watermark; therefore, the estimation approximately only depends on the value t_0 and the CRB increases with it.

- Then, the CRB reaches a local maximum and starts decreasing with t_0 when the induced structure of the pdf of Z appears to improve the estimation.

- Finally, after the CRB gets a local minimum, its value will increase if the effect of the modularization can not be discarded.

For low values of t_0, the CRB curves of our approximations using $f_{Z|T,K}^{\text{low-SNR}}(z|t,d)$ almost perfectly match the numerically obtained CRB curves as $\text{HQR} \gg 1$, $\text{SCR}(t_0) \ll 1$, and $\text{TNQR}(t_0) \gg 1$ are approximately verified for small values of t_0 ($t_0 \in [0, 0.2]$ for $\alpha = 0.25$ and $t_0 \in [0, 0.75]$ for $\alpha = 0.75$). For the CRB curve based on the approximation $f_{Z|T,K}^{\text{low-SNR},2}(z|t,d)$, the width of these intervals, where the approximation is tight, is reduced in its lower endpoint due to the required fulfillment of $\text{TNHR}(t_0) \ll 1$; specifically, for both $\alpha = 0.25$ and $\alpha = 0.75$ cases $\text{TNHR}(t_0) \ll 1$ holds approximately for $t_0 \geq 0.1$ (note that $\text{TNHR}(t_0)$ decreases with t_0). It is worth pointing out that in these examples, if t_0 takes low values, the CRB of the DPCE case will tend to the CRB of the variance-based estimator; this is reflected in the proposed CRB approximations for the low-SNR case as the second term in both approximations (4.6) and (4.7)



(a) CRB (b) Hypotheses

Figure 4.2: In (a), CRB vs. t_0 curves with our CRB approximations for low-SNR cases (the inverse of (4.6) and for $\text{TNHR}(t_0) \ll 1$ (4.7)) and for high-SNR cases (the inverse of (4.11)) are shown. In addition, the curves for the t_0 variance-based estimator (4.40) calculated in App. 4.A, and the CRB numerically calculated (CRB) are also depicted. DWR = 30 dB, WNR = 0 dB, and $\alpha = 0.75$. In (b), the corresponding values of the hypotheses are shown.

tends to zero, just remaining the part due to the variance-based estimator of t_0 that is calculated in App. 4.A.

For large values of t_0 and $\alpha = 0.25$, the CRB approximation $1/I_{\text{high-SNR}}(t_0)$ is closer to the numerically obtained curve of the CRB than $1/I_{\text{low-SNR}}(t_0)$ and $1/I_{\text{low-SNR},2}(t_0)$. For the $\alpha = 0.75$ scenario, the low-SNR and the high-SNR CRB approximations approximately match the actual CRB curve for large values of t_0; however, as $\text{TNQR}(t_0)$ decreases with t_0 for larger values of t_0, the high-SNR CRB approximation gets tighter than the other approximations.

In Fig. 4.3, the CRB curves and our approximations are shown as a function of t_0 for different values of DWR. As expected, whenever $\text{TNQR}(t_0)$ takes large values and, thus, there is no structure in the pdf of Z (this does not occurs for $t_0 \leq 0.1$ for DWR = 20 dB), the CRB is approximately independent of the value of the DWR. In addition, the structure appears almost for the same value of t_0 (i.e., around $t_0 = 0.3$); however, the achieved values of CRB for larger t_0 depend on DWR as one can deduce from om the expression of the high-SNR approximation to the Fisher information in (4.11): the larger the value of the DNR, the smaller the CRB. It is worth noting, accordingly to the the verification of the hypotheses, that for large values of t_0, the curves for the high-SNR case are better approximations than the low-SNR ones.

Fig. 4.4 is similar to Fig 4.3, but fixing the value of the DWR to 40 dB, modifying the value of the WNR $= -3, 0, 3$ dB, and $\alpha = 1$. As above, if there is no structure in the distribution of Z, for low values of t_0, then the CRB will

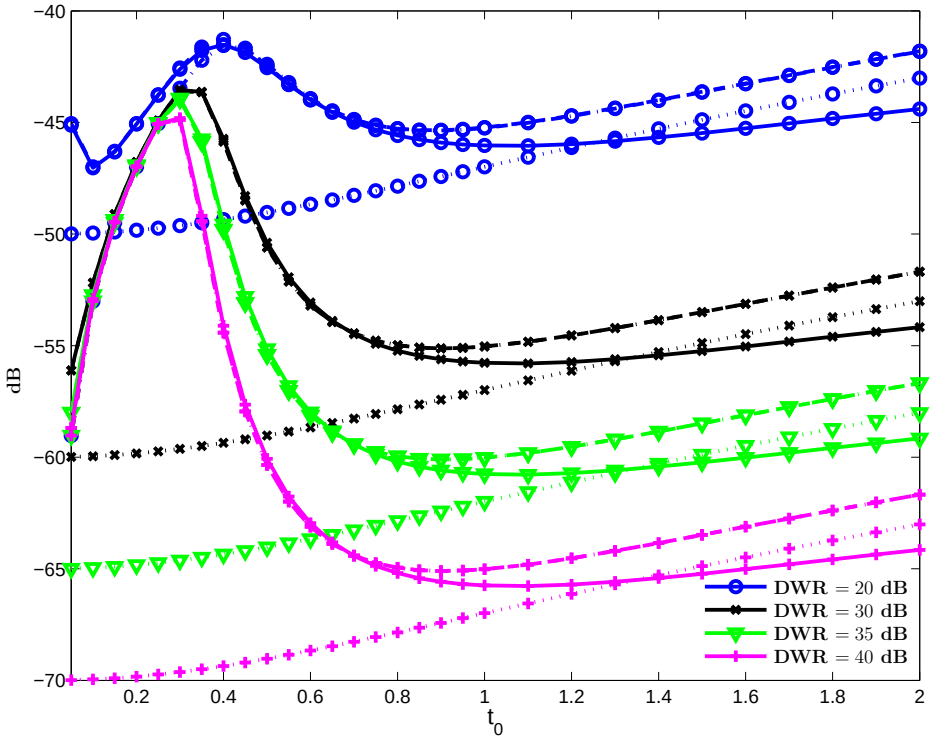


Figure 4.3: CRB vs. t_0 numerically obtained (solid lines), using the low-SNR CRB approximations (dashed lines for the inverse of (4.6) and dashdot lines for the inverse of (4.7)), and the high-SNR CRB approximation (dotted lines based on (4.11)). WNR = 0 dB , $\alpha = 0.5$, and $L = 10^3$.

not depend on the value of WNR. The value of t_0 for which there is structure in the pdf of Z depends on the WNR: the larger WNR, the smaller the required value of t_0. For large values of t_0, the CRB decreases with WNR as one can easily deduce from (4.11). Note that in comparison with the previous figure, the CRB asymptotically reaches the minimum for $t_0 \to \infty$ (approximately, $1/(L\text{DNR})$); while in the previous figure, after reaching a local minimum, the CRB increases with t_0 again. This is due to the fact that in this case there is no self-noise ($\alpha = 1$) while in Fig. 4.3 there is, which interferes in the estimation in addition to the channel noise.

The curves of the CRB as a function of the value of t_0 are shown in Fig. 4.5 for DWR = 30 dB, WNR = 0 dB, $L = 10^3$, and $\alpha = 0.5, 0.75, 1, \alpha_{\text{Costa}}$. In this figure only $1/I_{\text{low-SNR},2}(t_0)$ is depicted for the low-SNR case (i.e., $1/I_{\text{low-SNR}}(t_0)$ is not shown).

From the analysis of the numerical CRB with respect to α, one can conclude that the structure of the pdf of Z arises for larger values of t_0 as the value of α is increased, and also achieves a larger asymptotic precision whenever t_0 takes

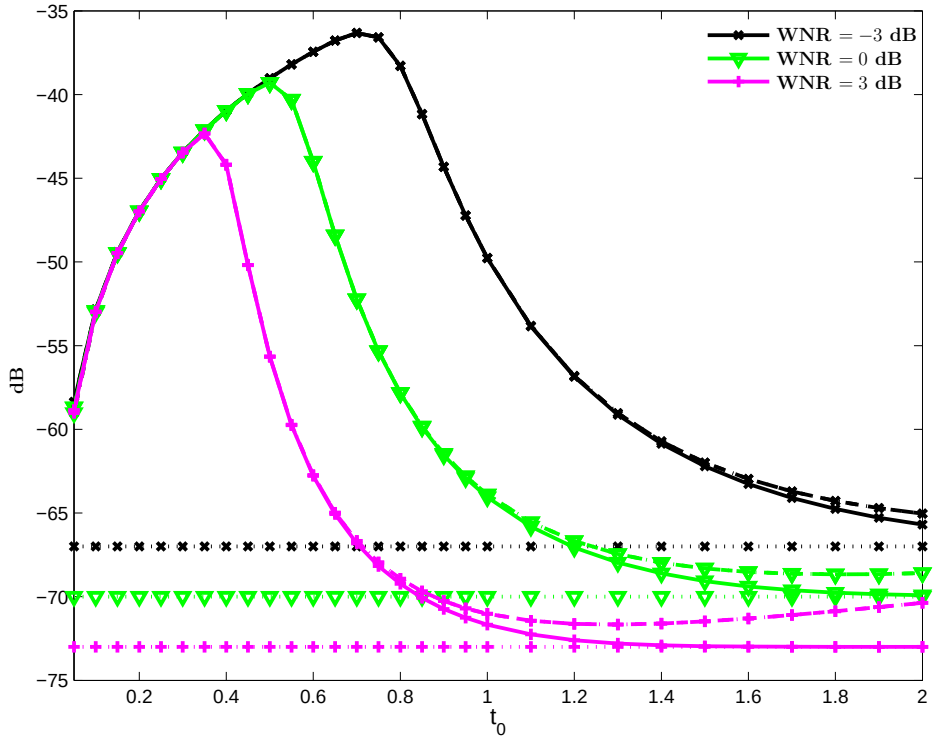


Figure 4.4: CRB vs. t_0 numerically obtained (solid lines), using the low-SNR CRB approximations (dashed lines using (4.6) and dashdot lines for (4.7)), and the high-SNR CRB approximation (dotted lines for the inverse of (4.11)). DWR $= 40$ dB , $\alpha = 1$, and $L = 10^3$.

large values. Therefore, there is a clear trade-off between the presence of the structure of the pdf of Z and the asymptotic value with large values of t_0. As stated above, the value $\alpha = \alpha_{\text{Costa}}$ was introduced due to its performance in digital watermarking. According to the shown results, the best performance is not achieved for this value of $\alpha = \alpha_{\text{Costa}}$ (as one might guess because of its optimality in digital watermarking) and, thus, this suggests that there is room for improvement in the performance of DPCE techniques with respect to α.

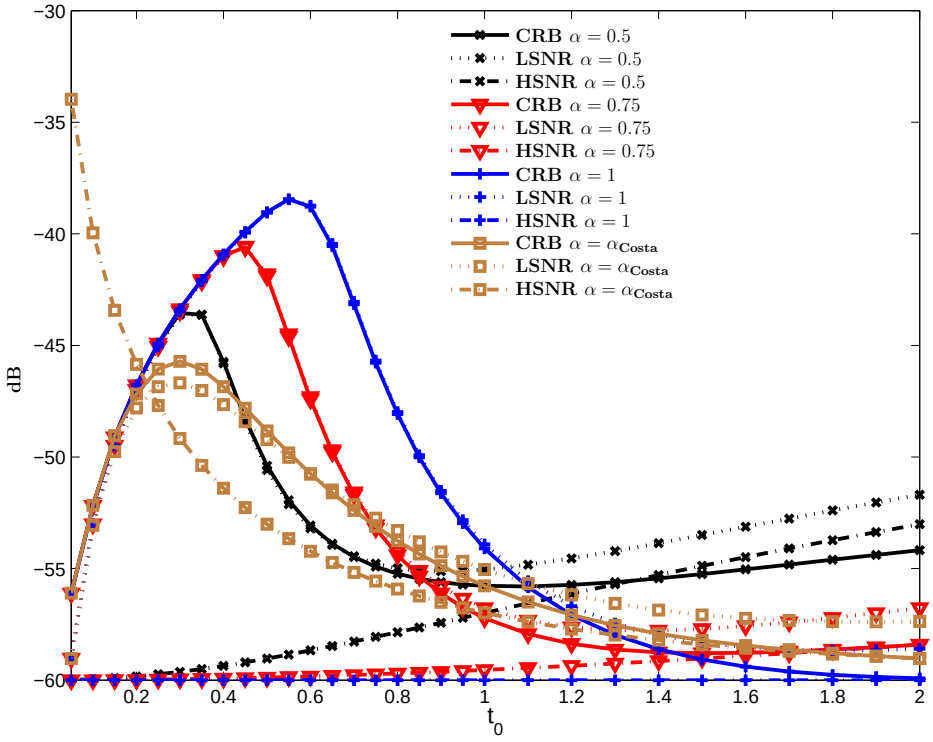


Figure 4.5: Cramér-Rao Bound vs. t_0 numerically obtained (CRB), using the low-SNR CRB approximation based on (4.7) (LSNR), and the high-SNR CRB approximation using (4.11) (HSNR). DWR = 30 dB , WNR = 0 dB, $\alpha = 0.5, 0.75, 1, \alpha_{\text{Costa}}$, and $L = 10^3$.

4.2 Information Theory: Mutual Information

As stated at the beginning of this chapter, the values taken by the gain of the channel are modeled by a Rayleigh distributed random variable T whose pdf is defined as:

$$f_T(t) = \begin{cases} \frac{te^{-\frac{t^2}{2\sigma_T^2}}}{\sigma_T^2} & \text{, if } t \geq 0 \\ 0 & \text{, elsewhere} \end{cases}.$$

It is worth recalling that the differential entropy of a continuous random variable B is defined as [11]

$$h(B) = -\int_{-\infty}^{\infty} f_B(\beta)\log(f_B(\beta))d\beta;$$

therefore, the differential entropy of T is given by

$$h(T) = 1 + \log(\sigma_T/\sqrt{2}) + \gamma/2,$$

where γ stands for the Euler's constant, i.e., $\gamma \triangleq \lim_{n\to\infty}\left[(\sum_{k=1}^{n}\frac{1}{k}) - \log(n)\right]$.

For the sake of simplicity, we will focus on the scenario where the considered signals are scalar, so the mutual information between Z and T given K, i.e. $I(Z;T|K)$, will be used to quantify the information about T acquired from the observation of Z by an agent aware of the secret key K.

This analysis will deal with two scenarios:

- The receiver knows the inserted symbol and the secret key (corresponding to the scenario where a pilot symbol is used).
- The receiver does not know the embedded symbol but indeed knows the secret key (traditional data hiding problem).

First, both cases are theoretically analyzed by assuming a noiseless scenario. Then, the channel noise is introduced in the framework for each case and the mutual information is approximated by taking advantage of the mathematical tractability of the proposed closed-form expressions of the pdfs of Z developed in Sect. 3.1.

4.2.1 Inserting Pilot Symbol

The mutual information between Z and T when the embedded symbol B is known can be expressed as

$$I(Z;T|K,B) = h(Z|K,B) - h(Z|T,K,B), \tag{4.12}$$

where in the previous expression $h(Z|K,B)$ stands for the differential entropy of Z given K and B, while $h(Z|T,K,B)$ similarly denotes the differential entropy of Z knowing K, B, and also T, which can calculated by

$$h(Z|T,K,B) = \int_0^\infty f_T(\tau) h\left(Z|T=\tau,K,B\right) d\tau. \tag{4.13}$$

4.2.1.1 No Channel Noise

In absence of the channel noise, Z can be expressed as $Z = T \cdot Y$; therefore, the expression (4.13) can be simplified

$$h(Z|T,K,B) = \int_0^\infty f_T(\tau) h(\tau Y|K,B) d\tau;$$

moreover, since $h(tY|K,B) = h(Y|K,B) + \log(t)$ [11], we can equivalently express the mutual information as

$$\begin{aligned} I(Z;T|K,B) &= h(Z|K,B) - h(Y|K,B) - \int_0^\infty \log(\tau) f_T(\tau) d\tau \\ &= h(Z|K,B) - h(Y|K,B) - \frac{1}{2}\left(\log(2\sigma_T^2) - \gamma\right). \end{aligned} \tag{4.14}$$

For both Add-SS/SIT and DPCE, we were not able to achieve a closed-form expression for $I(Z;T|K,B)$, there is not a closed-formula for the exact pdf of Z; however, the asymptotic values of $I(Z;T|K,B)$ when DWR goes to ∞ dB and to $-\infty$ dB can indeed be analyzed. Let us mention that we are aware that the set of applications of the case DWR $\to -\infty$ dB is smaller than DWR $\to \infty$ dB case due to the embedding distortion; however, we still find it interesting to study in order to obtain the complete asymptotic characteristics of the proposed problem.

4.2.1.1.1 Additive Spread Spectrum and Superimposed Training

For the analyzed Add-SS embedding [12] and SIT techniques [58], Y can be expressed as

$$Y = X + (-1)^{B+1} S,$$

where, on one hand, S models the so-called spreading sequence in Add-SS, which is a deterministic function of K; therefore, for the sake of notational simplicity we will write $K = s$ implying that s is the only key-dependent component in Y. On the other hand, $(-1)^{B+1}S$ models the superimposed training sequence following the SIT approach; therefore the obtained results are also valid for SIT. We will assume, without loss of generality, that the embedded bit is $B = 1$, yielding the following pdf for Y

$$f_{Y|K,B}(y|s,1) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{(y-s)^2}{2\sigma_x^2}},$$

and consequently $h(Y|K,B) = \frac{1}{2}\log(2\pi e\sigma_X^2)$.

In order to calculate $I(Z;T|K,B)$, $f_{Z|K,B}(z|s,1)$ is also needed. This distribution can be obtained as the marginal distribution of Z from the joint pdf of Z and T ($f_{Z,T|K,B}(z,t|s,1)$)

$$f_{Z|K,B}(z|s,1) = \frac{1}{\sigma_T^2\sigma_X\sqrt{2\pi}}\int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w-s)^2}{2\sigma_X^2}} dw. \tag{4.15}$$

DWR$\to\infty$ dB

For large values of DWR, it is reasonable to assume that pdf of Y given K can be approximated, dropping the mean value s, as

$$f_{Y|K}(y|s) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma_X^2}},$$

so one can approximate (4.15) as

$$f_{Z|K,B}(z|s,1) \simeq \frac{1}{\sigma_T^2\sigma_X\sqrt{2\pi}}\int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w)^2}{2\sigma_X^2}} dw = \frac{e^{-\frac{|z|}{\sigma_T\sigma_X}}}{2\sigma_X\sigma_T},$$

i.e., the pdf of Z can be approximated by a Laplacian distribution with variance $2\sigma_X^2\sigma_T^2$, being the corresponding differential entropy

$$h(Z|K,B) = \log(2e\sigma_T\sigma_X).$$

Thus, using (4.14) with the previous expression under the high DWR assumption, $I(Z;T|d)$ can be approximated by

$$I(Z;T|K,B) \simeq \frac{1}{2}\left(1+\gamma-\log\left(\pi\right)\right) \triangleq \kappa \simeq 0.216. \tag{4.16}$$

DWR$\to-\infty$ dB

In this case, the distribution of Y is very narrow compared with the value of s and, thus, the pdf of Y can be approximated, in order to calculate $f_{Z|K,B}(z|s,1)$, by a delta function centered at $\pm s$, i.e., $\delta(y\pm s)$. Consequently, the distribution of Z is the distribution of T scaled by $|s|$, and possibly inverted, i.e., for $z>0$

$$f_{Z|K,B}(z|s,1) \simeq \frac{ze^{-\frac{z^2}{2s^2\sigma_T^2}}}{s^2\sigma_T^2}\frac{1}{\sigma_X\sqrt{2\pi}}\int_{-s}^\infty e^{-\frac{\eta^2}{2\sigma_X^2}} d\eta \simeq \begin{cases} \frac{ze^{-\frac{z^2}{2s^2\sigma_T^2}}}{s^2\sigma_T^2}, & \text{if } s>0 \\ 0, & \text{otherwise} \end{cases}.$$

Similarly, whenever $z < 0$ one can approximate

$$f_{Z|K,B}(z|s,1) \simeq \begin{cases} \frac{-ze^{-\frac{z^2}{2s^2\sigma_T^2}}}{s^2\sigma_T^2} & , \text{ if } s < 0 \\ 0 & , \text{ otherwise} \end{cases}.$$

Be aware that the obtained function for $z > 0$ and $s > 0$, or $z < 0$ and $s < 0$, corresponds to a Rayleigh distribution with parameter $s\sigma_T$. This allows one to approximate the entropy of Z given B and a particular value of K as,

$$h(Z|K = s, B) \simeq h(T) + \log(|s|) = 1 + \log(\sigma_T/\sqrt{2}) + \gamma/2 + \log(|s|),$$

so

$$h(Z|K, B) \simeq 1 + \log(\sigma_T/\sqrt{2}) + \gamma/2 + \int_{-\infty}^{\infty} f_K(s) \log(|s|) ds.$$

From the last formula, (4.14), and the previously calculated $h(Y|K, B)$, $I(Z; T|K, B)$ can be approximated as

$$I(Z; T|K, B) \simeq 1 - \log(2) + \gamma - \frac{1}{2} \int_{-\infty}^{\infty} f_K(s) \log\left(2\pi e \frac{\sigma_X^2}{|s|^2}\right) ds.$$

As a particular case of the previous formula we can study the schemes where the spreading sequence is just binary; concretely, wherever $P(K = -\sigma_W) = P(K = +\sigma_W) = \frac{1}{2}$, the last formula can be rewritten as

$$I(Z; T|K, B) \simeq 1 - \log(2) + \gamma - \frac{1}{2} \log\left(2\pi e \text{DWR}\right). \tag{4.17}$$

4.2.1.1.2 Dirty Paper Coding

Recalling (2.2) introduced in Sect. 2.3, Y, obtained using a distortion compensated quantizer, can be expressed as

$$Y = X + \alpha \left[\mathcal{Q}_\Delta \left(X - \Delta\left(B/M\right) - D\right) - \left(X - \Delta\left(B/M\right) - D\right)\right],$$

where in the previous expression $\mathcal{Q}_\Delta(\cdot)$ denotes a scalar uniform quantizer with step-size Δ, and B takes values in $\{0, \cdots, M-1\}$. We will focus on the binary case, i.e., $M = 2$, and the dither vector D, which is a deterministic function of the secret key K, is assumed to be uniformly distributed in $[-\Delta/2, \Delta/2]$, i.e., $D \sim \mathcal{U}(-\Delta/2, \Delta/2)$. Furthermore, for the sake of notational simplicity, and without loss of generality, we will assume that $B = 0$. Similarly to the Add-SS and SIT cases, we will write $K = d$ when meaning that the considered value of K yields the dither d. In this way

$$f_{Y|K,B}(y|d,0) = \begin{cases} \sum_{i=-\infty}^{\infty} \frac{1}{1-\alpha} f_X\left(\frac{y-\alpha(i\Delta+d)}{1-\alpha}\right), \\ \quad \text{if} \quad |y - d - i\Delta| < (1-\alpha)\frac{\Delta}{2} \\ \\ 0, \quad \text{otherwise} \end{cases}.$$

Moreover, the probability density function of Z can be obtained as

$$f_{Z|K,B}(z|d,0) = \frac{1}{\sigma_T^2}\int_0^\infty e^{-\frac{\tau^2}{2\sigma_T^2}} f_{Y|K,B}(z/\tau|d,0)d\tau. \tag{4.18}$$

DWR$\to \infty$ dB

When DWR $\to \infty$ dB, it is well-known that one can accurately model $W \sim \mathcal{U}(-\alpha\Delta/2, \alpha\Delta/2)$. Therefore, $h(Y|K,B)$ can be calculated by using the properties of the differential entropy, yielding

$$\begin{aligned} h(Y|K,B) &= h(Y|B) - I(Y;K|B) = h(Y) - h(K|B) + h(K|Y,B) \\ &\simeq h(X) - \log(\Delta) + \log((1-\alpha)\Delta) \\ &= h(X) + \log(1-\alpha), \end{aligned} \tag{4.19}$$

where in (4.19), we are considering that $h(K) = h(K|B)$, $K \sim \mathcal{U}(-\Delta/2, \Delta/2)$ and due to the DWR $\to \infty$ dB approximation, $h(Y) \simeq h(X)$ and $h(K|Y,B) \simeq \log((1-\alpha)\Delta)$.

Now, focusing on $h(Z|K,B)$, it can be seen that

$$f_{Z|K,B}(z) \simeq \frac{e^{-\frac{|z|}{\sigma_X\sigma_T}}}{2\sigma_T\sigma_X},$$

so

$$h(Z|K,B) = \log(2e\sigma_T\sigma_X).$$

Finally, we have that

$$I(Z;T|K,B) \simeq \frac{1}{2}\left(1 + \gamma - \log(\pi) - \log((1-\alpha)^2)\right). \tag{4.20}$$

DWR$\to -\infty$ dB

If DWR $\to -\infty$ dB, D becomes a critical parameter. Its study will be divided in three different cases: $D = 0$, $D = \pm\Delta/2$, and $0 < D < \Delta/2$ (equivalent to $-\Delta/2 < D < 0$).

- On one hand, if $D = 0$ and the value of Δ goes to ∞ (trying to achieve DWR$\to -\infty$ dB), almost all the probability of X is concentrated in the Voronoi region of the zero centroid. Therefore, $y \simeq x - \alpha x$, the minimal DWR is $1/\alpha^2$ (thus, in this case DWR$\to -\infty$ dB cannot be achieved) and

$f_{Y|K,B}(y|0,0) = 1/(1-\alpha) f_X\left(y/(1-\alpha)\right)$. Furthermore, $f_{Z|K,B}(z|0,0)$ can be approximated as

$$f_{Z|K,B}(z|0,0) \simeq \frac{e^{-\frac{|z|}{\sigma_T(1-\alpha)\sigma_X}}}{2(1-\alpha)\sigma_X\sigma_T},$$

which is obtained by following the same approach of Add-SS/SIT for DWR $\to \infty$ dB.

As a result, it is easy to see that

$$I(Z;T|K=0,B=0) \simeq \frac{1}{2}\left(1+\gamma-\log(\pi)\right) = \kappa.$$

- On the other hand, if $D = \Delta/2$ and DWR goes to $-\infty$ dB (increasing the size of the quantization step Δ), $f_{Y|K,B}(z|\Delta/2,0)$ can be approximated by

$$f_{Y|K,B}(y|\Delta/2,0) \simeq \begin{cases} \frac{1}{1-\alpha} f_X\left(\frac{y-\alpha\Delta/2}{1-\alpha}\right), & \text{if } \ y \geq \alpha\Delta/2 \\ \frac{1}{1-\alpha} f_X\left(\frac{y+\alpha\Delta/2}{1-\alpha}\right), & \text{if } \ y < -\alpha\Delta/2 \\ 0, & \text{otherwise} \end{cases},$$

so the corresponding differential entropy will be

$$h(Y|K=\Delta/2,B=0) = \frac{1}{2}\log\left(2\pi e\sigma_X^2\right) + \log\left(1-\alpha\right).$$

Note that $h(Y|K=\Delta/2,B=0)$ does not depend explicitly on Δ, and its dependence on α is just a shift by $\log(1-\alpha)$. Regarding $h(Z|K=\Delta/2,B=0)$, operating in a similar way to Sect. 4.2.1.1.1, it can be seen that

$$\begin{aligned} h(Z|K=\Delta/2,B=0) &\simeq h(T) + \log(2) + \log\left(\frac{\alpha\Delta}{2}\right) \\ &= 1 + \log\left(\frac{\sigma_T}{\sqrt{2}}\right) + \frac{\gamma}{2} + \log(\alpha\Delta). \end{aligned}$$

Be aware that in this case $h(Z|K=\Delta/2,B=0)$ depends on α and Δ through $\log(\alpha\Delta)$. Taking into account all these considerations about $h(Y|K=\Delta/2,B=0)$ and $h(Z|K=\Delta/2,B=0)$, it is straightforward to see that when $K=\Delta/2$ and DWR$\to -\infty$ dB $I(Z;T|K=\Delta/2,B=0)$ will be described by a family of parallel curves obtained for the different values of α, and it will go to infinity as Δ is increased, or equivalently, as the DWR is decreased. Formally,

$$I(Z;T|K=\Delta/2,B=0) \simeq 1-\log(2)+\gamma-\frac{1}{2}\log\left(2\pi e\frac{(1-\alpha)^2\sigma_X^2}{\alpha^2\Delta^2}\right).$$

Therefore, we can state that with $K=\Delta/2$ and $B=0$, the DWR can decrease boundlessly (be aware that the distortion will also increase boundlessly, which is generally not possible in real applications) and, at the same time, increase $I(Z;T|K=\Delta/2,B=0)$.

- In the scenario where D is in $(-\Delta/2, 0) \cup (0, \Delta/2)$ and $B = 0$, most of probability of Y corresponds to the centroid of $B = 0$ which is the closest one to the origin, i.e., at αd; hence, the pdf of Y is a shift and scaled version of the pdf of X,

$$f_{Y|K,B}(y|d,0) \simeq \frac{1}{1-\alpha} f_X\left(\frac{y-\alpha d}{1-\alpha}\right),$$

and therefore $h(Y|K=d,B=0) = \frac{1}{2}\log\left(2\pi e \sigma_X^2\right) + \log(1-\alpha)$.

In order to calculate the pdf of Z, the pdf of Y is approximated by $f_{Y|K,B}(y|d,0) \simeq \delta(y-\alpha d)$ and thus the distribution of Z becomes a scaled version of the original Rayleigh distribution, i.e., $f_{Z|K,B}(z|d,0) \simeq \frac{f_T(z/(\alpha d))}{\alpha d}$; where the corresponding differential entropy is $h(Z|K=d,B=0) \simeq 1 + \log(\sigma_T/\sqrt{2}) + \gamma/2 + \log(|\alpha d|)$.

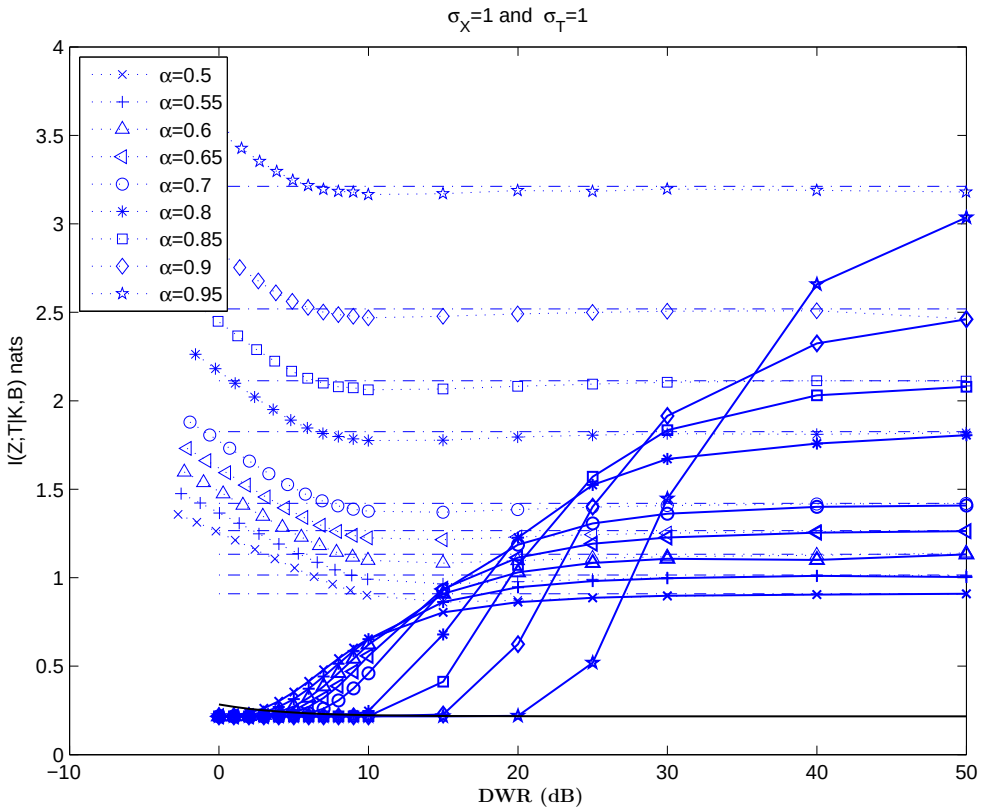


Figure 4.6: $I(Z;T|K,B)$ vs. DWR curves for both Add-SS (solid line without symbols) and DPCE cases for different α with $\sigma_X = 1$ and $\sigma_T = 1$. S follows an equiprobable binary antipodal distribution. In the DPCE case, the cases with $d = 0$ (solid lines with symbols) and $d = \Delta/2$ (dotted lines with symbols) are also depicted for $B = 0$. In addition, several asymptotic values are represented (dashdot lines without symbols).

Finally, the mutual information when $D \in (-\Delta/2, 0) \cup (0, \Delta/2)$ and DWR

goes to $-\infty$ dB is

$$I(Z;T|K=d,B=0) \simeq 1+\gamma-\log(2)-\log\left(\frac{(1-\alpha)}{\alpha}\frac{\sigma_X}{|d|}\right)-\frac{1}{2}\log(2\pi e). \tag{4.21}$$

Since D is uniformly distributed in $[-\Delta/2,\Delta/2]$; therefore, the mutual information between Z and T given K and B is obtained using the previous expression averaging with respect to D as

$$\begin{aligned} I(Z;T|K,B) &\simeq 1+\gamma-\log(2)-\log\left(\frac{(1-\alpha)\sigma_X}{\alpha}\right)-\left[\frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2}\log\frac{1}{|\tau|}d\tau\right] \\ &\quad -\frac{1}{2}\log(2\pi e) \\ &= 1+\gamma-\log(2)-\log\left(\frac{(1-\alpha)\sigma_X}{\alpha}\right)-\left[1+\log\left(\frac{2}{\Delta}\right)\right] \\ &\quad -\frac{1}{2}\log(2\pi e) \\ &= \gamma-\frac{1}{2}\log(2\pi e)-\log\left(\frac{4\sigma_X(1-\alpha)}{\alpha\Delta}\right) \\ &= \gamma-\frac{1}{2}\log(2\pi e\mathrm{DWR})+\log\left(\frac{\sqrt{3}}{2(1-\alpha)}\right). \end{aligned} \tag{4.22}$$

4.2.1.1.3 Numerical Results

Fig. 4.6 depicts $I(Z;T|K,B)$ as a function of the DWR, for both Add-SS and DPCE (for $K=0$ and $K=\Delta/2$, $B=0$, and different values of α) scenarios with the following parameters: $\sigma_X=1$ and $\sigma_T=1$. Moreover, the asymptotic values when DWR $\to -\infty$ dB and DWR $\to \infty$ dB are also represented. In Fig. 4.7 the mutual information curves for different dither sequences (without $\Delta/2$ and 0) and the corresponding approximations for the DPCE are represented when DWR$\to -\infty$ dB. Furthermore, Fig. 4.8 shows $I(Z;T|K,B)$ when the Add-SS embedding technique is used with the DWR$\to \pm\infty$ dB approximations.

By analyzing Figs. 4.6-4.8, we can observe that the theoretical and the numerical results almost perfectly match, showing the accuracy of our analysis.

Focusing on the asymptotic case DWR$\to \infty$ dB, the curves show that the DPC-based technique outperforms the Add-SS/SIT approach. Theoretically, this can be verified in their mutual information approximations (4.20) for DPC-based techniques and (4.16) for Add-SS/SIT, where that obtained for our technique is the mutual information for Add-SS/SIT plus the term $-\log((1-\alpha)^2)$ that takes positive values if $\alpha>0$ (note that for $\alpha=0$, there is not watermark) due to $\alpha\leq 1$ and, therefore, $\log((1-\alpha)^2)$ is negative. These results indicate that for our

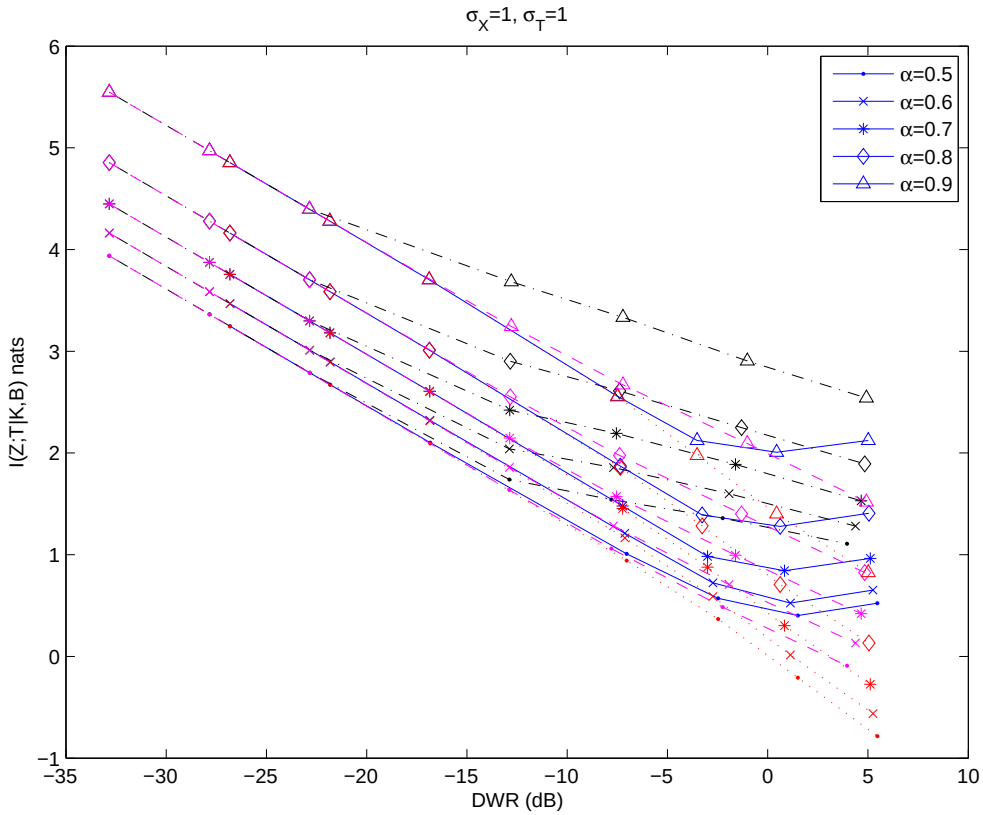


Figure 4.7: $I(Z;T|K=d,B=0)$ vs. DWR curves using DPCE embedding for different α and dither sequences with $\sigma_X = 1$ and $\sigma_T = 1$. The solid lines correspond to the empirically obtained values of $I(Z;T|K=d,B)$ for $d=-0.2\Delta$ and dashdot lines when $d=0.4\Delta$; the dotted lines are the approximation (4.21) when DWR$\to -\infty$ dB for $d=-0.2\Delta$ and dashed lines if $d=0.4\Delta$.

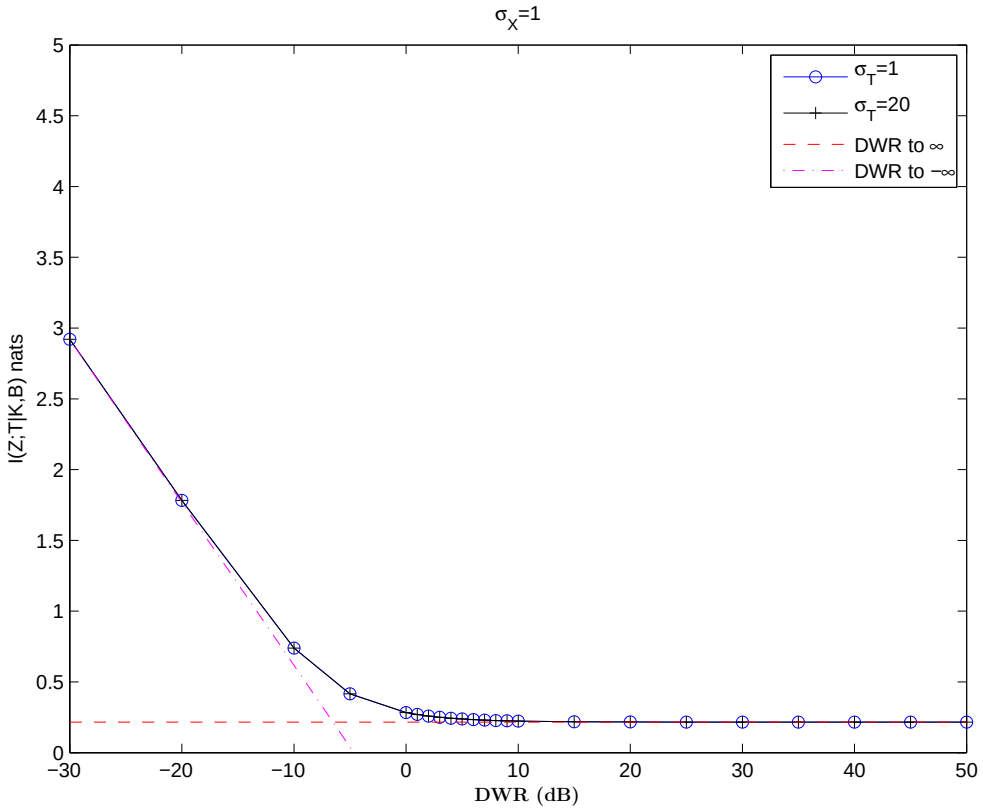


Figure 4.8: $I(Z;T|K=s,B)$ vs. DWR curves (solid lines) for Add-SS/SIT embedding and the DWR$\to \pm\infty$ dB approximations. S follows a equiprobable binary antipodal distribution.

technique α should take the largest possible value in order to obtain the largest mutual information value in the noiseless case.

Now we consider the asymptotic case DWR$\rightarrow -\infty$ dB. By comparing their expressions, (4.17) for Add-SS/SIT and (4.22) for DPC-based estimation techniques, and according to Figs. 4.6-4.8, if the dither sequence is uniformly distributed in $[-\Delta/2, \Delta/2]$, then the DPC-based estimation will outperform Add-SS/SIT. This conclusion can be easily reached by comparing the corresponding expressions for mutual information. If $\log(\sqrt{3}/(1-\alpha)) > 1$, then DPCE will outperform Add-SS/SIT which is verified for $\alpha > 1 - \sqrt{3}/e \approx 0.36$ (in shown curves α takes values larger than this bound). Again, the mutual information for our technique indicates that α must take the largest possible value, which is coherent with the noiseless case for digital watermarking, where the optimal $\alpha \rightarrow 1$ whenever WNR$\rightarrow \infty$.

4.2.1.2 With Channel Noise

By assuming that the distribution of Z is mainly due to variability of X and T (i.e., neglecting the influence of the watermark and the channel noise on that distribution, which is reasonable under the HQR $\gg 1$ and TNHR$(t_0) \ll 1$ assumptions) one can write, both for Add-SS/SIT and our DPC-based approach,

$$f_Z(z) \approx \int_0^\infty \frac{te^{-t^2/(2\sigma_T^2)}}{\sigma_T^2} \frac{e^{-z^2/(2t^2\sigma_X^2)}}{\sqrt{2\pi t^2\sigma_X^2}} dt = \frac{e^{-|x|/(\sigma_T\sigma_X)}}{2\sigma_T\sigma_X},$$

i.e., Z is Laplacian distributed with zero-mean and variance $2\sigma_T^2\sigma_X^2$; consequently, its differential entropy is

$$h(Z|K,B) \approx 1 + \log(2\sigma_T\sigma_X). \tag{4.23}$$

4.2.1.2.1 Additive Spread Spectrum and Superimposed Training

Under the hypotheses HQR $\gg 1$ and TNHR$(t_0) \ll 1$, the scenario is analogous to the noiseless case with DWR $\rightarrow \infty$ dB introduced above; thus, if $T = t$, then $h(Z|T=t,K) = \frac{1}{2}\log(2\pi e t^2\sigma_X^2)$, whose average over $f_T(t)$ is

$$h(Z|T,K,B) \approx \log(2) + \frac{1}{2}\left[1 - \gamma + \log(\pi\sigma_T^2\sigma_X^2)\right].$$

Consequently, the mutual information between Z and T for superimposed pilots is

$$I(Z;T|K,B) = h(Z|K,B) - h(Z|T,K,B) \approx \frac{1}{2}\left[1 + \gamma - \log(\pi)\right] = \kappa. \tag{4.24}$$

Note that for superimposed pilots the considered mutual information does not depend on the host signal variance σ_X^2, the channel variance σ_N^2, or σ_T^2.

4.2.1.2.2 Dirty Paper Coding

In order to obtain the closed-form expressions for $I(Z;T|K,B)$ in scenarios with channel noise, the approximations to the pdf of Z obtained in Sect. 3.1 are used for the two analyzed scenarios: low-SNR and high-SNR. In order for the analysis to be valid, the hypotheses for the deterministic gain of each case must be verified in the intervals of the support set of T whose probability is not negligible.

Low-SNR Case

For the sake of simplicity, we use in our analysis the simplest approximation of the pdf of Z for low-SNR cases (3.9) obtained assuming that HQR $\gg 1$, SCR$(t_0) \ll 1$, TNQR$(t_0) \gg 1$, and TNHR$(t_0) \ll 1$.

