A unified IF-detection theory for practical coherent optical system design

van

Pieter W. Hooijmans

Delft, 15 februari 1993
Coerente optische detectie is een treffende illustratie van het duale golf-deeltjes karakter van licht; de heterodyne menging is slechts te verklaren vanuit het golfkarakter terwijl de fotodetectie een duidelijk gequantiseerd proces is.


Fotodetectie is geen kwadratisch maar een lineair proces, met een rechtevriendige relatie tussen het optische vermogen en de elektrische stroom.

Dit proefschrift, paragraaf 1.1.2.

De overeenkomst tussen de introductie van elektrische heterodyne in het eerste kwart van deze eeuw en de huidige ontwikkeling van coerente optische systemen is dat er in feite hetzelfde gebeurt, alleen bij frequenties die ruwweg miljoen maal groter zijn.

Dit proefschrift, paragraaf 1.3.

Deze historische overeenkomst maakt het aannemelijk dat coerente optische systemen een zelfde wijdverbreide toepassing zullen vinden als de radio, de televisie en de andere elektrische heterodyne systemen.

Voor hoge-snelheid coerente optische transmissie is CPFSK fasediversiteitsontvangst met polarisatiecontrole de ideale oplossing met betrekking tot bandbreedteefficientie, ontvangergevoeligheid en lijnbreedtetolerantie.

Dit proefschrift, paragraaf 7.4.

Het analyseren van een niet-synchrone ASK-ontvanger zonder daarbij de beslis-drempel te berekenen is zinloos.

Dit proefschrift, paragraaf 4.2.

Het is verrassend dat mensen die coerente optische systemen afwijzen vanwege de (on)betrokkenheid van een additionele locale oscillator laser in de ontvanger, wel bereid zijn directe detectie systemen met optische versterkers, bij voorkeur bidirectioneel gepompt, te aanvaarden.
VII
Zowel vonkzenders als optische systemen met directe detectie ontroeren door de botte kracht waarmee ze toch nog tot opmerkelijke prestaties weten te komen, zij dienen dan ook náast elkaar op een opvallende plaats in het museum te staan.

VIII
De fouten in de berekenende beslissdrempel ten gevolge van het gebruik van de verkeerde typen kansdichtheidsfuncties (bijvoorbeeld Gaussische) worden treffend geïllustreerd door de resultaten van Jacobsen en Garrett.


IX
Zowel de markt als de industrie zijn conservatief als het gaat om de introductie van nieuwe technologieën.

X
Wie beweert dat directe detectie en/of golflengte multiplexing dezelfde functionaliteit en betrouwbaarheid kunnen verschaffen, tegen vergelijkbare kosten, als coherente optische technieken, baseert zich niet op technisch-economische argumenten.

XI
Hoogfrequente analoge elektronica is geen vak maar een ambacht, waarbij praktische ervaring met ongewenste oscilaties, stukjes blik en aardingsproblemen de enige leerschool is. Gezien het geringe aantal hoogfrequent-ambachtslieden dat door het Nederlandse onderwijsbestel wordt afgeleverd kan men gerust stellen dat de overblijvers de ‘klompenmakers’ onder de elektronici zijn geworden.

XII
De overeenkomst tussen een Bourgondië ‘premier cru’ en een proefschrift is dat zij beide jaren van rijping vereisen.

XIII
De geboorte van een kindje tijdens de schrijffase van een proefschrift is bevorderlijk voor een strak werkschema en kan deswege aan alle promovendi worden aanbevolen.
A unified IF-detection theory for practical coherent optical system design
On the front cover:
A Philips 140 Mbit/s FSK Coherent Multi-Channel (CMC) optical receiver as built for the RACE-1010 CMC-demonstrator. The signals symbolise, from top to bottom, the FSK-modulated optical lightwave carrier; the optical heterodyned signal in the receiver, with the FSK-modulation converted to the IF amplitude beat signal; and the digital output signal after proper channel selection, detection and regeneration.
A unified IF-detection theory
for practical
coherent optical system design

Proefschrift

ter verkrijging van de graad van
doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus,
prof.drs. P.A. Schenck,
in het openbaar te verdedigen ten overstaan van een commissie
aangewezen door het College van Dekanen
op maandag 15 februari 1993 te 16.00 uur.

door

Pieter Werner Hooijmans

geboren te Voorburg,
elektrotechnisch ingenieur.
A unified IF-detection theory for practical coherent optical system design
Pieter Werner Hooijmans - [S.l. : s.n.]-III.
Proefschrift Delft. - Met lit. opg. en index.
Met samenvattingen in het Nederlands en Frans.
ISBN 90-74445-02-0
Trefw.: optische communicatie / modulatie / demodulatie.

The work described in this thesis has been carried out at the Philips Research Laboratories, Eindhoven as part of the Philips Research Programme during the period May, 1987, to May, 1992.

©Philips Electronics N.V. 1993

All rights are reserved. Reproduction in whole or in part is prohibited without the written consent of the copyright owner.

The copyright does not pertain to figure 3-17 and the data of figure 7-2, which have been reprinted with permission of The Institute of Electrical and Electronics Engineers.
To Christine and the boys

III
Preface

This thesis presents a unified theory of Intermediate Frequency (IF) detection in coherent optical receivers. Since the lasers that are used in coherent optical systems exhibit a considerable linewidth (or phase noise), optical heterodyne receivers with wide IF bandpass filters are required. Due to the phase noise of the signals, only non-coherent or differentially coherent IF-detection can be used. In practice this is equivalent to wide IF bandpass filters in combination with quadratic peak detectors or delay-line discriminators. Proper post-detection lowpass filtering is required for reducing the excess noise caused by the large IF bandwidth.

The analysis presented covers the complete IF and baseband section of a coherent optical receiver. By using only probability density functions, results are fully analytical. The laser linewidth is assumed to be negligible, and is consequently not included in the analysis. However, simple relations between the required IF bandwidth or frequency deviation and the actual linewidth are derived. When these relations are satisfied, the assumption of negligible linewidth holds.

For single-branch systems, the influence of the relative IF bandwidth and the signal-to-noise ratio on the sensitivity is taken into account. A very compact and accurate approximation for the decision threshold setting has been derived. For multi-branch (diversity) systems the effects of additional branches and imbalances between the branches are also included. The analysis covers ASK, FSK, CPFSK and DPSK digital modulation methods, and can easily be extended for evaluation of alternative modulation and detection schemes.

Results from the theoretical analysis are verified by evaluation of two practical coherent optical systems, operating at 140 Mbit/s and 1 Gbit/s. Differences between theory and practice are found to be minute.
Contents

Preface .................................................. V

Introduction ............................................. 1
  Optical fibre communication .......................... 1
  Coherent optical transmission ........................ 3
  System evaluation ...................................... 5
  Outline of the thesis .................................. 6

1 Coherent optical communication ...................... 9
  1.1 Principles ......................................... 9
    1.1.1 Signal definition ................................ 9
    1.1.2 Photo detection ................................ 10
  1.2 Signal-to-Noise Ratio ............................. 14
  1.3 How coherent is coherent? ........................ 16

2 Review of system aspects ............................. 21
  2.1 Noise sources ..................................... 22
    2.1.1 Electronic noise sources ........................ 22
    2.1.2 Photo detection noise .......................... 23
  2.2 Laser characteristics ............................. 24
    2.2.1 Laser linewidth ................................ 24
    2.2.2 Linewidth effects .............................. 26
    2.2.3 Intensity noise ................................ 27
  2.3 Digital modulation and detection ................. 30
    2.3.1 Amplitude Shift Keying ........................ 30
    2.3.2 Frequency Shift Keying ........................ 30
    2.3.3 Phase Shift Keying ............................. 33
  2.4 Optical detection schemes ........................ 35
    2.4.1 Heterodyne detection ........................... 35
    2.4.2 Homodyne detection ............................. 36

VII
2.4.3 Phase diversity .................................................. 38
2.5 Polarisation handling ........................................... 41
  2.5.1 Control ......................................................... 41
  2.5.2 Diversity ....................................................... 41
2.6 System qualification ............................................ 43
  2.6.1 Coherent IF-detection ....................................... 43
  2.6.2 Differentially coherent IF-detection ...................... 44
  2.6.3 Non-coherent IF-detection .................................. 45

3 Non-coherent IF-detection ........................................ 47
  3.1 Receiver model .................................................. 48
    3.1.1 Circuits .................................................... 48
    3.1.2 IF signals .................................................. 51
    3.1.3 Signal-to-Noise Ratio ...................................... 54
  3.2 Probability density functions ................................ 55
    3.2.1 PDF's of the IF signal .................................... 55
    3.2.2 Moments of the distributions ............................... 56
  3.3 IF-detection .................................................... 58
    3.3.1 Diode detectors ............................................ 58
    3.3.2 Spectral effects ............................................ 61
  3.4 Post-detection filtering ....................................... 65
    3.4.1 Integrate-and-dump filters ................................ 65
    3.4.2 Inter-symbol interference .................................. 70
  3.5 Decision theory ................................................ 71
  3.6 Linewidth effects ............................................. 73
    3.6.1 Theory ....................................................... 73
    3.6.2 Linewidth floor ............................................. 74
    3.6.3 IF bandwidth ............................................... 79
    3.6.4 Procedure ................................................... 82
  3.7 Summary ......................................................... 82

4 Single-filter detection .......................................... 83
  4.1 Reference case \(m=1\) ........................................... 84
    4.1.1 IF-detection ............................................... 84
    4.1.2 Decision threshold ......................................... 86
    4.1.3 The Bit-Error Rate ......................................... 91
    4.1.4 Sensitivity ................................................. 94
  4.2 Non-optimum IF bandwidth \(m>1\) ................................ 99
    4.2.1 Post-detection filtering ................................... 99
    4.2.2 The decision threshold for quadratic detection ........ 100
    4.2.3 The Bit-Error Rate, quadratic detection ................. 105
  4.3 Sensitivity and penalties \(m>1\) ................................ 108
CONTENTS

4.3.1 Numerical evaluation ........................................... 108
4.3.2 Comparison with other results ......................... 110
4.4 Linear IF-detection ............................................. 111
  4.4.1 Post-detection filtering ............................... 111
  4.4.2 Characteristic functions method .............. 112
4.5 Summary ......................................................... 115

5 Dual-filter detection ........................................ 117
  5.1 Receiver model ............................................ 118
    5.1.1 A second IF branch .................................. 118
    5.1.2 FSK decision theory .................................. 119
    5.1.3 IF signals ............................................. 121
    5.1.4 Post-detection filtering ............................ 121
    5.1.5 Linewidth effects .................................. 122
  5.2 IF-detection ................................................ 124
    5.2.1 Probability density functions .................. 124
    5.2.2 The Bit-Error Rate .................................. 127
  5.3 Sensitivity and penalties ................................ 129
  5.4 FSK-ASK sensitivity improvement ..................... 129
  5.5 Summary ..................................................... 132

6 Diversity reception ........................................... 133
  6.1 Polarisation diversity .................................... 135
    6.1.1 Receiver model ...................................... 135
    6.1.2 BER for ASK polarisation diversity .............. 138
    6.1.3 SNR imbalance ....................................... 142
    6.1.4 BER for FSK polarisation diversity .............. 144
    6.1.5 SNR imbalance ....................................... 148
  6.2 ASK phase diversity ...................................... 150
    6.2.1 Receiver model ...................................... 150
    6.2.2 2x2 phase diversity receivers .................... 151
    6.2.3 2x2 phase diversity with IF imbalance .......... 153
    6.2.4 3x3 phase diversity receivers ..................... 155
    6.2.5 3x3 phase diversity with IF imbalance .......... 155
  6.3 Conclusions .................................................. 158