To approximate $h(Z|T,K,B)$, first $h(Z|T=t,K,B)$ is calculated using the pdf approximation as

$$\begin{aligned} h(Z|T=t,K,B) &= -\int_{-\infty}^{\infty} f_{Z|T,K}^{\text{low-SNR},2}(z|t,d)\log\left(f_{Z|T,K}^{\text{low-SNR},2}(z|t,d)\right)dz \\ &= -\int_{-\infty}^{\infty} f_{Z|T,K}^{\text{low-SNR},2}(z|t,d)\log\left(\frac{e^{-\frac{z^2}{2t^2\sigma_X^2}}}{\sqrt{2\pi t^2\sigma_X^2}}\right)dz \\ &\quad -\int_{-\infty}^{\infty} f_{Z|T,K}^{\text{low-SNR},2}(z|t,d)\log\left(1+2e^{-\frac{2\pi^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2t^2}{12}\right)}{\Delta^2t^2}}\cos\left(\frac{2\pi z}{\Delta t}-\frac{2\pi d}{\Delta}\right)\right)dz. \end{aligned} \tag{4.25}$$

The result of the first integral in (4.25) averaged with respect to D is $-1/2\log\left(2\pi\sigma_X^2t^2\right)-1/2$. The average with respect to D of the second integral in (4.25) is $2\exp\left\{-1/3\pi^2(1-\alpha)^2-(4\pi^2\sigma_N^2)/(\Delta^2t^2)\right\}$, where we have used that if $|x|\ll 1$ then $\log(1+x)\approx x$. With these two approximations, (4.25) becomes

$$h(Z|T=t,K,B)\approx\frac{1}{2}\left(-4e^{-\frac{1}{3}\pi^2(1-\alpha)^2-\frac{4\pi^2\sigma_N^2}{\Delta^2t^2}}+\log\left(2\pi\sigma_X^2t^2\right)+1\right).$$

In order to obtain the conditional differential entropy, the previous expression is averaged with respect to T as

$$\begin{aligned} h(Z|T,K,B) &\approx \int_0^{\infty} f_T(\tau)\frac{1}{2}\left(-4e^{-\frac{1}{3}\pi^2(1-\alpha)^2-\frac{4\pi^2\sigma_N^2}{\Delta^2\tau^2}}+\log\left(2\pi\sigma_X^2\tau^2\right)+1\right)d\tau \\ &\approx \frac{1}{2}\left(-\frac{8\sqrt{2}\pi e^{-\frac{1}{3}\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{2\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T}+2\log(\sigma_T\sigma_X)-\gamma+1+\log(4\pi)\right), \end{aligned} \tag{4.26}$$

where $K_{1,\mathrm{B}}(\cdot)$ stands for the modified Bessel function of the second kind [1, p. 376] and, in this case, it can be expressed in its integral form as

$$K_{1,\mathrm{B}}(z) \triangleq z \int_1^\infty e^{-z\tau}\sqrt{\tau^2-1}d\tau. \tag{4.27}$$

Therefore, by subtracting (4.26) from (4.23), it is straightforward to obtain the mutual information approximation for the low-SNR case as

$$\begin{aligned} I(Z;T|K,B) &\approx \frac{1}{2}\left(\frac{8\sqrt{2}\pi e^{-\frac{1}{3}\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{2\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} + 1 + \gamma - \log(\pi)\right) \\ &= \frac{4\sqrt{2}\pi e^{-\frac{1}{3}\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{2\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} + \kappa. \end{aligned} \tag{4.28}$$

Since the argument of $K_{1,\mathrm{B}}(\cdot)$ is positive in the previous expression, it is straightforward to realize from (4.27) that $K_{1,\mathrm{B}}(\cdot)$ will also take positive values; thus

$$\frac{4\sqrt{2}\pi e^{-\frac{1}{3}\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{2\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} > 0.$$

Therefore, $I(Z;T|K,B)$ takes values larger than or equal to the ones obtained for Add-SS/SIT, as one can conclude comparing (4.28) and (4.24). In addition, the division into two summands of the mutual information (4.28) allows us to identify that the term containing $K_{1,\mathrm{B}}(\cdot)$ corresponds to the contribution due to the introduced structure in the distribution of Z by DPCE.

High-SNR Case

In the high-SNR scenario, we can approximate

$$\begin{aligned} h(Z|T=t,K) \approx{}& \frac{1}{2}\log(2\pi e\sigma_X^2) - \log(\Delta) \\ &+ \frac{1}{2}\log\left(2\pi e\left[\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right]\right), \end{aligned} \tag{4.29}$$

where $\frac{1}{2}\log(2\pi e\sigma_X^2) - \log(\Delta)$ is the discrete entropy of the centroid used at the embedder when HQR $\gg 1$ and $\frac{1}{2}\log(2\pi e[\sigma_N^2 + (1-\alpha)^2\Delta^2 t^2/12])$ stands for the differential entropy of Z given the used centroid. Note that under the high-SNR assumption $\mathrm{SCR}(t_0) \ll 1$, so the sum of channel noise and self-noise can be approximated as Gaussian, and $\mathrm{TNQR}(t_0) \ll 1$, and consequently the modulo-lattice reduction of the total noise can be neglected.

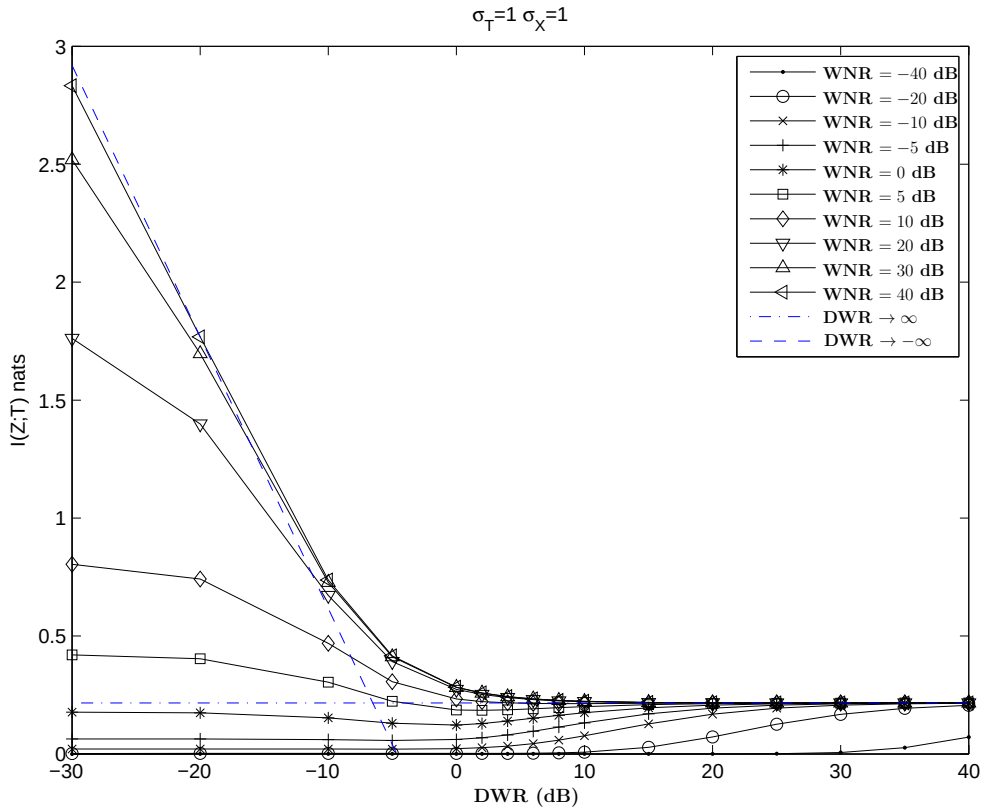


Figure 4.9: $I(Z;T|K,B=0)$ vs. DWR employing Add-SS/SIT techniques when the embedded bit is known and S follows an equiprobable binary antipodal distribution for different values of WNR (expressed in dB). The dashdot line corresponds with the approximation for DWR $\to \infty$ dB. The dashed line corresponds to the noiseless case when DWR $\to -\infty$ dB. $\sigma_X = 1$ and $\sigma_T = 1$.

The expression in (4.29) can be averaged over T, yielding

$$h(Z|T,K) \approx \frac{1}{2}\log(2\pi e\sigma_X^2) - \log(\Delta) + \frac{1}{2}\left(-e^{\frac{6\sigma_N^2}{(\alpha-1)^2\Delta^2\sigma_T^2}}\,\mathrm{Ei}\left(-\frac{6\sigma_N^2}{(\alpha-1)^2\Delta^2\sigma_T^2}\right) + 2\log(\sigma_N) + 1 + \log(2\pi)\right);$$

where, $\mathrm{Ei}(x) \triangleq -\int_{-x}^{\infty}(e^{-t}/t)dt$ stands for the exponential integral function [1, p. 228]. Therefore, the mutual information in this case can be approximated as

$$I(Z;T|K) \approx \frac{1}{2}e^{\frac{6\sigma_N^2}{(\alpha-1)^2\Delta^2\sigma_T^2}}\,\mathrm{Ei}\left(-\frac{6\sigma_N^2}{(\alpha-1)^2\Delta^2\sigma_T^2}\right) + \log\left(\frac{\Delta\sigma_T}{\pi\sigma_N}\right); \qquad (4.30)$$

the limit of the previous expression when the value of WNR (defined as in Chap. 2) goes to the infinity for a finite value of σ_T is

$$\lim_{\mathrm{WNR}\to\infty} I(Z;T|K) = \frac{1}{2}\left(\log(6) + \gamma - \log(\pi^2) - \log((1-\alpha)^2)\right). \qquad (4.31)$$

This expression is compared with the counterpart for the case without the channel noise (4.20) showing the same increasing tendency with α discussed above. In addition, the result of subtracting (4.31) from (4.20) is $1/2(1-(\log(6)-\log(\pi))) \approx 0.176$ which is the gap between the two approximations for this asymptotic case.

4.2.1.2.3 Numerical Results

Fig. 4.9 shows empirical curves of the mutual information when the Add-SS/SIT technique is used for several scenarios with channel noise. In addition, the approximation for DWR$\to\infty$ dB is depicted and the approximation in absence of channel noise is also depicted for DWR$\to-\infty$ dB. It can be shown that the empirical curves asymptotically converge to the respective approximations.

Figs. 4.10 and 4.11 depict the curves of $I(Z;T|K,B)$ obtained numerically for our algorithm considering the noisy scenario. As one could guess, the larger the value of WNR and DWR, the closer these curves to their corresponding approximations of the noiseless scenario. Moreover, the approximations of the mutual information for the analyzed cases are also depicted. In both cases, the approximations get tighter by increasing α in their hypotheses verification scenarios. The approximation for high-SNR cases when WNR goes to infinity (4.31) indicates, in accordance to the numerical results with noise, the mutual information increases with α (note that the value taken by α cannot be 1).

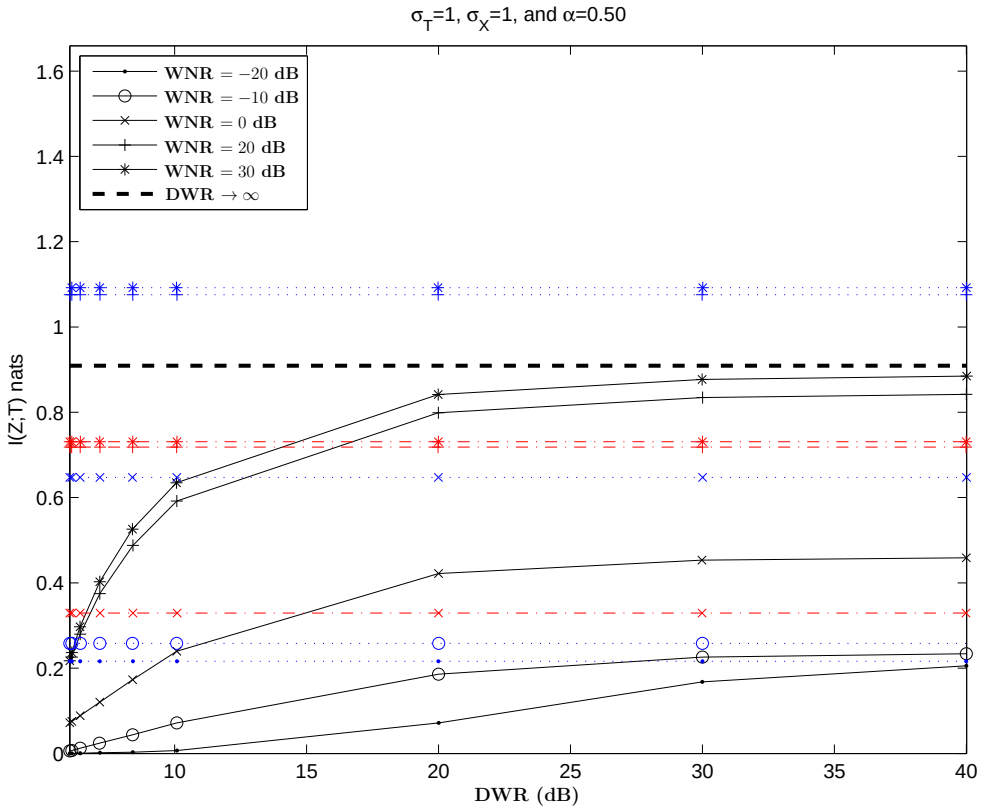


Figure 4.10: $I(Z;T|K=0,B=0)$ vs. DWR curves using DPCE and $\alpha = 0.5$ numerically obtained (solid lines), with the approximation for low-SNR cases (4.28) (dotted lines), and (4.30) for high-SNR cases (dashdot lines). The dashed line corresponds with the noiseless case for DWR$\to\infty$ dB.

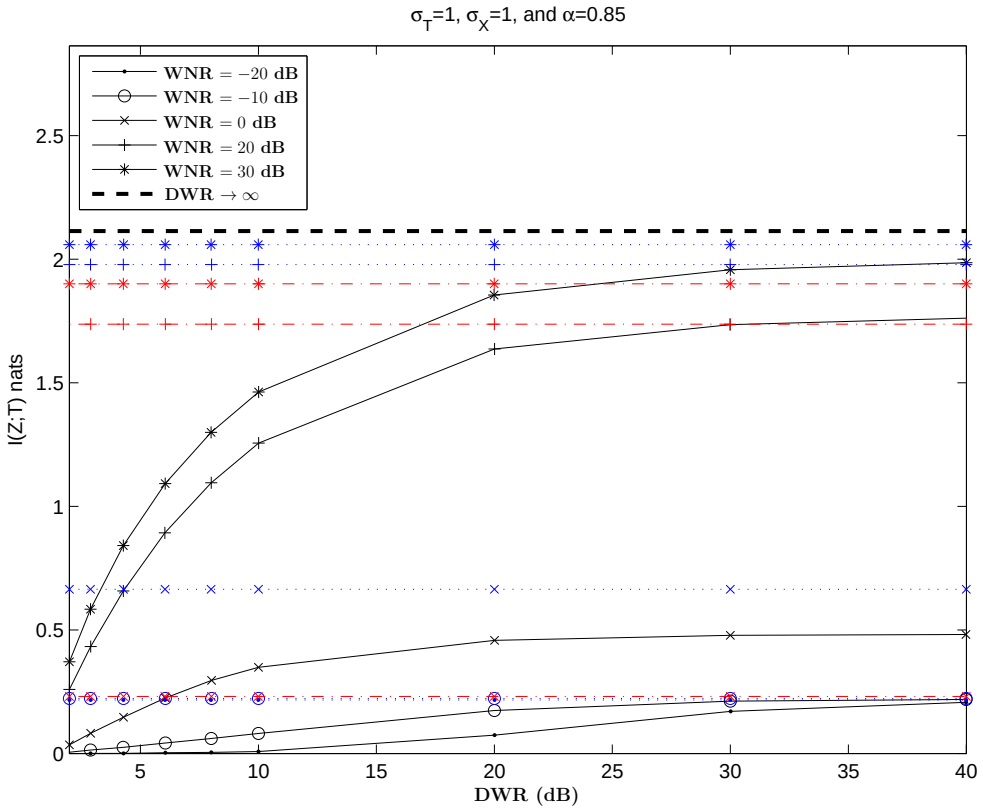


Figure 4.11: $I(Z;T|K=0,B=0)$ vs. DWR curves using DPCE and $\alpha = 0.85$ numerically obtained (solid lines), with the approximation for low-SNR cases (4.28) (dotted lines), and (4.30) for high-SNR cases (dashdot lines). The dashed line corresponds with the noiseless case for DWR$\rightarrow \infty$ dB.

Focusing on the curves for WNR $= -20$ dB, one can see that the mutual information tends to κ (i.e., approximately 0.2162); as discussed above, this verifies that the mutual information is always larger for DPCE than for Add-SS/SIT when DWR$\rightarrow \infty$ dB, being only equal when WNR tends to $-\infty$ dB.

4.2.2 No Pilot Symbols

In this section, the mutual information between Z and T is studied assuming that the secret key is known, but the embedded bit is not known. In this case,

$$I(Z;T|K) = h(Z|K) - h(Z|T,K),$$

where the previous expression is not conditioned by B as (4.12) does.

4.2.2.1 No Channel Noise

If there is not channel noise, $I(Z;T|K)$ can be expressed as

$$\begin{aligned} I(Z;T|K) &= h(Z|K) - h(Y|K) - \int_0^\infty \log(\tau) f_T(\tau) d\tau \\ &= h(Z|K) - h(Y|K) - \frac{1}{2}\left(\log(2\sigma_T^2) - \gamma\right). \end{aligned}$$

As in Sect. 4.2.1.1, we have not been able to find a closed-formula for $I(Z;T|K)$; however, an analysis of both techniques (Add-SS and DPCE) when DWR$\rightarrow \pm\infty$ dB is carried out next.

4.2.2.1.1 Additive Spread Spectrum and Superimposed Training

For the sake of comparison, the Add-SS/SIT techniques described in Sect. 4.2.1.1.1 are used in this section. Considering that $B = 0$ and $B = 1$ are equiprobable, the pdf of Y is

$$f_{Y|K}(y|s) = \frac{1}{2\sigma_x\sqrt{2\pi}} e^{-\frac{(y-s)^2}{2\sigma_x^2}} + \frac{1}{2\sigma_x\sqrt{2\pi}} e^{-\frac{(y+s)^2}{2\sigma_x^2}}. \tag{4.32}$$

Concerning the pdf of Z,

$$f_{Z|K}(z|s) = \frac{1}{2}\frac{1}{\sigma_T^2\sigma_X\sqrt{2\pi}}\left[\int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w-s)^2}{2\sigma_X^2}} dw + \int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w+s)^2}{2\sigma_X^2}} dw\right].$$

DWR$\to \infty$ dB

When DWR $\to \infty$, Y can be approximated by a zero-mean Gaussian distribution, as $|z/w| \gg |s|$ with probability asymptotically close to 1. Therefore, $f_{Z|K}(z|s)$ can be approximated by,

$$\begin{aligned} f_{Z|K}(z|s) &\simeq \frac{1}{2}\frac{1}{\sigma_T^2\sigma_X\sqrt{2\pi}}\left[\int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w)^2}{2\sigma_X^2}} dw + \int_0^\infty e^{-\frac{w^2}{2\sigma_T^2}} e^{-\frac{(z/w)^2}{2\sigma_X^2}} dw\right] \\ &= \frac{e^{-\frac{|z|}{\sigma_T\sigma_X}}}{2\sigma_T\sigma_X}, \end{aligned}$$

where $f_{Z|K}(z|s)$ becomes a Laplacian distribution. Thus, the entropies of Y and Z are respectively

$$\begin{aligned} h(Y|K=s) &\simeq \frac{1}{2}\log(2\pi e\sigma_X^2), \\ h(Z|K=s) &\simeq \log(2e\sigma_T\sigma_X), \end{aligned}$$

and the resulting mutual information between Z and Y is

$$I(Z;T|K=s) \simeq h(Z) - h(Y) - \frac{1}{2}\left(\log(2\sigma_T^2) - \gamma\right) = \frac{1}{2}\left(1 + \gamma - \log(\pi)\right) = \kappa;$$

note that this expression is equal to (4.16). Be aware that in this case $h(Y|K=s)$, $h(Z|K=s)$, and $I(Z;T|K=s)$ are all independent of the particular value of K (as it will negligible compared to z/t), so one could equivalently consider $h(Y|K)$, $h(Z|K)$, and $I(Z;T|K)$.

DWR$\to -\infty$ dB

If DWR $\to -\infty$ dB, the two peaks of $f_{Y|K}(y|s)$ (centered at $\pm s$) are far enough to consider that the tails of each bell of the distribution do not overlap, thus $h(Y|K=s) \simeq h(X) + \log(2)$. Moreover, in order to calculate the pdf of Z, a similar reasoning to that in Sect. 4.2.1.1.1 can be used, obtaining that Z is nothing but T scaled by $\pm s$, i.e.

$$f_{Z|K}(z|s) \simeq \begin{cases} \frac{f_T(z/|s|)}{2|s|} & \text{if } \ z \geq 0 \\ \frac{f_T(-z/|s|)}{2|s|} & \text{otherwise} \end{cases}.$$

Using this assumption,

$$h(Z|K=s) \simeq 1 + \log\left(\frac{\sigma_T}{\sqrt{2}}\right) + \frac{\gamma}{2} + \log(2) + \log(|s|),$$

yielding

$$I(Z;T|K=s) = 1 - \log(2) + \gamma - \frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2}{|s|^2}\right).$$

Therefore,

$$I(Z;T|K) = 1 - \log(2) + \gamma - \int_{-\infty}^{\infty} f_K(s)\frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2}{|s|^2}\right)ds,$$

and in the particular case where $P(K=-\sigma_W) = P(K=+\sigma_W) = \frac{1}{2}$, the last formula can be rewritten as

$$I(Z;T|K) = 1 - \log(2) + \gamma - \frac{1}{2}\log\left(2\pi e\mathrm{DWR}\right),$$

coinciding with its equivalent expression in Sect. 4.2.1.1.1, i.e., (4.17).

4.2.2.1.2 Dirty Paper Coding

In order to calculate $I(Z;T|K)$ when our DPC-based technique is used, $f_{Y|K}(y|d)$ is required; as we are assuming that the two possible symbols are equiprobable and $\alpha \geq 0.5$, $f_{Y|K}(y|d)$ can be written, for $\alpha \geq 0.5$, as

$$f_{Y|K}(y|d) = \begin{cases} \sum_{i=-\infty}^{\infty} \frac{1}{2}\frac{1}{1-\alpha}\sum_{k=0}^{1} f_X\left(\frac{y-\alpha(i\Delta+d+k\Delta/2)}{1-\alpha}\right) \\ \qquad \text{if } |y-(i\Delta+d+k\Delta/2)| < (1-\alpha)\Delta/2 \\ 0 \qquad \text{otherwise} \end{cases}.$$

In order to obtain the pdf of Z, the expression (4.18) is used, replacing $f_{Y|K,B}(y|d,0)$ by $f_{Y|K}(y|d)$.

DWR$\rightarrow \infty$ dB

On one hand, if D follows a uniform distribution from $-\Delta/2$ to $\Delta/2$, W can accurately be approximated by $U(-\Delta/2, \Delta/2)$. When K is known, $h(Y|K)$ can be calculated as

$$\begin{aligned} h(Y|K)) &= h(Y) - I(Y;K) = h(Y) - h(K) + h(K|Y) \\ &\simeq h(X) - \log(\Delta) + \log(\Delta(1-\alpha)) + \log(2) \qquad (4.33) \\ &= h(X) + \log\left(2(1-\alpha)\right), \end{aligned}$$

where in (4.33), since DWR$\rightarrow \infty$ is assumed, $h(Y) \simeq h(X)$. In addition, $h(K|Y) \simeq \log(\Delta(1-\alpha)) + \log(2)$, these components are respectively due to the uncertainty of the self-noise and the ignorance of the embedded symbol. The

difference between the previous expression and (4.19) is $\log(2)$, as in the current case two shifted versions of the estimation error of K given Y must be considered.

As in Sect. 4.2.1.1.2, if DWR$\rightarrow\ \infty$ dB, Z can be approximated by a zero-mean Laplacian distribution with standard deviation $\sqrt{2}\sigma_T\sigma_X$ and, thus, $h(Z|K) \simeq \log(2e\sigma_T\sigma_X)$. Using both $h(Y|K)$ and $h(Z|K)$ approximations, the mutual information between Z and T knowing the secret key K is

$$I(Z;T|K) \simeq \frac{1}{2}\left(1+\gamma-\log(\pi)\right)-\log(2(1-\alpha)).$$

Notice that the previous expression and (4.20) just differ in the aforementioned term $\log(2)$.

DWR$\rightarrow -\infty$ dB

In this case, $f_{Z|K}(z|d)$ is calculated by the average of the distribution obtained by embedding the symbol $B=1$ and $B=0$, when the dither vector is equal to d, i.e., $K=d$. Here, there are at least two peaks in the pdf, whose distance increases as the DWR goes to $-\infty$ dB, and therefore $I(Z;T|K)$ increases.

This scenario requires to divide the analysis into two cases:

- In the first one, the pdf of Y when $K=0$ (also comprising the analogous cases $K=\pm\Delta/2$) is approximated by

  $$f_{Y|K}(y|0) \simeq \frac{f_X(y/(1-\alpha))}{2(1-\alpha)}+\frac{f_X(\frac{y-\alpha\Delta/2}{1-\alpha})}{2(1-\alpha)}U(y-\alpha\Delta/2)+\frac{f_X(\frac{y+\alpha\Delta/2}{1-\alpha})}{2(1-\alpha)}U(-y+\alpha\Delta/2),$$

  where in the previous expression $U(\cdot)$ denotes the Heaviside step function and the differential entropy of Y is $h(Y|K=0) \simeq h(X)+\log(2)+\log(1-\alpha)$.

  In order to obtain the distribution of Z, the following approximation to the pdf of Y is used

  $$f_{Y|K}(y|0) \simeq \frac{\delta(y-\alpha\Delta/2)}{4}+\frac{\delta(y+\alpha\Delta/2)}{4}+\frac{1}{2(1-\alpha)}f_X(y/(1-\alpha)).$$

  In this way, the resulting pdf of Z is

  $$f_{Z|K}(z|0) \simeq \frac{1}{4\alpha\Delta/2}f_T\left(\frac{z}{\alpha\Delta/2}\right)+\frac{1}{4\alpha\Delta/2}f_T\left(\frac{-z}{\alpha\Delta/2}\right)+\frac{1}{2}\frac{e^{\frac{-|z|}{(1-\alpha)\sigma_T\sigma_X}}}{2\sigma_T\sigma_X(1-\alpha)}.$$

  Assuming that the three components of the previous expression are not overlapped, one can approximate the differential entropy of Z by

  $$\begin{aligned} h(Z|K=0) \simeq & \frac{1}{2}\left(1+\log(\sigma_T/\sqrt{2})+\gamma/2+\log(\alpha\Delta)+\log(2)\right) \\ & +\frac{1}{2}\left(\log(2e\sigma_X\sigma_T)+\log(1-\alpha)+\log(2)\right). \end{aligned}$$

Finally, the mutual information between Z and Y when $K = 0$ (and also $K = \pm\Delta/2$) is

$$I(Z;T|K=0) \simeq \frac{1}{2} + \frac{3\gamma}{4} + \frac{1}{2}\log\left(\frac{\alpha\Delta}{(1-\alpha)}\right) - \frac{\log(8\pi^2)}{4} - \frac{\log(\sigma_X)}{2}.$$

- However, if D is in $(-\Delta/2, 0) \cup (0, \Delta/2)$, the pdf of Y can be approximated by

$$f_{Y|K}(y|d) \simeq \begin{cases} \frac{f_X(\frac{y-\alpha d}{1-\alpha})}{2(1-\alpha)} + \frac{f_X(\frac{y-\alpha d-\alpha\Delta/2}{1-\alpha})}{2(1-\alpha)} & \text{if } \frac{-\Delta}{2} < d < 0 \\ \frac{f_X(\frac{y-\alpha d}{1-\alpha})}{2(1-\alpha)} + \frac{f_X(\frac{y-\alpha d+\alpha\Delta/2}{1-\alpha})}{2(1-\alpha)} & \text{if } 0 < d < \frac{\Delta}{2} \end{cases},$$

the differential entropy of Y $h(Y|K=d) \simeq \log(2\pi e\sigma_X^2)/2 + \log(2) + \log(1-\alpha)$.

Moreover, since the two peaks of the pdf of Y are very narrow compared to Δ, we can apply the same considerations of the previous case to calculate the distribution of Z, i.e., the pdf of Z can approximated by a scaled version of the pdf of T, namely

$$h(Z|K=d) \simeq 1 + \log(\sigma_T/\sqrt{2}) + \gamma/2 + \log(2) + \frac{1}{2}\left(\log(\alpha^2|d|\cdot||d|-\Delta/2|)\right).$$

Therefore, the mutual information becomes

$$I(Z;T|K=d) = \frac{1}{2} + \gamma - \log(1-\alpha) - \frac{\log(\pi\sigma_X^2)}{2} + \frac{\log\left(\frac{1}{8}\alpha^2|d|\cdot||d|-\Delta/2|\right)}{2}.$$

Since the logarithm is a monotonically increasing function, the first derivative of the argument of the logarithm function of the fifth addend of the previous expression with respect to d for $0 < d < \Delta/2$, where this argument is a continuous function, is obtained to study how this parameter affects the mutual information. The first derivative is $\alpha^2(\Delta - 4d)/16$; it has one root at $d = \Delta/4$ and it straightforward to demonstrate that the second derivative with respect to d is negative; therefore, $I(Z;T|K=d)$ has a maximum at $d = \Delta/4$ if $d \in (0, \Delta/2)$. A similar analysis for $d \in (-\Delta/2, 0)$ reveals $I(Z;T|K=d)$ has a maximum at $d = -\Delta/4$, being intuitively consistent with the symmetry of the problem for this scenario.

Finally, $I(Z;T|K=d)$ is integrated over the distribution of K in order to obtain $I(Z;T|K)$, yielding

$$I(Z;T|K) \simeq -\frac{1}{2} + \gamma - \log(1-\alpha) - \frac{1}{2}\log\left(\frac{32\pi\sigma_X^2}{\alpha^2\Delta^2}\right).$$

Although the mutual information expressions were independently obtained in this section, it is worth pointing out that in some cases they can be systematically obtained from the expressions derived in Sect. 4.2.2.1. The basic idea behind

this systematic approach is that given a random variable U, if one can write its pdf as $f_U(u) = f_{U_1}(u)/2 + f_{U_2}(u)/2$, i.e., as the equiprobable mixture of two pdfs, where the support sets of f_{U_1} and f_{U_2} are disjoint, then it is easy to see that $h(U) = h(U_1)/2 + h(U_2)/2 + \log(2)$. In the scenarios where this property is asymptotically verified, it is also possible to calculate the expressions of the mutual entropy without knowing the inserted symbol from the expressions when the inserted pilot is known.

4.2.2.1.3 Numerical Results

Fig. 4.12 shows $I(Z;T|K)$ as a function of the DWR and the approximations of $I(Z;T|K)$ for DWR$\to \pm\infty$ dB. Note that this figure matches with Fig. 4.8 as it was expected from the analysis. Furthermore, these curves show independence with respect to the actual value of σ_T, verifying what the approximations pointed out.

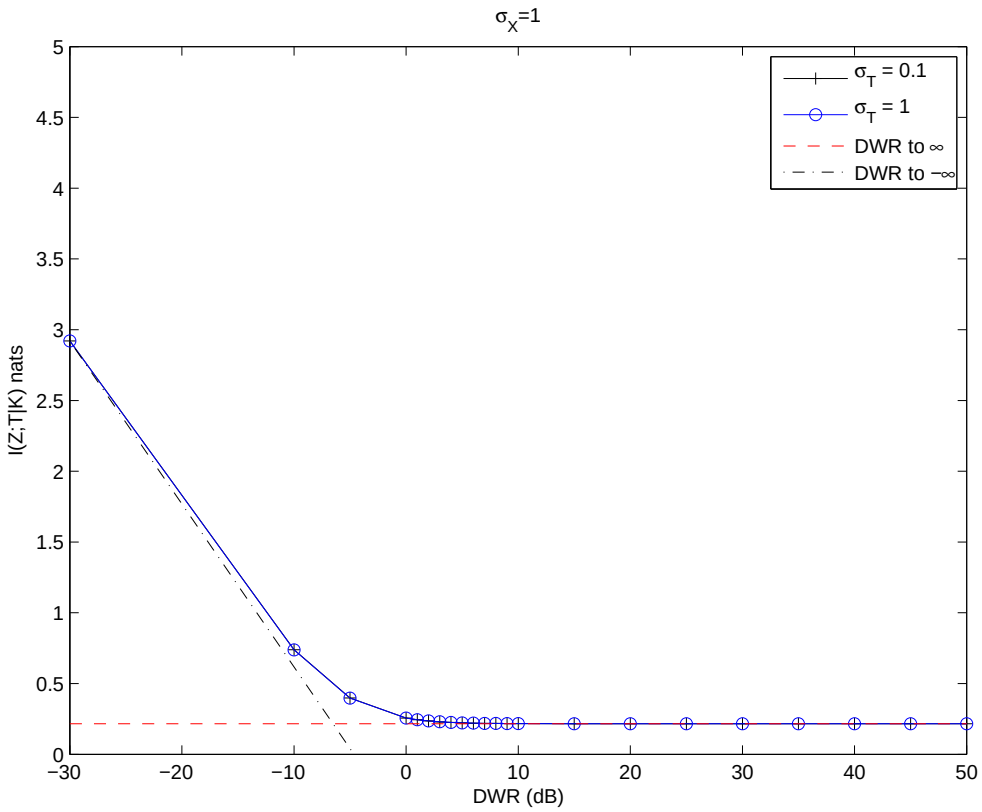


Figure 4.12: $I(Z;T|K)$ vs. DWR curves (solid lines) for different σ_T using the analyzed Add-SS/SIT technique with the corresponding approximations for DWR$\to \infty$ dB (dashed line) and for DWR$\to -\infty$ dB (dashdot line). $\sigma_X = 1$ and S following an equiprobable binary antipodal distribution.

On the other hand, Fig. 4.13 depicts $I(Z;T|K = d)$ as a function of the value

of DWR for several values of α and different values of d, as well as the approximation of $I(Z;T|K)$ when the DWR goes to infinity. As it was previously stated, when the DWR goes to infinity $I(Z;T|K)$ differs from $I(Z;T|K,B)$ (plotted in Fig. 4.6) by log(2).

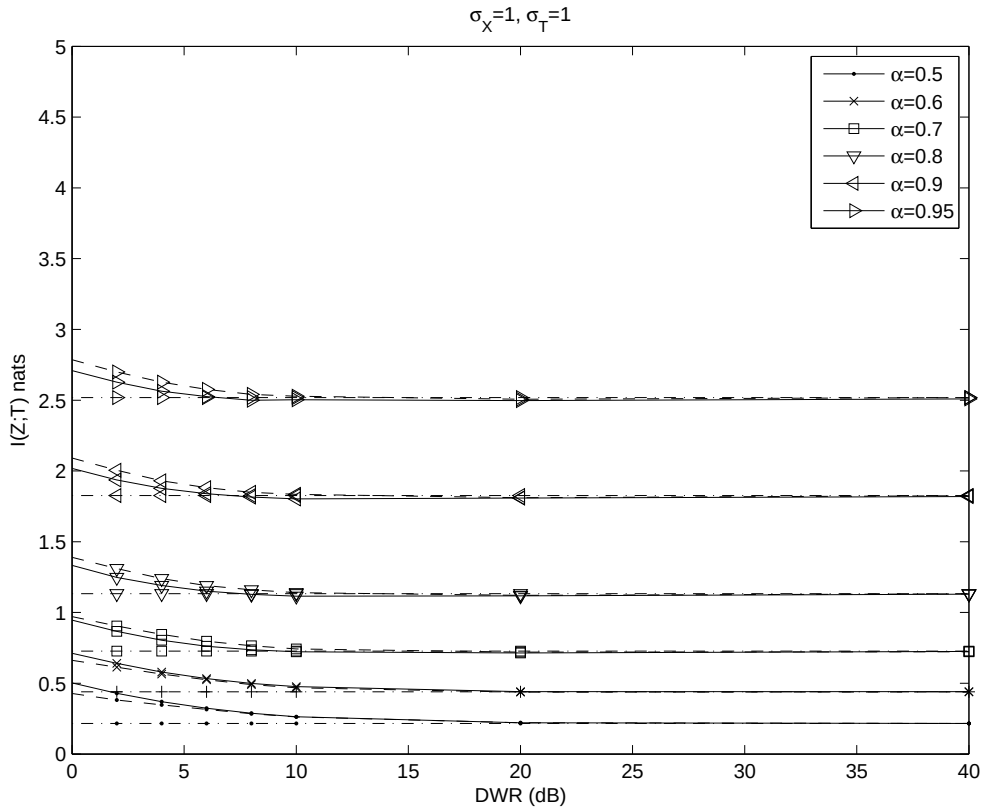


Figure 4.13: $I(Z;T|K=d)$ vs. DWR curves for different α and d (solid lines for representing $d=-0.3\Delta$, while dashed lines for $d=0.5\Delta$) using DPCE with their approximations for DWR$\rightarrow\infty$ dB (dashdot lines) with $\sigma_X=1$ and $\sigma_T=1$.

Fig. 4.14 shows the experimental and the approximation curves of $I(Z;T|K=d)$ when DWR $\rightarrow -\infty$ dB. It is worth pointing out that, as discussed above, $I(Z;T|K=d)$ reaches the maximum if $d=\Delta/4$, i.e., the positive and the mirrored Rayleigh pdfs are scaled by the same factor (i.e., $\alpha\Delta/4$).

4.2.2.2 With Channel Noise

Under the same considerations stated in the case of embedding pilot sequences with channel noise described in Sect. 4.2.1.2, i.e., in order to analyze the distribution of Z, one can discard the influence of the watermark and the channel noise. Therefore, as approximated above, one can model Z as a Laplacian distribution with zero-mean and variance $2\sigma_T^2\sigma_X^2$; thus, $h(Z|K)$ can be approximated by

$$h(Z|K)\approx 1+\log(2\sigma_T\sigma_X). \tag{4.34}$$

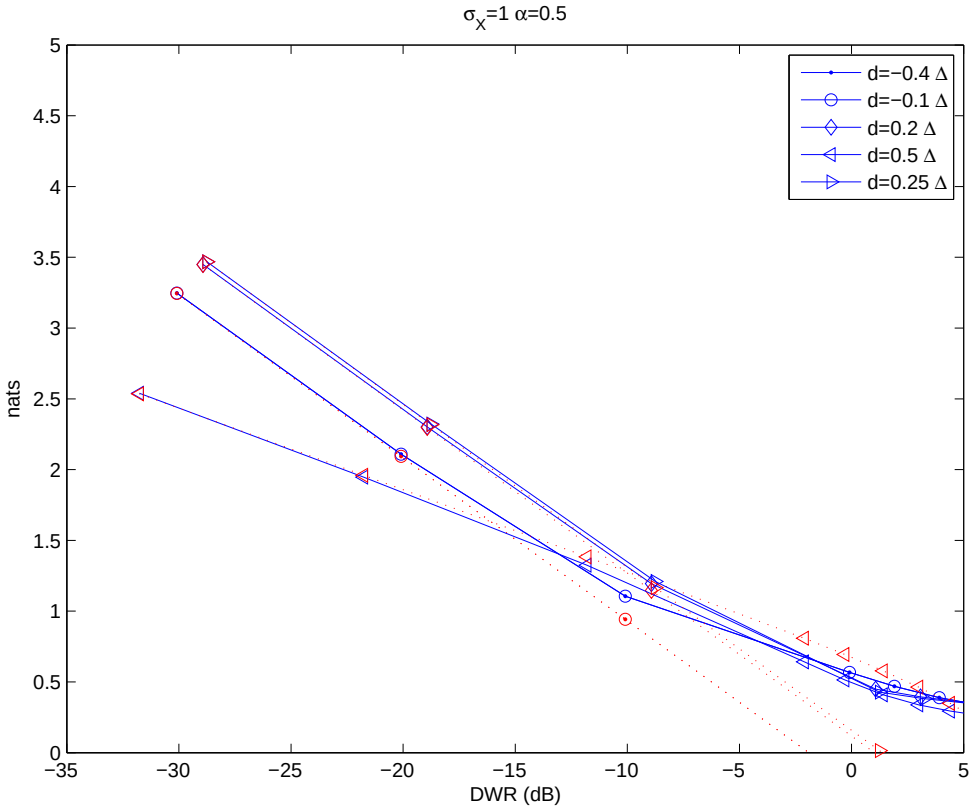


Figure 4.14: $I(Z;T|K=d)$ vs. DWR experimental curves (solid lines), the corresponding approximations (dotted lines) when DWR goes to $-\infty$ dB using DPCE with the following parameters: $\alpha = 0.5$, $\sigma_X = 1$ and $\sigma_T = 1$.

4.2.2.2.1 Additive Spread Spectrum and Superimposed Training

Under the hypotheses of high-SNR and from the distribution of Y for Add-SS/SIT (4.32), the pdf of Z given t can be approximated by

$$\frac{1}{2\sigma_x t\sqrt{2\pi}} e^{-\frac{(y-s)^2}{2\sigma_x^2 t^2}} + \frac{1}{2t\sigma_x\sqrt{2\pi}} e^{-\frac{(y+s)^2}{2\sigma_x^2 t^2}};$$

therefore, assuming that the two Gaussian pdfs centered at $\pm s$ are not overlapped, it is straightforward to approximate the differential entropy of Z given t as

$$h(Z|T=t,K) \approx \log(2) + \frac{1}{2}\left(2\pi e t^2 \sigma_X^2\right),$$

and by averaging with respect to the distribution T, the conditional differential entropy can be approximated as

$$h(Z|T,K) \approx \log(2) + \frac{1}{2}\left[1 - \gamma + \log(\pi\sigma_T^2\sigma_X^2)\right].$$

Therefore, the mutual information between Z and T knowing K in this scenario as the result of subtracting the previous expression from (4.34) is

$$I_{\text{imposed}}(Z;T|K) \approx \frac{1}{2}\left[1 + \gamma - \log(\pi)\right] = \kappa.$$

4.2.2.2.2 Dirty Paper Coding

As in Sect. 4.2.1.2.2, the approximations of the pdf of Z are used in order to obtain a closed-form expression of $I(Z;T)$ when the secret key is known but, in this case, the embedded symbol is unknown. In this scenario, the position of the active centroids of the pdf of Y is separated by $\Delta/2$ instead of Δ as whenever the embedded symbol is known at the estimator. This difference is taken into account to properly modify the approximations of the required pdfs. Note that also the hypotheses must be adapted; specifically, the definition of $\text{TNQR}(t_0)$ is modified considering that the distance of contiguous centroids is divided by 2 and, thus, the second moment of the lattice is $\Delta^2/(12\cdot 4)$.

Low-SNR Case

It is straightforward to adapt the pdf of Z obtained for the low-SNR case (3.9) in Sect. 3.1.1 by taking into account that, under the defined hypotheses, only the structured part of the pdf (i.e., the part within the outer parentheses) is affected by dividing by two the distance between active centroids, obtaining

$$f_{Z|T,K}^{\text{low-SNR},2'}(z|t,d) \approx \frac{e^{-\frac{z^2}{2t^2\sigma_X^2}}}{\sqrt{2\pi t^2\sigma_X^2}}\left(1 + 2e^{-\frac{2\pi^2\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)}{(\Delta/2)^2 t^2}}\cos\left(\frac{2\pi z}{\Delta/2t} - \frac{2\pi d}{\Delta/2}\right)\right).$$

By following the same approach done in the corresponding analysis carried out in Sect. 4.2.1.2.2, the approximation of $h(Z|T,K)$ becomes

$$h(Z|T,K) \approx -\frac{8\sqrt{2}\pi e^{\frac{1}{3}(-4)\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{4\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} + \log(\sigma_T\sigma_X) + \frac{1}{2}(-\gamma+1+\log(4)+\log(\pi)).$$

And since, as reasoned above, $h(Z|K)$ can be approximated considering that Z follows a Laplacian distribution, the mutual information between Z and T without inserting pilot symbols can be approximated by

$$\begin{aligned} I(Z;T|K) &\approx \frac{8\sqrt{2}\pi e^{\frac{1}{3}(-4)\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{4\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} + \frac{1}{2}(1+\gamma-\log(\pi)) \\ &= \frac{8\sqrt{2}\pi e^{\frac{1}{3}(-4)\pi^2(1-\alpha)^2}\sigma_N K_{1,\mathrm{B}}\left(\frac{4\sqrt{2}\pi\sigma_N}{\Delta\sigma_T}\right)}{\Delta\sigma_T} + \kappa. \end{aligned} \tag{4.35}$$

High-SNR Case

In order to take into account that the separation between contiguous active centroids of the distribution of Y is reduced by 2 in the high-SNR case, the entropy of Z given t can be approximated by

$$\begin{aligned} h(Z|T=t,K) &\approx \frac{1}{2}\log(2\pi e\sigma_X^2) - \log(\Delta/2) \\ &\quad + \frac{1}{2}\log\left(2\pi e\left[\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right]\right), \end{aligned} \tag{4.36}$$

where the difference between the previous expression and (4.29) is the change of $\log(\Delta/2)$ by $\log(\Delta)$ in the discrete entropy of the used centroid at the embedder. Then, (4.36) is averaged over the distribution of T to obtain the conditional differential entropy $h(Z|T,K)$

$$\begin{aligned} h(Z|T,K) \approx& \frac{1}{2}\left(-e^{\frac{6\sigma_N^2}{(1-\alpha)^2\Delta^2\sigma_T^2}}\,\mathrm{Ei}\left(-\frac{6\sigma_N^2}{(1-\alpha)^2\Delta^2\sigma_T^2}\right) + 2\log(\sigma_N) + 1 + \log(2\pi)\right) \\ &- \log\left(\frac{\Delta}{2}\right) + \frac{1}{2}\log\left(2\pi e\sigma_X^2\right); \end{aligned}$$

thus, the mutual information between Z and T can be approximated for this case as

$$I(Z;T|K) \approx \frac{1}{2}\left(e^{\frac{6\sigma_N^2}{(1-\alpha)^2\Delta^2\sigma_T^2}}\,\mathrm{Ei}\left(-\frac{6\sigma_N^2}{(1-\alpha)^2\Delta^2\sigma_T^2}\right)\right) + \log\left(\frac{\Delta\sigma_T}{2\pi\sigma_N}\right). \tag{4.37}$$

4.2.2.2.3 Numerical Results

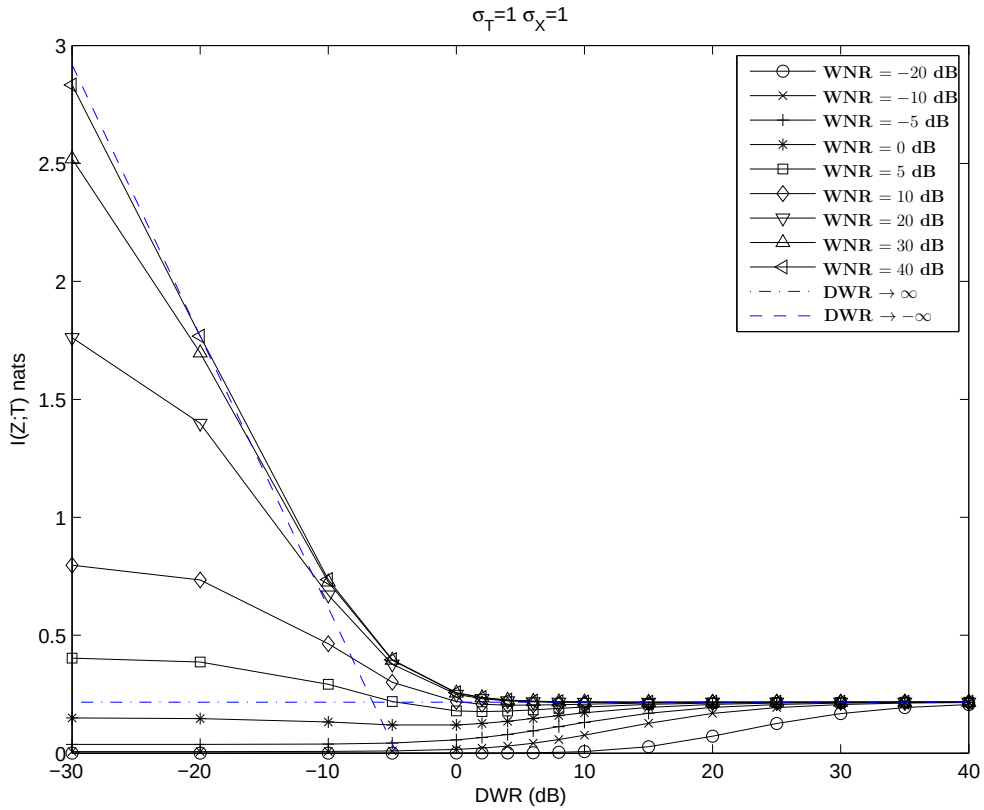


Figure 4.15: $I(Z;T|K)$ vs. DWR using Add-SS/SIT when the embedded bit is unknown, $\sigma_T = 1$ and S follows an equiprobable binary antipodal distribution. The dashdot line corresponds to the noiseless case when DWR $\to \infty$ dB, while the dashed line corresponds to DWR $\to -\infty$ dB.

Fig. 4.15 shows empirical curves of the mutual information when Add-SS/SIT is used in several scenarios with noise when a pilot is not embedded. In addition, the approximations for DWR $\to \pm\infty$ dB of the noiseless case are depicted. It can be shown that the empirical curves asymptotically converge to the respective approximations for DWR $\to \infty$ and in accordance to DWR $\to -\infty$ dB if WNR $\to \infty$, which corresponds to the noiseless case.

Fig. 4.16 and Fig. 4.17 depict $I(Z;T|K)$ obtained numerically vs. DWR for several values of WNR using DPCE considering the noisy scenario for $\alpha = 0.5$ and $\alpha = 0.95$, respectively. In addition to the approximation for DWR$\to \infty$ of the noiseless case, the approximations of the mutual information calculated in this section are also shown.

By analyzing these two figures, one can realize that, confirming the intuition, the larger the value of WNR, the closer the numerical results to the corresponding approximation of the noiseless scenario. Apart from this, our approximations get close to the numerical results whenever the corresponding hypotheses hold. On these scenarios, one can state that the introduced approximations require large values of DWR; in addition, on one hand, the low-SNR approximation needs WNR ≤ 0 dB; while, on the other hand, the high-SNR approximation requires large values of WNR.

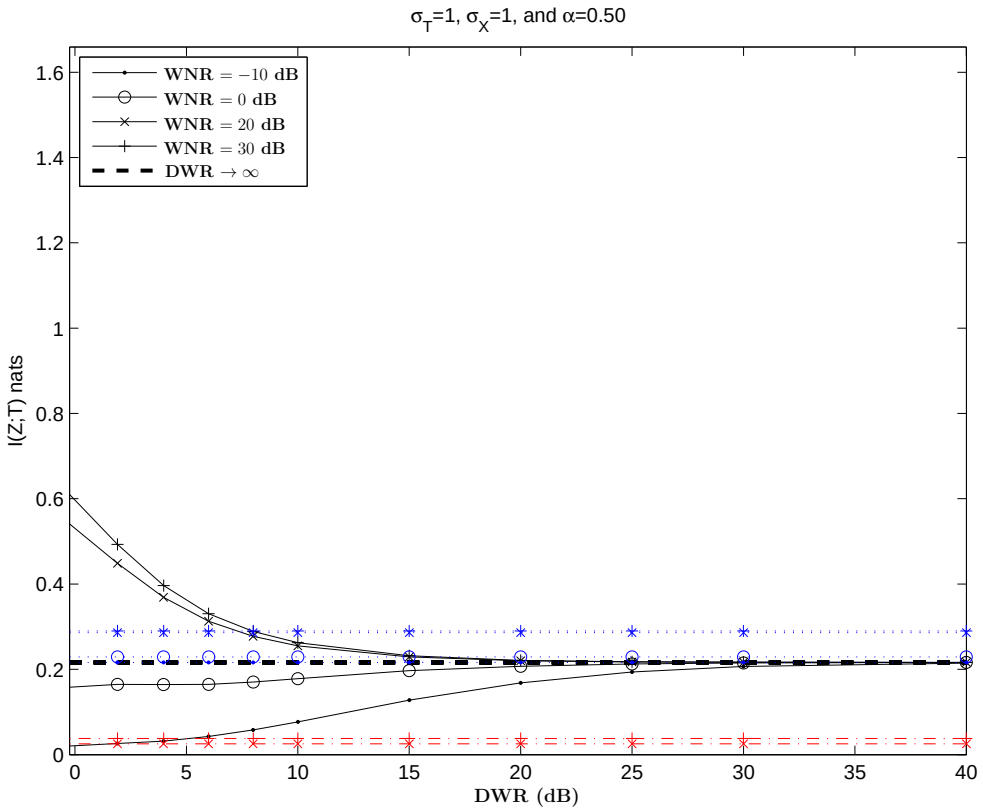


Figure 4.16: $I(Z;T|K=0)$ vs. DWR curves using DPCE and $\alpha = 0.5$ when the embedded symbol is unknown when $I(Z;T|K=0)$ obtained numerically (solid lines), with the approximation for low-SNR cases (4.35) (dotted lines), and (4.37) for high-SNR cases (dashdot lines). The dashed line corresponds to the noiseless case for DWR$\rightarrow \infty$ dB.

Focusing on the curves for WNR $= -10$ dB of Fig 4.16, one can see that the mutual information tends to κ when DWR $\rightarrow \infty$; this verifies that the mutual information for DPCE is always, as long as $\alpha > 0.5$, larger than or equal to that of Add-SS/SIT when DWR$\rightarrow \infty$ dB, being only equal if WNR tends to $-\infty$ dB or $\alpha = 0.5$ since in that case there is not structure in the pdf of Z.

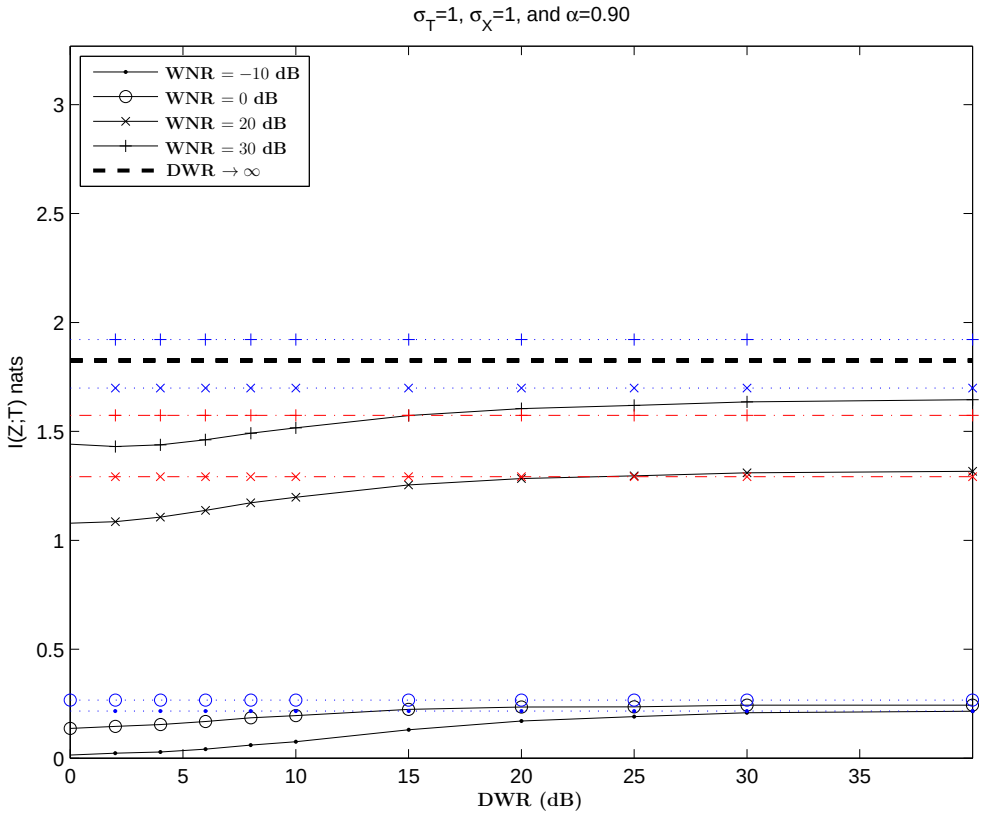


Figure 4.17: $I(Z;T|K=0)$ vs. DWR curves using DPCE and $\alpha = 0.9$ when the embedded bit is unknown obtained numerically (solid lines), with the approximation for low-SNR cases (4.35) (dotted lines), and (4.37) for high-SNR cases (dashdot lines). The dashed line corresponds to the noiseless case for DWR$\rightarrow \infty$ dB.

Appendix

4.A Analysis of the Gaussian Distributed Z Approximation

Assuming that the dither sequence $\boldsymbol{d}$ is unknown, X is a zero-mean Gaussian distributed random variable, and DWR takes large values, the pdf of Z can be approximated assuming that the involved signals are Gaussian distributed, yielding

$$f_{Z|T}^{\text{Gauss}}(z|t_0) \approx \frac{e^{-\frac{z^2}{2\left(\left(\sigma_X^2+\sigma_W^2\right)t_0^2+\sigma_N^2\right)}}}{\sqrt{2\pi\left(\left(\sigma_X^2+\sigma_W^2\right)t_0^2+\sigma_N^2\right)}}. \tag{4.38}$$

4.A.1 Cramér-Rao Bound

It is worth calculating the CRB corresponding to this scenario for the sake of comparison with other methods introduced in this work. The CRB is calculated as

$$\begin{aligned}\text{CRB} &= -\left[\text{E}\left\{\frac{\partial^2 \log\left(f_{\boldsymbol{Z}|T}^{\text{Gauss}}(\boldsymbol{z}|t)\right)}{\partial t^2}\right\}\right]^{-1} \\ &= -\left[L\text{E}\left\{\frac{\partial^2 \log\left(f_{Z|T}^{\text{Gauss}}(z|t)\right)}{\partial t^2}\right\}\right]^{-1}, \end{aligned} \tag{4.39}$$

where the second equality in (4.39) can be set because the components of $\boldsymbol{Z}$ are i.i.d.. The second derivative of the logarithm of the pdf of Z with respect to t_0 is obtained

$$\begin{aligned}&\frac{\partial^2 \log\left(f_{Z|T}^{\text{Gauss}}(z|t_0)\right)}{\partial t_0^2} \\ &= \frac{\left(\sigma_X^2+\sigma_W^2\right)\left(-\sigma_N^4+\left(\sigma_X^2+\sigma_W^2\right)^2 t_0^4+\left(\sigma_N^2-3\left(\sigma_X^2+\sigma_W^2\right)t_0^2\right)z^2\right)}{\left(\sigma_N^2+\left(\sigma_X^2+\sigma_W^2\right)t_0^2\right)^3},\end{aligned}$$

and its expectation becomes

$$\mathrm{E}\left\{\frac{\partial^2 \log\left(f_{Z|T}^{\mathrm{Gauss}}(z|t_0)\right)}{\partial t_0^2}\right\} = -\frac{2\left(\sigma_X^2+\sigma_W^2\right)^2 t_0^2}{\left(\sigma_N^2+\left(\sigma_X^2+\sigma_W^2\right)t_0^2\right)^2}.$$

Using the previous expression in (4.39),

$$\mathrm{CRB}_{\mathrm{Var}}^{'} = \frac{\left(\sigma_N^2+\left(\sigma_X^2+\sigma_W^2\right)t_0^2\right)^2}{2L\left(\sigma_X^2+\sigma_W^2\right)^2 t_0^2}.$$

This expression can be simplified considering that $\sigma_X^2 \gg \sigma_W^2$ obtaining

$$\mathrm{CRB}_{\mathrm{Var}} = \frac{\left(\sigma_N^2+\sigma_X^2 t_0^2\right)^2}{2L\sigma_X^4 t_0^2}. \tag{4.40}$$

In addition, if $t_0^2\sigma_X^2 \gg \sigma_N^2$, then

$$\mathrm{CRB}_{\mathrm{Var}} \approx \frac{t_0^2}{2L}.$$

4.B Expectation of the Second Derivative of $\log\left(f_{Z|T,K}^{\textbf{low-SNR},b}(z|t,d)\right)$

The second derivative of (4.4) with respect to t is given by

$$\frac{\partial^2 2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+(1-\alpha)^2\Delta^2t^2/12\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi\sigma_X^2tz}{\Delta(\sigma_N^2+\sigma_X^2t^2)}-\frac{2\pi d}{\Delta}\right)}{\partial t^2}=\frac{2e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\pi\sigma_X^2}{9\Delta^4\left(\sigma_N^2+\sigma_X^2t^2\right)^4}$$
$$\times\Bigg[\pi\Bigg(144\pi^2\sigma_N^4\sigma_X^6t^2+(1-\alpha)^2\Delta^4\sigma_N^2\left(-3\sigma_N^4+\left(6+(1-\alpha)^2\pi^2\right)\sigma_N^2\sigma_X^2t^2+9\sigma_X^4t^4\right)$$
$$+\Big(12\Delta^2\sigma_X^2\left(3\sigma_N^6-2\left(3+(1-\alpha)^2\pi^2\right)\sigma_N^4\sigma_X^2t^2-9\sigma_N^2\sigma_X^4t^4-3\left(\sigma_N^2-\sigma_X^2t^2\right)^2z^2\right)\Big)\Bigg)$$
$$\times\cos\left(\frac{2\pi\left(d-\frac{\sigma_X^2tz}{\sigma_N^2+\sigma_X^2t^2}\right)}{\Delta}\right)+12\Delta\sigma_X^2t\Bigg(12\pi^2\sigma_N^2\sigma_X^2(\sigma_N-\sigma_Xt)(\sigma_N+\sigma_Xt)+$$
$$\Delta^2\left(-\left(9+(1-\alpha)^2\pi^2\right)\sigma_N^4+\left(-6+(1-\alpha)^2\pi^2\right)\sigma_N^2\sigma_X^2t^2+3\sigma_X^4t^4\right)\Bigg)$$
$$\times z\sin\left(\frac{2\pi\left(d-\frac{\sigma_X^2tz}{\sigma_N^2+\sigma_X^2t^2}\right)}{\Delta}\right)\Bigg].$$

Therefore, using the previous expression, the expectation of the second derivative can be expressed as

$$\mathrm{E}\left\{\frac{\partial^2 2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+(1-\alpha)^2\Delta^2t^2/12\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi\sigma_X^2tz}{\Delta(\sigma_N^2+\sigma_X^2t^2)}-\frac{2\pi d}{\Delta}\right)}{\partial t^2}\right\}$$
$$\approx\int_{-\infty}^{\infty}\frac{\partial^2 2e^{-\frac{2\pi^2\sigma_X^2\sigma_N^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi\sigma_X^2t\tau}{\Delta(\sigma_N^2+\sigma_X^2t^2)}-\frac{2\pi d}{\Delta}\right)}{\partial t^2}f_{Z|T,K}^{\text{low-SNR}}(\tau|t,d)d\tau,$$

which is

$$
\begin{aligned}
&\frac{2e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\pi^2\sigma_X^2}{9\Delta^4\left(\sigma_N^2+\sigma_X^2t^2\right)^4}\Bigg(144\pi^2\sigma_N^4\sigma_X^6t^2\\
&\times\left(e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}+e^{-\frac{2\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi d}{\Delta}\right)+e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+48\sigma_X^2\right)t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\right.\\
&\left.\times\cos\left(\frac{4\pi d}{\Delta}\right)\right)+(1-\alpha)^2\Delta^4\sigma_N^2\left(-3\sigma_N^4+\left(6+(1-\alpha)^2\pi^2\right)\sigma_N^2\sigma_X^2t^2+9\sigma_X^4t^4\right)\\
&\times\left(e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}+e^{-\frac{2\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi d}{\Delta}\right)+e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+48\sigma_X^2\right)t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\right.\\
&\left.\times\cos\left(\frac{4\pi d}{\Delta}\right)\right)+12\Delta^2\sigma_X^2\Big(\left(3\sigma_N^6-2\left(3+(1-\alpha)^2\pi^2\right)\sigma_N^4\sigma_X^2t^2-9\sigma_N^2\sigma_X^4t^4\right)\\
&\times\left(e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+(1-\alpha)^2\Delta^2t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}+e^{-\frac{2\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi d}{\Delta}\right)+e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+48\sigma_X^2\right)t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\right.\\
&\left.\times\cos\left(\frac{4\pi d}{\Delta}\right)\right)-\frac{3\left(\sigma_N^2-\sigma_X^2t^2\right)^2}{\Delta^2}\Bigg(e^{-\frac{2\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\left(-4\pi^2\sigma_X^4t^2+\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)\right)\cos\left(\frac{2\pi d}{\Delta}\right)\\
&+e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+48\sigma_X^2\right)t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\Bigg(\Delta^2e^{\frac{8\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\left(\sigma_N^2+\sigma_X^2t^2\right)\\
&+\left(\Delta^2\sigma_N^2+\sigma_X^2(\Delta-4\pi\sigma_X)(\Delta+4\pi\sigma_X)t^2\right)\cos\left(\frac{4\pi d}{\Delta}\right)\Bigg)\Bigg)\Bigg)\\
&+24\sigma_X^4t^2\Bigg(12\pi^2\sigma_N^2\sigma_X^2(\sigma_N-\sigma_Xt)(\sigma_N+\sigma_Xt)\\
&+\Delta^2\left(-\left(9+(1-\alpha)^2\pi^2\right)\sigma_N^4+\left(-6+(1-\alpha)^2\pi^2\right)\sigma_N^2\sigma_X^2t^2+3\sigma_X^4t^4\right)\Bigg)\\
&\times\left(-e^{-\frac{2\pi^2\sigma_X^4t^2}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\cos\left(\frac{2\pi d}{\Delta}\right)-2e^{-\frac{\pi^2\sigma_X^2\left(12\sigma_N^2+\left((1-\alpha)^2\Delta^2+48\sigma_X^2\right)t^2\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\sin\left(\frac{4\pi d}{\Delta}\right)\right)\Bigg).
\end{aligned}
$$

Chapter 5

Practical Estimation Algorithms

As discussed in Sect. 3.2.3, the proposed ML-based cost-functions ((3.11) and (3.12) for the low-SNR case, and (3.15) for the high-SNR case) show several local maxima/minima. This makes standard optimization algorithms unsuitable. In addition, the application of brute-force techniques is computationally prohibitive. To tackle this issue, a set of *ad-hoc* estimation algorithms is proposed in order to obtain accurate estimations with affordable computational resources.

As illustratively shown in Fig. 5.1 for $t_0 \geq 0$, the proposed estimation techniques can be described in a modular way. First, based on the characteristic of the cost functions, an optimization search-interval $[t_-, t_+]$ is calculated. Then, the search-interval is sampled based on the statistical properties of the objective function obtaining a candidate set $\mathcal{T}$. Given this set of candidate values, a local optimization is carried out, yielding a set of locally optimal solutions $\mathcal{T}^*$; $\hat{t}_0(\boldsymbol{z})$ is finally selected as that element of $\mathcal{T}^*$ that minimizes the cost function. For each of these procedures, we propose several alternatives, which should be selected according to the requirements of the specific application scenario. In the end of this section, the performance of our techniques is analyzed and discussed with the support of extensive simulations.

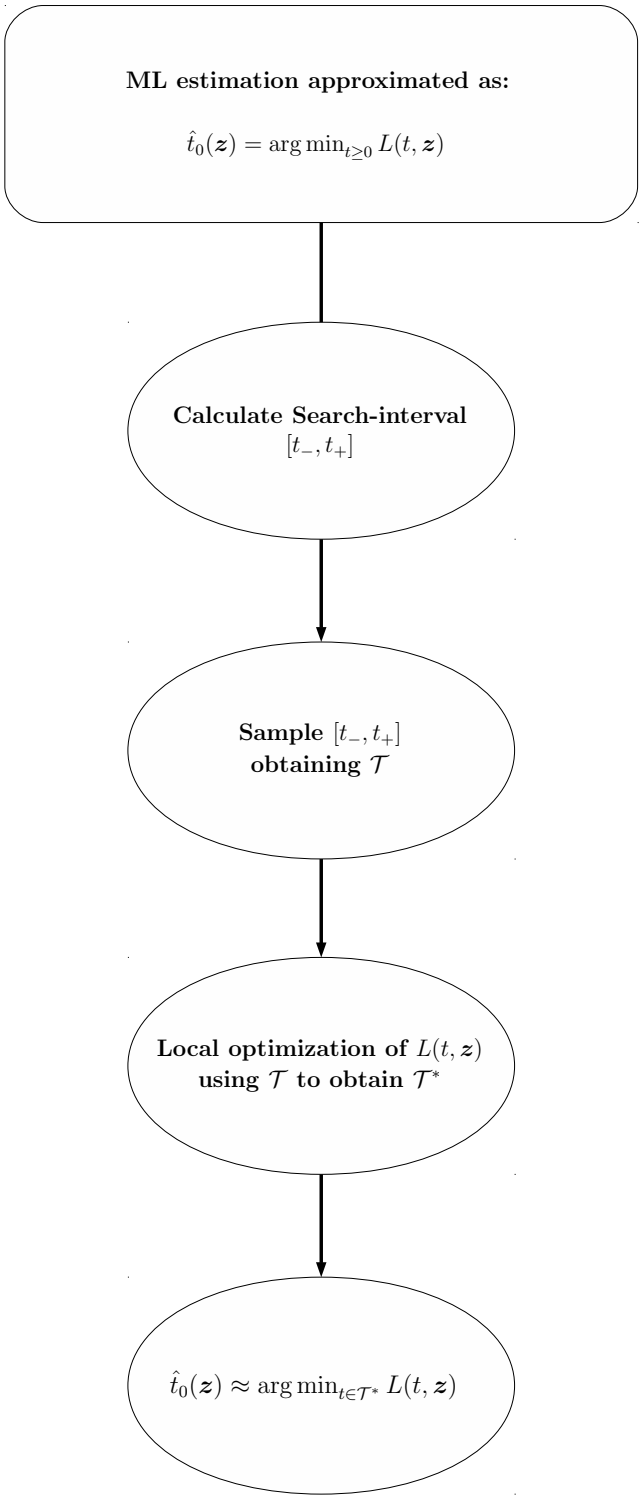


Figure 5.1: Modular Outline of the *Ad-hoc* Estimation Algorithm.

5.1 Search-Interval

Assuming that $t_0 \geq 0$, we propose to obtain $[t_-, t_+]$ following a statistical approach (this interval is denoted by $[t_-^{\mathrm{V}}, t_+^{\mathrm{V}}]$), following a deterministic approach (which is denoted by $[t_-^{\mathrm{D}}, t_+^{\mathrm{D}}]$), or by the intersection of these two intervals (i.e., $[t_-, t_+] = [t_-^{\mathrm{V}}, t_+^{\mathrm{V}}] \cap [t_-^{\mathrm{D}}, t_+^{\mathrm{D}}]$).

5.1.1 Statistical Interval

In order to obtain $[t_-^{\mathrm{V}}, t_+^{\mathrm{V}}]$, the unbiased variance-based estimator of t_0^2 is used

$$\hat{t_0^2}(\boldsymbol{z})_{\mathrm{var}} = \frac{\|\boldsymbol{z}\|^2/L - \sigma_N^2}{\sigma_X^2 + \alpha^2\Delta^2/12}, \tag{5.1}$$

which is the ML-estimator straightforwardly obtained using (4.38) from App. 4.A. For a value of L large enough, by using the CLT, the distribution of $\hat{t_0^2}(\boldsymbol{z})_{\mathrm{var}}$ can be accurately approximated by a Gaussian distribution centered at t_0^2 with variance $2(t_0^2(\sigma_X^2 + \alpha^2\Delta^2/12) + \sigma_N^2)^2/(L(\sigma_X^2 + \alpha^2\Delta^2/12)^2)$. The limits of the interval are obtained using this approximation as

$$(t_\pm^{\mathrm{V}})^2 = \max\left(\epsilon, \hat{t_0^2}(\boldsymbol{z})_{\mathrm{var}} \pm K_2\sqrt{\frac{2(\hat{t_0^2}(\boldsymbol{z})_{\mathrm{var}}(\sigma_X^2 + \alpha^2\Delta^2/12) + \sigma_N^2)^2}{L\left(\sigma_X^2 + \alpha^2\Delta^2/12\right)^2}}\right),$$

where in the previous expression, $\epsilon > 0$ guarantees that both $(t_-^{\mathrm{V}})^2$ and $(t_+^{\mathrm{V}})^2$ take positive values; in addition, $K_2 \geq 0$ determines the width of $[(t_-^{\mathrm{V}})^2, (t_+^{\mathrm{V}})^2]$ and also the probability that t_0^2 lies on this interval; this probability can be approximated by $P(t_0^2 \in [(t_-^{\mathrm{V}})^2, (t_+^{\mathrm{V}})^2]) \approx \mathrm{erf}(K_2/\sqrt{2})$ if $\hat{t_0^2}(\boldsymbol{z})_{\mathrm{var}} \approx t_0^2$. $[t_-^{\mathrm{V}}, t_+^{\mathrm{V}}]$ is obtained as the positive square roots of the endpoints of $[(t_-^{\mathrm{V}})^2, (t_+^{\mathrm{V}})^2]$. It is worth noting that in this problem $P\left(t_0 \in [t_-^{\mathrm{V}}, t_+^{\mathrm{V}}]\right) = P\left(t_0^2 \in [(t_-^{\mathrm{V}})^2, (t_+^{\mathrm{V}})^2]\right)$.

5.1.2 Deterministic Interval

The cost function is lower bounded by a function $L_2(t, \boldsymbol{z})$, which is defined in App. 5.A. In this section, for the sake of notational simplicity $L_2(t, \boldsymbol{z})$ denotes the corresponding function $L_2^{\text{low-SNR}}$, $L_2^{\text{low-SNR,2}}$, or $L_2^{\text{high-SNR}}$ depending on the used pdf of Z to obtain it. An initial guess of the gain t_{initial} is used (which is usually obtained applying a less complex estimator) to calculate $[t_-^{\mathrm{D}}, t_+^{\mathrm{D}}]$ based on the inequality $L(t_{\mathrm{initial}}, \mathbf{z}) \geq L(\hat{t}_0(\mathbf{z}), \mathbf{z}) \geq L_2(\hat{t}_0(\boldsymbol{z}), \mathbf{z})$ as

$$t_-^{\mathrm{D}} = \underset{t:t\leq t_L}{\arg\min} |L(t_{\mathrm{initial}}, \boldsymbol{z}) - L_2(t, \boldsymbol{z})| \tag{5.2}$$

$$t_+^{\mathrm{D}} = \underset{t:t\geq t_L}{\arg\min} |L(t_{\mathrm{initial}}, \boldsymbol{z}) - L_2(t, \boldsymbol{z})|, \tag{5.3}$$

where t_L is the location of minimum of $L_2(t, \boldsymbol{z})$ for $t \geq 0$, and is calculated in App. 5.A, where the increasing/decreasing tendency of $L_2(t, \boldsymbol{z})$ with respect to t for the low-SNR and high-SNR cases is also studied. For the low-SNR case and the high-SNR case with $\alpha \in [0, 1)$, the minimization problems defined in (5.2) and (5.3) have one and only solution since, on one hand, $L_2(t, \boldsymbol{z})$ is continuous and decreases in $t \in [0, t_L)$ and $\lim_{t\to 0} L_2(t, \boldsymbol{z}) = \infty$; on the other hand, $L_2(t, \boldsymbol{z})$ is continuous and increases for finite $t > t_L$ going to ∞ as $t \to \infty$. However for the high-SNR $\alpha = 1$ case, $L_2(t, \boldsymbol{z})$ decreases for $t \geq 0$; therefore, there exists a unique value of t where $L_2(t, \boldsymbol{z})$ crosses $L(t_{\text{initial}}, \boldsymbol{z})$, corresponding to t_-^{D}, while t_+^{D} is set to ∞.

To the best of author's knowledge, there is no closed-form expression for $[t_-^{\text{D}}, t_+^{\text{D}}]$; however, a numerical method, like bisection, can be used.

5.2 Candidate Set

Given $[t_-, t_+]$, the candidate set $\mathcal{T}$ is obtained by sampling the search-interval based on the properties of the used cost function. We would like to define $\mathcal{T}$ with a cardinality as small as possible, as the computational cost of the proposed procedure will increase with the cardinality of that candidate set. Nevertheless, simultaneously we would like to be able to guarantee that the result of running the proposed procedure is indeed the ML estimate. As stated at the beginning of this section, we will focus on the calculation of $\mathcal{T}$ assuming that $t_0 \geq 0$. However, if $t_0 < 0$, due to the symmetry of the problem with respect to $t_0 = 0$, the estimation algorithm is easily modified in order to take this case into account by setting $\mathcal{T} \triangleq \mathcal{T}^+ \cup -\mathcal{T}^+$, where in this expression $\mathcal{T}^+$ denotes the candidate set obtained for positive values of t_0.

5.2.1 Sampling Based on DC-QIM's Modulo-Lattice Reduction

From (3.12) and (3.15), it is clear that the main problem concerning the optimization of both target functions comes from their respective last terms, which are ultimately related to the modulo-lattice reduction of DC-QIM. Indeed, if a good estimate of the centroid used at the embedder were available, then the modulo operation could be mainly neglected and the gain estimation would be much simpler; from a more intuitive point of view, that would mean that we would be in the main lobe of the target function.

In order to find a point for which the modulo reduction is ineffective, we first notice that when the two following conditions hold simultaneously: 1) high-SNR

constraint, 2) t is close to t_0, then

$$(\boldsymbol{Z} - t\boldsymbol{D})\mathrm{mod}(t\Lambda) \approx \boldsymbol{Z} - tQ(\boldsymbol{X}), \tag{5.4}$$

as the modulo-lattice reduction of the total noise (i.e., equation (3.19) in App. 3.C) will be negligible.[1]

In such case, recalling that $\boldsymbol{z} = t_0\boldsymbol{y} + \boldsymbol{n}$, and using (2.2) and the law of large numbers, for large L we can write

$$||(\boldsymbol{z} - t\boldsymbol{d})\mathrm{mod}(t\Lambda)||^2/L \approx (t_0 - t)^2\sigma_X^2 + (t - \alpha t_0)^2\Delta^2/12 + \sigma_N^2 \triangleq \xi(t, t_0). \tag{5.5}$$

We emphasize that for (5.4) to be valid, t must be close to t_0. In fact, when we consider values of t away from t_0, the term $(t_0 - t)\boldsymbol{x}$ in (3.19) makes (5.4) (and consequently the presented variance approximation) no longer valid, as the effect of modulo-lattice reduction starts to be significant; indeed, when t is far from t_0 the distribution of $(\boldsymbol{Z} - t\boldsymbol{D})\mathrm{mod}(t\Lambda)$ will converge to a random variable uniformly distributed in the fundamental Voronoi region of $t\Lambda$, and consequently $||(\boldsymbol{Z} - t\boldsymbol{D})\mathrm{mod}(t\Lambda)||^2/L$ will converge to the second moment of $t\Lambda$, i.e., $t^2\Delta^2/12$.

This property could be useful for generating the elements of $\mathcal{T}$: to detect whether t belongs to the main lobe of the target function, one might think of checking whether $||(\boldsymbol{Z} - t\boldsymbol{D})\mathrm{mod}(t\Lambda)||^2/L$ is smaller than $t^2\Delta^2/12$ in a statistically significant sense. Unfortunately, this criterion would not be suitable, as it turns out that for small values of t, $\xi(t, t_0)$ will be always larger than $t^2\Delta^2/12$.

In order to solve this issue, we propose a criterion based on the increments of $\xi(t(l+1), t(l))$ with respect to $\xi(t(l), t(l))$ with $t(l), t(l+1) \in \mathcal{T}$. To this end, given $t(l)$, we hypothesize that $t_0 = t(l)$ and consider that for $t(l+1)$ sufficiently close to $t(l)$, (5.5) will hold with $t = t(l+1)$. Thus, we establish a threshold to $\xi(t(l+1), t(l)) - \xi(t(l), t(l))$ proportional to the second moment of the scaled lattice, namely, $K_1 t^2(l+1)\Delta^2/12$, which allows us to compute the next sampling point $t(l+1)$. The condition can be rewritten as

$$\xi(t(l+1), t(l)) = \xi(t(l), t(l)) + K_1 t^2(l+1)\Delta^2/12. \tag{5.6}$$

It can be shown that for K_1 sufficiently small, the solution in $t(l+1)$ to (5.6) when $t(l) \leq t_0 \leq t(l+1)$ is such that $\xi(t(l+1), t_0) < t^2(l+1)\Delta^2/12$; this guarantees that the main lobe of the target function is not missed. Furthermore, condition (5.6) can be satisfied for small values of $t(l+1)$, so the above problem with using $\xi(t(l+1), t(l))$ alone disappears.

In practice, there is an obvious tradeoff in the selection of K_1 used in (5.6). The larger K_1 is, the more likely is that the main lobe is missed due to the

[1] A similar reasoning is followed for (3.12), where the argument of the cosine, which is also proportional to $\boldsymbol{Z} - tQ(\boldsymbol{X})$, must be taken into account. Therefore, the same candidate set will be used for the low-SNR target function.

stochastic nature of $(\boldsymbol{z} - t\boldsymbol{d})\text{mod}(t\Lambda)$. This is more so for smaller L, as the approximation $||(\boldsymbol{z} - t\boldsymbol{d})\text{mod}(t\Lambda)||^2/L \approx \xi(t, t_0)$ is less accurate. On the other hand, the smaller K_1, the larger the cardinality $\mathcal{T}$, and therefore, the complexity.

The solution to (5.6) is

$$t(l+1) = \frac{t(l)\left(\alpha\Delta^2/12 + \sigma_X^2 + \Delta/\sqrt{12}\sqrt{\Delta^2/12\left((1-\alpha)^2 + K_1(2\alpha-1)\right) + K_1\sigma_X^2}\right)}{\sigma_X^2 + \frac{\Delta^2(1-K_1)}{12}}, \tag{5.7}$$

where $t(1) = t_-$, and the stopping condition is $t(l) \geq t_+$. We remark that although the proposed sampling strategy guarantees that at least a point in $\mathcal{T}$ is in the target function main lobe, it is conservative, in the sense that the search interval is finely sampled.

5.2.1.1 Adaptation for Unknown Variances

We consider now the case where both σ_X^2 and σ_N^2 are unknown, and moreover, to illustrate its use, we assume STDM (Sect. 3.4) in which projection of the L components onto L_{ST} dimensions is carried out. With these considerations, the search interval and the candidate set $\mathcal{T}$ must be modified accordingly. The new search interval is defined by:

- $t_- = \epsilon^*$, where $\epsilon^* > 0$ is a predefined threshold.
- $t_+ = \sqrt{\|\boldsymbol{z}\|^2/(L_{\text{ST}}\alpha^2\Delta^2/12)}$, where we have taken into account the estimate given in (5.1) that can be bounded as follows

$$\frac{\|\boldsymbol{z}\|^2/L - \sigma_N^2}{\sigma_X^2 + \alpha^2\Delta^2\frac{L_{\text{ST}}}{12L}} \leq \frac{\|\boldsymbol{z}\|^2/L}{\alpha^2\Delta^2\frac{L_{\text{ST}}}{12L}} = \frac{\|\boldsymbol{z}\|^2/L_{\text{ST}}}{\alpha^2\Delta^2/12},$$

  and where

$$\frac{\text{E}\{\|\boldsymbol{Z}\|^2\}}{L} = t_0^2\sigma_X^2 + t_0^2\alpha^2\Delta^2\frac{L_{\text{ST}}}{12L} + \sigma_N^2.$$

Concerning the candidate set, we exploit the monotonically decreasing nature of the right hand side of (5.7) with respect to σ_X^2 (analyzed in App. 5.B), and the fact that for $|t(l)| < t_0$, $\|\boldsymbol{z}\|^2/(Lt^2(l))$ is an upper-bound on σ_X^2, and consequently the sampling criterion that results after replacing σ_X^2 by $\|\boldsymbol{z}\|^2/(Lt^2(l))$ in (5.7) is valid (albeit conservative).

5.2.2 Sampling Based on the Mean of $L(t, \mathbf{z})$ of the low-SNR Case

In order to derive another sampling criterion of the search-interval, we take advantage of the introduced approximations of the pdf of Z for the low-SNR case (i.e., (3.8) and (3.9)), which allow us to straightforwardly obtain the distribution of $L(t, \boldsymbol{z})$, due to its good mathematical tractability, by applying the CLT assuming that $L \to \infty$ and that each component of $\boldsymbol{Z}$ is independent given t_0.

The proposed approach to obtain $\mathcal{T}$ is based on the expectation of $L(t, Z)$ $\mathrm{E}\{L(t, Z)\}$ using $f_{Z|T,K}^{\text{low-SNR}}(z|t, d)$ defined Sect. 3.1.1. Let us mention that in App 5.D, another method is outlined based on comparing the approximation of the distribution of $L(t, Z)$ whenever t is close t_0, with the approximation of the target function if t and t_0 are not close.

The expression of $\mathrm{E}\{L(t, Z)\}$ after some algebraic manipulations and the calculation of its expectation with respect to the dither, as detailed in App. 3.A, becomes

$$\mathrm{E}\{L(t, Z)\} \approx \frac{\sigma_N^2 + \sigma_X^2 t_0^2}{2\left(\sigma_N^2 + \sigma_X^2 t^2\right)} + \frac{1}{2}\log\left[2\pi\left(\sigma_N^2 + \sigma_X^2 t^2\right)\right] - 2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2 t^2}{\sigma_N^2+\sigma_X^2 t^2}+\frac{4\sigma_X^2 t t_0}{\sigma_N^2+\sigma_X^2 t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2 t t_0\right)^2}{\left(\sigma_N^2+\sigma_X^2 t^2\right)^2\left(\sigma_N^2+\sigma_X^2 t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2 t_0^2}{\sigma_N^2+\sigma_X^2 t_0^2}\right)}{\Delta^2}}. \quad (5.8)$$

By using $\mathrm{E}\{L(t, Z)\}$, some interesting properties can be determined on the asymptotic behavior of the cost function with $L \to \infty$, such as the number of stationary points, the position of the maxima/minima, etc.

As discussed in Sect. 3.2.1, $\mathrm{E}\{L(t, Z)\}$ for the low-SNR case only has a minimum at t_0 asymptotically as the hypotheses of the low-SNR case hold. In order to reach the minimum of the target function, one can argue that a gradient descent method taking as starting point t_- or t_+ can be used for large enough L; however, this solution is computational costly since this algorithm could require to perform a large number of iterations.

Our approach to deal with this issue is based on $\partial \mathrm{E}\{L(t, Z)\}/(\partial t)$; indeed, the maxima/minima of the first derivative with respect to t are obtained in order to guarantee a faster convergence of a gradient descent algorithm. Since the CRB (which the inverse of the Fisher information defined in (4.6)) is the lower bound (which is asymptotically achieved when $L \to \infty$) of the variance of the proposed ML estimator, the CRB can be used as a reference for the asymptotic properties of the estimator of t_0. As discussed in Sect. 4.1.1.3, there are two different parts of the CRB: one that has a behavior related to the variance-based estimator, and another that reduces the latter and that is due to the structure introduced by

the embedding. In order to analyze $\partial \mathrm{E}\{L(t,Z)\}/\partial t$ we will focus on the term induced by such structure.

The difference between $t(l+1)$ and $t(l)$ of $\mathcal{T}$ is defined in this method as the difference between the maximum and the minimum of the first derivative with respect to t of the term of (5.8) due to the structure of the pdf of Z, term which is denoted by M

$$M \triangleq -2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t^2}{\sigma_N^2+\sigma_X^2t^2}+\frac{4\sigma_X^2tt_0}{\sigma_N^2+\sigma_X^2t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2tt_0\right)^2}{\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t_0^2}{\sigma_N^2+\sigma_X^2t_0^2}\right)}{\Delta^2}}. \tag{5.9}$$

These minimum and maximum are calculated in App. 5.C, where the minimum are denoted by $t_l(t_0)$ and $t_r(t_0)$ respectively, and they are mathematically expressed as

$$\begin{aligned} t_l(t_0) &= \frac{9\sigma_N^2}{\sigma_X^2 t_0} + t_0 - \frac{\Delta t_0}{2\pi\sigma_X} \\ t_r(t_0) &= \frac{9\sigma_N^2}{\sigma_X^2 t_0} + t_0 + \frac{\Delta t_0}{2\pi\sigma_X}. \end{aligned} \tag{5.10}$$

Obviously, t_0 is not known in advance; however as in the method described in the previous section, in order to obtain $\mathcal{T}$, t_0 is assumed known to carry out the following recursive algorithm corresponding to the lth iteration: first, it is assumed that lth element of $\mathcal{T}$ is the lower endpoint of the interval as $t_l(t_0) = t(l)$, then the t_0 for the current iteration of the sampling algorithm is calculated from (5.10) (since $t_l(t_0)$ is assumed to be known), and finally with this obtained t_0, the $(l+1)$th element of $\mathcal{T}$ is calculated as $t(l+1) = t_r(t_0)$. In the generation algorithm of $\mathcal{T}$, the first element of $\mathcal{T}$ is $t(1) = t_-$, and this algorithm stops if $t(l) > t_+$.

5.2.2.1 Adaptation for Unknown Variances

Similarly to the discussion in Sect. 5.2.1.1, the search-interval is defined as $[\epsilon^*, \sqrt{\|\boldsymbol{z}\|^2/(L\alpha^2\Delta^2/12)}]$ with $\epsilon^* > 0$. On one hand, it is assumed that σ_N^2 can be neglected in comparison to $t_0^2\sigma_X^2$ if $\mathrm{TNHR}(t_0) \ll 1$ holds to compute $t(i) \in \mathcal{T}$ and, thus, the term $9\sigma_N^2/(\sigma_X^2 t_0)$ of the two expressions of (5.10) is discarded. On the other hand, σ_X^2 is upper-bounded by $\|\boldsymbol{z}\|^2/(t(i)^2 L)$.

5.3 Local Optimization

After the search interval is defined $[t_-, t_+]$, and sampled in order to obtain the candidate set $\mathcal{T}$, local optimizations are carried out to obtain $\mathcal{T}^*$, and select the

global minimum of the target function from those local solutions. It is worth pointing out that different techniques were studied in this thesis (e.g., Decision-aided, Bisection method, Newton method, approximation of the distribution of the cost function at t_0, etc.) as well as combinations of them. The best results were obtained using the Decision-Aided technique and the Bisection method.

5.3.1 Decision-Aided Optimization

Based on the criterion used for defining the sampling strategy in Sect. 5.2.1, i.e., that at least one of the values in $\mathcal{T}$ yields a good estimate of the embedder centroid, the estimate of t_0 will be improved by applying a local optimization algorithm. Formally, we compute

$$c_j = \mathcal{Q}_\Delta\left(z_j/t - d_j\right) + d_j,$$

$j = 1, \ldots, L$, for each $t \in \mathcal{T}$. Then, the Minimum Mean Square Error (MMSE) criterion is used for estimating t_0 from $\boldsymbol{c}$ and $\boldsymbol{z}$, i.e.,

$$t^* \triangleq \arg\min_{\xi} \|\boldsymbol{z} - \xi\boldsymbol{c}\|^2;$$

it is easy to check that the solution to this decision-aided optimization is $t^* = \boldsymbol{z}^T\boldsymbol{c}/\|\boldsymbol{c}\|^2$. This procedure is performed for each $t \in \mathcal{T}$, $\mathcal{T}^*$ contains the t^* values. Finally, the proposed approximated ML estimate is

$$\hat{t}_0(\boldsymbol{z}) = \arg\min_{t\in\mathcal{T}^*} L(t, \boldsymbol{z}).$$

The reduced computational cost arises as one of the advantages of this technique, as the target function is only evaluated once per initial candidate point of $\mathcal{T}$, the calculation of $\boldsymbol{c}$ and t^* have low computational requirements, and the overall computational resources required by this method do not depend on the desired precision (i.e., there is no threshold used to control a tradeoff between the number of required computations and the precision).

5.3.2 Optimization Based on the Bisection Method

We propose another approach for looking for $\hat{t}_0(\boldsymbol{z})$ based on the Bisection Method. Given two consecutive elements of $\mathcal{T}$, the sign of the first derivative with respect to t of the cost function is evaluated at $t(i)$ and $t(i+1)$ to decide if there is a local maximum/minimum within the interval $[t(i), t(i+1)]$; if their signs are different a bisection algorithm is used to obtain a candidate estimate $t^*(i)$ of the actual scaling factor. The following pseudocode defines this algorithm

Algorithm 1 Bisection Method-Based Optimization

```
for i = 1 to |T|−1 do
    t_lo = t(i);
    t_up = t(i + 1);
    s_0 = sign (∂L(t_lo, z)/(∂t));
    s_1 = sign (∂L(t_up, z)/(∂t));
    if (s_0 != s_1) then
        while (true) do
            dist = t_up - t_lo;
            s_0 = sign (∂L(t_lo, z)/(∂t));
            s_1 = sign (∂L(t_lo + dist/2, z)/(∂t));
            if s_0 = s_1 then
                t_lo = t_lo + dist/2;

            else
                t_up = t_up − dist/2;

            end
            if (t_up − t_lo) < ε_B then
                t*(i) = t_lo;
                L_value(i) = L(t*(i), z);
                break
            end
        end
    else
        if (L(t_lo, z) > L(t_up, z)) then
            t*(i) = t_up;
            L_value(i) = L(t*(i), z);

        else
            t*(i) = t_lo;
            L_value(i) = L(t*(i), z);

        end
    end
end
i* = arg min_i L_value(i);
t̂_0(z) = t*(i);
```

The main advantage of this technique is that it guarantees to achieve a local minimum. The parameter ϵ_{B} of the previous pseudocode frame defines the stopping condition of the algorithm which plays a key role in the computational load and precision of the method. Due to the modulo operation, the first derivative with respect to t of the cost function of the high-SNR case cannot be expressed in a closed-form expression; therefore, the derivative is approximated by computing

it numerically.

5.4 Performance Comparison

In this section, we compare the performance of the DPC-based estimators proposed in this work Dirty Paper Coding Estimation for Low-SNR Case (DPCEL), Dirty Paper Coding Estimation for High-SNR Case (DPCEH), the variance-based estimator denoted by Var and defined as (3.13) in Sect. 3.2.1.1, and the PDD algorithm [29]. The performance is analyzed by taking into account the accuracy, measured in terms of MSE, and the complexity of the algorithms, quantified by means of the average run time.