7 Differentially coherent IF-detection .................... 161
  7.1 Delay-line discrimination ................................ 162
    7.1.1 Receiver model ...................................... 162
    7.1.2 Linewidth effects ................................... 166
    7.1.3 Comparison of CPFSK and FSK linewidth requirements ... 171
  7.2 CPFSK ....................................................... 173
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2.1</td>
<td>IF-detection</td>
<td>173</td>
</tr>
<tr>
<td>7.2.2</td>
<td>The Bit-Error Rate</td>
<td>177</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Sensitivity and penalties</td>
<td>180</td>
</tr>
<tr>
<td>7.3</td>
<td>Polarisation diversity CPFSK</td>
<td>184</td>
</tr>
<tr>
<td>7.3.1</td>
<td>The Bit-Error Rate</td>
<td>184</td>
</tr>
<tr>
<td>7.3.2</td>
<td>SNR imbalance</td>
<td>184</td>
</tr>
<tr>
<td>7.4</td>
<td>CPFSK phase diversity</td>
<td>186</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Receiver model</td>
<td>186</td>
</tr>
<tr>
<td>7.4.2</td>
<td>IF-detection</td>
<td>188</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Effects of SNR imbalance</td>
<td>191</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Compensation of IF phase mismatch</td>
<td>193</td>
</tr>
<tr>
<td>7.4.5</td>
<td>Linewidth effects</td>
<td>194</td>
</tr>
<tr>
<td>7.5</td>
<td>DPSK</td>
<td>194</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Receiver model</td>
<td>194</td>
</tr>
<tr>
<td>7.5.2</td>
<td>IF-detection</td>
<td>195</td>
</tr>
<tr>
<td>7.5.3</td>
<td>Non-optimum DPSK (de-)modulation</td>
<td>197</td>
</tr>
<tr>
<td>7.6</td>
<td>Orthogonal signalling</td>
<td>202</td>
</tr>
<tr>
<td>7.7</td>
<td>Summary</td>
<td>203</td>
</tr>
<tr>
<td>8</td>
<td>Experimental verification</td>
<td>205</td>
</tr>
<tr>
<td>8.1</td>
<td>FSK heterodyne system description</td>
<td>206</td>
</tr>
<tr>
<td>8.1.1</td>
<td>Balanced front ends</td>
<td>208</td>
</tr>
<tr>
<td>8.1.2</td>
<td>IF AGC-amplifiers</td>
<td>209</td>
</tr>
<tr>
<td>8.1.3</td>
<td>Frequency discriminator</td>
<td>210</td>
</tr>
<tr>
<td>8.1.4</td>
<td>Baseband combining</td>
<td>214</td>
</tr>
<tr>
<td>8.1.5</td>
<td>Biphas line-coding</td>
<td>214</td>
</tr>
<tr>
<td>8.1.6</td>
<td>Post-detection filtering and regeneration</td>
<td>216</td>
</tr>
<tr>
<td>8.1.7</td>
<td>Construction of the receivers</td>
<td>221</td>
</tr>
<tr>
<td>8.2</td>
<td>FSK heterodyne system measurements</td>
<td>221</td>
</tr>
<tr>
<td>8.2.1</td>
<td>140 Mbit/s sensitivity measurements</td>
<td>221</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Analysis of sensitivity penalties</td>
<td>222</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Polarisation dependence</td>
<td>226</td>
</tr>
<tr>
<td>8.2.4</td>
<td>560 Mbit/s sensitivity measurements</td>
<td>227</td>
</tr>
<tr>
<td>8.2.5</td>
<td>Analysis of results</td>
<td>230</td>
</tr>
<tr>
<td>8.3</td>
<td>CPFSK phase diversity measurements</td>
<td>232</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Receiver configuration</td>
<td>232</td>
</tr>
<tr>
<td>8.3.2</td>
<td>Sensitivity measurement and analysis</td>
<td>236</td>
</tr>
<tr>
<td>8.4</td>
<td>Accuracy of system measurements</td>
<td>239</td>
</tr>
<tr>
<td>8.5</td>
<td>Results from literature</td>
<td>240</td>
</tr>
<tr>
<td>8.6</td>
<td>Concluding remarks</td>
<td>245</td>
</tr>
</tbody>
</table>
CONTENTS

9 Conclusions 247
  9.1 IF-detection theory ........................................ 247
  9.2 Design guidelines for coherent optical systems .......... 249

Appendices

A Review of probability density functions 257
  A.1 Summary of pdf characteristics ............................ 257
  A.2 Normal distribution ....................................... 261
  A.3 Chi-square distribution .................................. 262
  A.4 Chi distribution .......................................... 267

B Decision threshold calculation 273
  B.1 Exact threshold computation ............................... 273
  B.2 Iterative threshold approximation ......................... 275
  B.3 Analytical threshold approximation ....................... 276
  B.4 Comparison of the results ................................ 276

C BER of single-filter receiver 279
  C.1 The partial Mark error rate for m=1 ...................... 279
  C.2 Partial error rates for quadratic IF-detection ........ 280
  C.3 The BER for quadratic IF-detection ..................... 285
  C.4 Linear IF-detection ..................................... 286

D BER of dual-filter receiver 289
  D.1 Probability density functions ............................ 289
  D.2 Bit-error rate ........................................... 293

E List of symbols 295

F List of abbreviations 301

References 303

Summary 315

Samenvatting 317

Sommaire 319
Account of previous work related to this thesis 321
  List of publications by the author 321
  Contributions by others 324
  Connectivity matrix 326

Acknowledgements 327

Curriculum Vitae 329

Index 332
Introduction

Optical fibre communication

Nearly 25 years ago Kao and Hockham\(^1\) proposed the use of glass fibres for communication purposes. By the end of the seventies it became possible to install the first practical fibre-optic links based on on-off modulated transmitters and direct-detection receivers\(^2\). During that decade considerable advances had been made in order to reach the point where field-installation was possible. For example, the insertion loss of glass had been reduced from more than one thousand dB/km to about 1 dB/km for graded-index multi-mode glass fibres. Semiconductor light emitting diodes (LED) in the first infrared window (850 nm) were developed, including means to couple the light with acceptable efficiency into the 50-100 \(\mu\)m wide glass fibre core. The general advancement of silicon transistor technology made it possible to construct receivers with bandwidths up to 100 MHz. With these systems distances of about 10 km could be spanned at 100 Mbit/s, a considerable advancement compared to coaxial cable transmission systems. The principle limitation of these multi-mode systems was the modal dispersion in fibre, resulting in a theoretical upper limit of the bitrate-distance (BL) product of about 4 Gbit/s-km\(^3\).

Since then direct-detection systems have seen a steady improvement in performance, especially since the introduction of single-mode fibre (SMF) in 1981–1982. This opened two long-wavelength infra-red windows for transmission, at 1.3 and 1.55 \(\mu\)m, characterised by zero-dispersion and minimum loss respectively. Typical losses of standard fibres in these two windows have decreased to 0.35 and 0.18 dB/km nowadays. On the component side the introduction of single-mode fibre required the development of semiconductor lasers for transmission, since the amount of power that can be coupled from a LED into a SMF is very low. This led to the introduction of

---


INTRODUCTION

Fabry-Perot type multi-longitudinal mode laser diodes at 1.3 μm. For the receivers germanium avalanche photodiodes (APD) were developed, while bitrates increased to 560 Mbit/s. The theoretical BL-product of these systems is about 250 Gbit/s·km\(^4\), nearly two orders of magnitude higher than for multi-mode transmission.

In order to exploit the low-loss 1.55 μm window, the problem of dispersion had to be solved. One solution, dispersion-shifted fibre, did not really succeed in practice, apart from some applications like short undersea cables. For operation over highly dispersive standard single-mode fibres, the single-frequency Distributed Feed-Back (DFB) laser diode was therefore developed\(^5\). The combination of low-loss SMF, low-chirp DFB lasers, high-speed APD’s and the first GaAs integrated circuits has led to ever increasing performance. The bitrate and record transmission distance of commercial systems have increased to 2.5 Gbit/s and 150 km\(^6\), while the theoretical B\(^2\)L-limit is around 2500 (Gbit/s)\(^2\)·km\(^7\).

One of the latest innovations in direct-detection systems has been the introduction of optical amplifiers, often Erbium-Doped Fibre Amplifiers (EDFA)\(^8\), which can be used as booster-, in-line- or pre-amplifiers. These devices arrived just in time, since direct detection transmission in its present form is approaching the limits set by fibre dispersion and span loss. In particular at bitrates of 10 Gbit/s or higher - which are currently under development - acceptable transmission distances in the 1.55 μm window can only be obtained by using very low-chirp lasers or external modulation lasers, in combination with booster and pre-amplifier EDFA’s. Extremely high distances have been spanned experimentally using large series of in-line EDFA’s\(^9\), although this is in apparent contrast to the usual claim of optical communication that in-line repeaters are not required. In addition, at bitrates in excess of 10 Gbit/s integrated circuits are essential, but hard to manufacture.

In general it can be stated that direct-detection, although well-engineered by now,

\(^4\)Henry, loc. cit.
\(^7\)P.G. Mols and T. Geukes, Limits and possibilities in the development of 10 Gigabit per second direct-detection systems, NL-TN.262/92, Philips Research Laboratories, Eindhoven, 1992. The figure-of-merit has changed from BL into B\(^2\)L since the spectral width is also determined by the bitrate.
has reached its limits, especially with regard to the non-amplified transmission distance of high-bitrate systems. Due to the difficulties in making advanced multi-GHz components the receiver penalties are, relatively speaking, much higher than at bitrates up to 2.5 Gbit/s. An alternative option for the transmission of high-bitrate data, that overcomes the drawbacks of ever increasing synchronous bitrates while offering the same effective throughput, is parallel transmission by other forms of multiplexing. This is where coherent optical transmission comes in.

**Coherent optical transmission**

Right after the development of the laser by the end of the fifties, the use of optical heterodyning was proposed\textsuperscript{10} and indeed tested with gas lasers. The main problem with this technique was the required spatial overlap of the signal and local oscillator beams, so it never really came into use. However, soon after the proposal of fibre-optic communication by Kao and Hockham\textsuperscript{11} in 1966, DeLange\textsuperscript{12} proposed optical frequency division multiplexing (OFDM) in combination with heterodyne reception. Just like in radio reception, heterodyne detection means that the incoming (light-)signal is mixed with a local oscillator signal generated in the receiver. This generates an intermediate frequency (IF), yields a much higher receiver sensitivity than with direct-detection, and introduces selectivity. As a result it is possible to exploit the enormous bandwidth of the single-mode fibre, which is about 30.000 GHz for each of the two windows at 1.3 and 1.55 μm.

Originally, coherent optical techniques were purely seen as another step in the race to record transmission distance. The first system experiments were reported in 1980 by Saito et al.\textsuperscript{13}, a 200 Mbit/s Frequency Shift Keying (FSK) system. FSK in different forms has since then been the most widely used modulation scheme, mainly because it employs direct modulation of the laser. In fact the development of coherent optical systems has been closely tied to the rapidly advancing laser technology. In order to obtain ‘coherent light’ single-frequency optical sources are required. The Distributed Feedback (DFB) laser, the simplest single-frequency laser diode, has since become the ‘workhorse’ of coherent optical detection. The requirements of a laser for use in coherent optical systems are however very stringent. It must be single-frequency, with a high stability (low linewidth), give high output power and have flat modulation characteristic, high bandwidth, low sensitivity to optical reflections and, when used as local oscillator, a large tuning range. Another major problem in selecting devices

\textsuperscript{11}Kao and Hockham, *loc. cit*.
for coherent optical systems has always been the required wavelength match between transmitter and local oscillator.