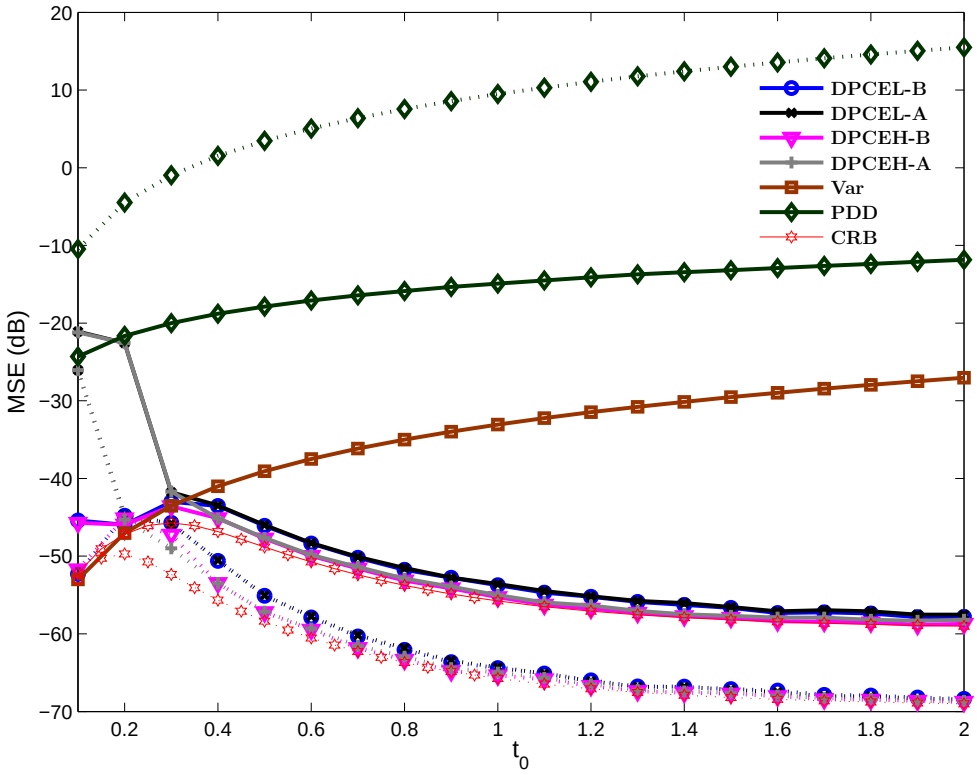


Figure 5.2: MSE vs. t_0 curves for the proposed techniques. The Bisection method technique for the Low-SNR (DPCEL-B) and for the High-SNR (DPCEH-B), and the Decision-Aided technique for the Low-SNR (DPCEL-A) and for the High-SNR (DPCEH-A). DWR = 30 dB (solid line), DWR = 40 dB (dotted line), WNR = 0 dB, $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also shown.

For the computation of the statistical interval we use $\epsilon = \epsilon^* = 10^{-2}$ and $K_2 = 10$. The parameters of the used local optimization techniques are $K_1 = 10^{-2}$

for the Decision-Aided technique, while for the Bisection method $K_1 = 1$ and $\epsilon_B = 10^{-4}$.

Concerning PDD [29], the estimation is based on projecting $\boldsymbol{z}$ onto a single dimension, and the so-called self-interference factor η (the value of η determines the power allocated for host interference cancellation) takes the value that minimizes the MSE; this optimization over η was implemented by using exhaustive search in $[1 - \sqrt{L/\mathrm{DWR}}, 1] \cap 10^{-2}\mathbb{N}$ (recall that $1 - \sqrt{L/\mathrm{DWR}}$ is the minimum value of η for which after partially canceling the host interference there is some remaining power available for transmitting the watermark).

5.4.1 Known Variances

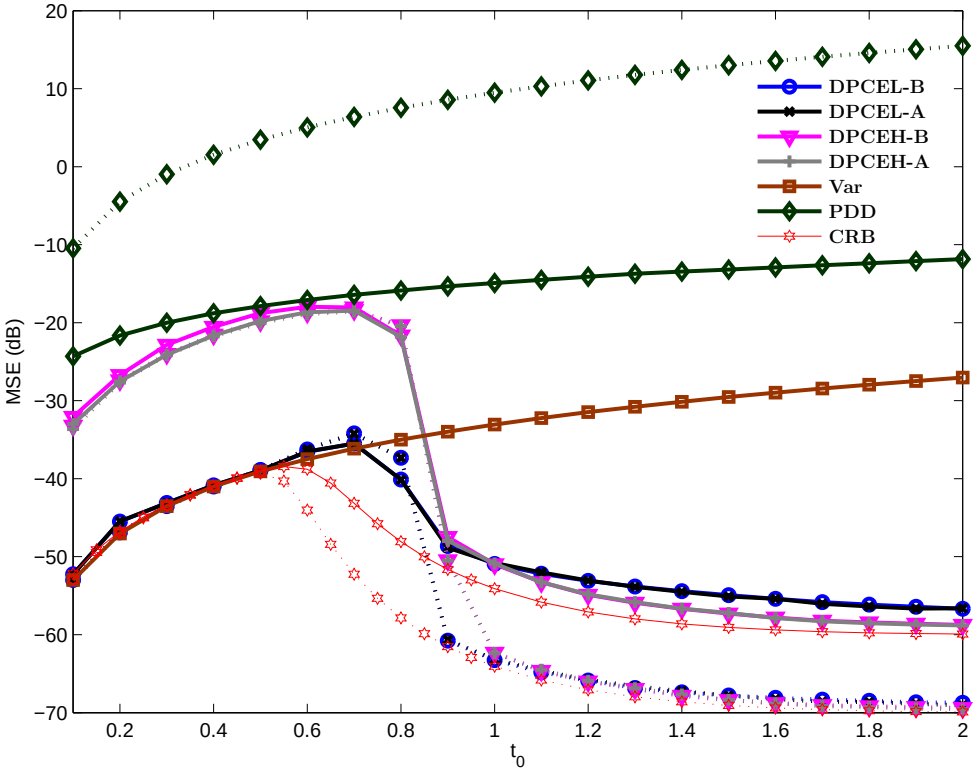


Figure 5.3: MSE vs. t_0 curves for the proposed techniques. The Bisection method for Low-SNR (DPCEL-B) and for High-SNR (DPCEH-B), and the Decision-Aided technique for low-SNR (DPCEL-A) and for high-SNR (DPCEH-A). DWR = 30 dB (solid line), DWR = 40 dB (dotted line), WNR = 0 dB, $\alpha = 1$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also shown.

In this section, the experiments were carried out assuming σ_X^2 and σ_N^2 are known. By default, to compute the search-interval the intersection of the two

proposed techniques (i.e., those used to calculate the statistical and deterministic intervals) is used, and the interval is sampled based on the DC-QIM Modulo-Lattice Reduction (introduced in Sect. 5.2.1). For the sake of simplicity, for DPCEL the cost function (3.12) is utilized.

5.4.1.1 Dependence on DWR

First of all, we compare the performance of the proposed DPCE-based strategy with respect to the variance-based estimator and PDD. In order to do so, the MSE as a function of t_0 for DWR $= 30, 40$ dB and WNR $= 0$ dB is shown in Fig. 5.2, where $L = 10^3$, $t_0 > 0$, and $\alpha = \alpha_{\text{Costa}}$.

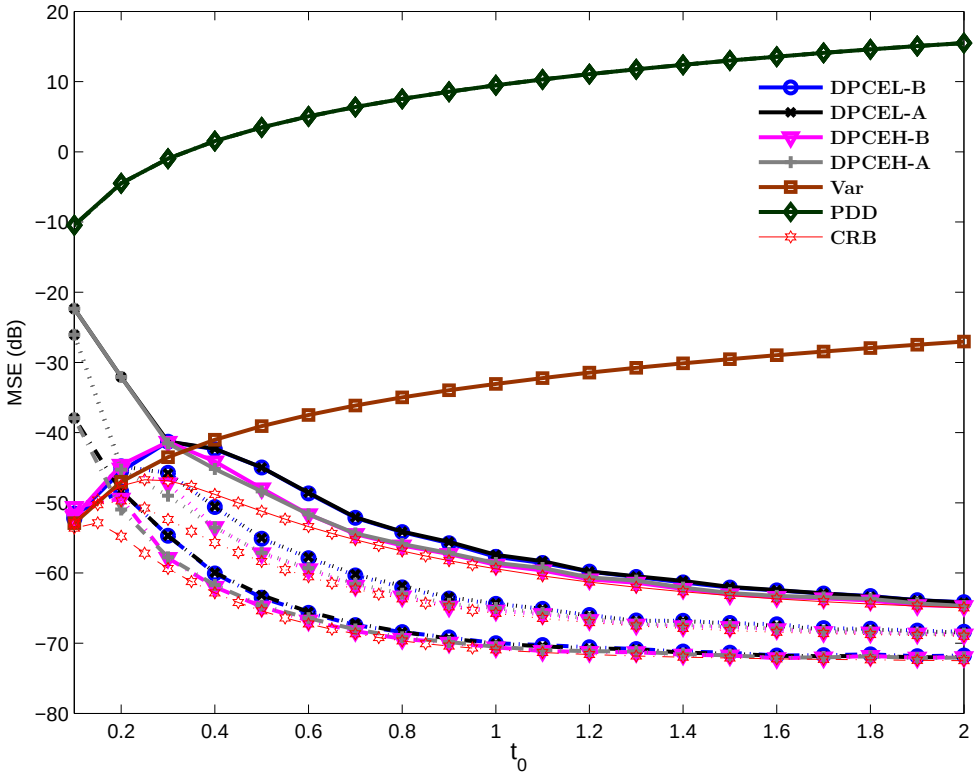


Figure 5.4: MSE vs. t_0 curves for the proposed techniques. The Bisection method for low-SNR (DPCEL-B) and for high-SNR (DPCEH-B), and the Decision-Aided technique for low-SNR (DPCEL-A) and for high-SNR (DPCEH-A). DWR $= 40$ dB, WNR $= -3$ dB (solid lines), WNR $= 0$ dB (dotted lines), WNR $= 3$ dB (dashdot lines), $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also depicted.

The behavior of CRB with t_0 can be summarized as follows: 1) for those values of t_0 where there is no structure in the distribution of $\boldsymbol{Z}$, the CRB curves increase

with t_0 until reaching a maximum (e.g., it is located at $t_0 \approx 0.2$ for DWR = 40 dB and WNR = 0 dB); 2) for those values of t_0 where the structure of $\boldsymbol{Z}$ arises , the CRB decreases with t_0; 3) for large values of t_0, the CRB converges to the inverse of (4.11). We can check that DPCE results for both target functions (namely, (3.12) and (3.15)) show a good agreement with the corresponding CRB for large values of t_0 (i.e., in the high-SNR scenario) and large DWR, with (3.15) yielding a slightly better performance; note that the proposed scheme takes advantage of the host variance in order to improve the estimator performance, i.e., the larger the DWR, the smaller the MSE (at the price of increasing the computational cost). On the other hand, for small values of t_0 (i.e., in the low-SNR scenario), the performance achieved by using (3.12) and decision-aided optimization does not agree with the CRB, as the constraints HQR $\gg 1$ and SCR$(t_0) \ll 1$ are not verified; however, the Bisection method obtains better results than the decision-aided optimization (even for DPCEH). This difference in the performance of the proposed local optimization techniques is explained by, on one hand, that the estimated position of the centroids for the decision-aided optimization is not accurate, which makes that the obtained $t^*(i)$ are not good estimates of t_0 and, as a consequence, substantial estimation errors are produced. On the other hand, for the Bisection method the search to obtain each $t^*(i)$ is limited by two consecutive elements of $\mathcal{T}$, therefore the error is bounded showing better performance than the other local optimization technique.

Contrarily to the observed monotonically decreasing nature of the MSE with respect to the DWR for the DPCE proposal, the MSE of PDD is monotonically increasing with DWR (as in that case the host signal interferes the estimate). For DWR = 30 dB, σ_W^2 is large enough to make possible the PDD host interference cancellation, while this is not the case for DWR = 40 dB. It is worth pointing out that the gain achieved by DPCE (for both low-SNR and high-SNR target functions) with respect to PDD is larger than 20 dB.

Fig. 5.3 is the counterpart of Fig. 5.2 but fixing $\alpha = 1$ instead of $\alpha = \alpha_{\text{Costa}}$. Although in this case the comments for the high-SNR case remain, in the low-SNR case (where it can be assumed that HQR $\gg 1$ and SCR$(t_0) \ll 1$ hold), DPCEL techniques outperform DPCEH independently of the local optimization technique. This difference appears because the sampling of $[t_-, t_+]$ is much finer when $\alpha = 1$, consequently $|\mathcal{T}|$ is larger, and the search of t_0 for the Bisection method is not so constrained.

5.4.1.2 Dependence on WNR

Fig. 5.4 is analogous to Fig. 5.2 but DWR = 40 dB with WNR $= -3, 0, 3$ dB. Here, the CRB curves evolve with t_0 like it was described in the previous section: they increase with t_0 until reaching a maximum (when the pdf of Z has no structure) and then decay until seemingly reaching an asymptotic minimum (also

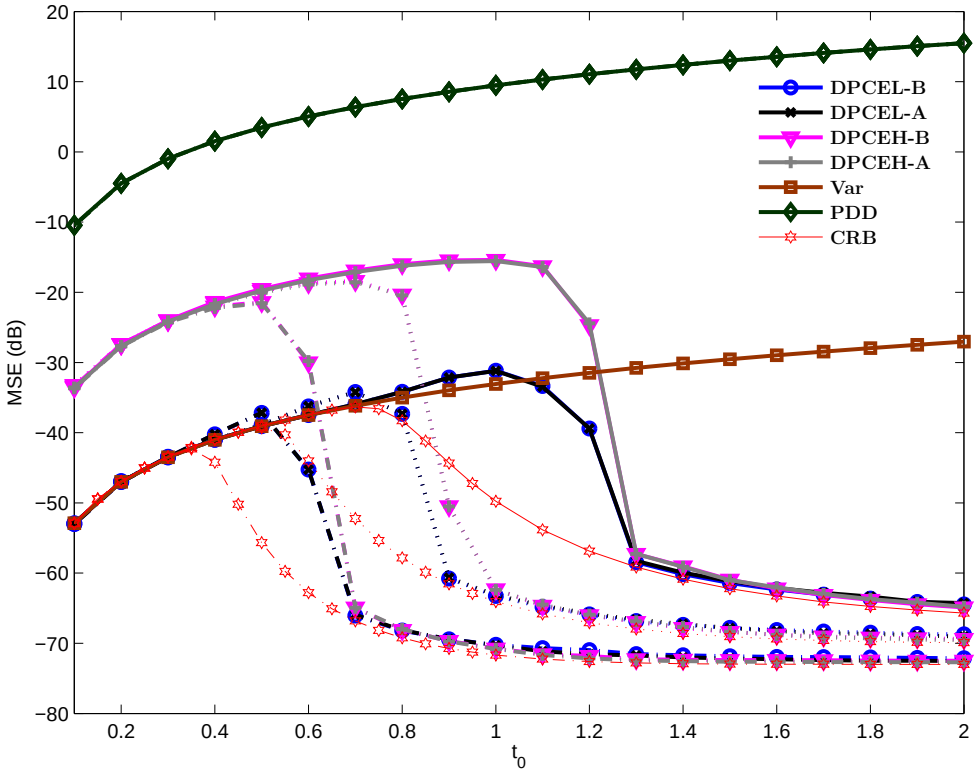


Figure 5.5: MSE vs. t_0 curves for the proposed techniques. The Bisection method for low-SNR (DPCEL-B) and for high-SNR (DPCEH-B), and the Decision-Aided technique for low-SNR (DPCEL-A) and for high-SNR (DPCEH-A). DWR = 40 dB, WNR = -3 dB (solid lines), WNR = 0 dB (dotted lines), WNR = 3 dB (dashdot lines), $\alpha = 1$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also depicted.

with the value $1/(\text{DWR}\cdot\text{WNR}\cdot L)$). Both DPCEL and DPCEH show a decreasing tendency with t_0 and improve the variance-based estimator for approximately $t_0 \geq 0.4$ for WNR = -3 dB. This crossing value of t_0 is reduced with WNR as the structure in the pdf of Z appears earlier. As in the previous section, for low-SNR scenarios, the techniques using the Bisection method show good results, while those obtained with the Decision-Aided technique differ from the CRB. In addition, the convergence of the obtained MSE curves to the CRB curves improves by increasing the value of WNR. As in the previous example, PDD (its MSE curves for WNR = $-3, 0, 3$ dB show almost the same performance) obtains the worst performance in this case.

The same scenario is considered in Fig. 5.5 with α fixed to 1. In this case, the difference caused by the appearance of structure in the pdf of Z is clearer. For example, by focusing on the CRB, one can state that the structure of the pdf of

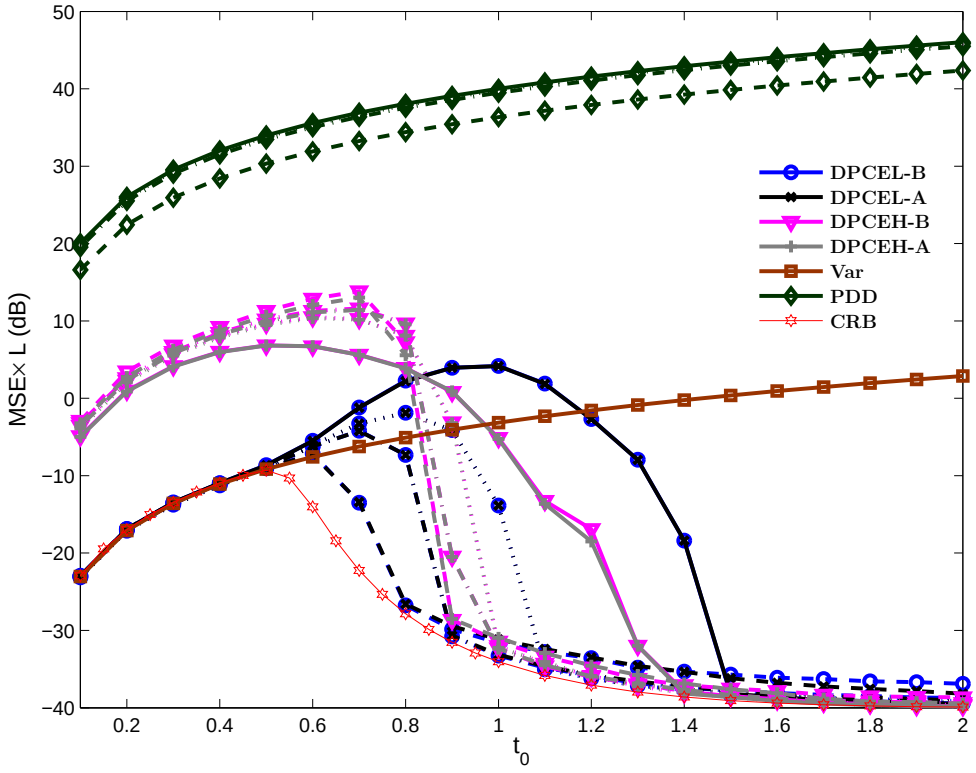


Figure 5.6: MSE as a function of t_0 for the proposed techniques. The Bisection method for low-SNR (DPCEL-B) and for high-SNR (DPCEH-B), and the Decision-Aided technique for low-SNR (DPCEL-A) and for high-SNR (DPCEH-A). DWR = 40 dB, WNR = 0 dB, $\alpha = 1$, $L = 10^2$ (solid lines), $L = 5 \cdot 10^2$ (dotted lines), $L = 10^3$ (dashdot lines), and $L = 5 \cdot 10^3$ (dashed lines). The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also shown.

Z approximately appears at about $t_0 = 0.4$ for WNR = 3 dB, and $t_0 = 0.8$ if WNR = -3 dB.

5.4.1.3 Dependence on L

Fig. 5.6 depicts MSE$\times L$ as a function of t_0 for $L = 10^2, 5 \cdot 10^2, 10^3, 5 \cdot 10^3$ in order to analyze the convergence of DPCE with respect to L, i.e., the efficiency of our estimators. On one hand, focusing on DPCEL, one can see the tendency of the performance of our algorithm to convergence to the CRB as L is increased although, as stated above, in the high-SNR scenarios, there is a little gap in comparison with DPCEH. DPCEH shows good results for the high-SNR scenarios, i.e., whenever there is structure in the pdf of Z; therefore, the algorithm cannot converge to the CRB even for increasing L if there is no structure, as one can

check in this figure that the performance shows an MSE abrupt decay for DPCEH around $t_0 = 0.9$.

5.4.1.4 Different Search-Interval and Sampling Criterion

By examining the results for the search-interval comparison shown in Fig. 5.7, one can realize that there is a difference in the small values of t_0 for the DPCEL case. Specifically, at $t_0 = 0.1$ there is a loss of 10 dB in the MSE; therefore, one can conclude that for DPCEL in low-SNR cases, the deterministic interval can improve the estimator in terms of MSE.

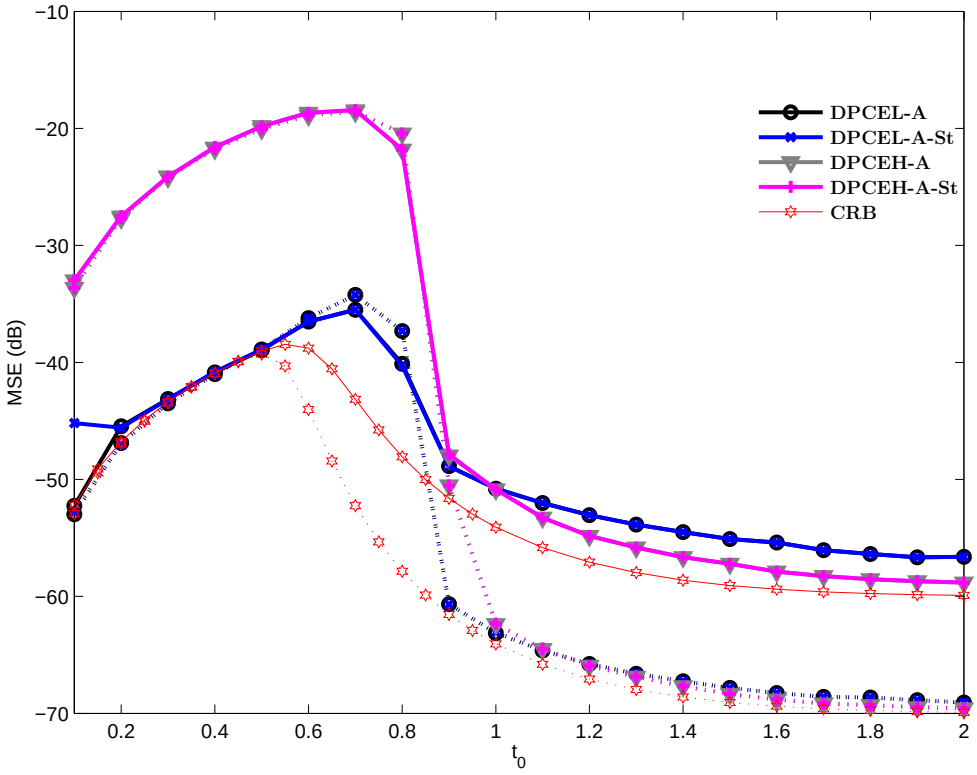


Figure 5.7: MSE vs. t_0 curves for the proposed techniques. The Decision-Aided technique for low-SNR (DPCEL-A) and for high-SNR (DPCEH-A) using the intersection of intervals of the two proposed methods and using only the statistical interval (with suffix -St in the legend). DWR = 30 dB (solid line), DWR = 40 dB (dashed line), WNR = 0 dB, $\alpha = 1$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB) are also shown.

A comparison among the DPCE estimators for the proposed search-interval sampling techniques is shown in Figs. 5.8-5.9 for $\alpha = \alpha_{\text{Costa}}$ and $\alpha = 1$, respectively. From these results, one can deduce that they are practically the same regardless of the sampling method, and that the behavior is coherent with those discussed above for these scenarios.

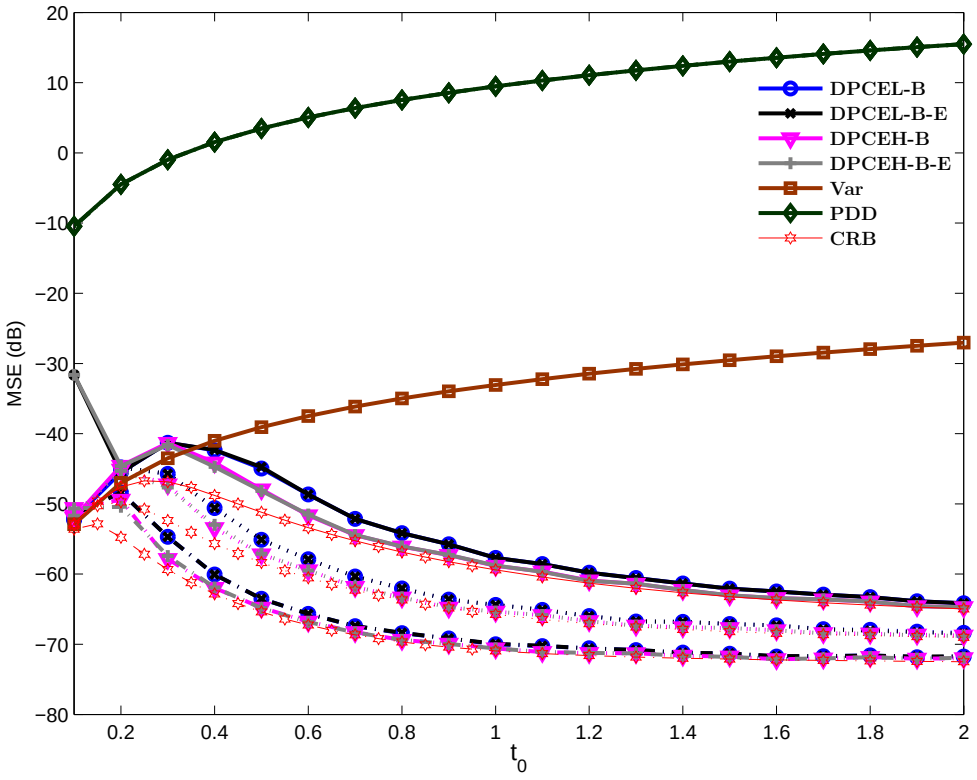


Figure 5.8: MSE vs. t_0 curves for the proposed techniques. The Bisection method for low-SNR (DPCEL-B) and for high-SNR (DPCEL-B) using the sampling based on the DC-QIM Modulo-Lattice Reduction and the sampling based on the mean of $L(t, \boldsymbol{z})$ (suffix -E in the legend) DWR = 40 dB, WNR = -3 dB (solid lines), WNR = 0 dB (dotted lines), WNR = 3 dB (dashdot lines), $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also depicted.

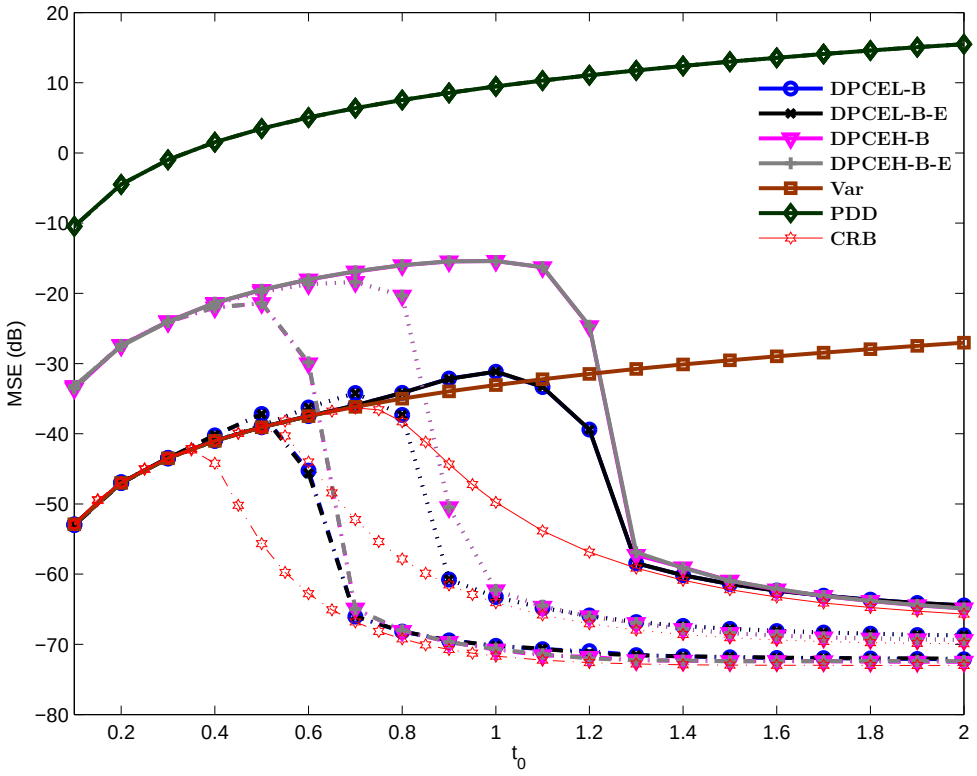
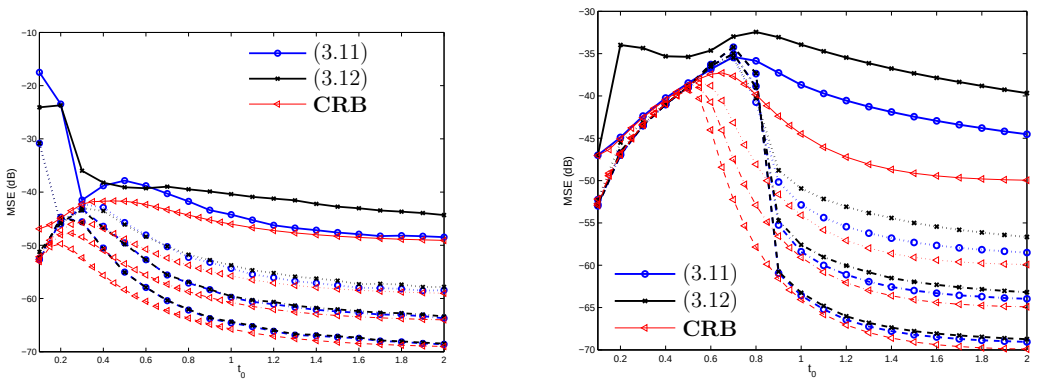


Figure 5.9: MSE vs. t_0 curves for the proposed techniques. The Bisection method for low-SNR (DPCEL-B) and for high-SNR (DPCEL-B) using the sampling based on the DC-QIM Modulo-Lattice Reduction and the sampling based on the mean of $L(t, \boldsymbol{z})$ (suffix -E in the legend) DWR = 40 dB, WNR = -3 dB (solid lines), WNR = 0 dB (dotted lines), WNR = 3 dB (dashdot lines), $\alpha = 1$, and $L = 10^3$. The corresponding numerically obtained CRB curves (CRB), the variance-based estimator curve (Var), and the Partially-Data-Dependent Superimposed Training curves (PDD) are also depicted.

5.4.1.5 DPCEL using (3.11) and (3.12)

So far in this section, we have applied (3.12) for DPCEL (we use the approximation to the pdf of Z in (3.9)). This is convenient as (3.12) is mathematically simpler than (3.11). Here, we compare the performance of the two DPCEL cost functions in Fig. 5.10 for $\alpha = \alpha_{\text{Costa}}$ (in the left pane of the figure) and $\alpha = 1$ (in the right pane of the figure) using the Bisection method. As expected (as (3.12) is a simplification of (3.11)), when the hypotheses HQR $\gg 1$, SCR$(t_0) \ll 1$, and TNQR$(t_0) \gg 1$ are verified, (3.11) shows equal or better performance than (3.12). The difference becomes obvious when the TNHR$(t_0) \ll 1$ is not verified (e.g., for small values of t_0 when DWR $= 20$ dB in Fig. 5.10 (b)). Besides that, it is worth mentioning that in high-SNR scenarios, there is a gap in the performance between those versions of DPCEL (indeed, (3.11) outperforms (3.12)) that decreases with DWR.



(a) $\alpha = \alpha_{\text{Costa}}$ (b) $\alpha = 1$

Figure 5.10: MSE vs. t_0 curves for the proposed techniques for low-SNR scenarios using approximations (3.8) and (3.9). DWR $= 20, 30, 35, 40$ dB (solid lines, dotted lines, dashdot lines, and dashed lines, respectively), WNR $= 0$ dB, and $L = 10^3$. The numerically obtained CRB curves (CRB) are shown.

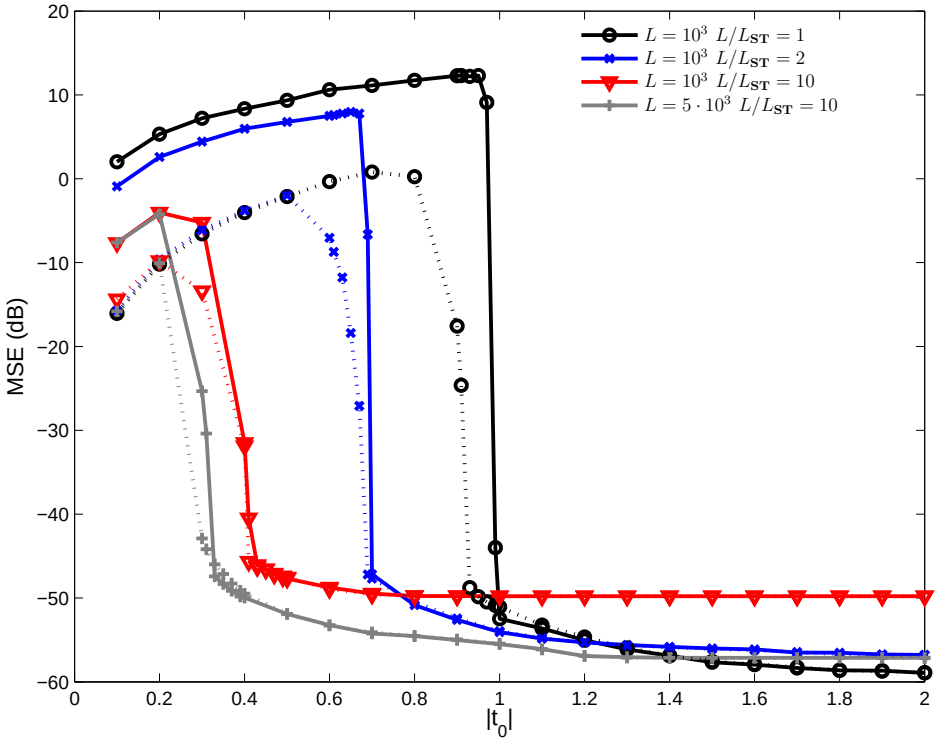


Figure 5.11: MSE vs. $|t_0|$ for ST-DM estimation in the high-SNR case DPCEH-A (dotted lines) and DPCE for unknown host and channel noise variances DPCEU (solid lines) for different L/L_{ST}. DWR = 30 dB , WNR = 0 dB, $\alpha = 1$, and $L = 10^3, 5 \cdot 10^3$.

5.4.2 Unknown Variance

In this section, we study the performance of the implementations of the Dirty Paper Coding Estimation for Unknown Variances (DPCEU) (check Sect. 3.3) using ST-DM-based estimation (proposed in Sect. 3.4). Throughout this section, the sampling of the search-interval is carried out using the DC-QIM Modulo-Lattice Reduction technique and the Decision-Aided technique is employed for the local optimization.

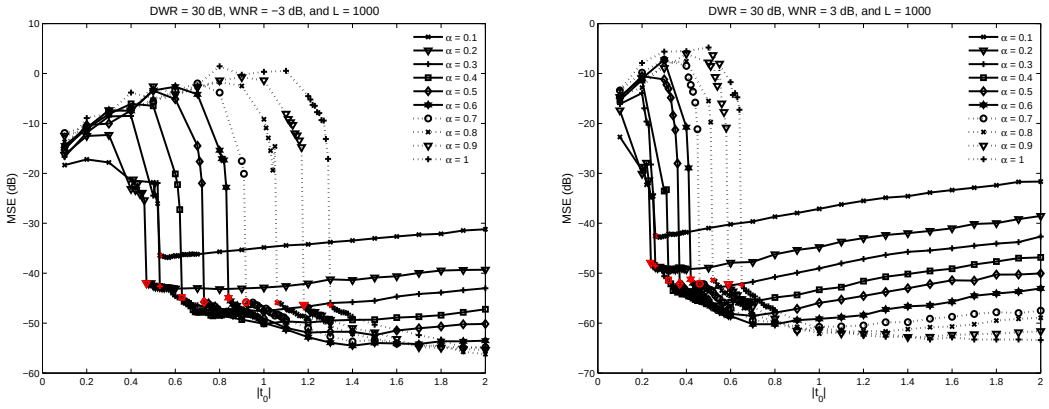
In Fig. 5.11, MSE vs. $|t_0|$ curves for DPCEU are depicted for different values of L/L_{ST} with DWR = 30 dB, WNR = 0 dB, $L = 10^3, 5 \cdot 10^3$, and $\alpha = 1$. The values of t_0 are estimated and take positive and negative values with equal probability. For the sake of comparison, the corresponding MSE curves for ST-DM-based DPCEH for this case are also shown. These results support the idea that by using ST-DM-based estimation, one can control (through the value of L_{ST}) the minimum value of $|t_0|$ that makes the structure of the pdf Z appear: the smaller the value of L_{ST}, the smaller the value of $|t_0|$ showing the structure of the pdf of $\boldsymbol{Z}^{ST}$. However, the cost to pay is the reduction of the number of samples

available to estimate and, therefore, a loss of accuracy of system (recall that the CRB inversely depends on the number of observations).

According to the experiments carried out, the minimum value of $|t_0|$ to correctly estimate using DPCE is related to the value of TNQR (defined in Chap. 2 that measures the ratio between the scaled self-noise plus the channel noise and the second moment of the scaled lattice). For example, the MSE curve for DPCEU with $L = 10^3$ and $L_{\mathrm{ST}} = L$ approximately has a minimum at $|t_0|= 1$, by projecting into $L_{\mathrm{ST}} = 5 \cdot 10^2$ dimensions the gain in terms of WNR is 3 dB; therefore, the minimum $|t_0|$ to obtain the same TNQR would be located at $1/\sqrt{2} \approx 0.71$, as approximately occurs in the figure. In addition, since DPCEU estimates with half of the samples, the MSE losses also approximately 3 dB for $|t_0|= 2$. However in the case $L = 10^3$ and $L/L_{\mathrm{ST}} = 10$, if one wants to attain the same value of TNQR as the $L/L_{\mathrm{ST}} = 1$ case, the minimum value of $|t_0|$ should be located at $1/\sqrt{10} \approx 0.32$ but the structure of the pdf of Z approximately appears at 0.41 in the figure. This mismatch is due to the number of available samples for the ML estimation, which in this case is not sufficient. In this way, by increasing the value of L_{ST} but fixing $L/L_{\mathrm{ST}} = 10$ as the curves for $L = 5{\cdot}10^3$ and $L/L_{\mathrm{ST}} = 10$ in the figure, the minimum value of $|t_0|$ where the structure of the pdf of Z appears is approximately at the predicted value for this L/L_{ST}.

By comparing the results for DPCEH and DPCEU, one can conclude that the DPCEH MSE curves decay slightly faster than DPCEU and their performance matches whenever the structure of the pdf Z appears; however, DPCEU does not require to know σ_X^2 and σ_N^2 to be used but demands more computational resources than DPCEH.

5.4.2.1 Analysis of TNQR



(a) WNR = −3 dB (b) WNR = 3 dB

Figure 5.12: MSE vs. $|t_0|$ for DPCEH considering different values of α, DWR = 30 dB, and $L = 10^3$. (a) WNR = −3 dB; (b) WNR = 3 dB.

The results obtained in the previous sections, as well as those shown in the theoretical analysis carried out in Chap. 4, indicate that our estimation algorithm outperforms other techniques (e.g., PDD, Add-SS, etc.) when the structure in the pdf of Z appears. Here, we study the relation between the values taken by TNQR and the appearance of structure in the distribution of the received sequence.

As an illustrative example, the curves of MSE as a function of the value of the gain are shown in Fig. 5.12 for different values of α with DWR = 30 dB, WNR $= -3,3$ dB, and $L = 10^3$. From these plots (as well others that are not shown, due to space limitations), the values of t_0 where the MSE abruptly decays, which is related to the appearance of structure in the received signal. We denote this value by $|t_{0,\min}|$ and by using them, TNQR($|t_{0,\min}|$) is calculated and shown for the case DWR = 30 dB in Fig 5.13 (similar results were obtained for DWR = 40 dB).

One can state that, according to these experiments, the obtained TNQR($|t_{0,\min}|$) does not vary significantly with DWR and WNR. In addition, without considering $\alpha = 0.1$ case, TNQR($|t_{0,\min}|$) takes values between 1 and 1.25, with DPCE being able to take advantage of the induced structure in pdf; in other words, the power of the scaled self-noise and the channel noise can be in the same order of magnitude as the second moment of the scaled lattice.

Let us assume that for DPCEU, TNQR must take values smaller than a critical constant $\text{TNQR}_{\text{Crit}}$ in order to guarantee that the received signal has structure in its pdf; therefore, $|t_{0,\min}| = \alpha/\sqrt{(\text{TNQR}_{\text{Crit}} - (1-\alpha)^2)\text{WNR}}$ ($\text{TNQR}_{\text{Crit}}$ must take values larger than or equal to 1 to obtain a real valued $|t_{0,\min}|$). In Fig. 5.14, the evolution of these theoretical $|t_{0,\min}|$ values with respect to α for $\text{TNQR}_{\text{Crit}} = 1.15$ in comparison with the value experimentally obtained $|t_{0,\min}|$ (denoted by $t_{0,\text{exp}}$) is shown for DWR $= 30, 40$ dB. These results indicate that $|t_{0,\min}|$ is close to $t_{0,\text{exp}}$ except for $\alpha = 0.1, 0.2$. Since this discrepancy appears for both DWR = 30 dB and DWR = 40 dB cases, one can conclude that this is not due to the lack of fulfillment of HQR $\gg 1$(notice that HQR = 20 dB for the DWR = 40 dB case) but $\text{SCR}(t_0) \ll 1$. From this discussion, we propose to assume that HQR $\gg 1$ and $\text{SCR}(t_0) \ll 1$ are fulfilled and fix the minimum value of TNQR to guarantee the arising of the structure of the pdf of Z to 1.

By fixing this threshold on TNQR, one can calculate the L_{ST} in order to have structure in the pdf of Z using the definition of TNQR as

$$L_{\text{ST}} = \min\left(L, \left\lfloor \frac{t_0^2 L \text{WNR}(1-(1-\alpha)^2)}{\alpha^2} \right\rfloor\right), \tag{5.11}$$

where in the previous expression, the min function is used to assure that $L_{\text{ST}} \leq L$. In addition, one must be aware that, as previously indicated, the value L_{ST} has to be large enough to guarantee that our ML-based estimation works properly. In Fig. 5.15, an example of how to decrease the smallest value of t_0 while keeping

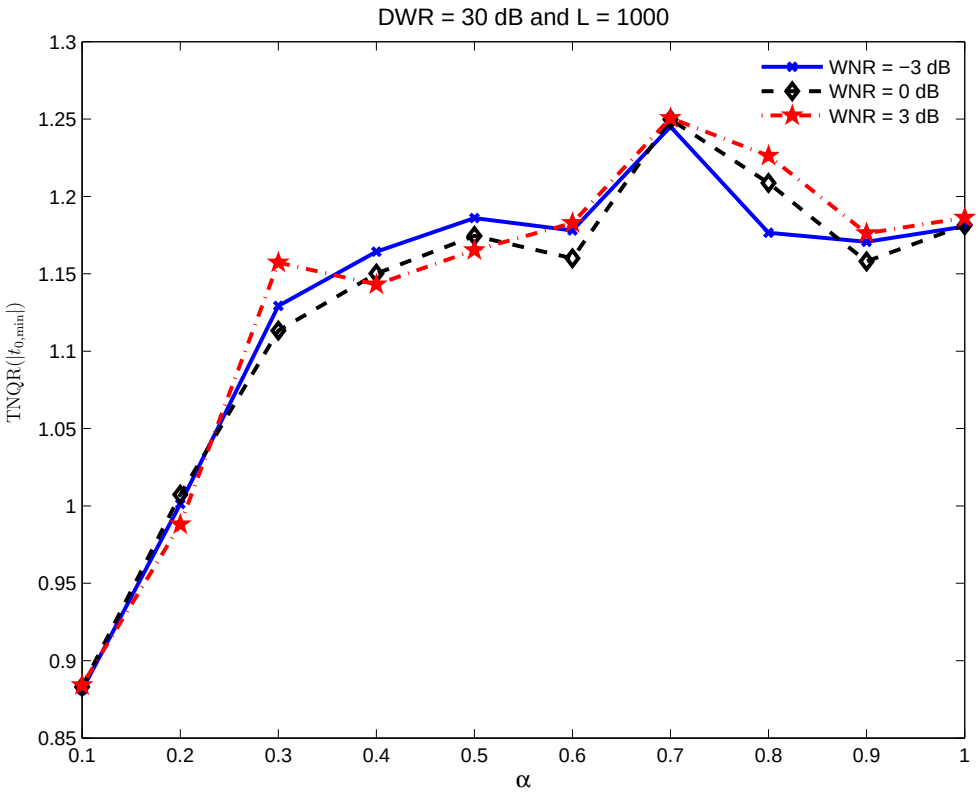
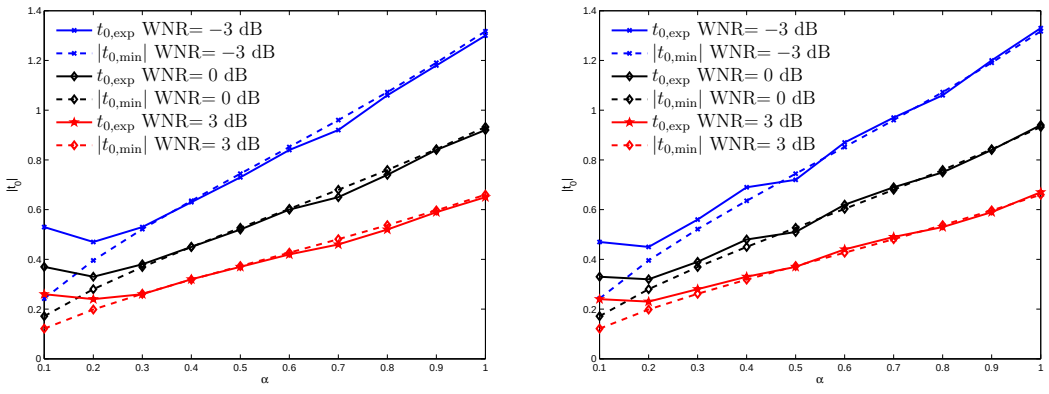


Figure 5.13: TNQR($|t_{0,\min}|$) vs. α with the experimentally obtained $|t_{0,\min}|$ for DPCEH with DWR $= 30$ dB, WNR $= -3, 0, 3$ dB, and $L = 10^3$.



(a) DWR $= 30$ dB (b) DWR $= 40$ dB

Figure 5.14: Experimentally obtained $|t_{0,\min}|$ and that obtained for $\text{TNQR}_{\text{Crit}} = 1.15$ with WNR $= -3, 0, 3$ dB, and $L = 10^3$. (a) for DWR $= 30$ dB and (b) for DWR $= 40$ dB.

structure in the pdf of Z is shown. For example, for WNR $= -3$ dB, the minimum value of t_0 to obtain a good estimate for DPCEU with is $t_0 \approx 1.3$ while for the projected version is $t_0 \approx 1$.

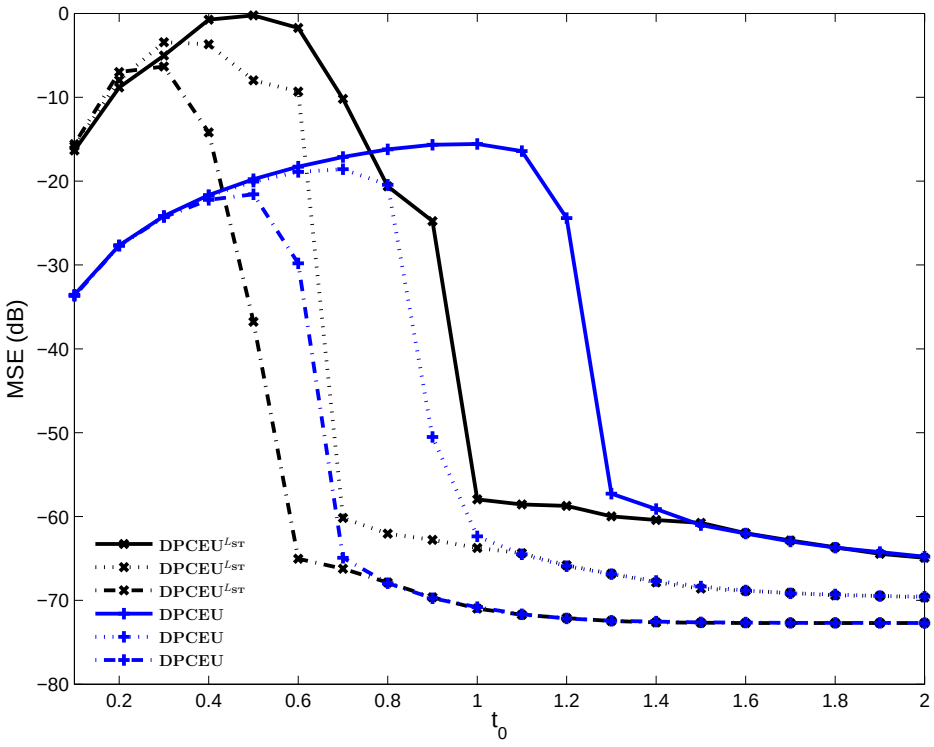


Figure 5.15: MSE vs. t_0 for DPCEU and DPCEU$^{L_{\mathrm{ST}}}$ (using (5.11)) with DWR $= 40$ dB, WNR $= -3, 0, 3$ dB (solid lines, dotted lines, and dashdot lines, respectively), $\alpha = 1$, and $L = 10^3$.

5.4.3 Computational Requirements

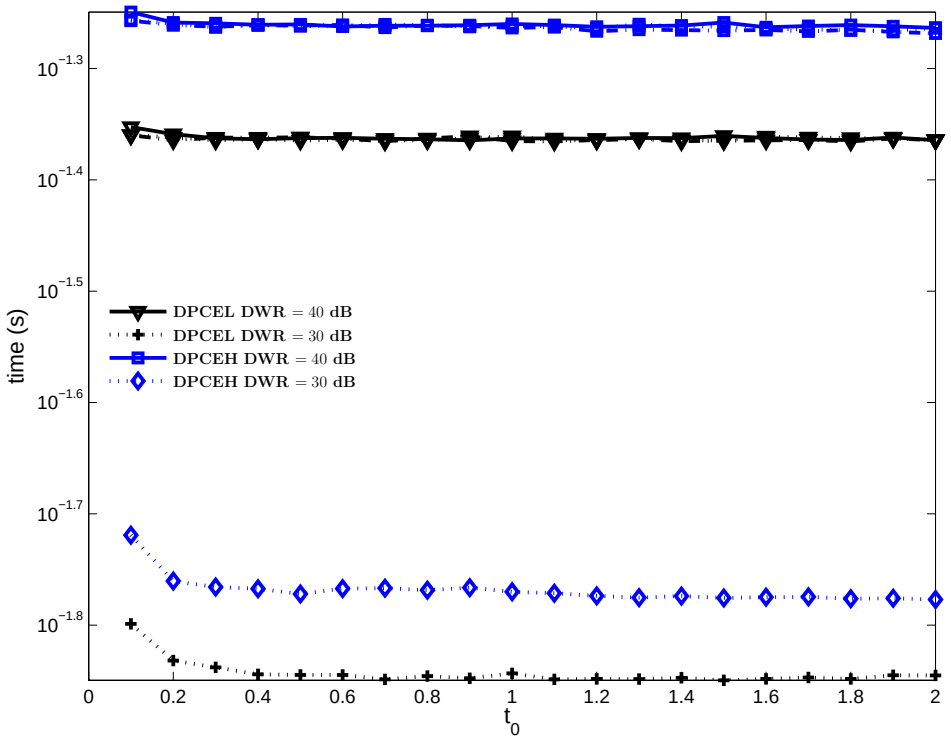


Figure 5.16: Time per estimate as a function of t_0 for DPCEL and DPCEH for DWR = 30 dB with WNR = 0 dB. DWR = 40 dB with WNR = $-3, 0, 3$ dB (solid lines, dotted lines, and dashdot lines, respectively). $\alpha = 1$, and $L = 10^3$.

In this section, the computational requirements of the DPCE are studied as the average time needed per estimate of t_0 using Monte Carlo runs. The simulations were carried out in a Matlab R2013a Core-i5-2500 3 GHz with 16 GB memory server.

In Fig. 5.16, a comparison is depicted of the time required for DPCEL and DPCEH for DWR = 30 dB and WNR = 0 dB and DWR = 40 dB and WNR = $-3, 0, 3$ dB with $L = 10^3$ and $\alpha = 1$ using the statistical interval, sampling based on modulo-lattice reduction, and the Decision-Aided technique. These results show that the required time to process 10^3 observations to estimate t_0 is less than $6 \cdot 10^{-2}$ s for any of the shown cases, indicating that our algorithms could be used in applications with severe time constraints. In addition, DPCEL takes slightly less time than DPCEH and it is worth noting that the required time increases with DWR, as one could expect since the search-interval is sampled more finely. The difference of the required time as function of the WNR for DWR = 40 dB is negligible.

It is worth noting that the required time approximately does not depend on

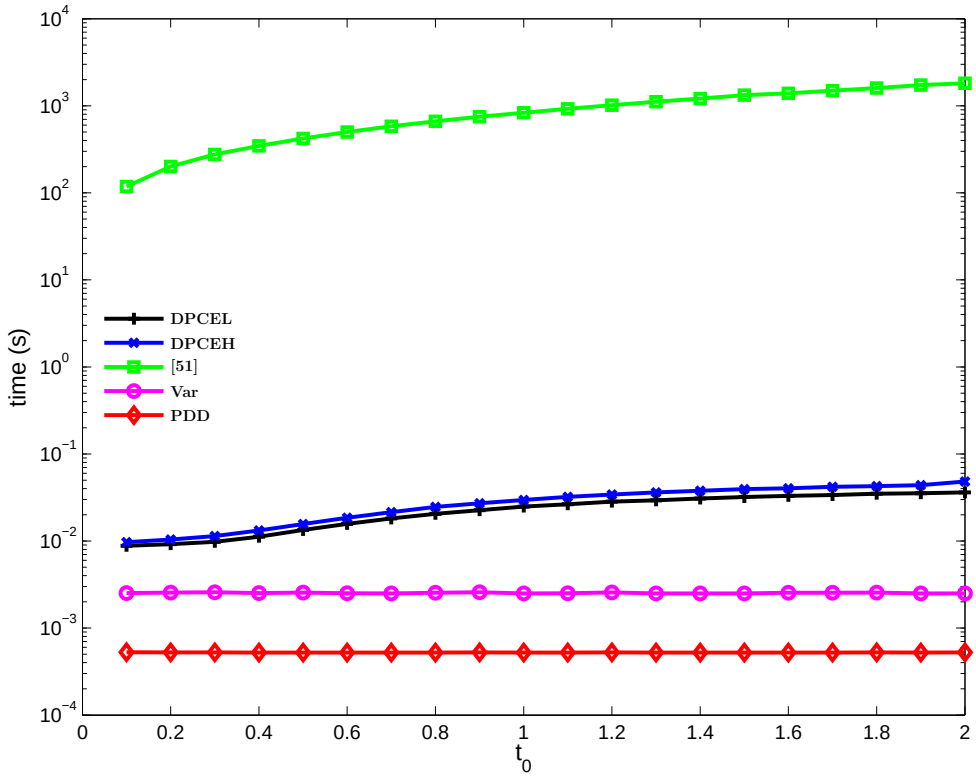


Figure 5.17: Time per estimate as function of t_0 for DPCEL, DPCEH, [51], variance-based estimator (Var), and PDD for DWR = 40 dB, WNR = 0 dB, $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$.

the value of t_0, although the width of the search-interval increases with t_0. This is explained because the cardinality of $\mathcal{T}$ does not depend on t_0 if the obtained t_- is not $\sqrt{\epsilon}$ as it occurs in this situation. In this case the iterative sampling process allows us to write

$$t_-\zeta^l \geq t_+,$$

where $|\mathcal{T}|$ is the lowest integer l verifying the previous inequality. In addition, in this expression, ζ denotes the right term of (5.7) divided by $t(l)$, i.e.,

$$\zeta \triangleq \frac{\left(\alpha\Delta^2/12 + \sigma_X^2 + \Delta/\sqrt{12}\sqrt{\Delta^2/12\left((1-\alpha)^2 + K_1(2\alpha-1)\right) + K_1\sigma_X^2}\right)}{\sigma_X^2 + \frac{\Delta^2(1-K_1)}{12}}.$$

After using the properties of the logarithmic function, l can be expressed as

$$\begin{aligned} l &\geq \frac{\log\left(t_+/t_-\right)}{\log\left(\zeta\right)} \\ &\approx \frac{\log\left(\sqrt{\left(1+K_2\sqrt{2/L}\right)/\left(1-K_2\sqrt{2/L}\right)}\right)}{\log\left(\zeta\right)}, \end{aligned}$$

where in approximation of the previous expression is assumed that $\hat{t_0^2}(\boldsymbol{z})_{\text{var}} \approx t_0^2$ and both σ_N^2 and $t_0^2\sigma_W^2$ can be neglected compared to $t_0^2\sigma_X^2$ in the calculation of $t_\pm$ of the statistical interval. It is obvious this approximation does not depend on t_0.

In Fig. 5.17, where the DPCE techniques use the configuration used in Fig. 5.16, the time required for the ML brute force technique introduced in [51], for the variance-based estimator and the PDD are also depicted. In this case, the analyzed framework is defined by DWR = 40 dB, WNR = 0 dB, $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$. On one hand, the required time for [51] is almost 4 orders of magnitude larger than for DPCE; therefore, its use for applications with time restrictions is severely limited. On the other, DPCE techniques require more time (the largest gap is located at $t_0 = 2$, more than an order of magnitude) than the variance-based estimator (note that a variance-based estimation is used to compute the search-interval) but the obtained MSE for the variance-based estimator is outperformed by DPCE techniques if the structure of the pdf of Z arises. PDD requires even less computational resources than Var, as PDD is based on first order statistics, while the variance-based estimator needs the second order statistics of the involved signals; however, the obtained MSE is worse than that achieved by the variance-based estimator.

Fig. 5.18 depicts a comparison of the required time for different values of L. For the $L = 5\cdot 10^2, 10^3, 5\cdot 10^3$ cases, the most time demanding for DPCEL and DPCEH is $L = 5\cdot 10^3$ which requires more operations for each evaluated point (this indicates that from a time perspective this effect dominates the reduction

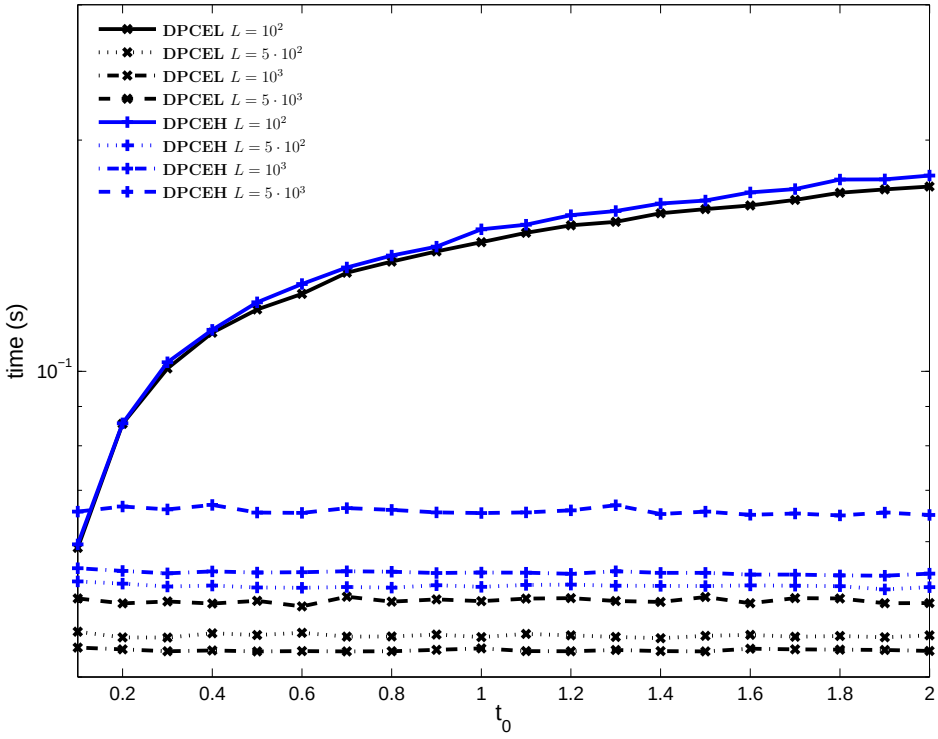


Figure 5.18: Time per estimate as function of t_0 for DPCEL and DPCEH for $L = 10^2, 5 \cdot 10^2, 10^3, 5 \cdot 10^3$. DWR = 40 dB, WNR = 0 dB, and $\alpha = 1$.

of the width of $[t_-, t_+]$ with L), while $L = 5 \cdot 10^2, 10^3$ require approximately the same time. For $L = 100$, one can realize from this figure that the required time increases with t_0, instead of being approximately flat as for the other values of L. This is a consequence of the dependence of the cardinality of the search-interval since, for these cases, $t_- = \sqrt{\epsilon}$ and the upper-bound increases with t_0. Apart from that, the cardinality of the search-interval is the largest, being almost 6 times the cardinality corresponding to $L = 5 \cdot 10^2$ with $t_0 = 2$.

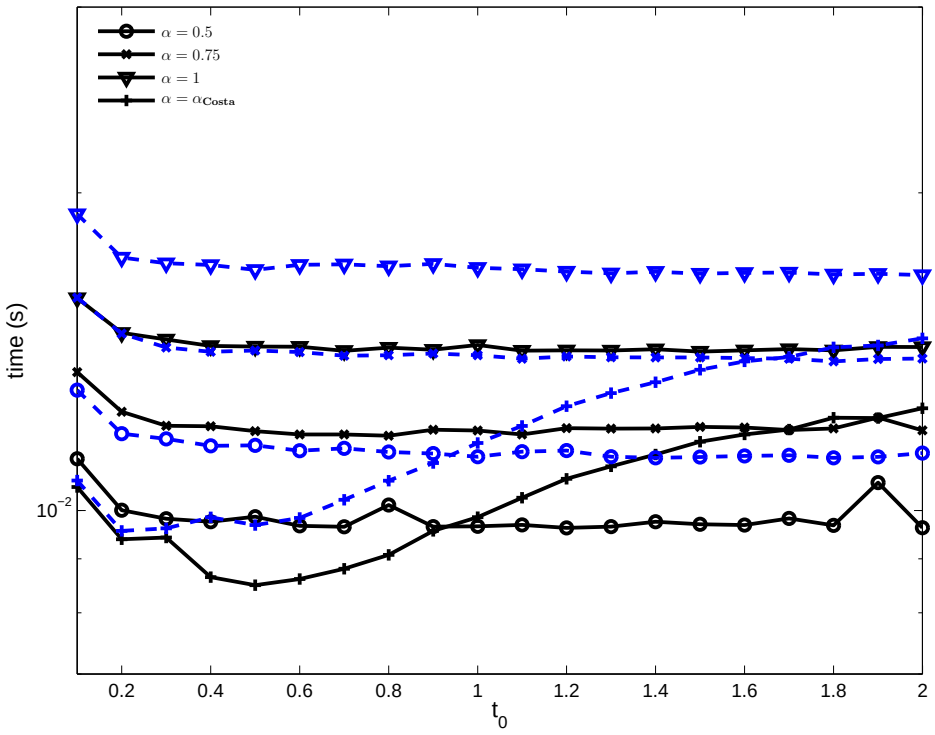


Figure 5.19: Time per estimate as function of t_0 for DPCEL (solid lines) and DPCEH (dashed lines) for DWR = 30 dB, WNR = 0 dB, $\alpha = 0.5, 0.75, 1, \alpha_{\text{Costa}}$, and $L = 10^3$.

Fig. 5.19 compares the time required by the DPCE techniques, also with the same configuration of Fig. 5.16, for different values of α (namely, $\alpha = 0.5, 0.75, 1, \alpha_{\text{Costa}}$) for DWR = 30 dB, WNR = 0 dB, and $L = 10^3$. The required time increases with α because for a given σ_W^2 if the value of α increases, then Δ will be reduced; therefore, the effect of the modulo reduction will increase and, consequently, the sampling of the search-interval will be finer. As in Fig. 5.17 for $\alpha = \alpha_{\text{Costa}}$, the required time varies as the value of α does. For example, for DPCEH when $t_0 = 2$, the corresponding value of α is 0.8, which is coherent with the time required since it needs slightly more time than that for $\alpha = 0.75$.

A comparison between the calculation of the search-interval considering the intersection of the two methods proposed in Sect. 5.1 and only using the

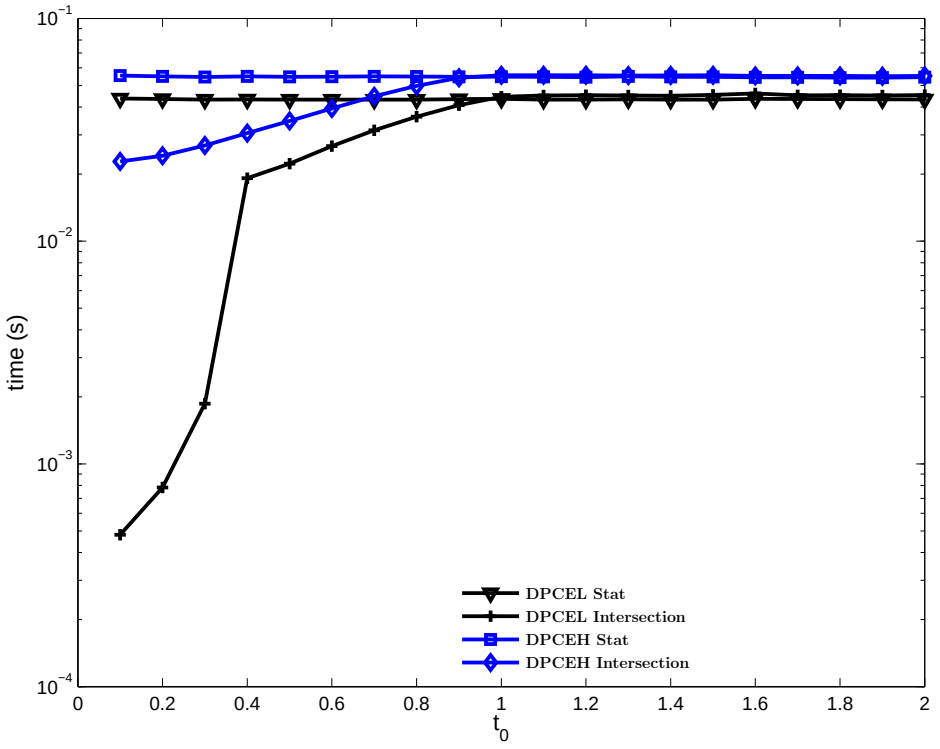
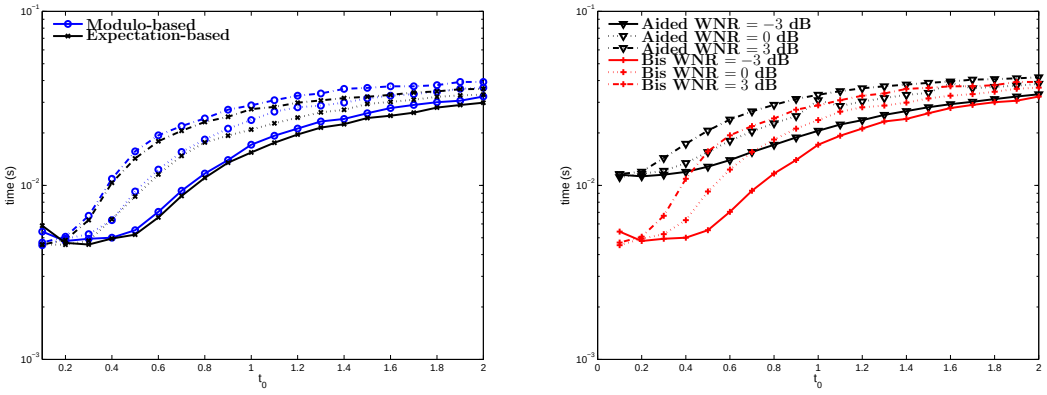


Figure 5.20: Time per estimate as function of t_0 for DPCEL and DPCEH for DWR = 40 dB and WNR = 0 dB obtaining the search-interval as intersection of the deterministic and the statistical intervals (Intersection), and using the statistical interval (Stat). $\alpha = 1$, and $L = 10^3$.

statistically-based approach defined in Sect. 5.1.1 is shown in Fig. 5.20. From these results, a clear conclusion can be drawn: the deterministically-obtained approach limits the search-interval for low values of t_0, i.e., whenever there is no structure in the pdf of Z, while when the structure of the pdf of Z arises, the statistical method limits the required time. This is coherent with the fact that the lower-bound of the cost function used for the deterministic interval is tight when there is no structure in the pdf of Z, while the statistical method is independent of the existence of such structure.

The left pane of Fig. 5.21 shows the comparison of the time required by DPCEL with the Bisection method using for sampling the search-interval both techniques proposed in Sect. 5.2. In this scenario, the obtained results indicate that such techniques require about the same time to calculate an estimate of t_0.



(a) Sampling Method (b) Local Optimization

Figure 5.21: Time per estimate as function of t_0 for DPCEL. DWR = 40 dB and WNR $= -3, 0, 3$ dB (solid lines, dotted lines, and dashdot lines, respectively) with $\alpha = \alpha_{\text{Costa}}$, and $L = 10^3$. For a) different sampling methods and b) Decision-Aided technique with $K_1 = 10^{-2}$ (Aided) and Bisection method with $K_1 = 1$ and $\epsilon = 10^{-4}$ (Bis).

The right pane of Fig. 5.21 compares the time required by the Decision-Aided technique and the Bisection method technique for DPCEL when DWR = 40 dB and WNR $= -3, 0, 3$ dB and $\alpha = \alpha_{\text{Costa}}$. For both optimization algorithms, the search-interval is obtained as the intersection of the two presented techniques; then, the search-interval is sampled using the DC-QIM's modulo-lattice reduction algorithm for Decision-Aided technique and Bisection method.

In order to obtain good results, the Decision-Aided technique can use a larger value of K_1 than the Bisection method technique and, thus, the cardinality of $\mathcal{T}$ is smaller; however, the time required to compute a local estimate when there exists a maximum/minimum within an interval limited by two consecutive elements of $\mathcal{T}$ is much longer than the time required for calculating any $t(i) \in \mathcal{T}$ using the Decision-Aided technique. This trade-off is better for the Bisection

method for low values of t_0 (i.e., whenever the cost function is smooth and there is no structure in the pdf of Z); however, when the structure of the pdf of Z appears this advantage disappears and, according to this figure, both techniques approximately require the same time.

In Fig. 5.22, the time required using Spread-Transform (ST) for the cases of DPCE for different ratios of L/L_{ST}, and where the variances of the involved signals are either known or unknown. First, if the performance of the algorithm for known variance and spreading is compared to the case where spreading is not used (in Fig. 5.16), one can conclude that the increment of the required time is significant, which is coherent with the increment of the number of computations required to estimate; indeed, for this case is more than 3 times the required time.

Second, the analysis of the two sets of curves reveals that there is a tendency in the required time to decrease with the ratio L/L_{ST} as the number of projected observations L_{ST} is reduced. Besides, the unknown variance case requires more computational resources than the variance-aware case specially because the search-interval is larger and since the value of σ_X^2 used to sample the search-interval is upper-bounded, the cardinality of the candidate-set is also much larger.

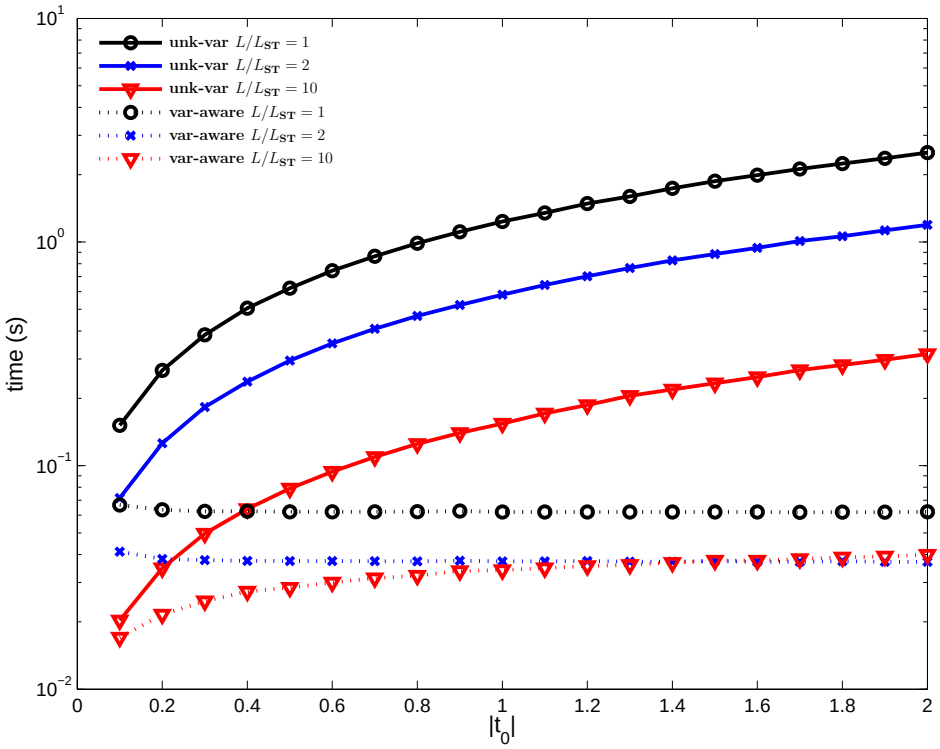


Figure 5.22: Time per estimate as function of t_0 with ST for variance-aware (var-aware) and unknown-variance (unk-var) case for different ratios L/L_{ST}. DWR = 30 dB, WNR = 0 dB, $\alpha = 1$, and $L = 10^3$.

Appendix

5.A Analysis of $L_2(t, \mathbf{z})$

The cost functions for the low-SNR case (3.11) and (3.12) are lower bounded by using that the cosine function takes values less than or equal to one in order to obtain respectively

$$L_2^{\text{low-SNR}}(t, \boldsymbol{z}) \triangleq \frac{1}{2}\left(\frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2 + \sigma_N^2} + L\log(2\pi(\sigma_X^2 t^2 + \sigma_N^2)) - 4Le^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)}{\Delta^2(\sigma_N^2 + \sigma_X^2 t^2)}}\right)$$

$$L_2^{\text{low-SNR},2}(t, \boldsymbol{z}) \triangleq \frac{1}{2}\left(\frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2} + L\log(2\pi\sigma_X^2 t^2) - 4Le^{-\frac{2\pi^2\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)}{\Delta^2 t^2}}\right); \tag{5.12}$$

where the first derivative with respect to t of (5.12) is

$$\frac{L}{t}\left(-\frac{8\pi^2\sigma_N^2 e^{-\frac{2\pi^2\left(\frac{(1-\alpha)^2\Delta^2 t^2}{12} + \sigma_N^2\right)}{\Delta^2 t^2}}}{\Delta^2 t^2} + 1 - \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2 L}\right), \tag{5.13}$$

where, by assuming that t is close to t_0, the first term of the outer parentheses can be neglected since the self-noise can be discarded compared to the channel noise because $\text{SCR}(t) \ll 1$, and $\exp(-2\pi^2\sigma_N^2/(\Delta^2 t^2))\sigma_N^2/(\Delta^2 t^2) \approx 0$ due to $\text{TNQR}(t) \gg 1$ holds for the low-SNR case. Based on this, for $t > 0$ the finite root of the approximation of the first derivative of $L_2^{\text{low-SNR},2}(t, \boldsymbol{z})$ with respect to t is approximately at $t_L = \sqrt{\|\boldsymbol{z}\|^2/(\sigma_X^2 L)}$. In addition, since $L_2^{\text{low-SNR},2}(t, \boldsymbol{z})$ is continuous, the simplified version of (5.13) tends to $-\infty$ as t tends to zero, and takes positive values for finite $t > t_L$, one can state that for finite $t \geq 0$, $L_2^{\text{low-SNR},2}(t, \boldsymbol{z})$ decreases until the unique finite minimum t_L is reached and increases for finite $t > t_L$.

The cost function for the high-SNR case (3.15) is lower bounded by

$$L_2^{\text{high-SNR}}(t, \boldsymbol{z}) \triangleq \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2} + L\log\left(2\pi\left(\sigma_N^2 + \frac{(1-\alpha)^2\Delta^2 t^2}{12}\right)\right),$$

whose zeros with respect to t are obtained by using the equivalent problem of analyzing $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ with respect to t^2. From the first derivative of $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ with respect to t^2, one can straightforwardly obtain the only positive root for $\alpha \in [0,1)$ located at $t_L^2 = (\|\boldsymbol{z}\|^2+\|\boldsymbol{z}\|\sqrt{\|\boldsymbol{z}\|^2+(4L\sigma_N^2\sigma_X^2)/((1-\alpha)^2\Delta^2/12)})/(2L\sigma_X^2)$; therefore, it is simple to prove that $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ has a unique zero at $t=t_L$. Taking this into account and additionally considering that $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ is continuous, and $L_2^{\text{high-SNR}}(t,\boldsymbol{z}) \to \infty$ as $t \to 0$ or $t \to \infty$, one can state that $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ only has a minimum at t_L, decreases for $t \geq 0$ until $t = t_L$, and increases for $t > t_L$.

In the high-SNR and $\alpha = 1$ case, the first derivative of $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ with respect to t is $-\|\boldsymbol{z}\|^2/\sigma_X^2 t^3$ and $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ is continuous; thus, $L_2^{\text{high-SNR}}(t,\boldsymbol{z})$ decreases with t for $t \geq 0$. In addition, note that $\lim_{t\to 0} L_2^{\text{high-SNR}}(t,\boldsymbol{z}) = \infty$ and $\lim_{t\to\infty} L_2^{\text{high-SNR}}(t,\boldsymbol{z}) = 0$.

5.B Analysis of (5.7) with Respect to σ_X

As described in Sect. 5.2.1, the elements $t(l)$ of the candidate set $\mathcal{T}$ are iteratively obtained as

$$t(l+1) = \frac{t(l)\left(\alpha\Delta^2/12 + \sigma_X^2 + \Delta/\sqrt{12}\sqrt{\Delta^2/12\left((1-\alpha)^2 + K_1(2\alpha-1)\right) + K_1\sigma_X^2}\right)}{\sigma_X^2 + \frac{\Delta^2(1-K_1)}{12}}. \tag{5.14}$$

The previous expression has a discontinuity at

$$\sigma_X = \pm\frac{\Delta\sqrt{K_1-1}}{2\sqrt{3}}, \tag{5.15}$$

which only takes real values for $K_1 \geq 1$. Taking this into account, it can be stated that on one hand (5.14) is continuous for $\sigma_X > 0$ if $K_1 < 1$ (verified in most of the real scenarios); on the other hand if $K_1 \geq 1$, (5.14) is continuous for $\sigma_X > 0$ except at $\sigma_X = \Delta\sqrt{K_1-1}/(2\sqrt{3})$ and

$$\lim_{\sigma_X \to \left(\Delta\sqrt{K_1-1}/(2\sqrt{3})\right)^-} t(l+1)/t(l) = -\infty,$$
$$\lim_{\sigma_X \to \left(\Delta\sqrt{K_1-1}/(2\sqrt{3})\right)^+} t(l+1)/t(l) = \infty.$$

The first derivative of (5.14) with respect to σ_X is

$$-\frac{12\Delta\sigma_X}{\left(\Delta^2(K_1-1)-12\sigma_X^2\right)^2\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}}$$
$$\times\Bigg[\Delta^2\left((4\alpha-3)K_1+2(\alpha-1)^2+K_1^2\right)$$
$$+2\Delta(\alpha+K_1-1)\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}+12K_1\sigma_X^2\Bigg]. \quad (5.16)$$

Assuming that $\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}$ is real (otherwise, the candidate set $\mathcal{T}$ would take imaginary values) and takes positive values, $12\Delta\sigma_X/\left(\Delta^2(K_1-1)-12\sigma_X^2\right)^2$ also takes positive values (except for the discontinuity); if the expression within brackets of (5.16) takes positive values, then (5.16) will take negative values. Based on this, we focus our analysis in the expression within brackets which can be rewritten as

$$\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}\left(\frac{\Delta^2(\alpha+K_1-1)^2}{\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}}\right.$$
$$\left.+\sqrt{\Delta^2\left((2\alpha-1)K_1+(\alpha-1)^2\right)+12K_1\sigma_X^2}+2\Delta(\alpha+K_1-1)\right). \quad (5.17)$$

By dismissing the square root of the previous expression since it takes real and positive values, as indicated in the previous paragraph, a positive root with respect to σ_X of the expression of the outer parentheses is located at $\sigma_X=\frac{\Delta\sqrt{K_1-1}}{2\sqrt{3}}$ (i.e., discontinuity of (5.14) for $\sigma_X>0$ shown in (5.15)), therefore:

- For $K_1<1$, there are not real roots of (5.14) with respect to σ_X; therefore, (5.14) is continuous and takes positive values for $\sigma_X\to\infty$, then (5.14) takes positive values for $\sigma_X>0$. In addition, the first derivative of (5.14) with respect to σ_X is negative.
- If $1\leq K_1$, there is a discontinuity of (5.14) at (5.15) for real values of σ_X; however, for values of σ_X different of (5.15), the three terms of the outer parentheses of (5.17) take positive values and, therefore, the first derivative (5.14) with respect to σ_X takes negatives values.

5.C Maximum/minimum of the Derivative of M

If the four hypotheses hold for t close to t_0, the element of the expectation of the cost function $\mathrm{E}\{L(t,Z)\}$ due to the structure M defined in (5.9) can be approximated as

$$M\approx -2e^{-\frac{\pi^2\left(t\left((1-\alpha)^2\Delta^2t^2+6\sigma_X^2(t-t_0)^2\right)t_0^2+6\sigma_N^2\left(t^3+2t^2t_0+5tt_0^2+2t_0^3\right)\right)}{3\Delta^2t^3t_0^2}}.$$

In order to obtain the maxima/minima of the first derivative of the previous expression with respect to t, its second derivative with respect to t is calculated

$$\begin{aligned}\frac{\partial^2 M}{\partial t^2} \approx & -\frac{8\pi^2 e^{-\frac{\pi^2\left(t\left((1-\alpha)^2\Delta^2t^2+6\sigma_X^2(t-t_0)^2\right)t_0^2+6\sigma_N^2\left(t^3+2t^2t_0+5tt_0^2+2t_0^3\right)\right)}{3\Delta^2t^3t_0^2}}}{\Delta^4t^8t_0^2} \\ & \times\Big(4\pi^2\left(\sigma_X^2tt_0^2(-t+t_0)+\sigma_N^2\left(t^2+5tt_0+3t_0^2\right)\right)^2 \\ & +\Delta^2t^3t_0\left(\sigma_X^2t(2t-3t_0)t_0^2-\sigma_N^2\left(2t^2+15tt_0+12t_0^2\right)\right)\Big). \end{aligned} \tag{5.18}$$

The roots of the previous expression correspond to the roots of the expression within the outer parentheses, i.e.,

$$\begin{aligned}&\Big(4\pi^2\left(\sigma_X^2tt_0^2(-t+t_0)+\sigma_N^2\left(t^2+5tt_0+3t_0^2\right)\right)^2 \\ &+\Delta^2t^3t_0\left(\sigma_X^2t(2t-3t_0)t_0^2-\sigma_N^2\left(2t^2+15tt_0+12t_0^2\right)\right)\Big). \end{aligned} \tag{5.19}$$

As stated above, it is assumed that t and t_0 are close, we propose to approximate that expression locally by its second order Taylor expansion around t_0, yielding

$$\begin{aligned}&324\pi^2\sigma_N^4t_0^4-\Delta^2t_0^6\left(29\sigma_N^2+\sigma_X^2t_0^2\right) \\ &+(t-t_0)\left(-2\Delta^2t_0^5\left(53\sigma_N^2+\sigma_X^2t_0^2\right)+72\pi^2\left(7\sigma_N^4t_0^3-\sigma_N^2\sigma_X^2t_0^5\right)\right) \\ &+\frac{1}{2}(t-t_0)^2\left(4\Delta^2t_0^4\left(-73\sigma_N^2+\sigma_X^2t_0^2\right)+8\pi^2\left(67\sigma_N^4t_0^2-32\sigma_N^2\sigma_X^2t_0^4+\sigma_X^4t_0^6\right)\right). \end{aligned} \tag{5.20}$$

A comparison between the expression within the outer parentheses of (5.18) and its approximation is shown in Fig. 5.23, which illustrates the resemblance between them.

Therefore, one can solve the following equation to determine the maxima/minima of M

$$\begin{aligned}&324\pi^2\sigma_N^4t_0^4-\Delta^2t_0^6\left(29\sigma_N^2+\sigma_X^2t_0^2\right) \\ &+(t-t_0)\left(-2\Delta^2t_0^5\left(53\sigma_N^2+\sigma_X^2t_0^2\right)+72\pi^2\left(7\sigma_N^4t_0^3-\sigma_N^2\sigma_X^2t_0^5\right)\right) \\ &+\frac{1}{2}(t-t_0)^2\left(4\Delta^2t_0^4\left(-73\sigma_N^2+\sigma_X^2t_0^2\right)+8\pi^2\left(67\sigma_N^4t_0^2-32\sigma_N^2\sigma_X^2t_0^4+\sigma_X^4t_0^6\right)\right)=0. \end{aligned}$$

The previous equation can be simplified by employing HQR $\gg 1$ and TNHR$(t_0) \ll 1$, yielding

$$t_0^4\left(-\Delta^2\sigma_X^2t_0^4+4\pi^2\left(9\sigma_N^2+\sigma_X^2t_0(-t+t_0)\right)^2\right)=0,$$

whose two solutions with respect to t are

$$\begin{aligned}t_l(t_0) &\triangleq \frac{9\sigma_N^2}{\sigma_X^2t_0}+t_0-\frac{\Delta t_0}{2\pi\sigma_X} \\ t_r(t_0) &\triangleq \frac{9\sigma_N^2}{\sigma_X^2t_0}+t_0+\frac{\Delta t_0}{2\pi\sigma_X}. \end{aligned}$$

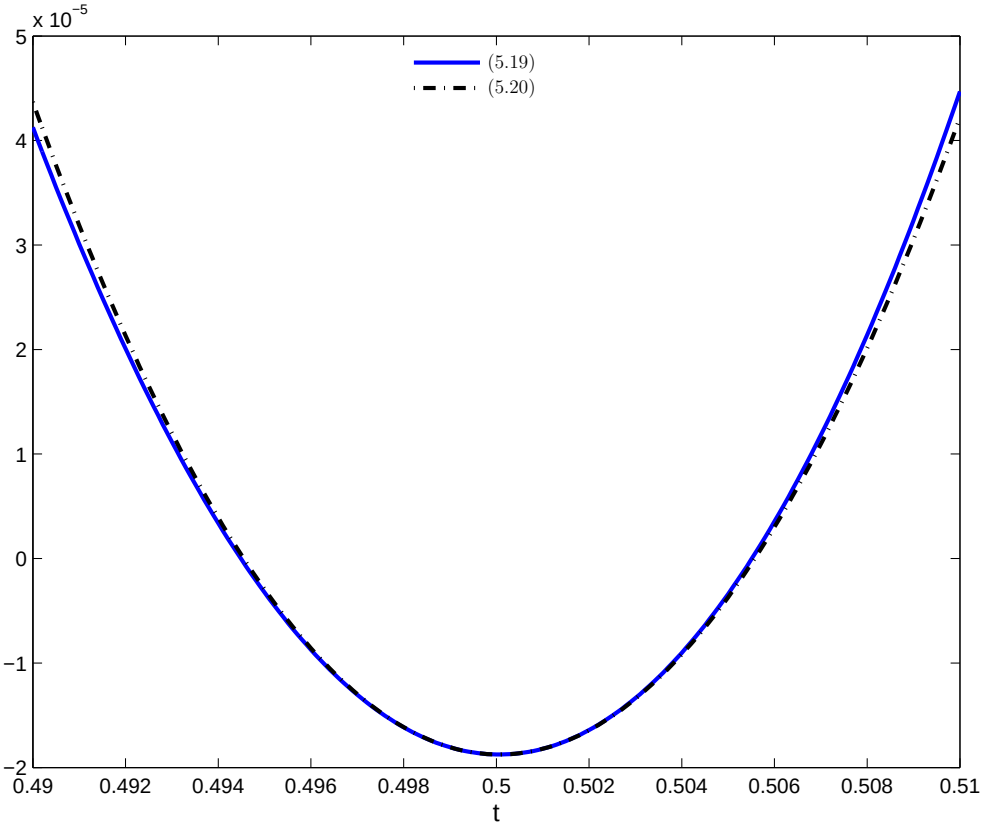


Figure 5.23: Comparison of (5.19) and its approximation (5.20) as a function of t. DWR $= 20$ dB, WNR $= 0$ dB, $\alpha = 0.5$, and $t_0 = 0.5$.

It is worth pointing out that both $t_l(t_0)$ and $t_r(t_0)$ asymptotically tend to t_0 as HQR $\gg 1$ and TNHR$(t_0) \ll 1$ are verified.

5.D Sampling Based on the Distribution of $L(t, \mathbf{z})$

In order to derive a sampling criterion of the search-interval, we will take into account that, as it has been already mentioned, the local minima of (3.11) (obtained in Sect. 3.2.1) are due to the term of induced structure, i.e.,

$$J(t, \boldsymbol{z}) \triangleq -\sum_{i=1}^{L} 2e^{-\frac{2\pi^2\sigma_X^2\left(\sigma_N^2+\frac{(1-\alpha)^2\Delta^2t^2}{12}\right)}{\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}} \cos\left(\frac{2\pi\sigma_X^2tz_i}{\Delta\left(\sigma_N^2+\sigma_X^2t^2\right)} - \frac{2\pi d_i}{\Delta}\right). \quad (5.21)$$

Therefore, we will focus our analysis on finding the set $\mathcal{T}$ that can locate the local minima of (5.21). It is worth pointing out that if the sampling of the search interval guarantees the convergence of the local optimization algorithm to the

local minima, then the local minima of $L(t, \mathbf{z})$ will be also reached using the same approach because the first and second terms, which are not considered in (5.21), are smooth functions that can be included without changing the essence of the procedure here presented.

In order to pinpoint the local minima of (5.21), we find it useful to have an approximation to the pdf of $J(t, \boldsymbol{z})$ when t_0 is given, as well as such pdf for the case where t is not close to t_0. Here, the value t is considered not close to t_0, if the resulting quantizer step size using t produces a set of centroids that are not compatible with the corresponding set of centroids for t_0 (i.e., the embedder and the estimator are not synchronized).

The CLT is used to get the distribution of $J(t, \boldsymbol{z})$ assuming that $L \to \infty$ and each component of $\boldsymbol{Z}$ is independent. If t and t_0 are close, the distribution of $J(t, \boldsymbol{Z})$ is denoted by V. First, the expectation of V is obtained by using $f_{Z|T,K}^{\text{low-SNR}}(z|t, d)$ as approximation to the pdf of Z

$$\mu_V(t) \approx L \int_{-\infty}^{\infty} f_{Z|T,K}^{\text{low-SNR}}(\tau|t, d)\, (J(t, \tau))\, d\tau,$$

which was calculated in App. 3.A. Then, we obtain

$$\begin{aligned}&\mu_V(t)\\ &= -L2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t^2}{\sigma_N^2+\sigma_X^2t^2}+\frac{4\sigma_X^2tt_0}{\sigma_N^2+\sigma_X^2t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2tt_0\right)^2}{\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t_0^2}{\sigma_N^2+\sigma_X^2t_0^2}\right)}{\Delta^2}}.\end{aligned} \tag{5.22}$$

As analyzed in Sect. 3.2.1, $\mu_V(t)$ has only a minimum, while its maxima are at $t = 0$ and $t \to \infty$, where it is null.

Secondly, the variance of V is approximated in App. 5.E by

$$\begin{aligned}\sigma_V^2(t) = {} & L2e^{-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}\\ & -L\left(2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t^2}{\sigma_N^2+\sigma_X^2t^2}+\frac{4\sigma_X^2tt_0}{\sigma_N^2+\sigma_X^2t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2tt_0\right)^2}{\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t_0^2}{\sigma_N^2+\sigma_X^2t_0^2}\right)}{\Delta^2}}\right)^2;\end{aligned} \tag{5.23}$$

therefore, the distribution of V can be approximated by $\mathcal{N}\left(\mu_V(t), \sigma_V^2(t)\right)$.

As stated above, the distribution of $J(t, \boldsymbol{Z})$ when t is not close to t_0, which is denoted by U, is required, which is approximated using the CLT by considering that $L \to \infty$. In this way, since t and t_0 are not close, one can assume that the embedder and the estimator are not synchronized; therefore, the distribution of

the cosine function of $J(t, \boldsymbol{z})$ in (5.21) can be modeled by a uniform distribution in $[-\Delta/2, \Delta/2]$, and, thus, $\mu_U(t) \approx 0$. Taking this into account in (5.23) by neglecting its second term, $\sigma_U^2(t)$ can be approximated as

$$\sigma_U^2(t) = L2e^{-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}.$$

Therefore, the distribution of U can be approximated by $\mathcal{N}(0, \sigma_U^2(t))$.