Nevertheless, since the first coherent optical system experiments in 1980 the advancement of laser technology and system performance have been considerable. The bitrate of state-of-the-art systems has increased to 2.5 Gbit/s, while exceptionally 10 Gbit/s has also been reported\(^\text{14}\). Laser linewidths have been decreased by more than two orders of magnitude, from around 100 MHz in 1980 to values in the sub-MHz region. This makes it possible to use more advanced modulation and detection schemes, e.g. optical phase-lock loops (PLL)\(^\text{15}\). Many coherent optical systems have obtained bitrate-distance products in excess of the best direct-detection systems\(^\text{16}\), and the first field trials have been conducted\(^\text{17}\).

Although the high receiver sensitivity associated with coherent optical detection certainly is an advantage, this can not be the single reason for introducing such a new technology. Especially the introduction of EDFA's has taken away the sensitivity-advantage, although it should be stressed that these amplifiers can be used in combination with coherent optical systems as well\(^\text{18}\). However, optical heterodyne techniques generally feature superior performance due to the multichannel capacity improvement and added flexibility. This is often referred to as Coherent Multi-Channel (CMC) transmission. One possible application would be the distribution of many coherent optical (TV-)channels in OFDM, to many subscribers over a passive optical star network\(^\text{19}\). In this way the enormous fibre bandwidth could be used (by tens of optical channels), while the large power budget would allow distribution to thousands of subscribers. Channel selection takes place with the tuneable coherent optical receivers\(^\text{20}\). Another


\(^{15}\)J.M. Kahn and B.L. Kasper, PSK homodyne lightwave transmission using semiconductor lasers, Proc. ECOC '89, Gothenburg, pp. 413–416.


\(^{18}\)Saito 1, loc. cit.

\(^{19}\)M. Shibutani, S. Yamazaki, N. Shimosaka, S. Murata and M. Shikada, Ten-channel coherent optical FDM broadcasting system, Proc. OFC '89, Houston, p. 140.


application exploiting the wavelength flexibility of heterodyne detection would be the use in optical cross-connected nodes in the highest level of the broadband transmission network. Here the high capacity is used to increase the effective throughput of a fibre link to values in excess of anything feasible with high-speed direct-detection systems. For example, 16 coherent 2.5 Gbit/s channels in OFDM yield an effective throughput of 40 Gbit/s. The tuneability of the receivers provides flexibility in the channel allocation (the cross-connect function), and added functionality such as protection switching and back-up routing.

System evaluation

After ten years of research, coherent optical systems have been engineered to a high level, meaning that they now show a high and consistent performance both in- and outside the laboratory. Many practical systems perform close to the shot noise limit, at which the minimum amount of photons are utilised for the information transmission. This entails that these systems be evaluated and analysed with a sufficiently high accuracy, preferably in the order of tenths of a dB. As in all communication systems it is essential to identify the factors that most degrade the system performance. In digital optical communication the system performance is expressed by the receiver sensitivity, the minimum optical input power that is required for obtaining a certain Bit-Error Rate (BER). The difference between the actual receiver sensitivity and the theoretical shot-noise limit for the particular modulation scheme considered is called the sensitivity penalty. It is a task of the system engineer to explain all significant individual penalties contributing to the overall sensitivity penalty. That is precisely what is treated in this thesis.

In general there are two large groups of penalties associated with coherent optical systems. The first consists of penalties related directly or indirectly to the optics and the lasers. Since many of these penalties introduced new system aspects compared to the simpler direct-detection systems, much work has been devoted to analysing them. Fortunately these penalties can easily be measured in practical systems, allowing verification of the theoretical results. The second, equally important group consists of the penalties related to the electronics. When analysing system experiments these penalties are usually covered by the vague, although literally correct, term 'non-ideal electronics'. In this thesis, I will evaluate the penalties related to the Intermediate Frequency (IF) detection process in coherent optical receivers. The two main conclusions are that these penalties can be calculated analytically and are certainly not negligible, even if 'ideal electronics' are used.

The basis of coherent optical detection is the availability of single-frequency semiconductor lasers, which make it possible to use heterodyning techniques in the optical domain. Although from the point of view of optics and laser technology the achievements in realising such sources are tremendous, from an electronic point of view the
results are still modest, to say the least. This leads to the paradoxical situation that the superior optical coherent detection is, in many cases, followed by the simplest of all electronic detection methods: non-coherent IF-detection. Indeed, single- and dual-filter receivers with amplitude detectors are still the most widely used schemes. Only recently has more interest evolved in differentially coherent IF-detection using CPFSK modulation in combination with delay-line discriminators, made possible by the availability of lasers with improved linewidth performance.

The analysis presented in this thesis is application-driven; results are intended to be directly applicable to the analysis and design of practical high-performance coherent optical systems. This is the main difference with the many analyses published so far, which deal mainly with the influence of a finite laser linewidth upon the system performance. A non-zero linewidth introduces a floor in the BER-characteristics of a system, giving rise to rapidly increasing sensitivity penalties. Although mathematically interesting due to the complicated subject, this is of less practical importance: observable floors are not accepted in high-performance systems, but can be avoided by proper selection of the way in which the IF-detection is performed. At this point the analysis presented in this thesis can take over, assuming that the system has been designed in such a way as to avoid linewidth effects, for example, by using IF bandpass filters with a sufficiently large bandwidth. The laser may then be effectively modelled as a stable oscillator with zero linewidth, and the key effort may be dedicated to the analysis of the IF-detection process.

Outline of the thesis

The thesis is arranged as follows. In Chapter 1 a short overview will be given of the principles of coherent optical detection in digital communication systems. Chapter 2 reviews most of the (optical) system aspects and options that will influence the system configuration. This includes the noise sources, the laser linewidth and the different modulation, detection and polarisation handling schemes. From this it will be concluded that, depending upon the laser linewidth, three IF-detection schemes can be used; coherent-, differentially coherent- and non-coherent IF-detection. The first method, which makes use of optical PLL’s, will not be evaluated in this thesis since it is based on a completely different mechanism.

Chapter 3, on non-coherent IF-detection, introduces the receiver model and the analysis method that will be used from then on. The analysis is based on the probability density functions (pdf) of the signals involved. This is the best model of the real detection processes in the receiver, taking into account both the additive and multiplicative noise sources and mechanisms. Results from literature on the effects of a finite laser linewidth are summarised. However, since these results are usually not in a form fit for practical system design, they are translated into compact expressions that are directly applicable for calculating the required IF bandwidth.
INTRODUCTION

After these general and introductory chapters, different types of coherent receivers will be analysed in order of increasing complexity: Amplitude Shift Keying (ASK) single-filter detection (Chapter 4), Frequency shift Keying (FSK) dual-filter detection (Chapter 5), ASK and FSK polarisation diversity and ASK phase diversity receivers (Chapter 6). For every scheme the pdf's will be derived, leading to analytical expressions for the Bit-Error Rate (BER). From this result the sensitivity at a pre-defined BER (usually $10^{-9}$) can be obtained, as well as the sensitivity penalties. The most important parameter in these analyses will be the ratio of the operationally required IF filter bandwidth and the bitrate, denoted by $m$. Chapter 7 introduces differentially coherent IF-detection using delay-line discriminators. This method can be used for Differential Phase Shift Keying (DPSK) as well as Continuous Phase FSK (CPFSK). Throughout the analysis the different options will be compared with each other for better assessment of the advantages of each. In relation to practical system measurements, the effects of imbalances in the systems are also investigated. In diversity receivers, especially, differences between the IF-branches can easily occur, degrading the performance of the receiver. In the case of CPFSK and DPSK reception, penalties are introduced when deviations in the proper delay time or center frequency give non-optimum detection. These penalties can be expressed in very simple formulae as well.

Chapter 8 checks the validity of the theoretical results obtained in the previous chapters. Results from both experiments and literature are used, proving the accuracy of the analysis. The practical results are obtained from two different coherent optical systems, built by the author and his colleagues. The first is a 140 Mbit/s polarisation diversity multichannel FSK system, with dual-filter non-coherent IF-detection. The second system uses CPFSK phase diversity reception with differentially coherent IF-detection at a bitrate of 1 Gbit/s. After a description of the essential system components, detailed sensitivity measurements are presented, followed by a thorough identification of the different sensitivity penalties.

Chapter 9 summarises the theoretical results and gives practical guidelines for system design, based on the results obtained from the theoretical analysis and practical system experiments.

To assist the reader the Appendices include reviews of general and specific mathematical tools, and lists of symbols and abbreviations.
INTRODUCTION
Chapter 1

Coherent optical communication

1.1 Principles

1.1.1 Signal definition

In a coherent optical system the frequencies - and possibly the phases - of different light sources are related. This implies that these sources must be single frequency, otherwise frequency and phase cannot be defined. The light generated by a single-frequency laser and propagating through a single mode fibre (SMF) can be fully defined by its amplitude, frequency, phase and state of polarisation (SOP). This is in complete analogy with radio waves, since light can in most cases be regarded as an electro-magnetic field. Ideal single mode fibres have a perfectly circular core, giving circularly symmetrical fields perpendicular to the direction of propagation. In practice however, all fibres feature small ellipticities, inhomogeneities or internal stress, leading to differences in propagation of the fields aligned with the main axes of the ellipse. This phenomenon is called birefringence and effectively introduces different propagation velocities along two main polarisation axis. In optics these two axis are usually referred to as parallel (denoted by ||) and perpendicular (\perp). In general, the SOP of a signal that has propagated over more than a few meters of SMF can thus be regarded as being random; the two E-fields aligned with the main polarisation axes are decorrelated. When we assume that only the fundamental mode may propagate, the resulting electrical fields are given by Jeunhomme [60] according to the plane-wave approximation:
\[ E_\parallel = \frac{Z_0}{n_{eff}} H_\perp \]  
\[ E_\perp = \frac{Z_0}{n_{eff}} H_\parallel \]  
\[ E_z = H_z = 0 \]

In these formulae \( Z_0 \) is the free-space impedance and \( n_{eff} \) the effective refractive index of the fibre, determined by the material properties, the structure of the fibre and the wavelength. Since the E and H-fields of a given mode are coupled, from here on only the E-fields will be used. Each electrical field is defined as the real part of a complex vector defined by its amplitude \( E_s \), its angular frequency \( \omega_s \) and its phase \( \theta_s \):

\[ E = \text{Re} \{ E_s \exp(j\omega_s t + j\theta_s) \} = E_s \cos(\omega_s t + \theta_s) \]

The complete signal electrical field at some position along the direction of propagation in a fibre can now be described with an E-field vector. For convenience the main axis of the polarisation ellipse is aligned with the vector coordinate axis.

\[ \overline{E_s} = \begin{pmatrix} E_\parallel \\ E_\perp \end{pmatrix} = \begin{pmatrix} E_s \cos \phi \cos(\omega_s t + \theta_s) \\ E_s \sin \phi \sin(\omega_s t + \theta_s) \end{pmatrix} \]

The second part of this formula introduces \( \phi \), which is a measure for the ellipticity of the polarisation. When \( \phi \) is equal to any multiple of \( \frac{\pi}{2} \) the SOP will be linear along either of the two axes. \( E_s \) is the peak value of the E-field in those situations. Note that for \( 0 \leq \phi \leq \pi \) and \( \pi \leq \phi \leq 2\pi \) the direction of rotation of the E-field is opposite. In this way right-handed and left-handed (circularly) polarised waves can be defined.