Given the distribution of V, the distribution of U, and the value of t_0, the interval of values of t guaranteeing that the local method converges to t_0 is obtained following a probabilistic approach. This approach is based on two thresholds functions $\gamma_U(t)$ and $\gamma_V(t)$, for U and V respectively. On one hand $\gamma_U(t)$ is defined as the threshold that is exceeded by U with probability $p_{\gamma,U}$, i.e.,

$$1 - \mathcal{Q}\left(\frac{\gamma_U(t)}{\sigma_U(t)}\right) = p_{\gamma,U},$$

where throughout this section $\mathcal{Q}(\cdot)$ denotes the $\mathcal{Q}$-function defined as

$$\mathcal{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{u^2}{2}}du.$$

On the other hand for V, $\gamma_V(t)$ is defined as the function of t that indicates the value that is not exceed by V with probability $p_{\gamma,V}$, i.e.,

$$\mathcal{Q}\left(\frac{\gamma_V(t) - \mu_V(t)}{\sigma_V(t)}\right) = p_{\gamma,V}.$$

$\gamma_U(t)$ is negative and strictly decreases with t as shown in App. 5.F. Furthermore, in order to analyze the expression of $\gamma_V(t)$

$$\gamma_V(t) = \sqrt{L}\sigma_V(t)\mathcal{Q}^{-1}(p_{\gamma,V}) + L\mu_V(t), \tag{5.24}$$

where $\mathcal{Q}^{-1}(\cdot)$ is the inverse of the $\mathcal{Q}$-function defined above. We first focus on the term depending on $\sigma_V(t)$, where $\sigma_V^2(t)$ is defined in (5.23), where we assume that in the relevant application scenarios with t close to t_0. The first component of (5.23) is smoother than the other component of this variance, i.e., the negative of the square of $\mu_V(t)$ (i.e., $\mu_V^2(t)$). Specifically, $\mu_V(t)$ has been previously studied showing that asymptotically only has a minimum at $t = t_0$ and, thus, $-\mu_V^2(t)$ only has a minimum at t_0 (due to the monotonically increasing nature of the square function); therefore, since the square root of $\sigma_V^2(t)$ maintains its maxima/minima (since the square root is also a monotonically increasing function for $t \geq 0$), one can accurately approximate that the first term of (5.24) has asymptotically a minimum at t_0. Coherently, the second component of (5.24), which contains $\mu_V(t)$, only has a minimum at $t = t_0$. Bearing this in mind, one can conclude that $\gamma_V(t)$ has a minimum at $t = t_0$ when t is close to t_0.

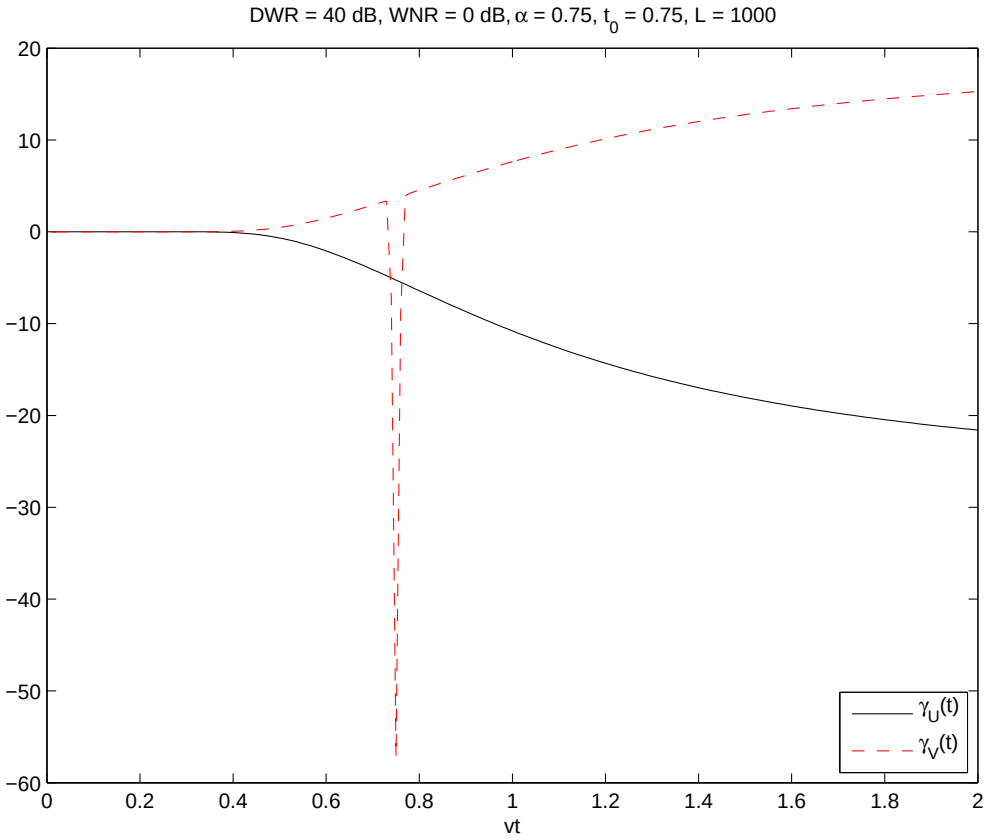


Figure 5.24: Example of curves $\gamma_U(t)$ (solid line) and $\gamma_V(t)$ (dashed line) vs. t. $t_0 = 0.75$, DWR $= 40$ dB, WNR $= 0$ dB, $\alpha = 0.75$, and $L = 10^3$.

Given this, one has to calculate where $\gamma_U(t)$ and $\gamma_V(t)$ cross each other to identify the range of values of t, where it is possible to know with a given probability that t is close to t_0. This can obtained as the solution of the following equation with respect to t

$$\gamma_V(t) = \gamma_U(t). \tag{5.25}$$

Note that for $p_{\gamma,V} \leq 0.5$, both functions cross for an arbitrarily large value of L due to the linear dependency of the expectation of V with L while the standard deviation linearly depends on $\sqrt{L}$. The overall probability of making a mistake in the differentiation of V and U can be approximated by $1-(1-p_{\gamma,U})(1-p_{\gamma,V})$. Given a value of L, if $p_{\gamma,U}$ or/and $p_{\gamma,V}$ are decreased, and there exists a solution for (5.25) (otherwise, L is not large enough), then obtained width of the range of t will be also decreased.

Fig. 5.24 depicts $\gamma_U(t)$ and $\gamma_V(t)$ with $t_0 = 0.75$, DWR $= 40$ dB, WNR $= 0$ dB, $\alpha = 0.75$, and $L = 10^3$, where one can check that $\gamma_U(t)$ and $\gamma_V(t)$ cross at two points. These points would constitute the endpoints of the interval where t is close to t_0.

So far, it was assumed that the real t_0 is known; however, this is the value that is needed to be estimated. In order to generate the values of $\mathcal{T}$, the following process is carried out using one of the search intervals defined in the previous section. First, it is assumed that $t_0 = t_-$ and the algorithm described above is used to obtain $t(1)$, next $t_0 = t(1)$ and $t(2)$ is obtained using our algorithm iteratively. This process is repeated until t_+ is exceeded by $t(i)$.

5.E Derivation of $\sigma_V^2(t)$

The variance of the third term of the cost function can be calculated as

$$\sigma_V^2(t) = L\mathrm{E}\left\{\left(J(t,Z)\right)^2\right\} - L\left(\mathrm{E}\left\{J(t,Z)\right\}\right)^2.$$

As the case of the expectation calculated in the previous appendix, the approximation of the pdf of Z used is $f_{Z|T,K}^{\text{low-SNR}}(z|t,d)$.

On one hand,

$$\begin{aligned}
\mathrm{E}\left\{\left(J(t,Z)\right)^2\right\} &= 2\left(\cos\left(\frac{4\pi d}{\Delta}\right)e^{-\frac{\pi^2\sigma_X^2\left(\left(\sigma_N^2+\sigma_X^2t^2\right)\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)+24\sigma_X^2t^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)^2}}\right.\\
&+\cos\left(\frac{2\pi d}{\Delta}\right)\left(2e^{-\frac{\pi^2\sigma_X^2\left(\sigma_N^2\left((1-\alpha)^2\Delta^2\left(2t^2+t_0^2\right)+12\sigma_X^2\left(t^2+3t_0^2\right)\right)+3\sigma_X^2t^2t_0^2\left((1-\alpha)^2\Delta^2+4\sigma_X^2\right)+36\sigma_N^4\right)}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)\left(\sigma_N^2+\sigma_X^2t_0^2\right)}}\right.\\
&+\exp\left(-\frac{\pi^2\sigma_X^2}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}\left(\sigma_N^4\left((1-\alpha)^2\Delta^2\left(2t^2+t_0^2\right)\right.\right.\right.\\
&+12\sigma_X^2\left(8t^2-4tt_0+3t_0^2\right)\Big)+2\sigma_N^2\sigma_X^2t\left((1-\alpha)^2\Delta^2t\left(t^2+2t_0^2\right)+6\sigma_X^2\left(t^3-4t^2t_0+12tt_0^2-4t_0^3\right)\right)\\
&\left.\left.+3\sigma_X^4t^2t_0^2\left((1-\alpha)^2\Delta^2t^2+4\sigma_X^2(t-2t_0)^2\right)+36\sigma_N^6\right)\right)+\cos\left(\frac{6\pi d}{\Delta}\right)\\
&\times\exp\left(-\frac{\pi^2\sigma_X^2}{6\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}\left(\sigma_N^4\left((1-\alpha)^2\Delta^2\left(2t^2+t_0^2\right)+12\sigma_X^2\left(8t^2+4tt_0+3t_0^2\right)\right)\right.\right.\\
&+2\sigma_N^2\sigma_X^2t\left((1-\alpha)^2\Delta^2t\left(t^2+2t_0^2\right)+6\sigma_X^2\left(t^3+4t^2t_0+12tt_0^2+4t_0^3\right)\right)\\
&\left.\left.+3\sigma_X^4t^2t_0^2\left((1-\alpha)^2\Delta^2t^2+4\sigma_X^2(t+2t_0)^2\right)+36\sigma_N^6\right)+\exp\left(-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}\right)\right);
\end{aligned}$$

and, on the other hand $\left(\mathrm{E}\left\{J(t,Z)\right\}\right)^2$ can be accurately approximated by the square of (5.22). Although $\boldsymbol{d}$ is given, we average over the dither to obtain the expectation taking into account that $\boldsymbol{Z}$ is not an identically distributed random

vector. The resulting expression of the variance of $J(t, \boldsymbol{Z})$ is

$$\sigma_V^2(t) = L2e^{-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}} - L\left(2e^{\frac{2\pi^2\sigma_X^2\left(-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t^2}{\sigma_N^2+\sigma_X^2t^2}+\frac{4\sigma_X^2tt_0}{\sigma_N^2+\sigma_X^2t^2}-\frac{\sigma_X^2(t+t_0)^2\left(\sigma_N^2+\sigma_X^2tt_0\right)^2}{\left(\sigma_N^2+\sigma_X^2t^2\right)^2\left(\sigma_N^2+\sigma_X^2t_0^2\right)}-\frac{\sigma_N^2+\frac{1}{12}(1-\alpha)^2\Delta^2t_0^2}{\sigma_N^2+\sigma_X^2t_0^2}\right)}{\Delta^2}}\right)^2.$$

5.F Analysis of $\gamma_U(t)$

The threshold $\gamma_U(t)$ is obtained as

$$\gamma_U(t) = \sqrt{L}\sigma_U(t)\mathcal{Q}^{-1}(1-p_{\gamma,U}),$$

thus, in order to analyze the maxima/minima of $\gamma_U(t)$ its first derivative with respect to t is obtained

$$\frac{\partial\gamma_U(t)}{\partial t} = \sqrt{L}\mathcal{Q}^{-1}(1-p_{\gamma,U})\frac{4\pi^2\sigma_N^2\sigma_X^2t\left(12\sigma_X^2-(1-\alpha)^2\Delta^2\right)e^{-\frac{\pi^2\sigma_X^2\left((1-\alpha)^2\Delta^2t^2+12\sigma_N^2\right)}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)}}}{3\Delta^2\left(\sigma_N^2+\sigma_X^2t^2\right)^2},$$

where in the previous expression, due to HQR $\gg 1$, $(1-\alpha)^2\Delta^2$ can be neglected in comparison to $12\sigma_X^2$, $\mathcal{Q}^{-1}(1-p_{\gamma,U})$ is negative (by assuming that $p_{\gamma,U} < 0.5$), therefore the first derivative of $\gamma_U(t)$ with respect to t is negative, implying that $\gamma_U(t)$ monotonically decreases.

Chapter 6

Applications

In the previous chapters of the thesis, an ML-based estimation technique was introduced for real-valued Gaussian distributed signals and real-valued channel gains. Its theoretical limits were studied following an estimation and information theoretical approach, a set of practical algorithms to rapidly obtain the estimate of t_0 was proposed, and, finally, its performance, measured in terms of accuracy and required time, was evaluated by carrying out a set of experiments.

In this chapter, with the aim of providing insights into the wide range of practical uses of DPCE, we present a set of applications of the proposed technique dealing with making SCS robust to gain attacks, equalizing the channel gain in real digital communication scenarios, and extending our algorithm to the case of complex Gaussian signals and complex gains.

6.1 Scalar Costa Scheme Robust to Gain Attacks

Watermarking schemes based on the DPC paradigm have been shown to achieve much higher rates than classical SS methods. However, in practice, the latter continue to be used due to their higher security and robustness. A simple but devastating special case for SCS (the most prevalent DPC method) is the fixed gain attack (a.k.a. linear valumetric attack), in which the channel simply multiplies the watermarked signal by a constant real number. Even such a simple channel has shown to have dramatic consequences on the decoding of SCS, yielding very large probabilities of decoding error. In this section, we follow the equalization approach, proposed in [18] and later developed in [4, 51], for which an estimate of the channel gain is needed and obtained by making use of DPCE.

6.1.1 Overview of SCS Data Hiding

In this section we will focus on the binary implementation of SCS, i.e., the case where two scalar quantizers (corresponding to the embedded bit) are used. The binary vector $\boldsymbol{m}$ is embedded by modifying the host signal $\boldsymbol{x}$ (we assume $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \sigma_X^2 I_{L\times L})$) to the watermarked signal $\boldsymbol{y}$, which is given by

$$y_i = x_i + \alpha\left(\mathcal{Q}_\Delta\left(x_i - d_i - m_i\frac{\Delta}{2}\right) - \left(x_i - d_i - m_i\frac{\Delta}{2}\right)\right),$$

for all $1 \leq i \leq L$. Note that the difference between the previous expression and its counterpart for DPCE (2.3) of Sect. 2.3 lies in that in this case a message is embedded into the host and the estimation is carried out taking advantage of such data embedding; while in the other case, the host is modified only for estimation purposes and, thus, there is not embedded message.

Here, we propose a modification of the framework introduced in Chap. 2 by considering two independent noise signals. Indeed, under the fixed gain attack the received signal $\boldsymbol{z}$ is defined as

$$\boldsymbol{z} = t_0\left(\boldsymbol{y} + \boldsymbol{n}_1\right) + \boldsymbol{n}_2, \tag{6.1}$$

where t_0 is a true gain factor, $\boldsymbol{N}_1 \sim \mathcal{N}(\boldsymbol{0}, \sigma_{N_1}^2 I_{L\times L})$, $\boldsymbol{N}_2 \sim \mathcal{N}(\boldsymbol{0}, \sigma_{N_2}^2 I_{L\times L})$, $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ are mutually independent and also independent of $\boldsymbol{Y}$. Be aware that in [4, 51] $\boldsymbol{z}$ corresponds to $\boldsymbol{z} = t_0\left(\boldsymbol{y} + \boldsymbol{n}_1\right)$; therefore, the model considered here is slightly more general (its usefulness in modeling practical situations will be shown in Sect. 6.1.3).

The most extended implementation of the decoder estimates the ith embedded bit as

$$\hat{m}_i = \underset{m\in\{0,1\}}{\arg\min}\left|\mathcal{Q}_\Delta\left(z_i - d_i - m\frac{\Delta}{2}\right) - \left(z_i - d_i - m\frac{\Delta}{2}\right)\right|.$$

However, if $t_0 \neq 1$ the embedding and decoding codebooks will be misaligned with the consequence of significantly increasing the decoding error probability [47]. This problem could be easily solved if the gain factor t_0 were known; as this is not the case, it must be estimated from the received samples. To this end, it is possible to take advantage of the structure of the watermarked signal distribution (which is induced by SCS embedding). The decoder can exploit this estimate, denoted by $\hat{t}_0(\boldsymbol{z})$, to equalize the received samples before decoding. Specifically,

$$\hat{m}_i = \underset{m\in\{0,1\}}{\arg\min}\left|\mathcal{Q}_\Delta\left(\frac{z_i}{\hat{t}_0(\boldsymbol{z})} - d_i - m\frac{\Delta}{2}\right) - \left(\frac{z_i}{\hat{t}_0(\boldsymbol{z})} - d_i - m\frac{\Delta}{2}\right)\right|. \tag{6.2}$$

6.1.2 Gain Factor Estimation

Since *a priori* knowledge of t_0 is not available in general, we propose to obtain an approximation of the ML estimate of t_0. Following the same approach proposed in this thesis, due to the componentwise independence of $\boldsymbol{Z}$, the ML estimate is calculated as $\hat{t}_0(\boldsymbol{z}) = \arg\min_t L(t, \boldsymbol{z})$, where $L(t, \boldsymbol{z}) \triangleq -2\sum_{i=1}^{L} \log f_{Z|T,K}(z_i|t, d_i)$, and the embedded bits are modeled by an i.i.d. random variables that take the value 0 or 1 with equal probability.

Unfortunately, $L(t, \boldsymbol{z})$ is an involved function, so we propose to simplify the ML estimation by approximating the pdf of Z based on the approach followed in Sect. 3.1.1 to obtain the Z's pdf for low-SNR cases. Specifically, the approximation used here is based on the adjustment of the assumptions presented in Chap. 2: a) $\sigma_X^2 \gg \Delta^2/12$ (HQR $\gg 1$ that is verified for a wide range of real applications) in order to use the flat-host assumption (see [44]); b) the scaled self-noise variance [45] is much smaller than the total channel noise variance, i.e., $(1-\alpha)^2 t_0^2 \Delta^2/12 \ll t_0^2\sigma_{N_1}^2 + \sigma_{N_2}^2$ (a version of SCR$(t_0) \ll 1$); c) the variance of the total noise (self-noise plus total channel noise) is larger than the second moment of the scaled quantization lattice used at the decoder, i.e., $(1-\alpha)^2 t_0^2 \Delta^2/12 + t_0^2\sigma_{N_1}^2 + \sigma_{N_2}^2 > t_0^2\Delta^2/48$ (adjustment of TNQR$(t_0) \gg 1$) ; and d) the variance of the total noise is much smaller than the variance of the scaled host, i.e., $(1-\alpha)^2 t_0^2 \Delta^2/12 + t_0^2\sigma_{N_1}^2 + \sigma_{N_2}^2 \ll t_0^2\sigma_X^2$ (adaptation of TNHR$(t_0) \ll 1$). Under these hypotheses, $f_{Z|T,K}(z|t, d)$ can be approximated as

$$f_{Z|T,K}(z|t,d) \approx \frac{e^{-\frac{z^2}{2\sigma_X^2 t^2}}}{\sqrt{2\pi\sigma_X^2 t^2}} \times \left(1 + 2e^{-\frac{2\pi^2\left(\sigma_{N_2}^2 + t^2\left(\sigma_{N_1}^2 + \frac{(1-\alpha)^2\Delta^2}{12}\right)\right)}{(\Delta/2)^2 t^2}} \cos\left(\frac{2\pi z}{\Delta t/2} - \frac{2\pi d}{\Delta/2}\right)\right).$$

It is worth comparing the previous approximation of the pdf of Z with the counterpart introduced for the low-SNR case (3.8) in Sect.3.1.1. The difference comes from their respective different frameworks: a) there is a component due to the scaled channel noise and given by $t^2\sigma_{N_1}^2$, and b) the distance between contiguous centroids is $\Delta t/2$, instead of Δt as in (3.8) when the message is not embedded. Therefore, under the assumptions introduced above, $L(t, \boldsymbol{z})$ can be approximated as

$$L(t, \boldsymbol{z}) \approx \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2 t^2} + L\log(2\pi\sigma_X^2 t^2) - 4\sum_{i=1}^{L} e^{-\frac{2\pi^2\left(\sigma_{N_2}^2 + t^2\left(\sigma_{N_1}^2 + \frac{(1-\alpha)^2\Delta^2}{12}\right)\right)}{(\Delta/2)^2 t^2}} \cos\left(\frac{2\pi z_i}{\Delta t/2} - \frac{2\pi d_i}{\Delta/2}\right). \quad (6.3)$$

The same issues regarding how to reach the global minimum $L(t, \boldsymbol{z})$ described in the basic framework of this thesis appear here; therefore, we adapt one of the

algorithms proposed in the previous chapter for this particular scenario; specifically, the statistical search-interval proposed in Sect. 5.1.1, the sampling of the candidate set based on DC-QIM's Modulo-Lattice Reduction of Sect. 5.2.1, and the Decision-Aided Optimization described in Sect.5.3.1 are adapted and their particular modifications are developed below.

First, a search-interval for the absolute value of t_0 is obtained from the variance-based unbiased estimate of t_0^2. Specifically,

$$\hat{t}_0^2(\boldsymbol{z})_{\text{var}} = \frac{\frac{\|\boldsymbol{z}\|^2}{L} - \sigma_{N_2}^2}{\sigma_X^2 + \sigma_W^2 + \sigma_{N_1}^2}. \tag{6.4}$$

If L is large enough to use the Central Limit Theorem (CLT), then the distribution of $\hat{t}_0^2(\boldsymbol{z})_{\text{var}}$ can be approximated by $\mathcal{N}(t_0^2, 2(t_0^2(\sigma_X^2 + \sigma_W^2 + \sigma_{N_1}^2) + \sigma_{N_2}^2)^2/(L(\sigma_X^2 + \sigma_W^2 + \sigma_{N_1}^2)^2))$, and t_0^2 will be within $[t_-^2, t_+^2]$ with large probability, where

$$t_\pm^2 \triangleq \max\Big(\epsilon, \hat{t}_0^2(\boldsymbol{z})_{\text{var}} \pm K_2\sqrt{\frac{2(\hat{t}_0^2(\boldsymbol{z})_{\text{var}}(\sigma_X^2 + \sigma_W^2 + \sigma_{N_1}^2) + \sigma_{N_2})^2}{L\left(\sigma_X^2 + \sigma_W^2 + \sigma_{N_1}^2\right)^2}}\Big);$$

$\epsilon > 0$ guarantees that both t_-^2 and t_+^2 take positive values, and $K_2 \geq 0$ controls the probability with which $|t_0|$ lies in the interval $[t_-, t_+]$.

Once it is available, the search interval $[t_-, t_+]$ is sampled, producing a candidate set $\mathcal{T}^+$; this sampling must be fine enough to guarantee that if a sample is within the main lobe of the target function, then at least one of its neighbors in the sampling set will be also in the main lobe. Specifically, the sampling criterion is based on the factor in (6.3) defining the lobes, i.e., the cosine function argument. Indeed, we consider the variance of $(\boldsymbol{z} - t\boldsymbol{d}) \text{mod}(t\Delta/2)$ when t is in a neighborhood of t_0, and for $t = t_0$; the sampled points $t(l)$ are iteratively computed as $t(l+1) = \frac{t(l)\left(\alpha\frac{\Delta^2}{48} + \sigma_X^2 + \frac{\Delta}{2\sqrt{12}}\nu\right)}{\sigma_X^2 + \frac{\Delta^2(1-K_1)}{48}}$, where $\nu \triangleq \sqrt{\Delta^2/48\left((1-\alpha)^2 + K_1(2\alpha - 1)\right) + K_1\sigma_X^2}$, $t(1) = t_-$, and the iterative sampling stops when $t(l) \geq t_+$. Parameter K_1 is introduced to control the separation between consecutive points in $\mathcal{T}^+$ and, thus, the cardinality of such set; the larger K_1, the smaller $|\mathcal{T}^+|$ (less computational cost), but the more likely it will be that $\mathcal{T}^+$ misses the main lobe of the target function, with a consequent performance loss. Since t_0 can be negative, by symmetry we define $\mathcal{T} = \mathcal{T}^+ \cup -\mathcal{T}^+$.

The centroid used at embedding is estimated for each $t \in \mathcal{T}$; this is done by equalizing the received observation, i.e., $c_j = \mathcal{Q}_{\Delta/2}\left(z_j/t - d_j\right) + d_j$, $j = 1, \ldots, L$. Then, given $t \in \mathcal{T}$, the vector of centroids $\boldsymbol{c}$ is estimated, and from this choice the minimum mean square error gain factor, i.e., $t^* \triangleq \arg\min_t \|\boldsymbol{z} - t\boldsymbol{c}\|^2$, is computed, then $t^* = (\boldsymbol{z}^T\boldsymbol{c})/\|\boldsymbol{c}\|^2$. We will denote by $\mathcal{T}^*$ the set of local optimizers t^* thus obtained. Note that $|\mathcal{T}^*| \leq |\mathcal{T}|$. Since the sampling method guarantees that

at least one $t \in \mathcal{T}$ belongs to the main lobe, the ML estimate of t_0 is finally approximated by $\hat{t}_0(\boldsymbol{z}) \approx \arg\min_{t \in \mathcal{T}^*} L(t, \boldsymbol{z})$.

6.1.3 Adaptation to Filtered Images

An interesting application of the technique introduced in the previous section goes beyond a pure scaling and considers a watermarked image that is convolved with a linear filter. From the estimation result, the embedded bits must be reliably extracted. In this section we assume the embedding to be performed in the full-frame Discrete Cosine Transform (DCT)[1] domain, and the considered filters to be circularly symmetric; therefore, $\boldsymbol{x}$ will denote the coefficients in that domain of a gray level image $\boldsymbol{x}^S$ of size $N_r \times N_c$.

Typically, the energy of natural images is concentrated at the low frequencies, which are the most perceptually significant components. Therefore, an attacker could remove the high frequencies without a large semantic distortion; consequently, most robust watermarking schemes embed the messages at the low-middle frequencies, excluding the DC component (e.g., [5]).

After embedding, the full-frame Inverse DCT (IDCT) of $\boldsymbol{y}$ is calculated to obtain $\boldsymbol{y}^S$. The pixel values of the watermarked image are rounded to the nearest integer and clipped; this operation, which is modeled by the addition of $\boldsymbol{n}_1$ in (6.1), is denoted by $\mathrm{rclip}(\cdot)$

$$\mathrm{rclip}(y_i^s) = \begin{cases} \mathrm{round}(y_i^s) & \text{if} \quad y_i^s \in [0, 2^q - 1] \\ 0 & \text{if} \quad y_i^s < 0 \\ 2^q - 1 & \text{if} \quad y_i^s > 2^q - 1, \end{cases}$$

where $\mathrm{round}(\cdot)$ stands for the round to the nearest integer function, and q denotes the pixel depth. Then, the watermarked image is filtered (and subsequently rounded and clipped) in the spatial domain, yielding $\boldsymbol{z}^S = (\boldsymbol{y}^S + \boldsymbol{n}_1^S) * \boldsymbol{h}^S + \boldsymbol{n}_2^S$, where $*$ denotes the convolution operation (we consider $\boldsymbol{z}^S$ to have the size of $\boldsymbol{y}^S$ and $\boldsymbol{n}_1^S$), $\boldsymbol{h}^S$ is an $N_r^h \times N_c^h$-sized spatial filter, and $\boldsymbol{n}_2^S$ models the $\mathrm{rclip}(\cdot)$ operation after filtering.

Assuming $N_r \gg N_r^h$ and $N_c \gg N_c^h$, as customary, the filtering border effect is neglected in our analysis; the spatial domain filtering is approximated by a DCT domain frequency-dependent gain (although one must be aware that the filtering effect is not purely multiplicative). So, if one can estimate the gain factor corresponding to each frequency, then the SCS decoder in (6.2) may be used.

This gain estimate will be performed block-wise, relying on the assumption of the filter frequency response to be approximately constant within each block.

[1]The definition proposed in [33] is used.

Non-overlapped $N_B \times N_B$-sized blocks are used. If N_B were too large, then the frequency response could no longer be assumed constant within each block; on the other hand, if N_B were too small, then the estimate precision will be poor, due to the small number of samples.

We assume the AC full-frame DCT coefficients used for embedding to be i.i.d. zero-mean Gaussian distributed with known variance, and independent of the coefficients in other blocks. Furthermore, $\mathrm{rclip}(\cdot)$ is modeled in the spatial domain by both $\boldsymbol{N}_1^S$ (rounding and clipping due to the pixel domain transformation of the watermarked image, before filtering) and $\boldsymbol{N}_2^S$ (rounding and clipping due to the pixel domain casting of the filtered image) following independent $U([-1/2, 1/2]^L)$ distributions. If $N_r \cdot N_c$ is large enough, the CLT can be applied, and $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ can be approximated to be i.i.d. zero-mean Gaussian distributed with variance $1/12$.

6.1.4 Experimental Results

In this section we compare, by using synthetic signals, the performance of our proposed method with that of previous schemes in the literature; we also illustrate the application to filtered images. Throughout this section, the parameters of our method have been set to $K_1 = 10^{-3}$, $K_2 = 10$, and $\epsilon = 10^{-3}$. Here, for the sake of comparison it will be useful to define the effective WNR as $\mathrm{WNR}_e \triangleq t_0^2\sigma_W^2/(t_0^2\sigma_{N_1}^2 + \sigma_{N_2}^2)$.

First, assuming that $t_0 > 0$, we compare the performance in terms of the Bit Error Rate (BER), of the scheme described in Sect. 6.1.2 with that of Balado *et al.* [4]. The results are shown in Fig. 6.1, where the *turbo-code* used in [4] is employed, i.e., a 1/15 turbo code based on the *recursive systematic convolutional code* $\boldsymbol{g} = (31, 21, 25, 35, 23, 33, 27, 37)$ (octal coding) and interleaver size of 10^3 uncoded bits (yielding $L = 1.5 \cdot 10^4$) [3]. This coding is also considered for the results of the current approach shown in Fig. 6.1. It is worth noting that in order to reduce the complexity, our gain factor estimation algorithm does not explicitly exploit the code structure; in other words, for the results of the current method in Fig. 6.1 the code error correcting capabilities are employed solely for message decoding once the received signal is equalized by $\hat{t}_0(\boldsymbol{z})$. Hence, further improvements in the gain factor estimation would be afforded by exploiting the code underlying structure at the expense of a higher computational cost.

Fig. 6.1 shows that our scheme outperforms [4] for all the considered WNR_e's, except for $\mathrm{WNR}_e \approx 1.76$ dB, where no decoding errors were found for either.[2] This is not surprising as this WNR_e corresponds to $t_0 = 1$. Indeed, the large sensitivity of [4] to gain attacks even slightly different from 1 is shown by the authors

[2]Be aware that also no decoding errors were found for our method when $\mathrm{WNR}_e = 1$ and 3 dB.

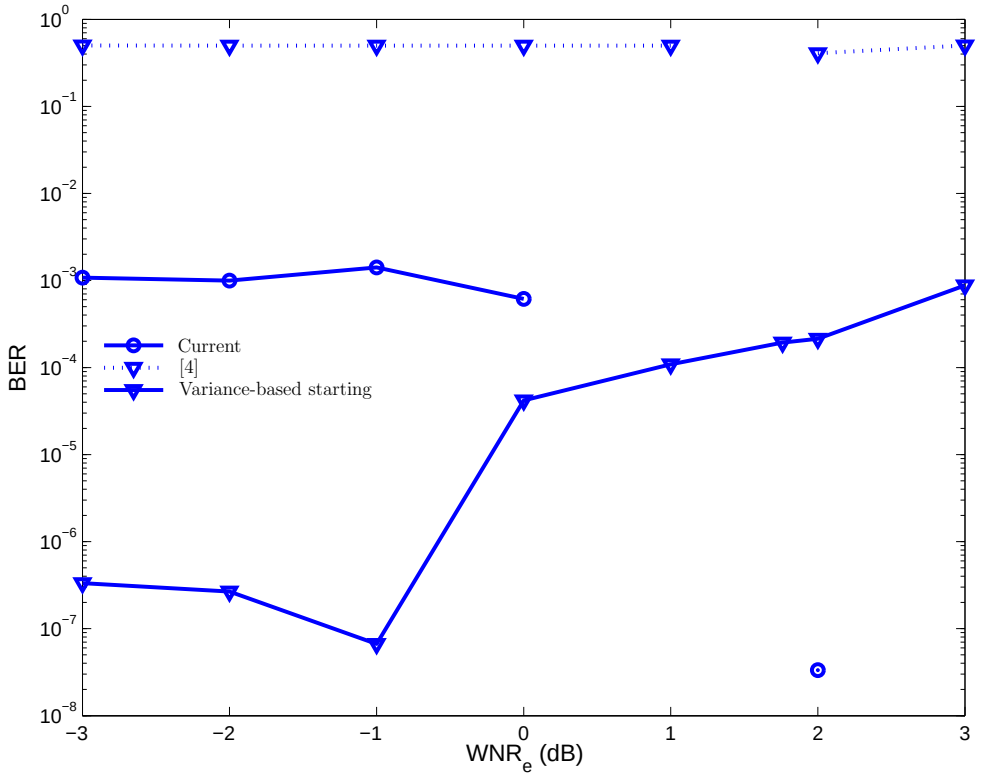


Figure 6.1: BER as a function of WNR_e for the method in [4], its variation when it is initialized by the variance-based estimate (6.4), and the current proposal. DWR = 30 dB, $L = 1.5 \cdot 10^4$, and $\alpha = \alpha_{\mathrm{Costa}}$.

in their original paper; for the sake of numerical illustration, in Fig. 6.1 the gains corresponding to $\mathrm{WNR}_e = 1$ and 2 dB are $t_0 \approx 0.850$ and $t_0 \approx 1.058$, respectively. Fig. 6.1 also shows the results obtained by initializing the scheme in [4] with the variance-based estimate introduced in (6.4); this initialization of Balado *et al.*'s method, newly proposed here, achieves the best results among all three methods for very small values of WNR_e (where the error in the variance-based estimate is very small), but it is clearly outperformed by the scheme described in Sect. 6.1.2 when larger WNR_e's are considered (corresponding to larger values of variance of the variance-based estimator).

Fig. 6.2 shows the BER as a function of WNR_e for [51], the variance-based estimate in (6.4), and our proposal when channel coding is not used, and $t_0 > 0$. [51] is carried out by sampling finely enough a search interval. Special attention was paid to reducing its computational cost as much as possible (e.g., precomputing the pdfs depending on a quantized version of the dither).

Since [51] uses the exact received signal pdf and exhaustive search, it was expected to provide the best results, as it is indeed the case. Furthermore, and similarly to Fig. 6.1, the variance-based estimate outperforms our proposal for

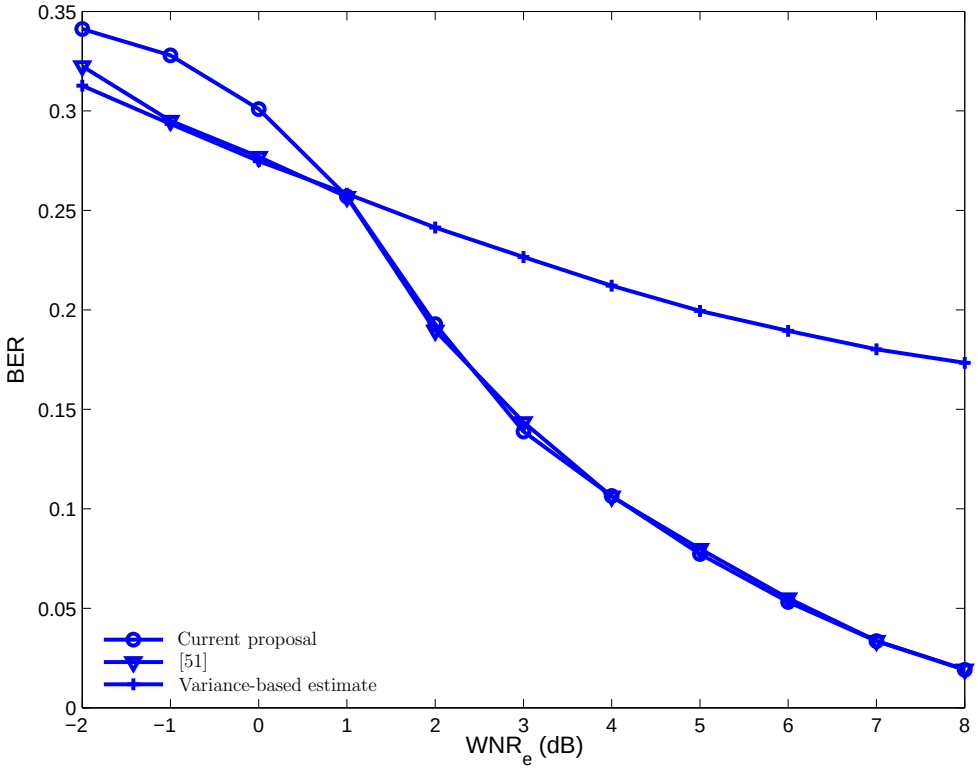


Figure 6.2: BER as a function of WNR_e for the method in [51], the variance-based estimate (6.4), and the current proposal. DWR = 30 dB, $L = 10^3$, and $\alpha = \alpha_{\mathrm{Costa}}$.

very low values of the WNR_e, as the structure on $f_{\boldsymbol{Y}}(\boldsymbol{y})$ induced by the watermark embedding is no longer observable; however, for larger WNR_e's such structure is made evident, and our scheme clearly improves the results of the variance-based estimate. It is also interesting to note that Shterev and Lagendijk's method behaves almost exactly as the best result among the variance-based estimate and our proposal, showing that both schemes are good choices (depending on the WNR_e) to be used as alternatives to the method proposed in [51], with a dramatic reduction in the computational cost over the latter. Specifically, each Monte Carlo trial of [51] for $\mathrm{WNR}_e = 6$ dB carried out in MatlabR2013b using a Core i5-2500 3.3GHZ 16 GB PC requires around 50 s, while our proposal approximately needs only 0.3 s.

Finally, Fig. 6.3 shows the results of the filtered-image-targeted adaptation proposed in Sect. 6.1.3 for a low-pass 5×5 spatial Gaussian filter with standard deviation 1, and a test set of 100 gray-converted 384×512-sized images pseudo-randomly selected from the UCID v2 image database [50]. For the reasons given in Sect. 6.1.3, only the first 10 zigzag-ordered DCT coefficient blocks of size 64×64 are used for hiding data.

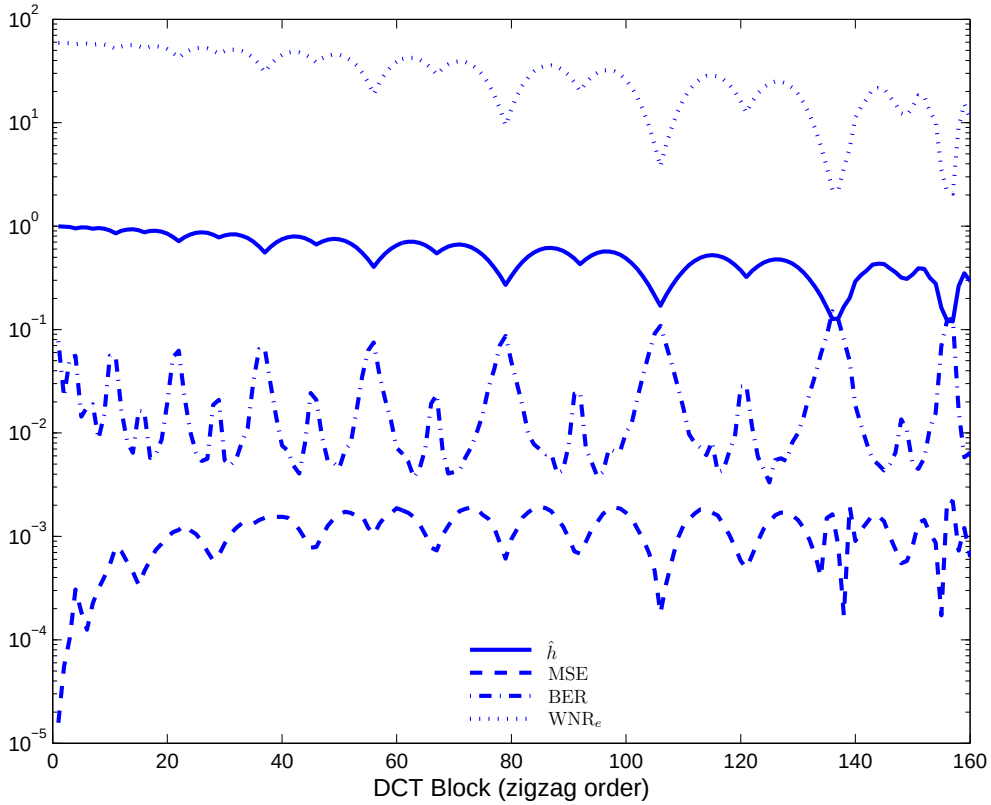


Figure 6.3: Estimate $\hat{h}$ of the block gain factor averaged over 100 images, the mean square estimation error of the gain factor, averaged BER per block, and WNR_e for each watermarked DCT block following the zigzag order. PSNR = 40 dB, $N_B = 16$, $\alpha = 1$, and Gaussian spatial filter of size 5×5 with standard deviation 1.

Fig. 6.3 shows the BER averaged over the test images for the considered blocks, when $N_B = 16$, $\alpha = 1$, and the PSNR, defined in this case as $255^2/\sigma_W^2$, is set to 40 dB. According to the shown results, the block BER approximately takes values between 10^{-1} and 10^{-2}, which illustrates that our scheme can be practically used in this demanding scenario. In addition, the BER seems to depend on the actual value of h (its estimate $\hat{h}$ is shown in this figure) as one would expect, since σ_W^2, $\sigma_{N_1}^2$, and $\sigma_{N_2}^2$ are approximately constant for all the watermarked blocks and, thus, the WNR_e only changes with h. These BER results are supported by the accuracy of the obtained estimates; specifically, in this example the mean square estimation error (MSE) of the gain factors takes values approximately around -30 dB in medium frequencies and less than -40 dB for low frequencies (where the energy of the images is concentrated).

6.2 Digital Communications: PAM Constellations

As an example of application of our techniques in digital communications, we propose to use DPCE on flat-fading channels in a digital communications framework. Since we mainly focus on the real gain case, the samples $\boldsymbol{x}$ are obtained from real constellations; indeed, we focus our attention in independent and uniformly distributed PAM constellation symbols. In this case, X is not Gaussian distributed and, in order to tackle this, the watermark embedding and the gain estimation are carried out in the ST domain. Thus, as explained in Sect. 3.4, by assuming that the value of L is large enough, the CLT guarantees that X^{ST} can be accurately modeled by a zero-mean Gaussian distribution and, therefore, DPCE techniques can be applied. Note that working in the ST domain allows us to choose the working point, i.e., we can reduce L_{ST} in order to reduce the TNQR (cf. Sect. 3.4); furthermore, this projection gain can not be achieved if one tries to estimate the channel scaling by looking at the PAM signal structure.

As in Chap. 2, the sent signal $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w};$$

however in this case, since $\boldsymbol{w} = V\boldsymbol{w}^{\mathrm{ST}}$ (V is the $L \times L_{\mathrm{ST}}$ orthonormal matrix used in the transformation), each component of $\boldsymbol{w}^{\mathrm{ST}}$ is zero-mean independently distributed, and if L_{ST} is assumed to be large enough, the distribution of W can be approximated by a zero-mean Gaussian distribution by applying the CLT. The decoder receives the scaled transmitted sequence $\boldsymbol{z}$ plus noise, i.e.,

$$\boldsymbol{z} = t_0\boldsymbol{y} + \boldsymbol{n} = t_0(\boldsymbol{x} + \boldsymbol{w}) + \boldsymbol{n}.$$

In this section, we consider that $t_0 \geq 0$ and $\boldsymbol{N}$ follows $\mathcal{N}(\boldsymbol{0}, \sigma_N^2 I_{L\times L})$. According to the embedding and due to the linear nature of ST, t_0 is calculated in the ST domain. It is worth pointing out that as DPCE is a ML-based technique, the length of the observed sequence in the ST domain L_{ST} must be large enough to obtain an accurate estimate of t_0.

After obtaining $\hat{t}_0(\boldsymbol{z})$, $\boldsymbol{z}$ is equalized as $\boldsymbol{z}/\hat{t}_0(\boldsymbol{z})$ then, the sent message is extracted. In this application, we use the BER to measure the performance. It is worth noting that, contrarily to SIT, the watermark is not removed at the decoder in our scheme; therefore, the watermark constitutes an additional source of noise with impact on the performance of the digital communications system. Let us mention that we are aware that there exist techniques to reduce the power of the watermark for scalar dirty paper coding techniques (e.g., [18] or [19]) but the obtained performance was almost the same compared to the case of not implementing those techniques and, therefore, we decided to dismiss them to avoid unnecessary complexity.

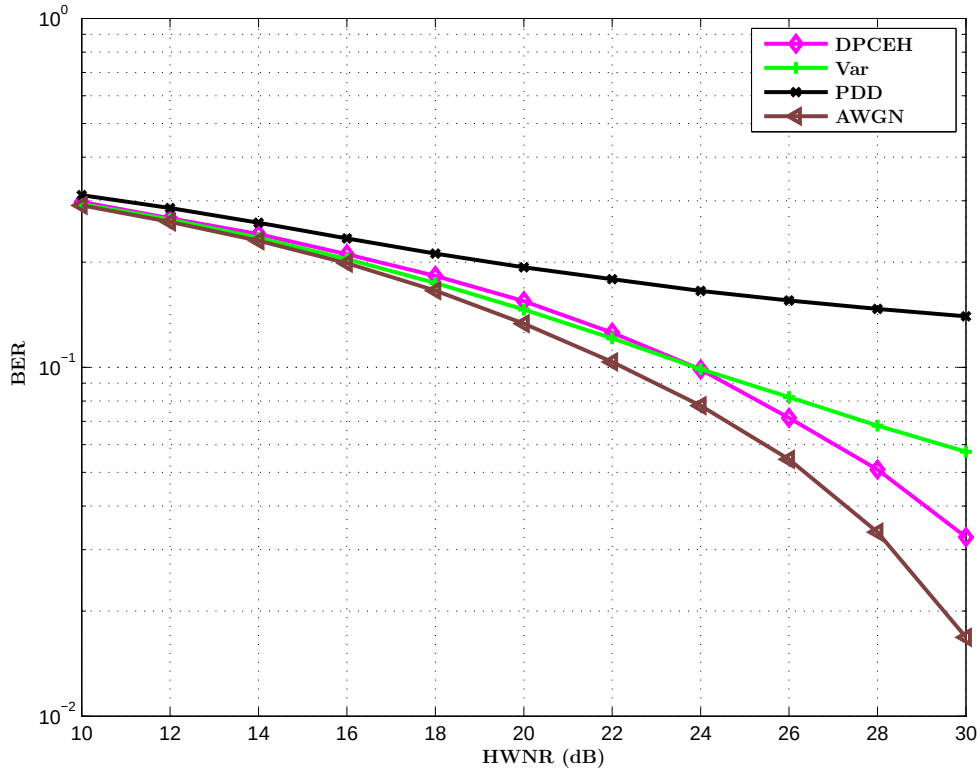


Figure 6.4: BER as function of HWNR for the optimal DPCEH, and for the optimal variance-based estimator (Var) are depicted. In addition, PDD, and for AWGN channel known t_0 are shown. $L = 10^2$, $L_{\text{ST}} = 10^2$, and $t_0 = 1$.