Every component of the field vector consists of both an amplitude and an optical carrier term, so it is convenient to split the expression for \( \overline{E_s} \) into a carrier-matrix \( \mathcal{M} \) and an amplitude vector \( \mathcal{E}_s \):

\[ \overline{E_s} = \mathcal{M} \mathcal{E}_s \\
= \begin{pmatrix} \cos(\omega_s t + \theta_s) & 0 \\ 0 & \sin(\omega_s t + \theta_s) \end{pmatrix} \begin{pmatrix} E_s \cos \phi \\ E_s \sin \phi \end{pmatrix} \]

This notation can be used to advantage in the next section when the photo detection process is analysed.

### 1.1.2 Photo detection

Photodiodes are used for conversion of optical signals into electrical signals. In coherent systems only PIN (from Positive-Intrinsic-Negative, in reference to the layer structure) diodes without an internal gain mechanism are employed. These devices can be considered as linear power detectors that convert optical power linearly into
1.1. **PRINCIPLES**

Electrical current

\[ i = R \cdot P \] (1-7)

The constant \( R \) is the responsivity of the diode and a measure for the effectiveness of the light-to-current conversion. Related to the definition of \( R \), light is usually modelled as photons with energy \( h \nu \), where \( h \) is Planck's constant and \( \nu \) the frequency of the light. The responsivity can then be written as

\[ R \triangleq \frac{\eta q}{h \nu} \ [A/W] \] (1-8)

The unit of \( R \) is \( A/W \) and \( \eta \) is defined as the quantum efficiency of the diode, being a measure for the fraction of photons actually converted into electrons with charge \( q \). The responsivity is an important indicator of the photodiode performance in a practical receiver.

In principle the photo detection process should be analysed using quantum-mechanics, but this is outside the scope of the thesis. A more classical analysis by Haus [34] yields the following expression:

\[ i_s = D \cdot E_s E_s^* = D \cdot |E_s|^2 \] (1-9)

Here \( * \) denotes complex conjugate, while \( D \) is the electrical field responsivity, with the dimension \( m^2/V \Omega \). It is clear that the carrier matrix \( M \) has disappeared from the expression, a consequence of the time-averaging effect of the photo detection process.\(^2\) This means that 'slow' changes of the amplitude vector will be detected by the photodiode whenever the frequency of these effects is within the bandwidth of the detector. Any 'fast' effects related to the optical carrier will not be seen in the electrical current.

\(^1\)A commonly made mistake is the assumption that photodiodes are quadratic detectors that convert the squared electrical field into a current. This would mean, however, that second harmonics of the input optical frequency are generated as well, which is physically impossible. The confusion is caused by the fact that, although PINs are diodes, they are used in a completely different regime compared to electrical diodes. These, when forward-biased close to the knee-voltage, yield an approximately square-law voltage-to-current conversion. Photo detectors on the other hand are reverse biased, which introduces a fully depleted I-layer that can be used for detection of incoming light.

\(^2\)Depending upon the mechanism used to describe the detection process this can be seen in two different ways. In all cases the incoming light generates electron-hole pairs. When the light is considered to be an EM-field, meaning that the energy passed to the electrons must be such that they can tunnel from the valence to the conduction band of the material. In order to reach this threshold the incoming power must be accumulated for some time. When the light is modelled as photons of energy \( h \nu \), the transfer from valence to conduction band will take place only if \( h \nu \) is approximately equal to the bandgap. However, in this case the frequency of the incoming light must be measured first, a process requiring several periods.
Coherent optical detection

When the light from two coherent (single frequency) sources is coupled into one fibre, the propagation axis and the Poynting vectors will overlap. Under some conditions interference can now take place, giving coherent optical detection. First the essential polarisation requirement will be derived. Since this requirement must be true for all SOP's of the incoming signal it will be derived for the simple case of linear signal and Local Oscillator (LO) polarisations. The general situation of an angular misalignment \( \Phi \) between two linear polarisations is illustrated in Figure 1-1. Defining the angular

![Figure 1-1: Polarisation angular misalignment between signal and local oscillator light coupled into the same fibre.](image)

frequency and phase of the LO-light as \( \omega_{lo} \) and \( \theta_{lo} \) the total E-field vector is given by

\[
\overline{E_{tot}} = \begin{pmatrix}
E_s \cos(\omega_s t + \theta_s) + E_{lo} \cos \Phi \cos(\omega_{lo} t + \theta_{lo}) \\
E_{lo} \sin \Phi \sin(\omega_{lo} t + \theta_{lo})
\end{pmatrix}
\]

(1-10)

In practical situations the amplitude of the LO optical power will be orders of magnitude higher than the signal, while the relative frequency difference between the two signals is extremely small\(^3\). Combining two (not necessarily optical) carriers with a small frequency difference will make them beat with a frequency equal to that difference. This difference frequency is defined, in analogy with classical radio techniques, as the Intermediate Frequency (IF). Figure 1-2 illustrates the generation of the IF. Using

\(^3\)In practice this is one of the most difficult aspects of coherent optical systems: obtaining lasers with nearly the same optical frequency. To give a numerical indication: at a wavelength of 1.55 \( \mu \)m the carrier frequencies are nearly 200,000 GHz, whereas the IF in most systems is somewhere between 1 and 5 GHz. This is only 5 to 25 ppm.
1.1. PRINCIPLES

![Figure 1-2: Interference between two optical carriers (left) and the resulting amplitude modulated optical signal (right).](image)

the cosine-rule in Figure 1-2 leads to the following expression for the length of the parallel component of the total field:

$$|E_{tot\|}| = \sqrt{E^2_{lo} \cos^2 \Phi + E^2_s + 2E_s E_{lo} \cos \Phi \cos(\omega_{IF}t + \theta_{IF}) \cdot \cos(\omega_{Io}t + \theta_{Io})} \quad (1-11)$$

where $\omega_{IF}$ and $\theta_{IF}$ are defined by

$$\omega_{IF} = |\omega_{lo} - \omega_s| \quad (1-12)$$

$$\theta_{IF} = |\theta_{IF} - \theta_s| \quad (1-13)$$

The total electrical field is now given by

$$\overline{E_{tot}} = \begin{pmatrix} \cos(\omega_{Io}t + \theta_{Io}) & 0 \\ 0 & \cos(\omega_{Io}t + \theta_{Io}) \end{pmatrix} \cdot \begin{pmatrix} \sqrt{E^2_{lo} \cos^2 \Phi + E^2_s + 2E_s E_{lo} \cos \Phi \cos(\omega_{IF}t + \theta_{IF})} \\ E_{lo} \sin \Phi \end{pmatrix} \quad (1-14)$$

The electrical current generated in an optical detector by this electrical field can be obtained with formula (1-9) and equals

$$i = D \{E^2_{lo} + E^2_s + 2E_s E_{lo} \cos \Phi \cos(\omega_{IF}t + \theta_{IF})\} \quad (1-15)$$

This formula is composed of three different terms. In the case of zero LO power only the component $DE_s^2$ is left, which is the direct detection signal current. However, in
the case of coherent detection this current may be neglected compared to the much larger dc-current \( DE_{i0}^2 \) caused by the LO power. It should be noted that the detected LO current is independent of the polarisation misalignment angle \( \Phi \). The third component is the IF signal, generated by the coherent optical detection mechanism. Since this term is still related to the amplitude, frequency and phase of the received signal, all three basic digital\(^4\) modulation schemes can be used in combination with coherent detection: Amplitude, Frequency and Phase Shift Keying (ASK, FSK and PSK). So, whatever the modulation used, the received optical signal has been frequency down- converted to an electrical IF signal, still containing all its modulation information. An essential condition is however that the intermediate frequency is lower than the bandwidth of the receiver. This provides the selectivity of the receiver, which is the main advantage of coherent optical detection. A second advantage is the gain obtained with the coherent detection mechanism, which can be seen by comparing the amplitudes of the direct detection signal component \( DE_s^2 \) and the IF component. The latter has an amplitude increased by the factor

\[
\frac{2E_{i0}\cos\Phi}{E_s}
\]

In most practical situations \( E_{i0} \) will be much larger than \( E_s \), implying that a significant gain can be obtained. The effect of this gain upon the SNR will be investigated in the next section.

The IF current in (1-15) still depends on the polarisation \( \Phi \). Maximum IF signal is obtained when \( \Phi \) is equal to \( k \cdot \pi \), in which case the signal and LO polarisations are aligned. In the situation of orthogonal polarisations \( (\Phi = k \cdot \frac{\pi}{2}) \) no IF signal is generated at all. In the general case of receiving an elliptically polarised signal one can deduce that the LO polarisation must be identical. This means that the main axis, the ellipticity and the direction of rotation of the two polarisations must coincide. Polarisation handling is therefore one of the more complicated aspects of coherent receivers, and it is directly related to the different system configurations analysed in this thesis.

### 1.2 Signal-to-Noise Ratio

The Signal-to-Noise Ratio (SNR) of the generated photocurrent determines the overall system performance to a large extend. Assuming an unmodulated IF carrier and

\(^4\)Coherent optical communication can of course also be used in combination with the analogue modulation schemes AM, FM and PM. However, in most system applications to be considered in this thesis, analogue modulation is of considerably inferior performance compared to digital keying, since it requires (depending upon the modulation method) much higher signal-to-noise ratios. Especially certain laser characteristics such as linewidth and RIN - to be discussed in sections 2.2.1 and 2.2.3 - make it difficult to obtain these high SNR-values. Analogue modulation will therefore not be treated in the analysis.
1.2. SIGNAL-TO-NOISE RATIO

Optimum polarisation alignment ($\Phi=0$), the signal power on a 1 $\Omega$ basis\(^5\) can be obtained directly from formula (1-15) as

$$S = \frac{1}{2} \cdot 4D^2 E_s^2 E_{lo}^2$$

(1-16)

At this point in the receiver two noise sources contribute to the total noise. The first is 'classical' receiver noise that also determines the SNR in direct detection systems. It will be denoted by $n_{rec}$. An extra noise source compared to direct detection is the shot noise due to the large LO-power. The statistical photo detection process in the photodiode introduces an amount of shot noise proportional to the dc-current flowing through the diode:

$$n_{lo} = 2qI_{\text{photo}}B_{eff} = 2qDE_{lo}^2B_{eff}$$

(1-17)

Here $B_{eff}$ is the effective noise bandwidth of the receiver. Since the shot noise and receiver noise are uncorrelated and have expected value 0, they can be summed for obtaining the total noise power. For the moment both noise sources are assumed to be white. This means that the 'channel' (which includes the transmitting laser, the fibre and the photodiode) may be modelled as an Average White Gaussian Noise (AWGN) channel. The analysis of the transmission system can then proceed in accordance with the classical theory, and the SNR is given by the expression

$$SNR = \frac{2D^2 E_s^2 E_{lo}^2}{2qDE_{lo}^2B_{eff} + n_{rec}}$$

(1-18)

For large LO-power the receiver noise can be neglected and equation (1-18) reduces to

$$SNR = \frac{DE_s}{qB_{eff}} = \frac{RP_s}{qB_{eff}}$$

(1-19)

The receiver sensitivity is the optical signal power $P_s$ that is necessary to achieve the required SNR:

$$P_s = \frac{qB_{eff}SNR}{R}$$

(1-20)

Finally, with $R$ substituted by (1-8) and $B_{eff}$ equal to the signalling rate for digital transmission $1/T$, the SNR can also be written as

$$SNR = \frac{T}{\eta h \nu}$$

(1-21)

$$P_s = \frac{SNR h \nu}{\eta T}$$

(1-22)

\(^5\)Throughout the theoretical analysis in this thesis, all electrical (noise) powers will be given on a 1 $\Omega$ basis. Only in Chapter 8 the powers are on the conventional 50 $\Omega$ basis.
CHAPTER 1. COHERENT OPTICAL COMMUNICATION

Since the shot noise is the limiting noise source here, the sensitivity $P_s$ is often referred to as the shot noise limit (when the quantum efficiency $\eta$ is 100% and the required SNR defined). It gives the theoretical limit of the sensitivity, since both signal and noise are determined by the same two optical sources. Since $h\nu$ is the energy of one photon and $T$ the bittime, the sensitivity of (coherent) optical systems can be expressed in photons per bit.

Sensitivity penalties

From equations (1-18), (1-20) and (1-22) two important, practical sensitivity penalties can be derived. First, any photodiode with a quantum efficiency of less than 100% will reduce the SNR and increase the required signal power $P_s$. In fact, each dB reduction of $\eta$ and $R$ will give a sensitivity penalty of the same amount\(^6\). The second sensitivity penalty is caused by the receiver noise, and should in practical receivers never be neglected. From (1-18) the expression for the penalty due to the receiver noise is derived

$$\Delta SNR = \frac{n_{lo}}{n_{lo} + n_{rec}} = \frac{2qRP_{lo}}{2qRP_{lo} + n_{rec}} \quad (1-23)$$

This penalty can be determined rather easily, since both the LO-power, the responsivity and the receiver noise can be measured in most practical systems. As a result one can say that up to the generation of the IF currents in the photodiodes the fundamental penalties in a coherent receiver\(^7\) are well known and can be determined quantitatively. In the following chapters these optical penalties will be addressed only briefly, while the analysis will concentrate on the more complicated electrical penalties.

1.3 How coherent is coherent?

In this thesis the electrical part of coherent optical systems will be analysed. Classical electronic communication systems and (coherent) optical communication systems have much in common, as reflected by many identical terms. On the other hand this can cause a considerable amount of confusion, especially since the same term may cover totally different concepts. A good example of this is the word 'coherent' in coherent optical systems. This indicates that there is a coherence between the optical frequencies of two light sources, which is an enormous improvement compared to

---

\(^6\)Many publications use as received power $\eta P_s$ instead of $P_s$, thereby correcting for the penalty due to the quantum efficiency and losses in the optical components. In this case one also speaks of effectively detected photo-electrons per bit instead of photons per bit. However, in my opinion this is an unacceptable way of camouflaging the deficiencies of the optical part of the receiver, and I will consequently not use these units.

\(^7\)Other penalties due to for example dispersion, optical reflections and laser intensity noise, are not included here. In a well-designed system they should have no influence.
1.3. HOW COHERENT IS COHERENT?

the 'multiple-frequency' Fabry-Perot lasers. On the other hand, in electronics the
term coherent is reserved for the situation of two oscillators which are locked both
in frequency and in phase. Using this stricter definition, only those optical systems
using an Optical Phase-Lock Loop (OPPL) for demodulation would be classified as
coherent. The confusion becomes even larger due to the fact that in all coherent
optical systems that will be analysed, the electrical IF-detection scheme will be either
non-coherent or differentially coherent. It will therefore be explicitly indicated whether
the point at issue is coherent optical detection or (non)-coherent IF-detection.

It should not be forgotten that what is happening at present in the domain of
optical communication is an almost perfect analogy of what happened in radio in
the early twenties of this century. At that time carrier-wave telegraph transmitters
started replacing spark transmitters (see the picture of Figure 1-3), the electrical
counterpart of Fabry-Perot lasers. The simultaneous invention of the triode resulted
in the introduction of heterodyne reception for radio telephony (Figure 1-4). However,
this analogy does not mean that the electronic principles, let alone the performance,
of radio and optical communication systems are completely identical. It will be shown
that the stability of the optical laser oscillators has a large impact on the design of
coherent optical systems. The next chapter will consider the details of the IF electronics
in coherent optical receivers, especially where these differ from conventional receivers.
Figure 1-3: When direct detection needed even bigger machines: “Transmitting hall at Malabar. View of the two great arc transmitters of 2400 kW”. (From Radiostation Malabar en overige stations op de Bandoengsche hoogvlakte, Gouvernements Post - Telegraaf en Telefoondienst in Nederlandsch-Indië, Bandoeng, Juni 1928).
Figure 1-4: One of the first long-distance coherent radio telephony transmitters, constructed in the Philips Research Laboratories in 1926–1927. Using the call-sign PCJJ, starting from March 11, 1927, it reached the Netherlands East Indies during experimental transmissions at a wavelength of 30.92 m. It was the first system in the world to span distances in excess of 10,000 mile (16,000 km). (From S. Derks, J.S. Haneveer, F. van der Put (ed.), Philips Honderd 1891–1991, Europese Bibliotheek, Zaltbommel, 1991).
CHAPTER 1. COHERENT OPTICAL COMMUNICATION
Chapter 2

Review of system aspects

Introduction

Different concepts may be used in digital coherent optical receivers. Some of the most elementary methods for modulation, demodulation and optical detection are therefore summarised in this chapter. Aspects directly related to coherent optical detection, especially with respect to laser linewidth, intensity noise and polarisation handling, are treated in more detail.

Figure 2-1 shows the elementary block diagram of a digital\(^1\) coherent optical receiver. In an optical combining network the light from the local oscillator (LO) laser and the incoming received light are mixed in order to generate the intermediate frequency (IF). The combined light is detected by one or more photodiodes, giving optoelectronic conversion. The resulting photo currents are filtered and detected in the IF section, which is finally followed by the section for baseband signal regeneration. The output of the regenerator is the original digital data signal from the transmitter, corrupted by errors. The proportion of errors in the regenerated random data signal is defined as the Bit Error Ratio (BER). In practice, clock extraction using a Phase Lock Loop (PLL) and control of the receiver (such as gain-, frequency- and polarisation-control) are critical elements. However, these control circuits are assumed to be ideal, which means that they will have no influence on the theoretical analysis and can be omitted accordingly.

The analyses in the next chapters will always start from a given IF signal with known Signal-to-Noise Ratio (SNR). Therefore this chapter will briefly consider the

\(^1\)Concerning the modulation formats it should be stressed that only binary transmission will be considered. Multi-level coherent optical systems are in principle possible and have been reported, but the analysis of these options follows directly from the results of the binary system analysis and is therefore omitted.
Figure 2-1: Block diagram of an elementary coherent optical receiver.

noise sources that determine the IF SNR: amplitude noise in the receiver and phase noise as well as intensity noise of the lasers.

2.1 Noise sources

2.1.1 Electronic noise sources

The number of noise sources to be considered in the analytical analysis of a communication system can be reduced considerably by using the following theorem:

**Theorem 2.1** In a well designed receiver the ratio of the signal and noise spectral densities in a unity spectral band (under normal conditions this ratio is proportional to the signal-to-noise ratio) can only change at locations where the signal undergoes a non-linear operation or where the signal amplitude approaches an unavoidable noise floor.

In the coherent optical receivers under consideration there are two locations where a non-linear operation occurs: in the IF detector and in the decision circuit/regenerator. All other circuits are supposed to be linear, so the SNR does not seriously degrade as long as the gain of the circuits is much larger than one. The only exception to this and

---

I was not able to trace this theorem down to any particular individual. However, after several years of practical system analyses my colleague Mark Tomesen and myself have 'reinvented' it, although I suppose that it has been stated somewhere before.
2.1. NOISE SOURCES

Therefore the third critical operation) is the photo detection in the photodiode. Here the signal level must approach the noise level as much as possible while maintaining the required SNR. This low signal level is defined as the receiver sensitivity (1-20). In all other circuits of well-designed receivers the SNR degradation can be kept relatively small and be easily calculated\(^3\). Of the three critical operations mentioned above the noise at the photo detector will be analysed in the next section, since it determines the SNR in the IF BPF. The noise conversion by the electrical IF-detection process is the main subject of this thesis and will consequently be treated in detail in the chapters to follow. The third operation, in the decision circuit, is considered to be ideal. When the signal input to the decision switch is more than ten to twenty times the 'sensitivity' of the circuit\(^4\) this is a reasonable assumption in most practical systems.

2.1.2 Photo detection noise

As indicated in section 1.2 there are two important noise sources at the photo detector in coherent optical systems. The first is related to the large and constant optical LO-power that illuminates the photodiode. Since the photo detection of the incoming photons is a statistical process it introduces shot noise proportional to the dc-current in the photodiode, with a uniform one-sided power spectral density equal to

\[
n_{lo} = 2q \cdot I_{\text{photo}}
\]  

(2-1)

The unit of the noise power spectral density \(n_{lo}\) is \(A^2/\text{Hz}\), but by taking the root of this one obtains an effective noise current in \(A/\sqrt{\text{Hz}}\) or, as is commonly used, \(\text{pA}/\sqrt{\text{Hz}}\). Ideally, the LO-generated shot noise dominates over other noise sources at the receiver input. In practice however the LO-power is usually limited to several milli-Watt at the most, which means that other noise sources cannot be neglected. These other sources are usually combined into one equivalent input noise source parallel to the photodiode and the shot noise source, which makes evaluation of the input SNR very easy. In a properly designed receiver front end, only the noise sources of the first stage will contribute to the equivalent input noise. Noise from following stages should be 'isolated' from the input by a sufficiently high gain of the first stage. The equivalent input noise is obtained with the inverse transfer function of each noise source back to the photodiode. If the transfer function of the \(i^{\text{th}}\) white noise source in an input stage (e.g. the collector current shot noise of a bipolar transistor or the thermal noise of a feedback resistor) to the output of the front end is given by \(H_i(f)\) and the transfer

---

\(^3\) This is for example the case in differentiating equalisation stages in optical receiver front ends, where the loss of such a stage increases with the frequency range of the equalisation. In extreme cases (e.g. minimum received power) the noise of the stage following the equaliser can sometimes not be neglected. However, with a good low-noise design of this stage the SNR reduction should be kept to a minimum.

\(^4\) The sensitivity of a decision circuit is defined as the minimum noise free peak-to-peak input signal required for obtaining a predetermined BER of (usually) \(10^{-9}\). The noise that limits the sensitivity is the circuit noise of the internal bi-stable elements in the circuit.
function from the photodiode signal current to the front end output is $H_{fe}(f)$, then the $i^{th}$ equivalent input noise is defined by:

$$n_{i,eq}(f) = \left| \frac{H_i(f)}{H_{fe}(f)} \right|^2 \cdot n_i \quad (2-2)$$

This should be done for every non-negligible noise source, after which all uncorrelated equivalent input noise sources can be summed. Note that, although the original internal noise sources are usually white, the equivalent input noise will be frequency dependent. Using such techniques as noise tuning of the input circuit it is nevertheless possible to obtain nearly flat noise spectral densities in the frequency regions of interest. This in turn means that the AWGN model of the transmission channel can be used, as well as the associated classical theory on optimal signal detection.