6.2.1 Experimental Results

In this section, we consider the transmission of independent and uniformly distributed 32-PAM symbols (which constitute the current host signal $\boldsymbol{x}$). The BER for both DPCE and PDD is minimized with respect to the splitting of the transmitter power budget (i.e., $\sigma_X^2 + \sigma_W^2$) into σ_X^2 and σ_W^2. Additionally, exhaustive search minimization is performed over α for DPCE, and over η for PDD (i.e., the self-interference parameter).

Once the estimation is performed for PDD, $\boldsymbol{z}/\hat{t}_0(\boldsymbol{z})$ is computed (as for DPCE), and the estimated watermark is removed from the signal; the result is the decoder input.

6.2.1.1 Known Variances

In Figs. 6.4-6.5, for the case of known the variances σ_X^2 and σ_N^2, the performance is measured in terms of BER vs. Host-plus-Watermark-to-Noise Ratio (HWNR) (defined as $(\sigma_X^2 + \sigma_W^2)/\sigma_N^2$) using the DPCEH techniques, the variance-based

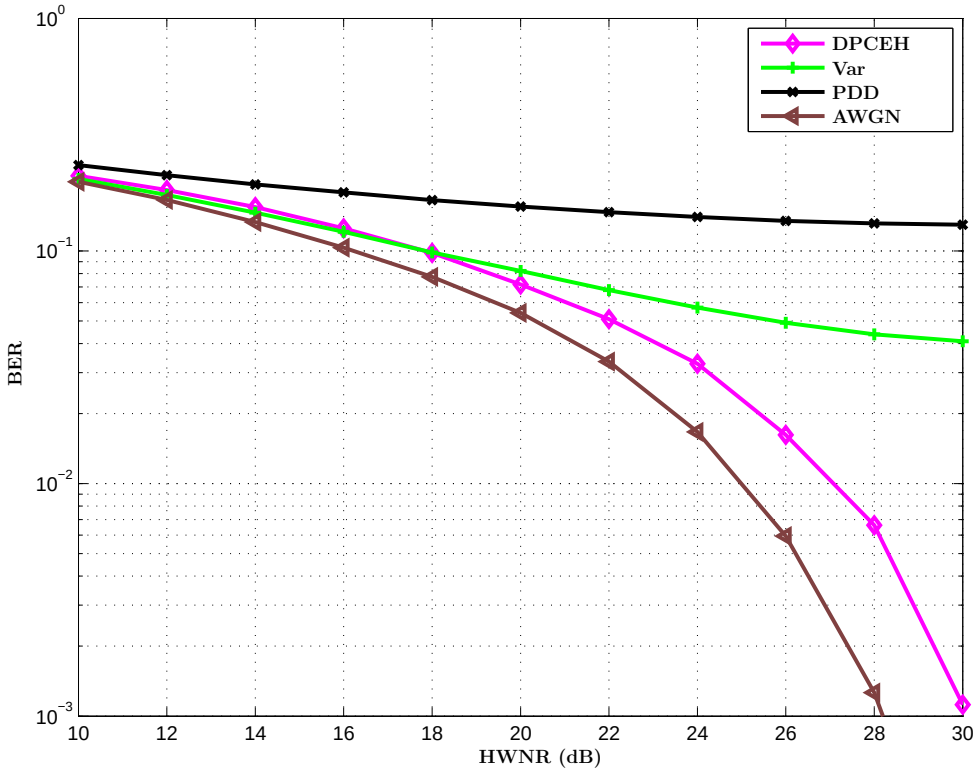


Figure 6.5: BER as function of HWNR for the optimal DPCEH, and for the optimal variance-based estimator (Var) are depicted. In addition, PDD, and for AWGN channel known t_0 are shown. $L = 10^2$, $L_{\mathrm{ST}} = 10^2$, and $t_0 = 2$.

estimator, and PDD. In these figures $L = L_{\mathrm{ST}} = 10^2$. In addition, the curve of known t_0 (in this case HWNR $= \sigma_X^2/\sigma_N^2$) is also shown as a reference. From these results, one can conclude that both the DPCEH and the variance-based estimator outperform PDD. In addition, approximately for values $\mathrm{HWNR_e} \leq 22$ dB (here, $\mathrm{HWNR_e} \triangleq t_0^2(\sigma_X^2 + \sigma_W^2)/\sigma_N^2$; therefore, this value approximately corresponds to HWNR = 22 dB in Fig. 6.4 and HWNR = 16 dB in Fig. 6.5), the variance-based estimator outperforms DPCEH, while for larger values of $\mathrm{HWNR_e}$, DPCEH shows better performance than the variance-based estimator. From this, one can conclude that the advantages of DPCE techniques (i.e., the host helps in the estimation of the gain) appear for large values of $\mathrm{HWNR_e}$.

6.2.1.2 Unknown Channel Noise Variance

Here we consider that σ_N^2 is unknown, the transmitter is power constrained, and the embedder and the decoder agree on the quantization step value Δ. In order to provide a realistic framework, these minimizations are performed by considering a fixed $\mathrm{HWNR_e}$ (we denote this scenario by Fixed $\mathrm{HWNR_e}$ Optimization, FHO), but the BER is evaluated for different values of $\mathrm{HWNR_e}$. Note that the

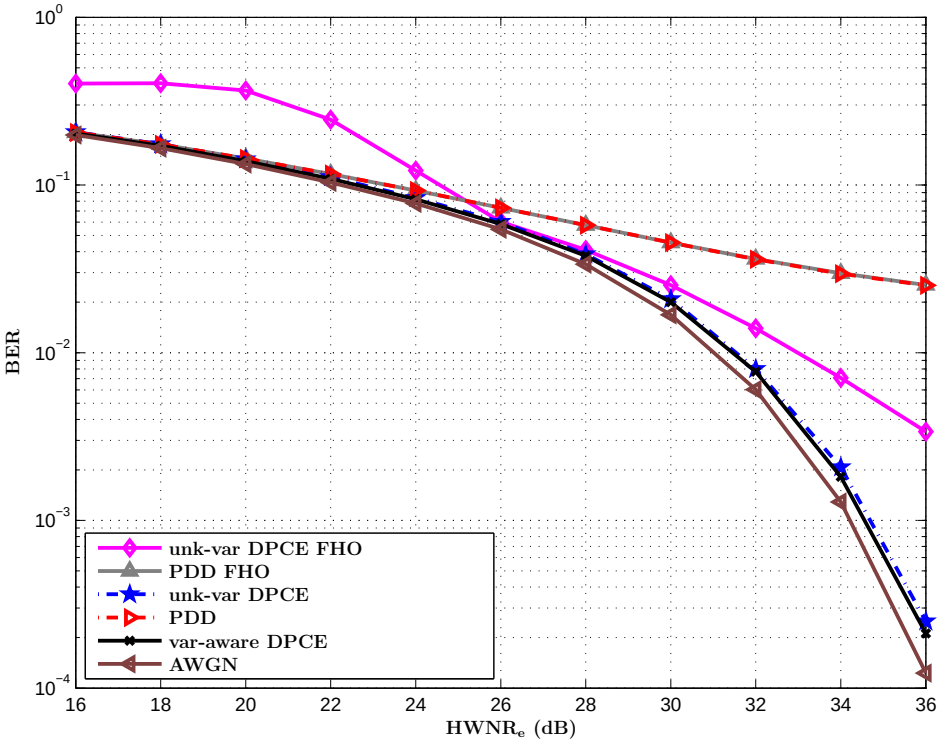


Figure 6.6: BER as a function of the HWNR$_e$ for the unknown-variance DPCE FHO, and PDD FHO when the optimization is performed for HWNR$_e$ = 26 dB. For the sake of comparison, we have also plotted the results for the variance-aware ST-DM-based DPCE (using the high-SNR target function), unknown-variance ST-DM-based DPCE, and PDD. The BER in these three plots was minimized for each HWNR$_e$ with respect to α (for DPCE plots), η (for PDD plot), and DWR. In addition, the curve of the AWGN case is also depicted. Independent uniformly distributed 32-PAM constellation, $L = 10^3$, and $t_0 = 0.9$.

system will be insensitive, in terms of BER, to changes on t_0 and σ_N^2 which verify $\frac{t_0^2}{\sigma_N^2}$ = constant, and consequently the BER is univocally defined as a function of HWNR$_e$. Note this can be also verified by comparing Fig. 6.4 and Fig. 6.5, where the latter corresponds to a shift of approximately 6 dB of the HWNR axis with respect to Fig. 6.4.

Fig. 6.6 shows the BER for $t_0 = 0.9$ as a function of the HWNR$_e$ for the unknown-variance ST-DM-based DPCE and PDD in FHO, where the optimization was performed for HWNR$_e$ = 26 dB, and using the constraint TNQR$(t_0) < 1$, $L_{\text{ST}} = \min(L, \max(1, \lfloor t_0^2 \text{WNR} L(1-(1-\alpha)^2)/\alpha^2 \rfloor))$. These results illustrate that variance-unknown DPCE FHO clearly outperforms PDD FHO for values of HWNR$_e$ larger than or equal to 26 dB; indeed, the gain of DPCE FHO with respect to PDD FHO increases with HWNR$_e$. As discussed above, this is a consequence of PDD spending part of the watermark power in reducing the host inter-

ference. Furthermore, PDD must achieve a good trade-off between σ_X^2 and σ_W^2: the larger σ_W^2, the more accurate the estimate, but also the lower σ_X^2, which is the power of the information transmitting signal, and consequently the BER might actually increase (i.e., t_0 estimation and data decoding are somehow competing). On the other hand, DPCE is not affected by host interference, but watermark removal is not implemented, and consequently the watermark interferes on information decoding; therefore, the larger the $\mathrm{HWNR_e}$, the larger the DWR, and consequently the more accurate the estimation of t_0 (as long as $\mathrm{TNQR}(t_0) < 1$), and the smaller the interference of $\boldsymbol{w}$ on $\boldsymbol{x}$, summing up both effects for reducing the BER.

For the sake of comparison, we also plot the results for the unknown-variance ST-DM-based DPCE, PDD, variance-aware ST-DM-based DPCE and the AWGN channel (where $\sigma_W^2 = 0$) when the optimization of the system parameters is performed for each particular $\mathrm{HWNR_e}$ (instead of being performed for a fixed $\mathrm{HWNR_e}$, as it is in FHO). Obviously, unknown-variance ST-DM-based DPCE FHO is outperformed by unknown-variance ST-DM-based DPCE, which in turn is outperformed by variance-aware ST-DM-based DPCE (although both plots are very similar), and the latter is outperformed by (but it is quite close to) the AWGN channel case. In contrast, the results for both PDD and PDD FHO are virtually the same, as both of them require complete host interference cancellation, yielding approximately the same DWR (around 15 dB).

6.3 Complex Gain Estimation

We now present how to adapt our technique to estimate real channel gains in order to address the scenarios of complex-valued signals and gains; specifically, we modify our basic framework (proposed in Chap. 2) to

$$\boldsymbol{z} = t_0(\boldsymbol{x} + \boldsymbol{w}) + \boldsymbol{n},$$

where in this case, $t_0 \in \mathbb{C}$, $\boldsymbol{X}$ and $\boldsymbol{N}$ are mutually independent random vectors and i.i.d. following a zero-mean circularly-symmetric complex Gaussian distribution.

Considering the two most widely used ways to multiply complex numbers, we propose also a pair of approaches: one based on using polar coordinates to multiply complex numbers in Sect. 6.3.1 and other focusing on the Cartesian coordinates, presented in Sect. 6.3.2

6.3.1 Polar Approach

Here, we exploit the nature of the complex product (multiplicative on the magnitude, additive on the phase) by considering a codebook defined in polar coordinates. By doing so, the estimation of t_0 will be decoupled into two simpler real estimation problems: first, an estimator $|\hat{t}_0(\boldsymbol{z})|$ of the magnitude is obtained, and this is then used to estimate the phase $\measuredangle\hat{t}_0(\boldsymbol{z})$. This decoupling will introduce some loss in performance, but, on the other hand, it will allow to significantly reduce the computational cost of the estimation.

6.3.1.1 Generation of the Transmitted Signal

The magnitude of x_i is modified as $|y_i| = |x_i| + \alpha\left(\mathcal{Q}_\rho\left(|x_i| - \varrho_i\right) - \left(|x_i| - \varrho_i\right)\right)$, where $i = 1, \ldots, L$, $\mathcal{Q}_\rho(\cdot)$ denotes a uniform scalar quantizer with step-size ρ, and $\boldsymbol{\varrho}$ stands for a dither sequence which is uniformly distributed in $[-\rho/2, \rho/2]^L$. It is worth noting that since the real and imaginary components of X follow independent zero-mean Gaussian distributions with variance σ_X^2, then $|X|$ will be Rayleigh distributed, with scale parameter σ_X.

In order to control the distortion introduced by the estimation aiding signal (i.e., the watermark signal), and at the same time provide a phase detection error probability similar to that of the magnitude detection, the quantization step applied to $\measuredangle x_i$, $i = 1, \ldots, L$, is chosen to yield an Euclidean distance between neighboring complex centroids sharing the same magnitude (i.e., those centroids

only distinguished by their phase), nearly equal to ρ.[3] Specifically, the quantization step used for quantizing the phase coordinate of the ith sample is calculated as

$$\phi_i = \begin{cases} 2\pi \left(\left\lceil \pi \left[\cos^{-1} \left(\frac{\sqrt{\left(\mathcal{Q}_\rho(|y_i|-\varrho_i)^2-(\rho/2)^2\right)}}{\mathcal{Q}_\rho(|y_i|-\varrho_i)} \right) \right]^{-1} \right\rceil \right)^{-1} \\ \qquad \text{if } \mathcal{Q}_\rho\left(|y_i|-\varrho_i\right) \neq 0, \\ 2\pi, \qquad \text{otherwise} \end{cases}, \tag{6.5}$$

where the arccosine function $\cos^{-1}(\cdot)$ takes values in $[-\pi, \pi)$, $\lceil \cdot \rceil$ stands for the ceil function, and we have used the relationships between Cartesian and polar coordinates.

Consequently, the modified phase will be obtained as $\measuredangle y_i = \measuredangle x_i + \alpha \left[\mathcal{Q}_{\phi_i} \left(\measuredangle x_i - \phi_i \varphi_i \right) - \left(\measuredangle x_i - \phi_i \varphi_i \right) \right]$, where $\boldsymbol{\varphi}$ is uniformly distributed in $[-1/2, 1/2]^L$. Note that the magnitude quantization step does not depend on i, but the phase quantization step does; in fact, the larger the magnitude of the ith sample, the smaller the used phase quantization step, which makes sense in order to achieve the target of controlling the estimation aiding signal power.

6.3.1.2 Magnitude Estimation

Since, in general, *a priori* information on $|t_0|$ is not available, as for the real case, the Maximum Likelihood (ML) estimator will be used. In order to obtain a mathematically tractable expression of $f_{|Z||T|,K}\left(|z_i||t|, \varrho_i\right)$, pdf of $|Z|$ given $|t|$ and ϱ_i, an approximation of that pdf is proposed based on three hypotheses (note that these are versions of the high-SNR hypotheses presented in Sect. 2.4 adjusted for estimating $|t_0|$): **1)** the variance of $|X|$ is much larger than the second moment of the quantization lattice (i.e., $\rho^2/12$). Therefore, the probability of the transmitted centroid given $|t_0|$ can be approximated by $\rho f_{|X|}(z/|t_0|)$ (this is the counterpart of HQR $\gg 1$), **2)** the variance of the scaled self-noise (i.e., $|t_0|^2(1-\alpha)^2\rho^2/12$) is much smaller than the variance of the channel noise σ_N^2 (version of $\mathrm{SCR}(t_0) \ll 1$). Therefore, the Gaussian channel noise dominates the total noise distribution, **3)** the square distance between scaled centroids (which we will quantify by using $|t_0|^2\rho^2/12$) is much larger than the variance of the total noise ($\sigma_N^2 + |t_0|^2(1-\alpha)^2\rho^2/12$) that corresponds to a modification of $\mathrm{TNQR}(t_0) \ll 1$. Therefore, the noise distribution is negligible outside of the quantization region of the transmitted centroid.

[3] In general that distance can not be exactly ρ, as the phase quantization step is required to be an integer divider of 2π, in order to verify the phase periodicity constraint.

By jointly considering these assumptions, one can approximate

$$f_{|Z||T|,K}\left(|z|||t|,\varrho\right) \approx \frac{|z|\rho e^{-\frac{|z|^2}{2\sigma_X^2|t|^2}}}{\sigma_X^2|t|} \frac{e^{-\frac{((|z|-\varrho|t|)\bmod\rho|t|)^2}{2\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)}}. \tag{6.6}$$

If the three hypotheses do not simultaneously hold, the validity of this pdf approximation and the accuracy of our estimator can no longer be guaranteed. Intuitively, the leftmost fraction of the previous expression approximates the probability of the centroid corresponding to $|z|$, while the rightmost fraction approximates the distribution of $|z|$ given that centroid. Note this split of the pdf is coherent with the corresponding real-valued high-SNR case (3.10) formulated in Sect. (3.1.2); indeed, in spite of using the magnitude in this expression, the difference of both pdfs is the distribution of the leftmost term, corresponding to a Gaussian distribution in the real-valued case and the Rayleigh distribution here. From (6.6), and given that the components of $\boldsymbol{z}$ are mutually independent, the ML estimation can be approximated as

$$\begin{aligned}|\hat{t}_0(\boldsymbol{z})| \approx \underset{|t|\geq 0}{\arg\min} \Bigg(& \frac{\|\boldsymbol{z}\|^2}{\sigma_X^2|t|^2} + \frac{\|(|\boldsymbol{z}|-\boldsymbol{\varrho}|t|)\bmod\rho|t|\|^2}{\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)} \\ & + L\log\left(|t|^2\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)\right)\Bigg).\end{aligned} \tag{6.7}$$

In order to limit the search-space of (6.7), a search-interval $[|t|_-, |t|_+]$ will be calculated by using an adaptation of the variance-based estimator (explained in Sect. 5.1.1) but for $|t_0|^2$, i.e., $|\hat{t}_0(\boldsymbol{z})|^2_{\text{var}} = \frac{\frac{\sum_{i=1}^{L}|z_i|^2}{L-1} - \left(\frac{\sum_{i=1}^{L}|z_i|}{L-1}\right)^2 - \sigma_N^2}{\sigma_{|X|}^2+\sigma_W^2}$, where $\sigma_{|X|}^2 = (4-\pi)\sigma_X^2/2$ and σ_W^2 denotes the variance of the magnitude of the estimation aiding signal (i.e., $\sigma_W^2 \approx \alpha^2\rho^2/12$). It can be shown that $|\hat{t}_0(\boldsymbol{z})|^2_{\text{var}}$ is an unbiased estimator of $|t_0|^2$; consequently, if L is large enough to apply the CLT, the distribution of $|\hat{t}_0(\boldsymbol{z})|^2_{\text{var}}$ can be approximated by a Gaussian distribution with mean $|t_0|^2$ and variance $2(|t_0|^2(\sigma_{|X|}^2+\sigma_W^2)+\sigma_N^2)^2/[(L-1)(\sigma_{|X|}^2+\sigma_W^2)^2]$. Therefore, if $|\hat{t}_0(\boldsymbol{z})|^2_{\text{var}} \approx |t_0|^2$, then $|t_0|^2$ will lie with approximated probability $\text{erf}(K_2/\sqrt{2})$ in the interval defined by $|t|^2_{\pm} = \max\left(\epsilon, |\hat{t}_0(\boldsymbol{z})|^2_{\text{var}} \pm K_2\sqrt{2\eta/(L-1)}\right)$, where $\epsilon > 0$ guarantees that $|t|^2_+$ and $|t|^2_-$ take positive values, and $\eta \triangleq \frac{(|\hat{t}_0(\boldsymbol{z})|^2_{\text{var}}(\sigma_{|X|}^2+\sigma_W^2)+\sigma_N^2)^2}{(\sigma_{|X|}^2+\sigma_W^2)^2}$. By applying the square root to those values, we obtain the interval we were looking for.

Here, based on the sampling technique based on the modulo-lattice reduction of the real valued case introduced in Sect. 5.2.1, we propose to sample the search interval $[|t|_-, |t|_+]$ finely enough to guarantee that two consecutive sampled points will be in the main lobe of the target function, which is indeed convex. The

sampling criterion is based on setting the total noise variance to be a multiple of the square quantization step-size, iteratively assuming that the considered magnitude value $|t(l)|= |t_0|$, so

$$|t(l+1)|= \frac{|t(l)|}{\mathrm{E}\{|X|^2\} + \frac{\rho^2(1-K_1)}{12}} \times \left(\alpha \frac{\rho^2}{12} + \mathrm{E}\{|X|^2\} + \frac{\rho}{\sqrt{12}} \sqrt{\frac{\rho^2}{12}\left((1-\alpha)^2 + K_1(2\alpha-1)\right) + K_1 \mathrm{E}\{|X|^2\}} \right),$$

where $\mathrm{E}\{|X|^2\} = \sigma^2_{|X|} + \sigma^2_X \pi/2$, $|t(1)|= |t|_-$, and the iterative sampling stops when $|t(l)|\geq |t|_+$. The parameter K_1 is introduced to control the separation between two consecutive elements of $\mathcal{T}$ and, thus, the cardinality of that set.

The Matlab optimization toolbox function `fminbnd` (which implements a bounded optimization algorithm based on golden section search and parabolic interpolation) is run once for each interval defined by two consecutive elements of $\mathcal{T}$; in this way a set $\mathcal{T}^*$ (of cardinality $|\mathcal{T}|-1$) with the corresponding optimization solutions, is built. Finally, the approximated ML estimate is that point in $\mathcal{T}^*$ which minimizes the target function in (6.7).

6.3.1.3 Phase Estimation

Assuming that $|\hat{t}_0(\boldsymbol{z})|$ obtained following the scheme described in the previous section is an accurate approximation of $|t_0|$, the normalized observation $|z_i|/(|\hat{t}_0(\boldsymbol{z})|)$, which is approximately equal to $|y_i|$, is used to estimate the phase quantizer step-size $\hat{\phi}_i$ as in (6.5).

Under the hypotheses introduced in the previous section, the distribution of Z given t_0 and the transmitted centroid, can be approximated by an i.i.d. Gaussian distribution centered at the transmitted centroid multiplied by t_0, and with variance equal to the sum of the noise channel variance and the self-noise variance scaled by $|t_0|^2$. Analogously to the magnitude estimation, the pdfs of neighboring phase centroids are approximately not overlapped. Therefore, the resulting ML estimator of $\measuredangle t_0$ can be approximated as

$$\measuredangle \hat{t}_0(\boldsymbol{z}) = \underset{t\in[-\pi,\pi)}{\arg\min} \sum_{i=1}^{L} \left| \mathcal{Q}_\rho\left(\frac{|z_i|}{|\hat{t}_0(\boldsymbol{z})|} - \varrho_i\right) - \left(\frac{|z_i|}{|\hat{t}_0(\boldsymbol{z})|} - \varrho_i\right) e^{j\left((\measuredangle z_i - \hat{\phi}_i \varphi_i - t) \bmod \hat{\phi}_i\right)} \right|^2,$$

where the modulo operation is used to measure the phase difference between the received samples and their closest centroids. For this algorithm, the previous optimization is carried out by exhaustive-search.

6.3.2 Cartesian Approach

We propose another approach to deal with a complex gain AWGN channel based on dividing the product into the real and the imaginary parts; as opposed to the polar approach introduced above, the nature of this approach does not provide an easy way to decouple the 2D estimation problem into two 1D estimation problems. However, this approach is more coherent with the proposed technique for the real gain proposed in Chap. 5 of this thesis, as will become apparent during its explanation.

6.3.2.1 Generation of the Transmitted Signal

The real and imaginary part of x_i is altered as

$$\begin{aligned}\mathrm{Re}(y_i) &= \mathrm{Re}(x_i) + \alpha\left(\mathcal{Q}_\Delta\left(\mathrm{Re}(x_i) - \mathrm{Re}(d_i)\right) - \left(\mathrm{Re}(x_i) - \mathrm{Re}(d_i)\right)\right)\\ \mathrm{Im}(y_i) &= \mathrm{Im}(x_i) + \alpha\left(\mathcal{Q}_\Delta\left(\mathrm{Im}(x_i) - \mathrm{Im}(d_i)\right) - \left(\mathrm{Im}(x_i) - \mathrm{Im}(d_i)\right)\right)\\ y_i &= \mathrm{Re}(y_i) + j\,\mathrm{Im}(y_i),\end{aligned}$$

where $i = 1, \ldots, L$, and $\mathrm{Re}(\cdot)$ stands for real part of the argument while $\mathrm{Im}(\cdot)$ denotes its imaginary part. In order to simplify the technique, the values taken by α and Δ are the same for for both the real and imaginary parts; obviously, a modification of the algorithm can be proposed with different values for the real and the imaginary parts but the control of the embedding distortion or the verification of the hypotheses would be more difficult. Regarding the dither sequence $\boldsymbol{d}$, both vectors $\mathrm{Re}(\boldsymbol{d})$ and $\mathrm{Im}(\boldsymbol{d})$ are mutually independent and uniformly distributed in $[-\Delta/2, \Delta/2]^L$.

6.3.2.2 Estimation of the Complex Gain

As in the previous complex gain estimation technique, the MLE is used. First, an approximation of the pdf of $\boldsymbol{Z}$ given t_0 and $\boldsymbol{d}$ is proposed based on three hypotheses (again, these hypotheses are modifications of high-SNR hypotheses introduced for the real case): **1)** the variance of the host is much larger than the second moment of the quantization lattice $2\Delta^2/12$ (i.e., an adaptation of HQR $\gg 1$), therefore the probability of a transmitted centroid can be approximated by $(\Delta f_X(z/|t_0|))^2$; **2)** the variance of the scaled self-noise (i.e., $2(1-\alpha)\Delta^2|t_0|^2/12$) is much smaller than the channel noise σ_N^2 (based on $\mathrm{SCR}(t_0) \ll 1$), thus, the Gaussian channel noise dominates the total noise distribution; and **3)** the variance of the total noise ($\sigma_N^2 + 2(1-\alpha)\Delta^2|t_0|^2/12$) is much smaller than square distance between the scaled centroids (corresponding to $\mathrm{TNQR}(t_0) \ll 1$), therefore, the noise distribution can be disregarded outside the quantization region.

The approximation of pdf of $\boldsymbol{Z}$ is written as

$$f_{\boldsymbol{Z}|T,K}(\boldsymbol{z}|t_0,\boldsymbol{d}) = \frac{1}{\|t_0\|^2}\left(\frac{\Delta^2 e^{-\frac{\|\boldsymbol{z}\|^2}{2\|t_0\|^2\sigma_X^2/2}}}{2\pi\sigma_X^2/2}\right)\cdot\frac{e^{-\frac{\|t_0\|^2\|(\boldsymbol{z}/t_0-\boldsymbol{d}) \mod \Delta\|^2}{2\left((1-\alpha)^2\Delta^2/12+\sigma_N^2/(2\|t_0\|^2)\right)}}}{2\pi\left((1-\alpha)^2\Delta^2/12+\sigma_N^2/(2\|t_0\|^2)\right)};$$

where in this case $A \mod B \triangleq (\mathrm{Re}(A)-\mathcal{Q}_B(\mathrm{Re}(A)))+j(\mathrm{Im}(A)-\mathcal{Q}_B(\mathrm{Im}(A)))$, for $A\in\mathbb{C}$ and $B\in\mathbb{R}$. As in the previous approach, the leftmost term approximates the centroid given z, while the rightmost term corresponds to the distribution of Z given a centroid. Using this approximation of the pdf of Z and since the components of $\boldsymbol{z}$ are mutually independent, the ML-based cost function to optimize can be approximated as

$$L(t,\boldsymbol{z}) \approx \left(\frac{\|t\|^2\|(\boldsymbol{z}/t-\boldsymbol{d}) \mod \Delta\|^2}{\sigma_N^2+2(1-\alpha)^2\frac{\|t\|^2\Delta^2}{12}} + 2L\log\left(2\pi\left(\sigma_N^2+2(1-\alpha)^2\|t\|^2\Delta^2/12\right)\right)+\frac{\|\boldsymbol{z}\|^2}{\sigma_X^2\|t\|^2}\right).$$

The search-interval $[|t|_-,|t|_+]$ is calculated as the intersection of the statistical interval $[|t|_-^{\mathrm{V}},|t|_+^{\mathrm{V}}]$ and the deterministic interval $[|t|_-^{\mathrm{D}},|t|_+^{\mathrm{D}}]$ in this case. The computation of $[|t|_-^{\mathrm{V}},|t|_+^{\mathrm{V}}]$ is carried out using the variance-based estimator, the estimate of the magnitude of the scaling factor t_0 is obtained as

$$|\hat{t}_0|_{\mathrm{var}}(\boldsymbol{z}) = \sqrt{\frac{\|\boldsymbol{z}\|^2/L-\sigma_N^2}{\sigma_X^2+2\alpha^2\Delta^2/12}};$$

the interval $[|t|_-^{\mathrm{V}},|t|_+^{\mathrm{V}}]$ is obtained using the approximation to the CRB of the variance-based estimator of t_0 from App. 4.A when $|t_0|^2\sigma_X^2 \gg \sigma_N^2$

$$|t|_\pm^{\mathrm{V}} = \max\left(\sqrt{\epsilon},|\hat{t}_0|_{\mathrm{var}}(\boldsymbol{z})\pm K_2\frac{|\hat{t}_0|_{\mathrm{var}}(\boldsymbol{z})}{\sqrt{2L}}\right).$$

Following the deterministic approach to generate the search-interval, the computation of a deterministic search-interval to estimate $|t_0|$ is obtained by lower bounding the cost function by

$$L_2(t,\boldsymbol{z}) = 2L\log\left(2\pi(\sigma_N^2+2(1-\alpha)^2\|t\|^2\Delta^2/12)\right)+\frac{\|\boldsymbol{z}\|^2}{\sigma_X^2\|t\|^2}.$$

As explained in Sect. 5.1.2, given an initial approximation of $|t_0|$ a dichotomy algorithm is carried out in order to obtain $[|t|_-^{\mathrm{D}},|t|_+^{\mathrm{D}}]$.

For the sampling, we propose to adapt the technique based on DC-QIM modulo-lattice reduction presented in Sect. 5.2.1. In this case, $[|t|_-,|t|_+]$ limits the search of values of $|t_0|$; therefore, a ring of possible values of t_0 in the

complex plane is defined by $|t_0| \in [|t|_-, |t|_+]$. This ring is sampled to obtain the candidate set $\mathcal{T}$ according to

$$\|t(i+1) - t(i)\|^2 \sigma_X^2 + \|t(i+1) - \alpha t(i)\|^2 \frac{2\Delta^2}{12} + \sigma_N^2 \leq \|t(i)\|^2 (1-\alpha)^2 \frac{2\Delta^2}{12} + \sigma_N^2 + K_1 \|t(i+1)\|^2 \frac{2\Delta^2}{12}; \tag{6.8}$$

which is a modification of the sampling technique proposed in Sect. 5.2.1. In the previous expression.

In order to reduce the computational complexity of this technique, the sampling of the ring is carried out following a simplified version of the sampling. This technique focuses on the first quadrant of the complex plane $\mathrm{Re}(t_0) \geq 0$ and $\mathrm{Im}(t_0) \geq 0$, i.e., it is considered that the phase of t_0 is in the interval $[0, \pi/2]$. First, the imaginary part of the candidate values of t_0 is set to zero, $[|t|_-/\sqrt{2}, |t|_+]$ is sampled as in the real scaling factor case: by iteratively solving the following inequality with $\mathrm{Re}(t(1)) = |t|_-/\sqrt{2}$

$$\begin{aligned}&(\mathrm{Re}(t(i+1)) - \mathrm{Re}(t(i)))^2 \sigma_X^2 + (\mathrm{Re}(t(i+1)) - \alpha\,\mathrm{Re}(t(i)))^2 \frac{2\Delta^2}{12} + \sigma_N^2 \\ &\leq (\mathrm{Re}(t(i)))^2 (1-\alpha)^2 \frac{2\Delta^2}{12} + \sigma_N^2 + K_1 (\mathrm{Re}(t(i+1))^2 \frac{2\Delta^2}{12},\end{aligned}$$

and stopping when $\mathrm{Re}(t(l)) \geq |t|_+$. In this way, we obtain the set of the real parts of the candidate set $\mathcal{T}^{\mathrm{Real}}$ corresponding the first octant with the phase in $[0, \pi/4]$. For each element $\mathrm{Re}(t(i))$ of $\mathcal{T}^{\mathrm{Real}}$, the imaginary interval $[0, |t|_+/\sqrt{2}]$ is sampled fixing $\mathrm{Re}(t(i,j)) = \mathrm{Re}(t(i,j+1)) = \mathrm{Re}(t(i))$ and iteratively obtaining the imaginary part $\mathrm{Im}(t(i,j+1))$ as

$$\begin{aligned}&-\frac{1}{(-1+K_1)\left(\frac{\sqrt{2}\Delta}{\sqrt{12}}\right)^2 - \sigma_X^2} \left[\alpha \left(\frac{\sqrt{2}\Delta}{\sqrt{12}}\right)^2 \mathrm{Im}(t(i,j)) \right. \\ &+\sigma_X^2 \,\mathrm{Im}(t(i,j)) + \left(\frac{\sqrt{2}\Delta}{\sqrt{12}}\right) \left[K_1 \sigma_X^2 \left(\mathrm{Im}(t(i,j))^2 + \mathrm{Re}(t(i))^2\right) + \left(\frac{\sqrt{2}\Delta}{\sqrt{12}}\right)^2 \right. \\ &\left.\left. \times \left(\left((-1+\alpha)^2 + (-1+2\alpha)K_1\right) \mathrm{Im}(t(i,j))^2 - (-1+K_1)K_1 \,\mathrm{Re}(t(i))^2\right) \right]^{1/2} \right],\end{aligned}$$

where this expression comes from the solution of (6.8) for this case and $\mathrm{Im}(t(i,1)) = 0$ till $\mathrm{Im}(t(i,j)) \geq |t|_+/\sqrt{2}]$. An example of the resulting sampling points is shown in Fig. 6.7 for DWR = 30 dB, WNR = 0 dB, $\alpha = 1$, $L = 400$, and $K_1 = 1$. From this initial set, the candidate points are those verifying that their magnitude is within interval $[|t|_-, |t|_+]$, i.e., they are within the search-interval for the magnitude of t_0. An example of the candidate points (represented as

squares) corresponding to the experiment is depicted in Fig. 6.7.. The candidate set of the first quadrant $\mathcal{T}^{++}$ is calculated by covering the remaining region of the ring of the first quadrant by swapping the real and the imaginary parts of the obtained points for $[0, \pi/4]$. An analogous process is carried out to sam-

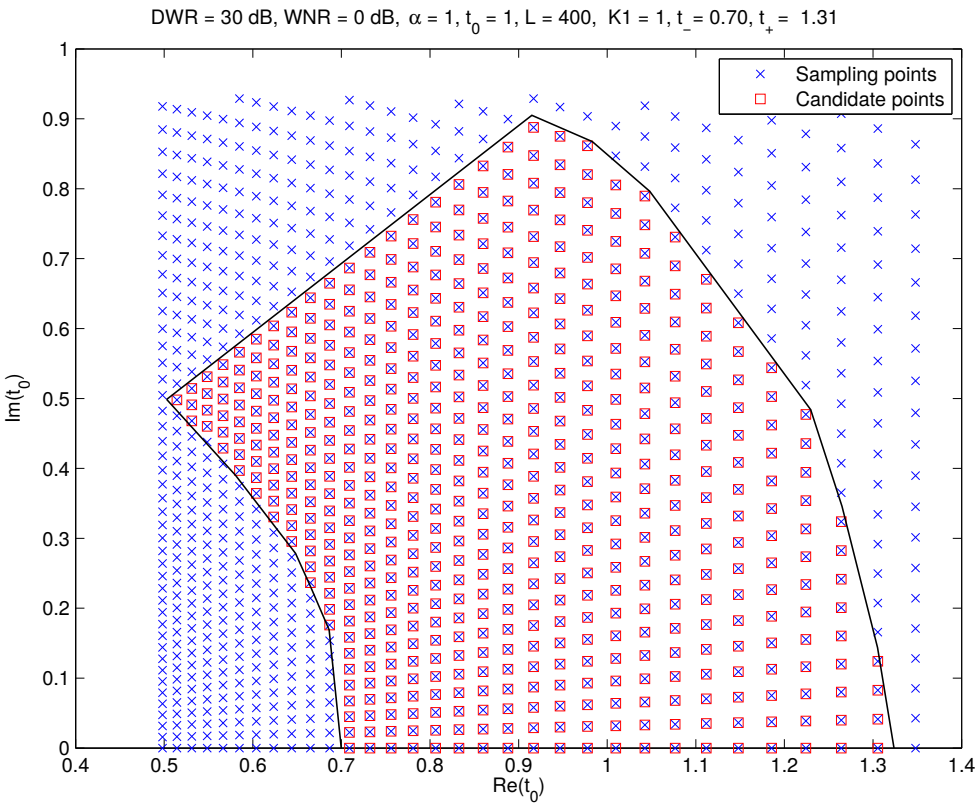


Figure 6.7: Example of the obtained sampling points (blue crosses) and the candidate points (red squares lying inside the polygon) that are in the ring $[|t|_-, |t|_+]$ for the first octant of the complex plane with phase in $[0, \pi/4]$. DWR = 30 dB, WNR = 0 dB, $\alpha = 1$, $L = 400$, and $K_1 = 1$.

ple the whole complex plain, i.e., given $\mathcal{T}^{++}$, $\mathcal{T}^{-+} = -\operatorname{Re}(\mathcal{T}^{++}) + j \operatorname{Im}(\mathcal{T}^{++})$, $\mathcal{T}^{+-} = \operatorname{Re}(\mathcal{T}^{++}) - j \operatorname{Im}(\mathcal{T}^{++})$, and $\mathcal{T}^{--} = -\operatorname{Re}(\mathcal{T}^{++}) - j \operatorname{Im}(\mathcal{T}^{++})$. Then, the candidate point set is obtained as the union of the candidate point sets of each quadrant, i.e.,

$$\mathcal{T} = \mathcal{T}^{++} \cup \mathcal{T}^{+-} \cup \mathcal{T}^{-+} \cup \mathcal{T}^{--}.$$

The estimate of t_0 given $\mathcal{T}$ is obtained using a complex version of the Decision-Aided technique presented in Sect. 5.3.1 as

$$\begin{aligned} \operatorname{Re}(c_j) &= \mathcal{Q}_\Delta\left(\operatorname{Re}(z_j/t(i)) - \operatorname{Re}(d_j)\right) + \operatorname{Re}(d_j) \\ \operatorname{Im}(c_j) &= \mathcal{Q}_\Delta\left(\operatorname{Im}(z_j/t(i)) - \operatorname{Im}(d_j)\right) + \operatorname{Im}(d_j), \end{aligned}$$

with $j = 1, \ldots, L$, $t^*(i) = \boldsymbol{c}^* \boldsymbol{z} / \|\boldsymbol{c}\|^2$ with $i = 1, \ldots, |\mathcal{T}|$ in order to obtain $\mathcal{T}^* = t^*(1), \ldots, t^*(L)$. Finally, the estimate is obtained as

$$\hat{t}_0(\boldsymbol{z}) = \arg \min_{t \in \mathcal{T}^*} L(t, \boldsymbol{z}).$$

6.3.3 Experimental Results

In this section, we compare the MSE of the DPC-based estimators proposed for complex scenarios, with that of PDD [29]. Figs. 6.8-6.11 show the MSE as a function of $|t_0| \in [0.1, 2] \cap 0.1\mathbb{Z}$, where the results for each of those points were obtained by using 10^3 Monte Carlo runs; for each run, $\measuredangle t_0$ was independently generated according to $U(-\pi, \pi)$. The DPC-based schemes $K_2 = 10$, $\epsilon = 10^{-3}$. For the polar approach $K_1 = 1$ and the exhaustive search performed in the estimate of $\measuredangle t_0$ considers $2 \cdot 10^4$ points uniformly located through $[-\pi, \pi)$. For the Cartesian approach $K_1 = 10^{-2}$.

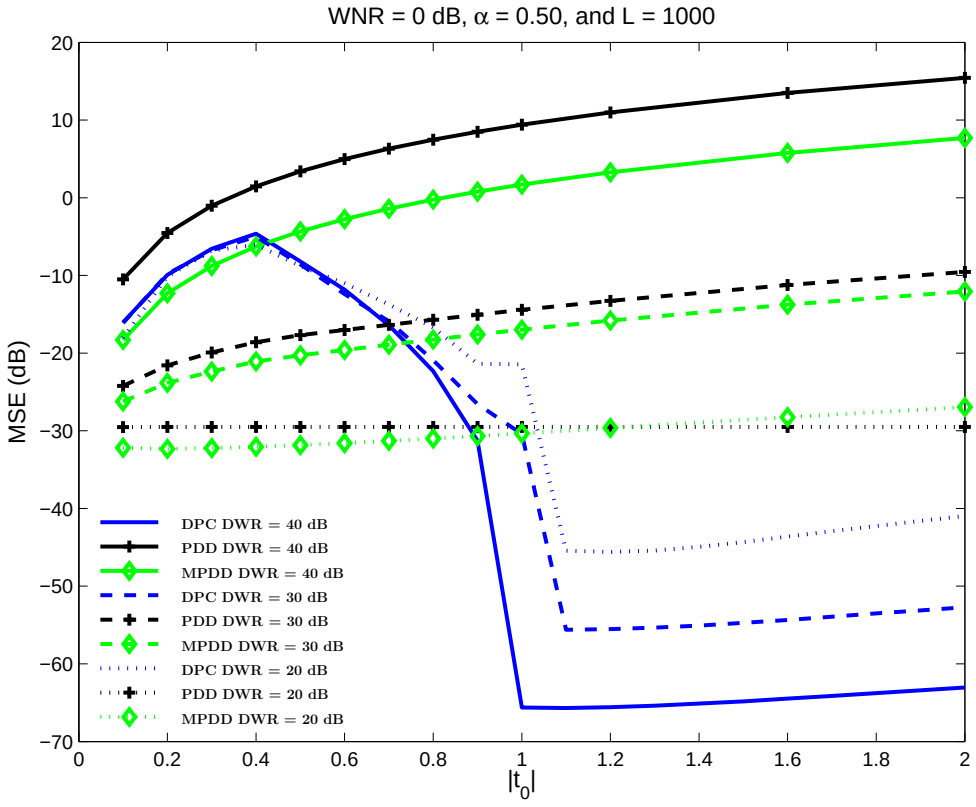


Figure 6.8: MSE vs. $|t_0|$ for the polar approach (DPC), PDD and Modification of PDD (MPDD). DWR $= 20, 30, 40$ dB, WNR $= 0$ dB, $\alpha = 0.5$, and $L = 10^3$.

Concerning the comparison with PDD, we consider the case where such scheme also deals with time invariant flat channels, even if it can be used in more general

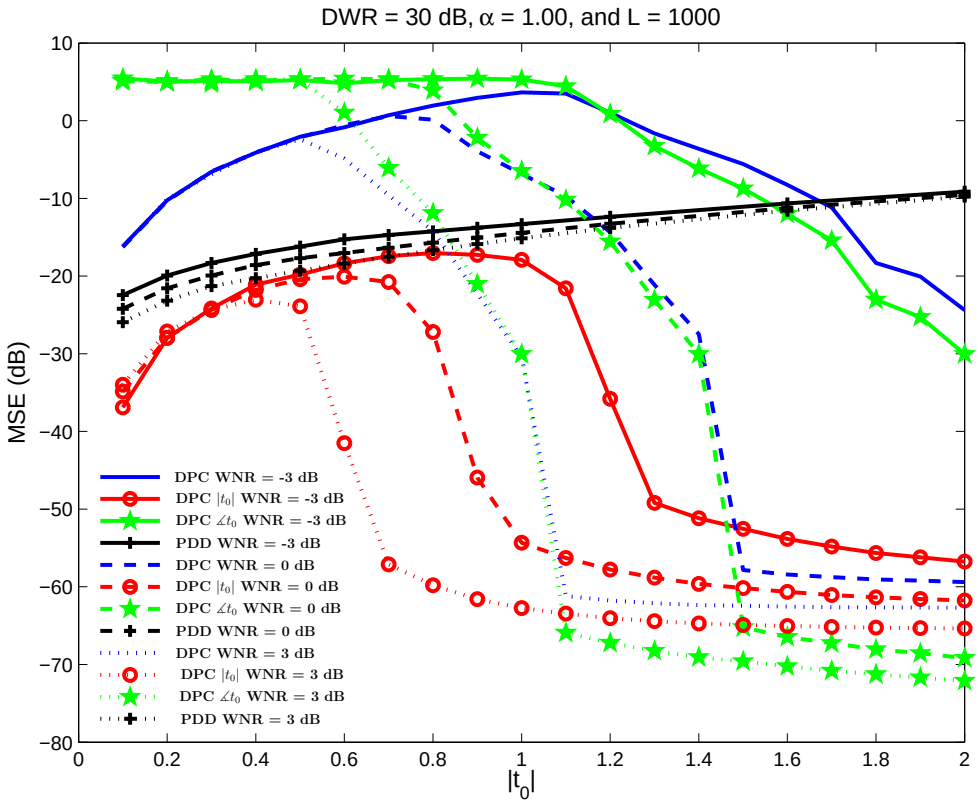


Figure 6.9: MSE vs. $|t_0|$ for our the polar approach (DPC), and PDD. WNR $= -3, 0, 3$ dB, DWR $= 30$ dB, $\alpha = 1$, and $L = 10^3$. MSEs of $|t_0|$ and $\measuredangle t_0$ are also provided.

frameworks; furthermore, the host-interference controlling parameter proposed in [29] is optimized in order to provide the best performance for that scheme. Note that the power of the distortion introduced on the host signal by PDD comprises both the power of the estimation aiding signal, and the power due to the reduction of the host interference. The channel estimator proposed in [29], once it is adapted to the complex flat fading case, is $\hat{t}_0(\boldsymbol{z}) = \boldsymbol{w}^*\boldsymbol{z}/\|\boldsymbol{w}\|^2$, i.e., it only uses the component of $\boldsymbol{z}$ in the direction of $\boldsymbol{w}$; consequently, the remaining $L-1$ components of $\boldsymbol{z}$ are disregarded. Since those $L-1$ components follow a $\mathcal{N}(0, |t_0|^2\sigma_X^2 + \sigma_N^2)$ distribution, they are indeed informative about $|t_0|$, and that dependence could be exploited. Therefore, we propose a suboptimal MPDD, where the $L-1$ components of $\boldsymbol{z}$ orthogonal to $\boldsymbol{w}$ are fed to a variance-based estimator, and the estimate of $\measuredangle t_0$ is $\measuredangle \boldsymbol{w}^*\boldsymbol{z}$.