Measuring the input noise sources

Both the LO generated shot noise and the equivalent receiver input noise have the advantage that they can be determined in a simple way. Shot noise can be calculated from the dc-current flowing through the illuminated photodiode. The receiver noise can be measured using a relative measurement method, where a known amount of light on the photodiode generates a known additional noise power.

The fact that both the 'wanted' LO shot noise and the 'unwanted' receiver noise can be determined, means that also the sensitivity penalty due to the reduced SNR can be calculated using equation (1-18). In fact this is about the only electronic penalty that can be calculated in a straightforward manner. Depending upon the receiver noise and the available LO-power, the sensitivity penalty in most coherent optical receivers lies between 1 and 6 dB.

2.2 Laser characteristics

2.2.1 Laser linewidth

Lasers are critical components in optical communication systems, especially in coherent optical systems. As already mentioned, only single-frequency lasers can be used in coherent optical systems due to the required coherence between the light from the transmitter and local oscillator sources. The most commonly used single-frequency source is the Distributed Feed-Back (DFB) laser. However, more advanced types such as multiple sections DFB's and Distributed Bragg Reflector (DBR) lasers are becoming available. Many parameters of the lasers in use directly influence the performance of the system in which they are used: emission wavelength, output power, amplitude and/or frequency modulation characteristics, bandwidth, tuning behaviour and sensitivity to optical reflections. In the following it will be assumed that all of
these characteristics are properly adapted for the system under consideration. Only two essential laser characteristics then have direct influence on the IF-detection mechanism: the phase/frequency noise and the amplitude noise. Both noise sources find their origin in the quantised spontaneous emission of the laser diode, which forms an essential part of the total transmitted power.

Laser phase noise

Laser phase/frequency noise is caused by the fact that spontaneous emission photons are not generated in phase with the stimulated emission photons, but with random phase. Furthermore, only the average rate at which these photons are generated is constant, the instantaneous number of spontaneous emission photons varies. In references [64, 90] detailed analyses of the mechanisms and statistics of laser phase noise can be found. Here I will only summarise the essential results that are of interest for the IF-detection.

Modifying equation (1-4) into an expression with a time-dependent phase \( \varphi(t) \) gives:

\[
E = E_\varphi \cos(\omega_\varphi t + \varphi(t))
\] (2-3)

The instantaneous phase and frequency fluctuations can be defined by

\[
\delta \varphi = \varphi - \langle \varphi \rangle \quad t
\]
\[
\delta \dot{\varphi} = \dot{\varphi} - \langle \dot{\varphi} \rangle
\] (2-4) (2-5)

where \( \delta \dot{\varphi} = (d/dt)\delta \varphi \) has zero mean and \( \langle \cdot \rangle \) indicates mean. The frequency noise \( \delta \dot{\varphi} \) can be modelled by the two-sided spectral density \( W_\varphi(\omega) \), which is in itself related to the autocorrelation function of \( \delta \dot{\varphi} \).

An important effect of the phase noise is the fact that the actual phase of the emitted light will exhibit random walk; after a delay \( \tau \), a random phase change \( \Delta \varphi \) has been accumulated:

\[
\Delta \varphi = \delta \varphi(t) - \delta \varphi(t - \tau)
\] (2-6)

Because \( \Delta \varphi \) is caused by many independent noise events (the generation of spontaneous emission photons) it can be shown through the central-limit theorem that \( \Delta \varphi \) has a Gaussian distribution.

In most cases it is valid to assume that the power spectral density of the frequency fluctuations is white. In that case the relation between the variance of the accumulated phase change and the spectral density is given by e.g. Petermann [90]:

\[
\langle \Delta \varphi^2 \rangle = W_\varphi(\omega) \mid \tau \mid = 2 \frac{\mid \tau \mid}{t_c}
\] (2-7)

The coherence time \( t_c \) is a measure for the stability of the laser. When \( t_c \) increases the corresponding noise spectral density will be lower. The coherence time also denotes
the maximum delay time \( \tau \) up to which two components of the optical field can stably interfere.

The power density spectrum of the electrical field can be derived by Fourier transformation of the autocorrelation function of the field vector, and yields \([90]\):

\[
W_E(\omega) = \frac{2t_c E_z^2}{1 + \{(\omega - \omega_s - \langle \phi \rangle)t_c\}^2}
\tag{2-8}
\]

i.e., a Lorentz-shaped spectrum centered at the frequency \( \omega_s + \langle \phi \rangle \). The linewidth is defined as the full width half maximum or \(-3\) dB bandwidth of the power density spectrum, given by

\[
\Delta \nu \triangleq \frac{1}{\pi t_c} = \frac{W_\phi}{2\pi}
\tag{2-9}
\]

### 2.2.2 Linewidth effects

The effects of the non-zero laser linewidth can be observed both in the time and frequency domain, as illustrated in figure 2-2. In the time domain the random walk of
the phase increases the probability that the phase of the optical signal (and thus of the IF after coherent optical detection) deviates from its starting value. Although the average phase deviation is 0, there is a finite probability that the deviation is large. In particular the tails of the pdf \( p(\Delta \phi, \tau) \) are of importance when the performance at low bit error rates of coherent optical receivers is analysed. In general the pdf \( p(\Delta \phi, \tau) \) is modelled as having a normal distribution with variance \( 2\pi \Delta \nu \tau \).

The effects in the frequency domain are of course the transform of the time domain effects. Again there is a small but finite probability that the frequency deviates much more than the actual linewidth from the central wavelength or frequency. Even the best DFB lasers can at present not attain linewidth values much lower than 1 MHz, which is extremely stable from the optical point of view (at 1.55 \( \mu \)m a quality factor of 2\( \times \)10\(^8\)) but not enough for optical phase recovery\(^5\).

### 2.2.3 Intensity noise

Apart from phase noise a semiconductor laser diode exhibits amplitude noise, often referred to as intensity noise. In principle the intensity noise is the optical equivalent of the electrical shot noise; the dc-current through the laser diode will generate white noise proportional to that current. In most cases however it is more convenient to write the intensity noise as being proportional to the optical output power. When the optical output power of a laser is written as

\[
P = \langle P \rangle + \delta P
\]

the Relative Intensity Noise (RIN) per unit bandwidth is defined by Petermann [90] by the ratio of the variance and the squared constant power:

\[
RIN = \frac{\langle \delta P^2 \rangle}{\langle P \rangle^2} = \frac{2W_p(\omega)}{\langle P \rangle^2}
\]

(2-11)

The spectral density \( W_p(\omega) \) in \( W^2/\text{Hz} \) is the optical analogon of the electrical shot noise spectral density \( q < I > \) and can thus be written as\(^6\) \( h\nu < P > \). When the optical power given by (2-10) is detected by a photodiode with responsivity \( R \) the photo current can be written as \( < I > + \delta I \). The first component will introduce a shot

---

\(^5\)To give an indication: using (2-9) and (2-7) one can easily calculate that with a 10 MHz linewidth and a bitrate of 1 Gbit/s the standard deviation of the phase within 1 bit is about 45\(^\circ\).

\(^6\)This is also the reason for the lower detection limit of the RIN; in order to be detected it must exceed the shot noise. So if the optical power \( < P > \) generates a current \( < I > \) the minimum detectable RIN is equal to \( 2q < I > / < I >^2 = 2q / < I > \). Typical RIN values are between \(-190\) and \(-130\) dB/Hz.

This brings us to the conclusion that the definition \( RIN = h\nu < P > \) is in principle not so good, since in this way we define the absolute minimum quantum level of the RIN, which will always be present. It would be better to define the RIN-degradation or RIN-increase in dB relative to this quantum level. This would also make this parameter power-independent.
noise $2q <I> = 2qR <P>$, while the second one gives a variance 
$<\delta I^2> = (R\delta P)^2 = R^2 <P>^2 \text{RIN}$. The expression for the SNR (1-18) can now be rewritten into

$$SNR = \frac{2R^2P_sP_{Io}}{2qRP_{Io}B_{eff} + \text{RIN} R^2P_{Io}^2B_{eff} + n_{rec}B_{eff}}$$

(2-12)

It is easy to show that this expression has its minimum value for the SNR at an LO-power equal to $\sqrt{n_{rec}/\text{RIN} R^2B_{eff}}$. For typical values of the receiver noise (10 pA/$\sqrt{\text{Hz}}$) and RIN (-150 dB/Hz), one obtains moderate LO-powers of 250 $\mu$W or -6 dBm. It is no use increasing the LO-power above this value since the SNR will only decrease (see also Figure 2-3). This problem is aggravated due to either optical reflections or when the laser does not exhibit single-mode behaviour, a consequence of which is mode-partitioning noise [90]. In that case, the optical noise spectral density will be larger than the quantum limit of $h\nu <P>$ as given above, so the RIN also increases.

Cancellation of the RIN

Fortunately it is not difficult to reduce the influence of the RIN. In all practical coherent optical receivers, the received light and the light from the LO will be combined in an
2.2. LASER CHARACTERISTICS

Figure 2-4: Principle of a balanced optical receiver for suppression of the RIN.

optical coupler, usually with splitting ratio 50:50. The generated IF signals in the two output branches are in antiphase, while the RIN signals are in phase in both branches (the RIN is not generated by coherent detection, like the IF signal, but comes from one of the sources and is simply split over both outputs). Use of a balanced optical receiver in which the photo currents from the two coupler output branches are subtracted, see Figure 2-4, can reduce the RIN considerably. If the receiver has a Common Mode Rejection Ratio (CMRR)\(^7\) of 20 dB (which is a practical value [37]) the optimum SNR will be attained at a 10 dB higher LO-power (see also Figure 2-3). The total receiver penalty due to the RIN and the receiver noise can now be written as a combination of (1-23) and (2-12). Since a balanced front end uses two photodiodes, the amount of LO-induced shot noise is doubled. Defining \(P_{LO,d}\) as the optical power per photodiode, the total sensitivity penalty can be written as\(^8\)

\[
\Delta SNR = \frac{2n_{lo}}{2n_{lo} + \frac{n_{RIN}}{CMRR} + n_{rec}}
\]

\(^7\)Here, the CMRR is defined as

\[
CMRR = \frac{i_A^2 + i_B^2}{(i_A - i_B)^2}
\]

where \(i_A\) and \(i_B\) are the photo currents in the two diodes of a balanced front end. The minimum value of the CMRR is 0.5 (-3 dB), when the two currents are in anti-phase.

\(^8\)Here it is assumed that each photodiode is connected to a separate receiver input, with its own receiver noise. When single-chip dual-PIN photodiodes are used only one receiver input is required, and in (2-14) one should replace \(2n_{rec}\) by \(n_{rec}\). The use of one single receiver does not necessarily mean that the equivalent input noise reduces, since the increased capacitance of the dual-PIN can easily compensate the gain in (thermal) receiver noise.
\[ \Delta \text{SNR} = \frac{4qRP_{LO,d}}{4qRP_{LO,d} + \frac{\text{RIN}}{\text{CMRR}} R^2 P_{LO,d}^2 + 2n_{rec}} \]  

(2-14)

A balanced front end gives both optimal detection of the generated IF signals generated in the 50:50 optical coupler, and maximum suppression of the RIN, and is therefore an essential part of all high-performance coherent optical receivers.

### 2.3 Digital modulation and detection

#### 2.3.1 Amplitude Shift Keying

Amplitude Shift Keying (ASK) is the simplest digital modulation scheme. An optical ASK signal can be generated by switching on and off a laser diode\(^9\). Maximum sensitivity is obtained using coherent IF-detection, yielding [98, 106]:

\[ P_e = \frac{1}{2} \text{erfc} \left( \frac{\sqrt{\rho}}{2} \right) \]  

(2-15)

\[ \approx \frac{1}{\sqrt{\pi \rho}} \exp \left( -\frac{\rho}{4} \right) \quad (\rho \gg 1) \]  

(2-16)

where \( P_e \) is the overall BER (in the case of equal Mark and Space probability) and \( \rho \) the SNR. For a BER of \( 10^{-9} \) a SNR \( \rho \) of 72.05 (or 18.58 dB) is required. The theoretical performance of ideal ASK receivers with non-coherent IF-detection (a bandpass filter centered at the IF carrier) is given as [98]

\[ P_e = \frac{1}{2} \exp \left( -\frac{\rho}{4} \right) \]  

(2-17)

For a BER of \( 10^{-9} \) a SNR of 80.12 (or 19.04 dB) is required, which in combination with (1-22) leads to an ASK shot noise limit of 80 photons/bit for a Mark. The average SNR in the case of equal Mark and Space probability is 40.06. The sensitivity penalty of non-coherent detection relative to ideal coherent IF-detection is thus 0.46 dB at a BER of \( 10^{-9} \).

#### 2.3.2 Frequency Shift Keying

Wide deviation FSK

---

\(^9\)The on and off switching of a laser diode introduces \textit{wavelength chirp}, i.e. when the current is reduced close to or below the threshold current of the laser the optical frequency exhibits a large variation. The resulting IF signal is then not a properly switched carrier (as schematically depicted in Figure 2-5) but exhibits a large frequency spread. This is especially true in the case of closely spaced optical frequency-multiplexed multichannel transmission, where this may cause interchannel interference. Therefore this scheme is rarely used in its pure form.
2.3. DIGITAL MODULATION AND DETECTION

Figure 2-5: Illustration of the different digital modulation schemes to be considered. At left typical IF-signals, at right the corresponding phasor diagrams. Crosses indicate a phase state at a sampling moment \( k \cdot T \) for a Space (S) and Mark (M). FSK is given for two different frequency modulation indexes \( M \), where \( M = 0.5 \) is a form of CPFSK often referred to as MSK.

Frequency Shift Keying (FSK) is the most widely used modulation scheme in coherent optical communication. The reason for this is the possibility to apply direct laser modulation, without the need for external modulators. The modulation index \( M \) of an FSK signal is defined as the peak-to-peak frequency deviation \( \Delta f \) relative to the bitrate:

\[
M = \frac{\Delta f}{1/T} = 2f_d T
\]  
(2-18)

\( f_d \) is the carrier-to-peak frequency deviation.
As with ASK detection, optimum performance is obtained using coherent IF detection. The corresponding expression for the BER is given by [98, 106] as

\[
P_e = \frac{1}{2} \text{erfc} \left( \sqrt{\frac{\rho}{2}} \right)
\]

\[
\approx \frac{1}{\sqrt{2\pi\rho}} \exp \left( -\frac{\rho}{2} \right) \quad (\rho \gg 1)
\]

yielding a SNR at \( P_e = 10^{-9} \) of 36.02 (15.57 dB).

FSK can be used in two different ways; as **wide deviation FSK** or as **Continuous Phase FSK (CPFSK)**. The first method uses a large frequency deviation in order to create a distinct separation between the *Mark* and *Space* signal peaks. This makes it possible to detect the two signals separately with two bandpass filters centered at the corresponding signal peaks. In fact this is equivalent to double ASK detection, both for the *Mark* and the *Space*. Since two times more signal and noise are detected the SNR improves by a factor of two as well (the signals add coherently, while the noise powers are summed non-coherently). The ideal performance of an FSK receiver with non-coherent IF-detection is thus given by the expression

\[
P_e = \frac{1}{2} \exp \left( -\frac{\rho}{2} \right)
\]

As expected, the required SNR for a BER of \( 10^{-9} \) is \( \rho = 40.06 \) (16.03 dB), leading to a shot noise limit of 40 photons/bit. The average power requirements for ASK and wide deviation FSK are thus the same, but the peak sensitivity is a factor of 2 (3 dB) better.

The disadvantage of wide deviation FSK is of course the broad IF spectrum. With a peak-to-peak frequency deviation of \( 2f_d \) and a bitrate of \( R \) the minimum required IF bandwidth for distortion-free reception is \( 2f_d + R \). In practical receivers the bandpass filters will be even broader, leading to a very large IF bandwidth. This is the reason why wide deviation FSK signals are often detected with one single filter, giving a receiver performance equal to ASK. However, with FSK modulation and single filter detection the laser modulation is much smoother than for pure ASK modulation, resulting in a more stable IF spectrum.

**CPFSK**

When the modulation index \( M \) is reduced to values below 2 the spectra centered around the two signal peaks overlap considerably. At \( M = 1 \) the two spectra have merged into one compact spectrum, from which it is not possible to extract the original modulating baseband signal by using two filters. One therefore needs delay-line discriminators as frequency detectors, which have the additional advantage that
2.3. **DIGITAL MODULATION AND DETECTION**

they not only use information contained by the frequency but also by the phase\(^\text{10}\). In fact the discriminator ‘compares’ the phases of the IF signal at times \(t\) and \(t - \tau\), where \(\tau\) is the delay time. For small modulation indexes, CPFSK thus becomes a form of phase modulation, of which the ultimate performance is given by [14, 98, 106]

\[
P_r = \frac{1}{2} \exp(-\rho)
\]

(2.22)

The required SNR for a BER of \(10^{-9}\) is \(\rho=20.03\) (13.02 dB), which is in turn 3 dB better than non-coherent FSK. From now on I will refer to IF-detection using delay-line discriminators as *differentially coherent IF-detection*.

In principle CPFSK is a subset of FSK, namely having phases without discontinuities\(^\text{11}\). Although narrow deviation frequency modulation with differentially coherent IF-detection requires CPFSK (for \(M \leq 2\) dual-filter IF-detection becomes impossible), continuous phase FSK and/or delay-line discrimination can also be used for large modulation indexes.

The one limiting form of FSK and CPFSK is obtained for \(M=0.5\), often called *Minimum Shift Keying (MSK)*. The phasor diagram in Figure 2-5 shows that the phase change accumulated during one bit period is only 90°, or differentially 180° between a *Mark* and a *Space*. This means that the difference between a *Mark* and *Space* is only half an oscillation in the IF, which implies that the frequency (and phase) stability must be extremely high. With decreasing \(M\) the bandwidth of a CPFSK signal reduces as well, obtaining a minimum for \(M=0.5\). However, MSK is not necessarily the ideal solution, especially when pulse distortion is taken into account. Kazovsky and Jacobson [67] have shown that the increasing sidelobe-level associated with low modulation indexes requires a larger IF bandwidth than obtained with Carson's rule \((BT=2+M)\). For \(M=0.5\) they found a required IF bandwidth of 3.7 times the bitrate, while a minimum of \(2.2/T\) was obtained for \(M=1.0\). This is the reason that MSK is rarely used in coherent optical communication, while most practical CPFSK systems use modulation indexes of 0.7–1.0.

### 2.3.3 Phase Shift Keying

**Pure PSK**

Phase Shift Keying is the most sensitive of the three elementary binary modulation schemes. For optimum demodulation the phase of the received IF is compared with the phase of a reference source. This source is usually locked to the received signal by means of a PLL and this detection scheme is therefore referred to as *coherent* or

---

\(^{10}\) This is a first intuitive explanation. A better explanation will be given in Chapter 7 using the results of the theoretical analysis. For the moment the intuitive approach suffices.

\(^{11}\) In theory an FSK signal can be generated by switching between two frequency sources that have no phase relation. This yields an FSK signal with - assuming instantaneous switching - random phase steps at bit transitions.
synchronous IF detection [98, 106]. The ultimate performance of a binary (PSK) communication system with coherent IF-detection is given by

$$P_e = \frac{1}{2} \text{erfc}(\sqrt{\rho})$$

(2-23)

$$\approx \frac{1}{2\sqrt{\pi \rho}} \exp(-\rho) \quad (\rho \gg 1)$$

(2-24)

For a BER of $10^{-9}$ an SNR $\rho$ of 18 (12.55 dB) is required, being equivalent to a sensitivity of 18 photons/bit.

The obvious advantage of PSK, high sensitivity, is countered by many practical problems involved with this modulation scheme. Apart from the complicated PLL it requires external optical phase modulators, since a semiconductor laser can normally not be properly phase-modulated\textsuperscript{12}. These modulators are difficult to make,\textsuperscript{13} require high driving voltages and exhibit high insertion losses. Furthermore the phase stability of the carriers in PSK systems must be extremely high, which means that narrow linewidth lasers should be used. Again this gives many practical problems, for example the required use of mechanically complicated and bulky external-cavity lasers. This makes that pure PSK systems are still in a very experimental stage.

DPSK

A way of overcoming the problems associated with pure PSK, while at the same time exploiting the advantages of phase modulation is Differential Phase Shift Keying. This requires that in the transmitter the original baseband signal is differentially encoded. The usual coding convention is that an original Space results in a change of state between two successive bits, while an original Mark gives no change. For demodulation in the receiver only the phase of two successive channel-bits then has to be compared. At the high bitrates involved in the systems considered, this requires a delay-line discriminator with a delay equal to the bittime $T$. At low bitrates optimum detection can be achieved using matched filters [106].

The advantage of differentially detected PSK is the much larger phase noise tolerance compared to coherent PSK detection; this will be elaborated at the end of this chapter. The ultimate performance of DPSK is identical to that of CPFSK, which also uses differentially coherent IF-detection, and is given by

$$P_e = \frac{1}{2} \exp(-\rho)$$

(2-25)

---

\textsuperscript{12}The feasibility of phase modulation of a laser is determined by the intrinsic phase stability of this source, being the linewidth as defined in section 2.2.1. At moderate bitrate and using lasers with very narrow linewidth it becomes possible to apply direct phase modulation to the laser [101].

\textsuperscript{13}Almost all available devices are made in Lithium-Niobate LiNbO$_3$, a complicated optical crystal. Required driving voltages are usually in the order of 8-15 V, which is very unpractical at high bitrates. Another problem is the high temperature dependence of the optical properties.
2.4. OPTICAL DETECTION SCHEMES

The required SNR $\rho$ for a BER of $10^{-9}$ is 20.03 (13.02 dB), giving a shot noise limit of 20 photons/bit. This shows that the sensitivity of DPSK (and also of CPFSK) is inferior to PSK by a factor $\frac{18}{20}$ or 0.46 dB, due to the differentially coherent IF-detection scheme.

2.4 Optical detection schemes

2.4.1 Heterodyne detection

Heterodyne detection is the most general form of coherent optical detection. As in classical radio it means that the intermediate frequency is not taken equal to 0. The IF current in this case is given by (1-15), which can also be written in the modified form (assuming optimum polarisation adjustment):

$$i = R \left\{ P_{io} + P_s + 2\sqrt{P_{io}P_s} \cos(\omega_IF t + \theta_IF) \right\}$$

(2-26)

The intermediate frequency $\omega_IF$ is the absolute value of the difference between $\omega_{io}$ and $\omega_s$, and can be generated with the LO-frequency either above or below the signal frequency (see Figure 2-6). However, when switching from upper to lower LO mixing,

![Diagram](image-url)

Figure 2-6: Generation of a heterodyne IF signal in a multi-channel environment. The upper diagram shows the channels in the optical frequency domain, the lower diagram in the (IF) electrical domain after heterodyning with $\omega_{LO}$. The dashed channel in the upper diagram is the image channel of selected channel #j and would give, when present, an overlapping IF spectrum.
the form of the IF spectrum remains the same\textsuperscript{14} but the baseband spectrum after demodulation (and thus the information content) will be inverted. This may cause problems with the clock extraction and/or data regeneration in practical receivers, while the stability of an Automatic Frequency Control (AFC) loop may be affected as well. Therefore either of the two methods of heterodyne IF-generation has to be selected.