Fig. 6.8 compares the MSE of the polar proposed scheme with that of PDD and MPDD as a function of $|t_0|$, for different values of DWR. One can observe that the larger the DWR (i.e., the larger the margin by which Hypothesis 1 in Sect. 6.3.1.2 is satisfied), the better the performance of the proposed scheme. Furthermore, in the proposed scheme a larger DWR helps to estimate t_0 (at the cost of increasing

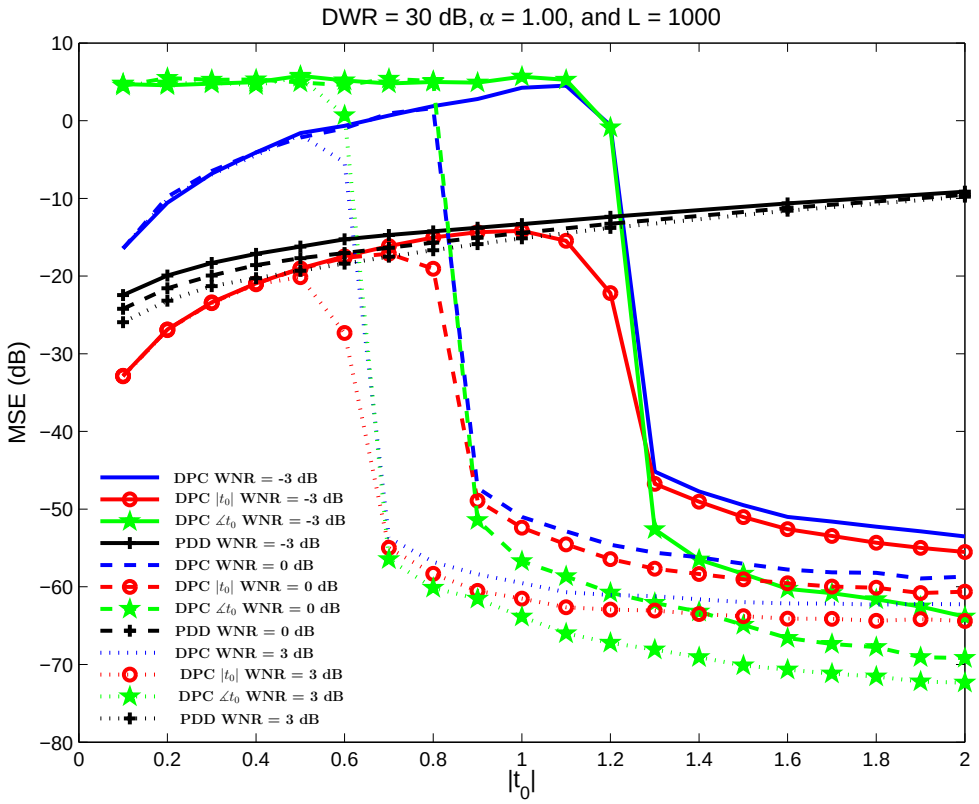


Figure 6.10: As Fig. 6.9 but for the Cartesian approach.

the estimation computational cost), contrarily to what happens with PDD and MPDD; indeed, in order for the DPC-based scheme to provide better results than PDD and MPDD, $|t_0|$ must take values larger than a DWR-dependent threshold; the larger the DWR, the smaller the $|t_0|$ value for the crossing point. It is worth mentioning that our method generally requires more computational resources than PDD or MPDD. Related to the comparison between PDD and MPDD, the larger the MPDD, the better MPDD is with respect to PDD; in that case PDD will not be able to cancel out the host interference on $\boldsymbol{w}$, and, as it was mentioned before, the estimator proposed in [29] does not take advantage either of the $L-1$ components of $\boldsymbol{z}$ orthogonal to $\boldsymbol{w}$ (as our proposed modification MPDD does).

Fig. 6.9 illustrates for the polar approach the contribution of $|t_0|$ and $\measuredangle t_0$ to the MSE of the estimate of t_0 for different values of WNR; again, the results for PDD are also plotted. Similarly to the discussion about Fig. 6.8, in this case we can check the effect of the margin by which Hypothesis 3 in Sect. 6.3.1.2 (i.e., $|t_0|^2\rho^2/12 \gg \sigma_N^2 + |t_0|^2(1-\alpha)^2\rho^2/12$) is satisfied on the performance of the estimator. Mainly, the larger the WNR, the better the provided approximation; of course, one must also take into account that a larger WNR will make easier the estimate, independently of the accuracy in the approximation of the pdf. Additionally, it must be noted that the DPCE MSE curves share a similar behavior

with respect to $|t_0|$: for small values of $|t_0|$, the value of the MSE increases with it; then, when the three hypotheses hold, it decreases with $|t_0|$. Furthermore, we can see that the main source of MSE seems to be the phase estimate; this is partially due to the fact that this estimator inherits the errors made by the magnitude estimator. Fig. 6.10 is the counterpart of Fig. 6.9 using the Cartesian approach. By comparing both graphs, one can easily conclude that the Cartesian approach shows better performance than the polar one. For example, the abruptly drop of the MSE appears around $|t_0|= 0.7$ for the Cartesian case while for $|t_0|\approx 1.1$ for the other approach. By analyzing Fig. 6.10, one can realize that the phase does not show the dominant effect as in the polar case does; indeed, the drop appears almost simultaneously in both cases. Furthermore, it is worth noting that the performance of the curves for high-SNR cases converges with the CRB for the real Gaussian case (i.e., the inverse of (4.11)) detailed in Sect. 4.1.2.

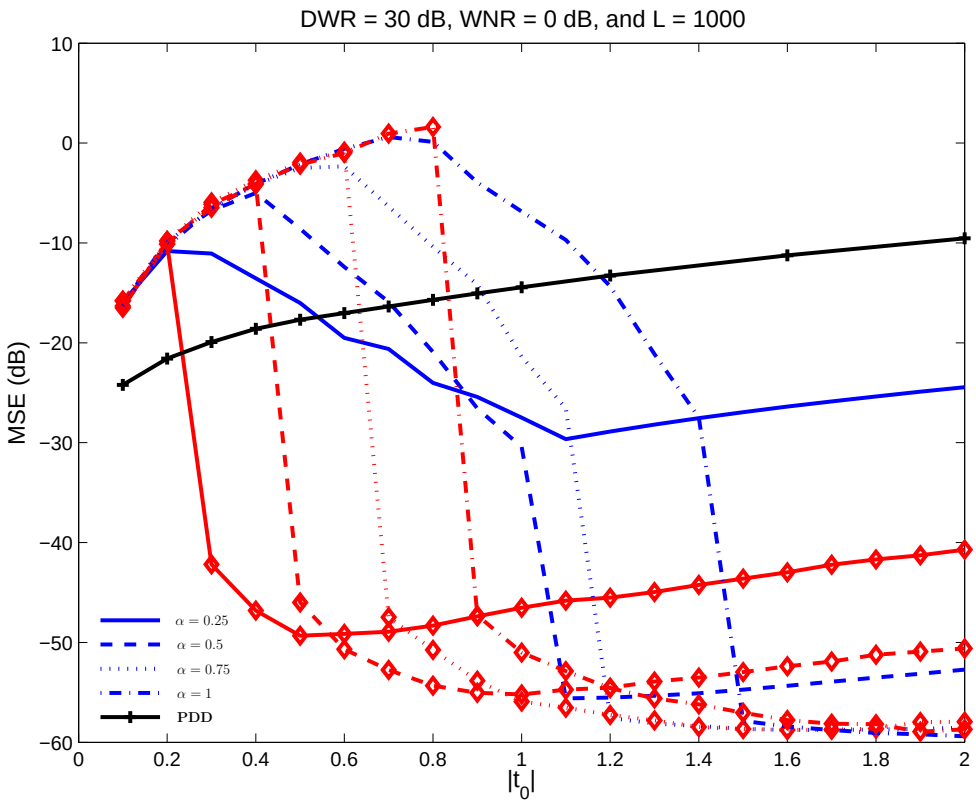


Figure 6.11: MSE vs. $|t_0|$ for our algorithm (DPC) (polar approach is represented without symbols while the Cartesian approach with $\diamond$), and PDD. $\alpha = 0.25, 0.5, 0.75, 1$, DWR $= 30$ dB, WNR $= 0$ dB, and $L = 10^3$.

Finally, Fig. 6.11 illustrates the behavior of the MSE as a function of the distortion compensation parameter α. According to these results, the performance of our schemes shows a trade-off between the value of $|t_0|$ at the crossing point with PDD, and the value of MSE when $|t_0|$ is increased. For example, for $\alpha = 0.5$ the Cartesian approach outperforms PDD for $|t_0|\geq 0.5$, and MSE ≈ -55 dB

for large values of $|t_0|$; on the other hand, for $\alpha = 0.75$ the crossing point is at $|t_0| \approx 0.65$, but MSE ≈ -59 dB for large values of $|t_0|$. It is straightforward to verify again in this figure that the drop of performance for the Cartesian approach appears significantly for smaller values of $|t_0|$ than for the polar case.

Chapter 7

Conclusions and Further Work

In this thesis, we have addressed the problem of gain estimation in a flat fading channel with Additive White Gaussian Noise (AWGN) using Dirty Paper Coding (DPC) estimation techniques, following the idea that, if the host interference rejection can be achieved in communications for digital watermarking, then this can be also obtained for estimation purposes.

We have proposed to use Maximum-Likelihood (ML) estimation. However, in order to deal with the real cost functions that are difficult to handle mathematically, we introduce several more tractable pdf approximations and, by using them, the corresponding cost functions of ML. In addition, we have also provided a modification of the technique whenever the variances of the original signal and the channel noise are unknown. Using an already established concept in digital watermarking, we have studied how to make full use of the Spread-Transform (ST) to estimate the channel gain.

In order to gain insight into the use of DPC for estimation, we have studied this scheme theoretically. From the perspective of estimation theory, we have developed some approximations of the Cramér-Rao Bound to determine the fundamental limits of the achievable accuracy; also, from an information theoretical perspective, we have studied our technique by means of mutual information in order to measure how much information the received sequence contains regarding the channel gain.

In order to propose practical estimation techniques requiring affordable computational complexity, we have proposed a set of techniques based on ML that makes full use of the statistical and deterministic analysis of the problem. Several of these techniques have been evaluated through experiments to verify and illustrate their effectiveness. In addition, our proposed schemes have been applied in different situations: robust digital watermarking to gain attacks, digital communications, and complex gain estimation.

Some conclusions drawn from the work carried out and presented in this thesis are that:

- The theoretical analysis indicates, and the experimental performance confirms, that DPCE is not only unaffected by the host, but in fact, the latter helps the estimation. Indeed, the results obtained asymptotically indicate that the resulting estimation accuracy is the same as if the available power were entirely devoted to sending a pilot signal.

- The structure of the pdf of the random variable modeling the received sequence has to become apparent to achieve host interference cancellation in the estimation. This means that the channel noise should not smear the structure created by the dirty paper code. Otherwise, the obtained results would be those achieved by the variance-matching techniques.

- Similar results can be attained whenever the variance of the host signal and the channel noise are not known, i.e., only making use of the induced structure on the pdf. This is in contrast to variance-based methods where the variance of the host and channel noise have to be known.

- Given the values of the WNR and the DWR, by using Spread-Transform in DPC estimation, one can control the effective WNR, and extending the range of WNRs for which the use of DPCE is feasible. This occurs at the expense of reducing the effective size of the vector of observations and, therefore, increasing the estimator variance.

- The proposed practical algorithms for DPC estimation require far less complexity than other brute-force estimation techniques. Indeed, our algorithm can be used in applications with strict time constraints.

- Our technique can be used in real digital watermarking applications to make the Scalar Costa Scheme become robust to scaling and filtering.

- The use of our estimation algorithms in real digital communication has been described in detail in this thesis. The performance of our techniques shows better results than the superimposed training techniques used in several real applications whenever the induced structure in the pdf can be used to estimate.

- Two approaches based on the proposed idea are introduced to deal with complex gains in AWGN channels. Both of them show similar asymptotic performance in terms of mean square error as the techniques proposed for the real gain case addressed in this thesis.

7.1 Future Lines of Research

The research carried out in this thesis leaves open several problems that we consider to be worthwhile addressing in the future:

- Our technique to estimate scalar gains in AWGN channels can be used to estimate more complex communication channels by working with subbands which, if they are sufficiently narrow, can roughly behave as flat channels.
- Our analysis was carried out by assuming zero-mean Gaussian distributed signals, so we propose to make our algorithm and its analysis independent of the distribution of the involved signals.
- In digital communications, the presence of the watermark acts as interference causing a reduction in performance; therefore, there is room for performance improvement if the watermark can be effectively removed at the decoder.
- Focusing on digital watermarking, we would like to study other relevant attacks (e.g., rotation, translation, quantization, etc.) and how DPC estimation can help to deal with them.
- We would like to extend our research to other uses including audio applications (e.g. room acoustic response estimation, active noise control, etc.), for digital forensic applications (e.g., filter estimation), for digital communications (e.g., Burst detector AGC, AGC in Satellite Communications Channel, SNR estimation), or for physical layer authentication.

Appendix A

Resumo

A.1 Introdución

A estimación de canle é un problema transversal en procesamento do sinal. Úsase en numerosas aplicacións, incluíndo comunicacións dixitais (p.ex., na estimación dos parámetros da canle, no control automático de ganancia, para a estimación da relación sinal a ruído, etc.), restauración de imaxes (p.ex., na deconvolución en imaxes), en forensia dixital (p.ex., para a estimación do filtro lineal usado no post-procesamento dunha imaxe), e acústica (p.ex., a estimación da resposta acústica dunha sala, da cancelación de eco, etc.).

Unha das propostas máis destacadas da estimación de canle é a estimación cega. Estas técnicas explotan certas propiedades subxacentes da canle e do sinal transmitido para estimar a canle usando unicamente o sinal recibido. Estas características poden ser estatísticas, como estadísticos de orde superior [14], ou deterministas, como nos algoritmos que usan o módulo constante [56] ou como o criterio de máxima verosimilitude determinista [32]. Unha das principais vantaxes da estimación cega é que non precisa modificar o sinal orixinal para estimar; polo tanto, selecciónase xeralmente a estimación cega para aplicacións con esa restrición (por exemplo, para a explotación de petróleo [37] usando procesamento do sinal sísmico). Desafortunadamente, os enfoques de estimación cego sofren de converxencia lenta (é dicir, é necesario un elevado número de mostras do sinal recibido), e é tamén posíbel que converxan incorrectamente [57].

Indiscutibelmente, a estimación baseada no uso de sinais piloto é a familia de técnicas de estimación de canle máis utilizada. Estes sistemas utilizan unha parte do orzamento total de enerxía para transmitir un sinal, que recibe o nome de piloto ou sinal de adestramento, que se coñece no receptor, de xeito que se pode empregar para inferir a resposta da canle. Na maioría dos casos, o sinal piloto transmítese nun subespazo ortogonal ao do sinal portador de información,

frecuentemente usando multiplexación no dominio temporal ou ben no dominio da frecuencia.

Os algoritmos baseados no uso de sinais piloto teñen unha serie de inconvenientes coñecidos [59, 29, 30]: **1)** nas canles que varían rapidamente, os sinais de adestramento deberán enviarse a miúdo, a fin de actualizar a información de estado da canle, perdendo así unha cantidade significativa de recursos [1], **2)** o sinal portador da información debe ser desactivado, o que require a implantación dunha lóxica adicional para sincronizar os *slots* de secuencia piloto (en calquera dominio no que se use), tanto no transmisor como no receptor, **3)** a estimación está baseada en lugares específicos das secuencias piloto (tipicamente en tempo e/ou frecuencia); polo tanto, necesítase frecuentemente a interpolación, a fin de obter as estimacións da canle noutras posicións temporais ou frecuenciais.

Aínda que sexan menos significativas que as dúas técnicas de estimación que se describiron anteriormente, queremos mencionar que existen técnicas de estimación chamadas estimación semi-cega que utilizan os estatísticos, como fai a estimación cega, e símbolos coñecidos como fan os algoritmos baseados no envío de pilotos [15]. Como vantaxe máis importante, estas técnicas precisan secuencias de adestramento máis curtas; con todo, aínda precisan usar parte da capacidade en tempo ou frecuencia para enviar secuencias de adestramento.

A.1.1 Ligazón *Superimposed Training* - Marcado de Auga Dixital

Recentemente, aínda que a idea básica foi orixinalmente proposta en 1996 por Farhang-Boroujeny [22], a chamada *superimposed training* gañou relevancia como unha alternativa ás técnicas de estimación anteriormente indicadas. En *superimposed training*, unha secuencia piloto coñecida (imos nomealo marca de auga debido ao paralelismo co marcado de auga, a primeira mención a esta relación aparece, ata onde sabemos, no traballo de Mazzenga [40]) engádese ao sinal portador da información (que tamén imos chamar *host*). Esencialmente, estas técnicas utilizan secuencias periódicas como marcas de auga para estimar a canle, a fin de tomar vantaxe da cicloestacionariedade provocada na secuencia enviada. Dado que ambos sinais son simplemente sumados (é dicir, son enviados á vez), a necesidade de determinar explicitamente os intervalos de tempo/frecuencia para adestramento non existe, en contraste cos métodos tradicionais de estimación que usan pilotos [58, 41, 59]. Con todo, partindo do principio de que o transmisor ten una potencia máxima fixa, o sinal portador de información sufrirá algunha perda de potencia, que adicionalmente será distorsionada polo sinal superposto. É interesante sinalar que esta é unha das técnicas de pre-codificación, que non

[1] En termos de aumento de ancho de banda ou perda na taxa de información concreto, a secuencia de adestramento en UMTS-TDD pode ser de ata o 20% da carga útil.

son novas na comunicación dixital xa que foron extensivamente estudadas despois de ser presentadas por Tomlinson-Harashima [53, 28] a fin de ter en conta a información lateral do estado da canle dispoñíbel no transmisor.

Por desgraza, en *superimposed training*, as secuencias do *host* e do piloto non son ortogonais; así, o primeiro vai interferir co sinal piloto. Este é un problema amplamente estudado no marcado de auga, onde se coñece como interferencia do *host*, e ocorre naqueles algoritmos nos que a marca de auga independentemente xerada do *host* engádese a esta última (como nos casos de espectro ensanchado aditivo [13]). En ambos os campos propuxéronse solucións que dedican parte da potencia dispoñíbel para cancelar parcialmente a interferencia do *host* na dirección da secuencia engadida. Estes esquemas foron desenvolvidos de forma independente por Malvar e Florêncio en 2003 [38] no campo do marcado de auga, e por He e Tugnait en 2008 [29] para estimación da canle (inspirado polo traballo presentado en 2005 para OFDM por Chen *et al.* [9], ata onde sabemos o primeiro traballo considerando cancelación total da interferencia do *host* para *superimposed training* foi proposto por Ghogho *et al.* en [26]), e foron denominados respectivamente ISS e PDD. De xeito interesante, e de novo, ata onde nós sabemos, esa conexión entre PDD e ISS non foi relatada antes que no noso traballo [16].

Tanto ISS como PDD unicamente cancelan parcialmente a interferencia do *host*, deixando así espazo para melloras. De feito, a cancelación da interferencia do *host* completa foi alcanzada na ocultación de datos a través da explotación do paradigma de DPC, inicialmente proposto por Costa [10]. Adaptando o código de construción de Costa, Chen e Wornell [8] propuxeron o uso de DC-QIM que, grazas á súa característica de rexeitamento do *host*, levou a melloras substanciais de rendemento en relación a ISS. As vantaxes de técnicas de DPC en marcado de auga dixital foron amplamente recoñecidas [10, 8, 18]. En concreto, os esquemas baseados en DPC poden alcanzar a capacidade da canle para canles aditivas con ruído branco Gaussiano [21].

Dado que DPCE é moi sensíbel aos ataques de ganancia (tamén coñecidos como ataques valumétricos lineais), a igualación de canle estúdase en marcado de auga como unas das posíbeis solucións a este problema. Neste caso, a canle simplemente multiplica o sinal por un número real constante, que pode ser estudada como unha canle de esvaecemento plano tradicional en comunicación dixital, obtendo grandes probabilidades de erro de descodificación. Debido á súa importancia, propuxéronse varias técnicas, en base a igualación da canle como con Balado *et al.* [4] onde se desenvolveu un método baseado en cuantificadores escalares uniformes e turbocódigos, que iterativamente calcula o factor de ganancia, compensa o efecto, e decodifica a mensaxe contida. Sen embargo, Shterev e Lagendijk [51] propuxeron unha implementación baseada na procura exhaustiva da estimación de máxima verosimilitude (ML das siglas en inglés) do factor de escalado; de novo, ese valor úsase para a igualación das observacións, e para realizar a descodificación co conxunto de palabras código orixinal. Con todo, o

custo computacional de [4, 51] é moi importante, deixando espazo para mellora. Esta cuestión foi abordada con éxito no noso traballo [17], onde propoñemos unha técnica que tamén usa o criterio de máxima verosimulitude, mais que esixe moito menos recursos computacionais que [51].

É importante sinalar por unha cuestión de completitude que existen outras técnicas que afrontan o problema da sensibilidade de DPC a ataques de ganancia dunha maneira que se pode chamar conxuntos de palabras código robustos. Neste caso, o conxunto de palabras códigos típicos de SCS [18] substitúese por un conxunto de palabras código implicitamente robusto contra o ataque ganancia [47, 2, 42]. Mentres que [42] propón a utilización de conxuntos de palabras código baseados na fase (en oposición aos baseados en magnitude), en que a información [2] insértase considerando a correlación máxima entre o sinal *host* e un conxunto de secuencias xerado pseudoaleatoriamente, en [47] úsase un conxunto de palabras código que depende das estatísticas empíricas do sinal marcado. Desafortunadamente, estas técnicas presentan diversas desvantaxes, como que a distorsión de inserción resulta difícil de controlar en técnicas baseadas na cuantización fase [42] e as técnicas DPC ortogonais [2] (que tamén son computacionalmente máis esixentes que o SCS), e o traballo [47] require unha memoria que debe ser cuberta antes da descodificación para poder realizarse de maneira robusta.

Nesta tese de doutoramento, propoñemos o estudo da estimación da canle de esvaecemento plano baseado na codificación de papel sucio, que será abordado usando a estimación de máxima verosimilitude. Propomos tamén un conxunto de algoritmos prácticos baseados en ML con rigorosas restricións de complexidade (a diferenza de [51] que emprega busca exhaustiva). Ademais, queremos analizar as prestacións da técnica, a fin de obter ideas máis claras sobre os seus límites fundamentais e poder determinar se a cancelación de interferencias do *host* de marcado de auga dixital podería ser alcanzada tamén para a estimación. As prestacións dos algoritmos propostos foron tamén verificadas para distintas condicións e comparadas con outras técnicas de estimación (por exemplo, con estimadores baseados nos estatísticos de segunda orde, como representante de estimadores cegos, e PDD como exemplo de *superimposed training*). Ademais da aplicación das nosas técnicas no esquema básico de estudo, preséntanse outras posíbeis aplicacións nunha variedade de campos tecnolóxicos para demostrar a súa versatilidade.

A.2 Formulación do Problema

Previamente mostramos o paralelismo entre o marcado de auga dixital e a estimación da canle. Como indicamos, *superimposed training* pódese considerar o análogo en estimación da canle á Add-SS. Ademais PDD sería o equivalente a

ISS. Sen embargo, non hai equivalente en estimación da canle para DC-QIM; esta tese ocupa ese oco, na que se propón un método baseado en DPC.

Aquí considérase que o sinal transmítese por unha canle con esvaecemento plano (e ganancia t_0) que tamén introduce ruído branco Gaussiano, así o sinal recibido pódese expresar como

$$\boldsymbol{z} = t_0\boldsymbol{y} + \boldsymbol{n};$$

onde $\boldsymbol{y}$ xérase usando unha versión escalar de DC-QIM

$$y_i = \mathcal{Q}_\Delta(x_i - d_i) + d_i + (1-\alpha)\left[x - \mathcal{Q}_\Delta(x_i - d_i) - d_i\right],$$

con $i = 1, \ldots, L$, $\mathcal{Q}_\Delta$ denotando un cuantificador escalar uniforme con escalón de cuantificación Δ. $\boldsymbol{n}$ e $\boldsymbol{x}$ son observacións do vector aleatorio independentemente e identicamente distribuído que segue unha distribución Gaussiana de media cero. $\boldsymbol{d}$ denota as observacións do vector aleatorio *dither* $\boldsymbol{D} \sim U([-\Delta/2, \Delta/2]^L)$.

Abordamos o estudo deste problema considerando que a varianza do sinal orixinal é moito maior que a varianza do erro de cuantificación. Ademais, asumimos que a varianza do ruído da canle é moito maior que a varianza do *self-noise* (que se define como $(1-\alpha)\left[x - \mathcal{Q}_\Delta(x_i - d_i) - d_i\right]$) escalado por t_0. Dividiremos a análise en dous escenarios ben diferenciados:

- O caso *low-SNR*, que precisa que a potencia de ruído total (isto é, ruído da canle e *self-noise* escalado) sexa moito maior que o segundo momento escalado do cuantificador. Ademais desta condición, tamén pode precisar que a potencia total de ruído sexa moito menor que a varianza do sinal orixinal escalada.
- O escenario *high-SNR* precisa ademais que a potencia do ruído total sexa moito menor que o segundo momento escalado do cuantificador.

A.3 Estimador de Máxima Verosimilitude

Para obter a estimación da ganancia t_0, usamos o criterio ML que procura o valor de t mais probábel das observacións $\boldsymbol{z}$ cando non hai información *a priori* do factor de escalado. Pódese obter o estimador ML como

$$\hat{t}_0(\boldsymbol{z}) = \arg\max_t f_{\boldsymbol{Z}|T,\boldsymbol{K}}(\boldsymbol{z}|t, \boldsymbol{d}),$$

onde na expresión anterior $f_{\boldsymbol{Z}|T,\boldsymbol{K}}(\boldsymbol{z}|t, \boldsymbol{d})$ denota a distribución conxunta de $\boldsymbol{Z}$ coñecido o factor de escalado e a secuencia *dither*.

En moitas aplicacións que requiren estimación de parámetros, selecciónase o criterio ML porque é un enfoque sistemático e ademais polas súas interesantes propiedades asintóticas [60]: a estimación ML é consistente, é asintoticamente eficiente e asintoticamente segue unha distribución Gaussiana con media t_0 e varianza a CRB[2].

Como se ve na expresión do estimador ML mostrada anteriormente, precísase a función de distribución de probabilidade de $\boldsymbol{Z}$ condicionada ao coñecemento da secuencia de *dither* e á ganancia. Desafortunadamente, esta pdf (función de densidade de probabilidade, das súas en inglés) é difícil de operar matematicamente. Por iso, desenvolvemos varias aproximacións desa pdf (que se poderá usar no estimador ML debido a que as compoñentes de $\boldsymbol{Z}$ son independentes) para os escenarios analizados: *low-SNR* e *high-SNR*. Estas aproximacións foron avaliadas por medio da KLD (diverxencia Kullback-Leibler, das súas siglas en inglés), que se pode considerar unha medida da distancia entre dúas distribucións (neste caso, a distancia entra as nosas aproximacións e distribución real), mostrando que as aproximacións son precisas nos escenarios para os cales foron deseñadas.

Usamos as aproximacións das pdfs obtidas para particularizar a función de custe de ML. Analizándoas determinouse que teñen dúas partes ben diferenciadas, unha coa parte da distribución do sinal orixinal e outra que aparece como resultado do proceso de marcado do sinal. Ademais, analizando as funcións de custe púidose comprobar que mostran varios mínimos e máximos locais que inhabilitan o uso de técnicas de optimización convencionais, facendo necesario deseñar as nosas propias técnicas *ad-hoc* de estimación da ganancia.

Finalmente, nesta tese propóñense varias alteracións da técnica orixinal que permiten a súa aplicación para diferentes escenarios. Concretamente para o caso no que se descoñezan as varianzas do sinal orixinal e de ruído da canle. Noutra versión, baseándonos nas ideas de *Spread-Transform* usadas en marcado de auga, conseguimos que se poida modificar o punto de traballo do sistema. Por exemplo, a relación efectiva entre a potencia do *host* e da marca de auga ou a relación entre a potencia da marca e do ruído da canle.

A.4 Análise Teórica: Teoría da Estimación e Teoría da Información

Nesta tese realizouse unha análise teórica co fin de obter os límites fundamentais das técnicas de estimación baseadas en DPC e tamén para entender como o seu funcionamento asintótico depende dos parámetros do esquema.

[2] A CRB é unha cota inferior da varianza da distribución de estimadores insesgados.

A análise teórica realízase seguindo un enfoque de teoría da estimación. En concreto, como en moitos traballos de investigación (por exemplo, [52], [24], etc.) estúdase a CRB, xa que determina a cota inferior da varianza do erro de estimación dos estimadores insesgados de t_0. Propóñense ademais varias aproximacións das expresións da CRB usando as aproximacións da pdfs de Z propostas, é dicir, aquelas desenvolvidas para o casos de *low-SNR* e *high-SNR*. É importante resaltar que, segundo o indicado máis arriba, as técnicas de estimación ML son asintoticamente eficientes cando L tende ao infinito. Polo tanto, comparando as prestacións do estimador baseado en ML (p.ex., por medio da MSE entre as estimacións e os valores reais de t_0) coa CRB, pódese avaliar a eficiencia dos nosos estimadores.

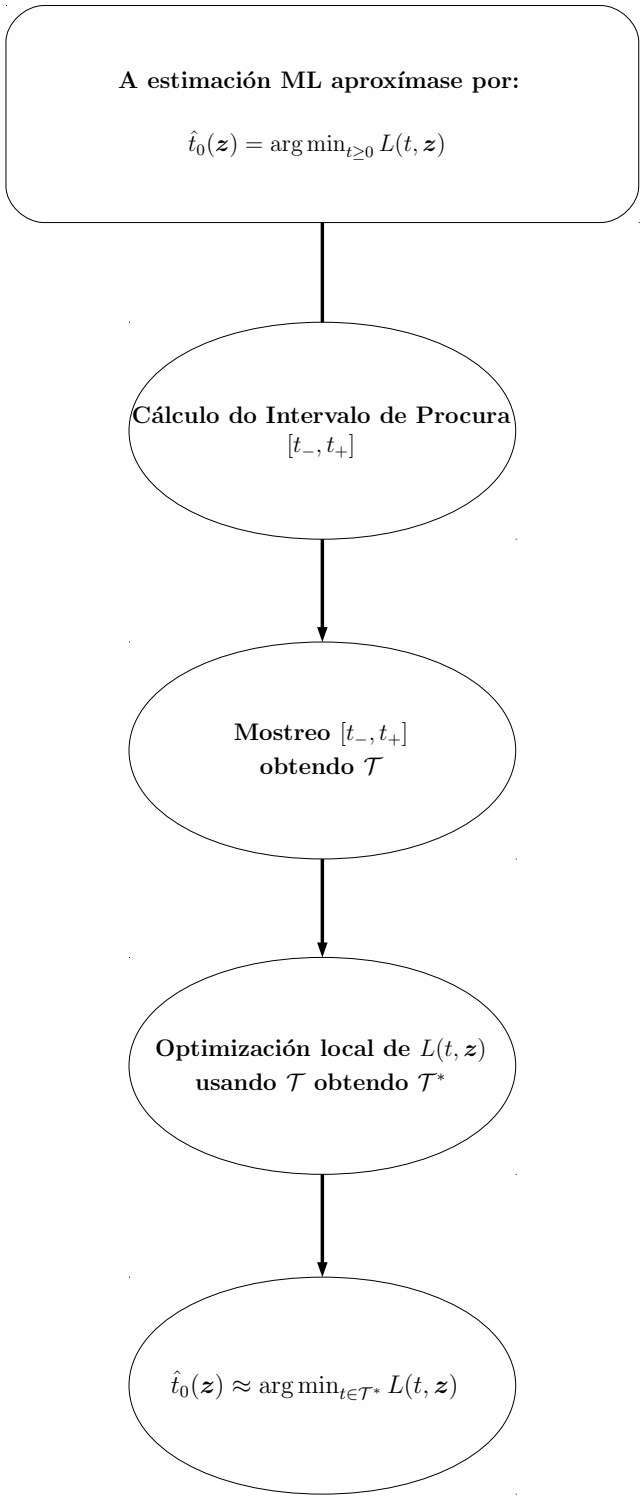
Analízase tamén o problema desde unha perspectiva de teoría da información. En concreto, comparamos as prestacións do esquema proposto coa obtida polas Add-SS/SIT en termos de información mutua entre entre Z e T dada a clave secreta, centrando a nosa análise, por unha cuestión de simplicidade, na caso de $L = 1$. Xa que a información mutua mide información que Z contén sobre a de T, a lóxica que sustenta a utilización desta métrica é que, a maior información mutua, maior será a información dispoñíbel sobre a ganancia. Aínda que no resto do traballo a ganancia é determinista, neste estudo T segue unha distribución Rayleigh con parámetro σ_T (esta distribución é amplamente utilizada para modelar a parte multiplicativa de canais con esvaecemento plano [39]).

Esta análise teórica indica que as prestacións da estimación usando pdf non só non son prexudicadas polo sinal *host*; senón que en realidade, este sinal axuda na estimación. De feito, os resultados obtidos indican que a asintótica da precisión da estimación é a mesma que se se enviara unha secuencia pilotos en lugar de información. Concretamente, a estrutura da pdf de Z provocada polo marcado de auga ten que aparecer (isto é, que o ruído da canle non domine a potencia de dita estrutura) para obter ditas prestacións nas que que o *host* axuda a estimar; se non, as prestacións obtidas serían as mesmas ás dun estimador que simplemente usase a relación entre as varianzas dos sinais do sistema.

A.5 Algoritmos Prácticos de Estimación

As funcións de custe ML baseadas nas pdfs de Z presentadas mostran varios máximos/mínimos locais. Isto provoca que moitos dos algoritmos de optimización tradicionais non se poidan usar. Ademais, a aplicación das técnicas de forza bruta resulta computacionalmente prohibitiva. Para resolver esta cuestión, propóñense un conxunto de algoritmos de estimación *ad hoc* que obteñen estimacións precisas requerindo uns custos computacionas aceptábeis.

Como mostra a Fig. A.1 de maneira ilustrativa para $t_0 \geq 0$, as técnicas de



$\hat{t}_0(\boldsymbol{z})$ **DPCE**

Figure A.1: Algortimo de Estimación *Ad-hoc*.

estimación propostas poden ser descritas de forma modular. En primeiro lugar, calcúlase, baseándose nas características das funcións de custo, un intervalo de procura para a optimización $[t_-, t_+]$. A continuación, o intervalo de procura mostréase explotando propiedades estatísticas da función obxectivo obtendo un conxunto de candidatos $\mathcal{T}$. Partindo deste conxunto de candidatos, lévase a cabo unha optimización local obténdose o conxunto de solucións locais $\mathcal{T}^*$; a estimación $\hat{t}_0(\boldsymbol{z})$ selecciónase como o elemento de $\mathcal{T}^*$ que minimiza a función de custo.

Para cada un destes procedementos, propoñemos varias alternativas, que deben ser seleccionadas de acordo co esixencias do escenario de aplicación específico. Concretamente, para a xeración do intervalo de procura deseñouse un método que usa as propiedades estatísticas do problema e outro que se fundamenta nas propiedades deterministas das funcións de custe. Para a xeración de $\mathcal{T}$ propóñense dous métodos, un baseado en tratar de estimar os centroides usados ao marcar e o outro na análise da media da función de custe. Finalmente, propóñense tamén dúas técnicas para a estimación local, unha asumindo que coñecemos os centroides usados en transmisión, minimiza a distancia das observacións e estes centroides escalados. A outra técnica de optimización local baséase na aplicación do método da bisección nas derivadas das funcións de custe con respecto á ganancia da canle para procurar o mínimo local da función de custe.

Os experimentos levados a cabo indican que os nosos algoritmos prácticos de estimación mostran unhas prestacións en termos de precisión próximas as da CRB cando se verifican as hipóteses e o número de observacións é suficientemente elevado.

Alen diso, móstranse como no caso de descoñecemento das varianzas do *host* e do ruído da canle obtéñense boas prestacións sempre que a estrutura da distribución de Z apareza. Verifícase como usando *Spread-Transform* no noso algoritmo podemos controlar o punto de traballo, incluso pudendo obter boas estimacións cando a potencia do ruído da canle é maior que a da marca de auga (sen usar esta modificación baseada en *Spread-Transform* non sería posíbel).

Usando experimentos que miden o tempo de computación, mostrouse que as nosas técnicas de estimación baseadas en DPC precisan moito menos tempo que outras baseadas en procura exhaustiva (p.ex. pasando de requerir cententas de segundo a décimas de segundo usando os nosos algoritmos). Polo tanto, as nosas técnicas poden ser usadas en aplicacións nas que se existen fortes restricións temporais.

A.6 Aplicacións

Co obxectivo de obter unha idea sobre a ampla gama de usos prácticos dos nosos algoritmos, esta tese presenta un conxunto de aplicacións fóra do escenario de aplicación básico (isto é, sinais reais que seguen unha distribución Gaussiana e ganancias reais), concretamente:

- Usamos os nosos algoritmos para facer o marcado de auga dixital baseado en DPC robusto para ataques de ganancia. Os resultados validan, usando sinais sintéticos e imaxes reais, a eficacia das nosas técnicas en tratar con tales ataques.
- Tamén amosamos, nun escenario de comunicacións da canle con esvaecemento plano, como igualar a ganancia estimada cos nosos algoritmos.Os resultados mostran que as nosas técnicas melloran as prestacións daquelas técnicas baseadas na varianza, así como as de *superimposed training*.
- Finalmente, tamén propoñemos como adaptar o noso algoritmo de estimación para o caso de sinais complexos e ganancias complexas. Os algoritmos propostos baséanse nas formas mais comúns de multiplicar dous números complexos, usando coordenadas polares ou usando coordenadas Cartesianas. As prestacións das técnicas indican que aproximadamente acándase os mesmos resultados que para o caso real e Gaussian: toda a potencia do *host* úsase para axudar na estimación.

Bibliography

[1] Milton Abramowitz and Irene Anne Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Number 55. Courier Corporation, 1964.

[2] Andrea Abrardo and Mauro Barni. Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding. *IEEE Transactions on Signal Processing*, 53(2):824–833, February 2005.

[3] Félix Balado. Personal communication, February 2014.

[4] Félix Balado, Kevin M. Whelan, Guénolé C. M. Silvestre, and Neil J. Hurley. Joint iterative decoding and estimation for side-informed data hiding. *IEEE Transactions on Signal Processing*, 53(10):4006–4019, October 2005.

[5] Mauro Barni, Franco Bartolini, Alessia De Rosa, and Alessandro Piva. Capacity of full frame DCT image watermarks. *IEEE Transactions on Image Processing*, 9(8):1450–1455, August 2000.

[6] F. Bartolini, A. Tefas, M. Barni, and I. Pitas. Image authentication techniques for surveillance applications. *Proceedings of the IEEE*, 89(10):1403–1418, October 2001.

[7] Centum-RT. Aeromark. `http://www.centum-rt.com`. Accessed: 2015-07-3.

[8] Brian Chen and Gregory W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4):1423–1443, May 2001.

[9] Ning Chen, G.T. Zhou, and A. Swami. Bandwidth and power efficiency considerations for optimal training in ofdm. In *2005 IEEE/SP 13th Workshop on Statistical Signal Processing*, pages 1364–1369, July 2005.

[10] Max H. M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.

[11] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[12] Ingemar J. Cox, Joe Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, December 1997.

[13] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom. *Digital Watermarking*. Morgan Kaufmann Publishers, 2002.

[14] Amod V. Dandawate and Georgios B. Giannakis. Asymptotic theory of mixed time averages and kth-order cyclic-moment and cumulant series. *IEEE Transactions on Information Theory*, 41(1):216–232, January 1995.

[15] Elisabeth De Carvalho and Dirk Slock. Blind and semi-blind fir multichannel estimation: (global) identifiability conditions. *IEEE Transactions on Signal Processing*, 52(4):1053–1064, 2004.

[16] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. Flat fading channel estimation based on dirty paper coding. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6479–6483, May 2014.

[17] Gabriel Domínguez-Conde, Pedro Comesaña, and Fernando Pérez-González. A new look at ML step-size estimation for Scalar Costa Scheme data hiding. In *Proceedings of the IEEE International Conference on Image Processing*, pages 4211–4215, Paris, France, September 2014.

[18] J.J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod. Inverse mapping of SCS-watermarked data. In *2002 11th European Signal Processing Conference*, pages 1–4, September 2002.

[19] J.J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod. Scalar Costa Scheme for information embedding. *IEEE Transactions on Signal Processing*, 51(4):1003–1019, April 2003.

[20] U. Erez and S. ten Brink. A close-to-capacity dirty paper coding scheme. *IEEE Transactions on Information Theory*, 51(10):3417–3432, October 2005.

[21] Uri Erez and Ram Zamir. Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. *IEEE Transactions on Information Theory*, 50(10):2293–2314, October 2004.

[22] B. Farhang-Boroujeny. Experimental study of semi-blind channel identification/equalization through pilot signals. In *1996 3rd International Conference on Signal Processing*, volume 1, pages 618–621 vol.1, October 1996.

[23] Hembrooke Emil Frank. Identification of sound and like signals, October 1961. US Patent 3,004,104.

[24] W. Gappmair, R. López-Valcarce, and C. Mosquera. Joint nda estimation of carrier frequency/phase and SNR for linearly modulated signals. *IEEE Signal Processing Letters*, 17(5):517–520, May 2010.

[25] Riccardo De Gaudenzi and Marco Luise. Analysis and design of an all-digital demodulator for trellis coded 16-qam transmission over a nonlinear satellite channel. *IEEE Transactions on Communications*, 43(2/3/4):659–668, February/March/April 1995.

[26] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami. Channel estimation and symbol detection for block transmission using data-dependent superimposed training. *IEEE Signal Processing Letters*, 12(3):226–229, March 2005.

[27] Gradiant. Shadow. `http://gradiant.org`. Accessed: 2015-07-5.

[28] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Transactions on Communications*, 20(4):774–780, August 1972.

[29] Shuangchi He and Jitendra K. Tugnait. On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences. *IEEE Transactions on Signal Processing*, 56(7):3214–3228, July 2008.

[30] Peter Hoeher and Fredrik Tufvesson. Channel estimation with superimposed pilot sequence. In *Proceedings of the IEEE Global Telecommunications Conference*, pages 2162–2166, 1999.

[31] Jaime Holguin. Actor sued over 'screener' leaks. `http://www.cbsnews.com/news/actor-sued-over-screener-leaks/`. Accessed: 2015-07-7.

[32] Y. Hua. Fast maximum likelihood for blind identification of multiple fir channels. *IEEE Transactions on Signal Processing*, 44(3):661–672, March 1996.

[33] A. K. Jain. *Fundamentals of Digital Image Processing*, pages 150–153. Prentice Hall, 1988.

[34] Steven M. Kay. *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, chapter 3. Pearson Education, 1993.

[35] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, May 1996.

[36] David C. Lay. *Linear Algebra and Its Applications*. Addison-Wesley, 2012.

[37] Hui Luo and Yanda Li. The application of blind channel identification techniques to prestack seismic deconvolution. *Proceedings of the IEEE*, 86(10):2082–2089, October 1998.

[38] Henrique S. Malvar and Dinei A. F. Florêncio. Improved spread spectrum: A new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 51(4):898–905, April 2003.

[39] T.L. Marzetta and B.M. Hochwald. Capacity of a mobile multiple-antenna communication link in rayleigh flat fading. *IEEE Transactions on Information Theory*, 45(1):139–157, January 1999.

[40] F. Mazzenga. Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence. *IEEE Transactions on Broadcasting*, 46(2):170–176, June 2000.

[41] A.G. Orozco-Lugo, M.M. Lara, and D.C. McLernon. Channel estimation using implicit training. *IEEE Transactions on Signal Processing*, 52(1):240–254, January 2004.

[42] Fabricio Ourique, Vinicius Licks, Ramiro Jordan, and Fernando Pérez-González. Angle QIM: A novel watermark embedding scheme robust against amplitude scaling distortion. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume II, pages 797–780, March 2005.

[43] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.

[44] Luis Pérez-Freire, Pedro Comesaña, and Fernando Pérez-González. Information-theoretic analysis of security in side-informed data hiding. In *Information Hiding*, pages 131–145. Springer, 2005.

[45] Luis Pérez-Freire, Fernando Pérez-González, and Pedro Comesaña. Secret dither estimation in lattice-quantization data hiding: a set membership approach. In *Electronic Imaging 2006*, pages 60720W–1–60720W–12. International Society for Optics and Photonics, January 2006.

[46] Luis Pérez-Freire, Fernando Pérez-González, and Sviatoslav Voloshynovskiy. An accurate analysis of scalar quantization-based data hiding. *IEEE Transactions on Information Forensics and Security*, 1(1):80–86, March 2006.

[47] Fernando Pérez-González, Carlos Mosquera, Mauro Barni, and Andrea Abrardo. Rational dither modulation: A high-rate data-hiding method invariant to gain attacks. *IEEE Transactions on Signal Processing*, 53(10):3960–3975, October 2005.

[48] Luis Pérez-Freire. *Digital Watermarking Security.* PhD thesis, Escola Técnica Superior de Enxeñeiros de Telecomunicación, Universidade de Vigo, 2008.

[49] Dominic Rushe. Us pressured spain to implement online piracy law, leaked files shows. `http://tinyurl.com/qxjrz87`. Accessed: 2015-07-7.

[50] Gerald Schaefer and Michal Stich. UCID - an uncompressed colour image database. In *SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia*, pages 472–480, January 2004.

[51] Ivo D. Shterev and Reginald L. Lagendijk. Amplitude scale estimation for quantization-based watermarking. *IEEE Transactions on Signal Processing*, 54(11):4146–4155, November 2006.

[52] Petre Stoica and Nehorai Arye. Music, maximum likelihood, and cramer-rao bound. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(5):720–741, May 1989.

[53] M. Tomlinson. New automatic equaliser employing modulo arithmetic. *Electronics Letters*, 7(5):138–139, March 1971.

[54] TRedess. CWS Software. `http://www.tredess.com`. Accessed: 2015-07-3.

[55] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory.* John Wiley and Sons, 1968.

[56] John R. Treichler and Brian G. Agee. A new approach to multipath corretion of constant modulus signals. *IEEE Transactions on Acoustisc, Speech, and Signal Processing*, 31(2):459–472, April 1983.

[57] Jitendra K. Tugnait, Lang Tong, and Zhi Ding. Single-user channel estimation and equalization. *IEEE Signal Processing Magazine*, 17(3):17–28, May 2000.

[58] J.K. Tugnait and Weilin Luo. On channel estimation using superimposed training and first-order statistics. *IEEE Communications Letters*, 7(9):413–415, September 2003.

[59] J.K. Tugnait and Xiaohong Meng. On superimposed training for channel estimation: performance analysis, training power allocation, and frame synchronization. *IEEE Transactions on Signal Processing*, 54(2):752–765, February 2006.

[60] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory-Part 1-Detection, Estimation, and Linear Modulation Theory.* John Wiley & Sons, 2001.

[61] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[62] W. Paul Webber. On the fundamental theorem of algebra. *Mathematics News Letter*, pages 9–13, 1933.