A second down-mixing problem associated with heterodyne detection is image-rejection. In order to avoid unwanted detection of an image channel in a multi-channel environment (e.g. channel \#i in Figure 2-6) it is necessary to keep the channel spacing in standard heterodyne receivers relatively high\textsuperscript{15}. This reduces the optical bandwidth efficiency and means that fewer channels can be received with a given tuning range of the LO. The problem aggravates for increasing IF and receiver bandwidth, which leads to the third and most important drawback of heterodyne detection.

At high bitrates (e.g. more than 1 Gbit/s) heterodyne detection requires a very large IF and receiver bandwidth. The problem becomes even bigger when also the IF and demodulated baseband spectrum are not allowed to overlap in order to avoid crosstalk. In particular in combination with FSK this leads to extremely high bandwidth receivers, relative to the bitrate. When the bandwidth efficiency of a receiver is defined as the ratio of the bitrate and the electrical receiver upper $-3$ dB band edge, this yields for an FSK heterodyne receiver approximately $1/(1+M+m)$. For heterodyne CPFSK receivers the bandwidth efficiency is usually of the order $1/(2+M+1)$, around $\frac{1}{3}$. Examples are a 300-2300 MHz bandwidth in a biphase-coded 140 Mbit/s (280 Mbaud) FSK system [39] and a 3-8 GHz IF band in 2.5 Gbit/s CPFSK systems [10], with bandwidth efficiencies of 1/16.4 (or 6.1%) and 1/3.2 (or 31%), respectively. Despite the drawbacks mentioned above, heterodyne detection is in most cases the preferred scheme due to its simplicity of operation. This will become clear when comparing it with other detection schemes, especially homodyne detection.

\subsection{2.4.2 Homodyne detection}

In the case of homodyne detection the optical frequencies of the LO and the received signal are identical, leading to a nominal IF of zero. The photo current (2-26) can then be modified into

\[ i = \mathcal{R} \left\{ P_\text{lo} + P_s + 2\sqrt{P_\text{lo}P_s} \cos \theta_{IF} \right\} \quad (2-27) \]

For PSK demodulation an Optical PLL (OPLL) is required, which is schematically depicted in Figure 2-7. In this case the optical coupler/front end-combination not

\textsuperscript{14}The envelop of the IF spectral density remains the same. However, since the IF signal is complex it can be derived that the envelop reamins the same but that the phase is inverted.

\textsuperscript{15}It is possible to construct optical image-rejection receivers [74], although at the cost of impractically high IF-values. Since the electronics are completely identical to those of normal heterodyne receivers, image-rejection receivers are not treated separately here.
only performs frequency down-conversion from the optical to the electrical domain, but also acts as an optical phase detector. Depending upon the transmitted bit the IF voltage in (2-27) is either $+2\sqrt{P_0 P_s}$ or $-2\sqrt{P_0 P_s}$, which means that the IF signal is equal to the desired baseband signal. No further IF filtering and detection are thus required, but it is a prerequisite that both the frequency and the phase of the LO can be accurately controlled. This is the reason why homodyne OPLL demodulation will not be analysed in this thesis; the detection mechanism is completely different compared to the (less coherent) systems using IF-detection.

![Figure 2-7: Schematic diagram of an Optical Phase Lock Loop. The loop contains branches for both frequency and phase control. The photodiodes, which act like an optical phase detector, already give the desired baseband signal.](image)

Since the photodetector output is the desired baseband signal, according to Nyquist's criterion [3] a minimum lowpass bandwidth equal to half the bitrate is required. This is a factor of two lower than the heterodyne IF bandwidth, which means that the sensitivity will increase by the same factor of two. PSK homodyne is therefore the ultimate demodulation method in (optical) communication, with a performance given by the expression

$$P_e = \frac{1}{2} \text{erfc}\left(\sqrt{2\rho}\right)$$

$$\approx \frac{1}{2\sqrt{2\pi\rho}} \exp(-2\rho) \quad (\rho \gg 1)$$

For a BER of $10^{-9}$ it yields the absolute shot noise limit in optical communication of 9 photons per bit.
Theoretically, ASK can be demodulated using an OPLL as well, but in practice this is a rarely used option. In principle it is more convenient to use a bandpass filter for demodulation, in analogy with the heterodyne case. Since this filter should be centered around 0 it becomes a lowpass filter, giving an improvement by a factor 2 compared to (2-17). However, in ASK the actual phases of the received signal and LO are not controlled, which means that \( \theta_{IF} \) in (2-27) can take any value between 0 and \( 2\pi \) with equal probability. So, independent of polarisation control and modulation content, the IF signal can fade to 0 due to, for example, phase noise.

2.4.3 Phase diversity

ASK and DPSK

Fading of the baseband signal, when ASK or DPSK are used in combination with homodyne reception, can be eliminated using phase diversity. In phase diversity receivers, the IF power is divided over multiple branches with different instantaneous IF phases. Because the total IF power within the receiver remains constant, it is possible to detect the modulation content under all phase conditions. Practical phase diversity systems can have two, three or four branches, while both ASK and DPSK modulation formats can be used. Since the analysis of these different options is almost identical, a three-branch ASK system will be evaluated as an illustrative example.

The symmetrical four-port or 2x2 coupler has already been introduced in section 2.2.3, in that case for the purpose of RIN suppression. The IF signals in both

![Figure 2-8: Principle of a three branch ASK phase diversity receiver using a 3x3 symmetrical 6-port optical coupler.](image-url)
2.4. **OPTICAL DETECTION SCHEMES**

output branches of a 2x2 coupler are of the same amplitude but in antiphase (a result of the power conservation law applied to this four-port). Similarly, one can easily show that the IF outputs of a symmetrical six-port 3x3 coupler will be of equal amplitude with 120° phase differences\(^{16}\). Taking one of the branches as the arbitrary reference the IF components of the three branches are given by

\[
i_A = 2R \sqrt{\frac{2P_{lo}P_s}{3}} \cos(\theta_{IF})
\]

\[
i_B = 2R \sqrt{\frac{2P_{lo}P_s}{3}} \cos(\theta_{IF} + \frac{2\pi}{3})
\]

\[
i_C = 2R \sqrt{\frac{2P_{lo}P_s}{3}} \cos(\theta_{IF} - \frac{2\pi}{3})
\]

These three IF signals have to be detected by square-law IF detectors, the combined outputs of which then becomes phase-independent because

\[
\cos^2(\theta_{IF}) + \cos^2 \left( \theta_{IF} + \frac{2\pi}{3} \right) + \cos^2 \left( \theta_{IF} - \frac{2\pi}{3} \right) = \frac{3}{2}
\]

This equality also represents the sum of the signal powers in the three branches and thus shows that the total IF power is indeed phase-independent. In this way all phase effects due to the linewidth can be cancelled while at the same time homodyne reception becomes possible. However, this is achieved at the expense of two additional IF branches, giving increased penalties. Another penalty is due to the LO-power, which is divided over three instead of two photodiodes (as in a standard receiver with a 2x2 coupler). This gives reduced shot noise in the photodiodes and a sensitivity reduction according to (1-23). In Chapter 6 it will be shown that the IF-detection penalties also increase.

**FSK**

FSK can also be combined with phase diversity, although the receiver becomes more complicated due to the frequency deviation of an FSK signal. The average IF of a received FSK-modulated signal can be made equal to 0, but this does not give homodyne detection because the peak IF will be located at the positive and negative frequencies \(\pm f_d\). Due to spectral folding of the negative frequencies the Mark and Space peaks of the IF power density spectrum will be at the same positive frequency \(f_d\). FSK phase diversity reception therefore requires a frequency discriminator capable of distinguishing between positive and negative intermediate frequencies.

\(^{16}\)In two-branch phase diversity receivers the phase difference between the two branches should be 90°. These are usually referred to as In-Phase (I) and Quadrature (Q) IF signals, which should not be confused with the in-phase and quadrature components of a complex phasor. The optical unit that generates these IF signals is often called optical 90° hybrid. A simple way of obtaining I and Q signals is by subtracting branch B and C of the 3x3 coupler presented above.
The principle of FSK phase diversity receivers is illustrated in Figure 2-9. The discriminator must have a sine-like characteristic, which can be obtained by multiplying a cosine and a delayed sine. It therefore requires In-phase (I) and Quadrature (Q) 90° out-of-phase IF signals, which must be detected by a delay-and-cross-multiply frequency discriminator. The characteristic of such a discriminator is also illustrated in the same figure. An additional advantage of this IF-detection scheme is that the second harmonics of $f_d$, generated in each of the two multipliers, are in anti-phase and cancel when subtracted. This gives low levels of IF-to-baseband crosstalk, which is very important since in principle the spectra of IF- and baseband signals overlap to a large extent. Although not literally homodyne (but ‘quasi-homodyne’), this demodulation scheme has the same advantage of minimum required IF bandwidth compared to heterodyne detection. This may be especially advantageous at high bitrates.
2.5 Polarisation handling

2.5.1 Control

The general expression for the generated IF signal in a coherent optical receiver as derived in section 1.1.2 contained the term \( \cos \Phi \), with \( \Phi \) the polarisation misalignment angle between signal and LO electrical fields. This misalignment should be 0 for maximum IF voltage. The most obvious way to accomplish this is direct control of the polarisation, with maximum IF as the control criterion.

However, polarisation control is rarely used due to the fact that good\(^{17}\) controllers are extremely difficult to make. The principle reason for this is the double-breaking of a fibre or other optically transparent material that is required for polarisation adjustments. Effectively, polarisation control means that the electrical field along one of the polarisation axes is forced to propagate faster or slower than the field along the other axis, thus compensating the natural birefringence in the glass-fibre trajectory. Since single-mode fibres under normal conditions are extremely symmetrical this requires unconventional control methods such as rotating, pulling or squeezing the fibre. Apart from obvious mechanical problems, such control methods have the disadvantage of a limited control range, so additional controllers and reset operations are needed in order to guarantee 'endless' polarisation control. Standard control units therefore contain at least four control elements, of which e.g. two are actively controlling and two are used for resets.

2.5.2 Diversity

A way of overcoming the problems associated with active polarisation control is polarisation diversity. As with phase diversity, the total IF signal will be split over multiple branches, depending upon the effective polarisation split angle \( \phi \). This angle is determined by the state of polarisation (SOP) of the received signal and can attain any value between 0 and \( 2\pi \). The idea behind polarisation diversity is to split both the received signal and the LO signal along two orthogonal polarisation axes, using polarisation beam splitters. The signal and LO outputs with identical polarisations (either parallel \( \parallel \) or perpendicular \( \perp \)) are then combined and received independently. In order to have equal LO-powers in both branches (for identical noise behaviour of the two front ends) the LO polarisation should be split evenly over both orthogonal polarisations.

\(^{17}\) By 'good' the following characteristics are meant: low insertion loss (<1 dB fibre-to-fibre), sufficient control bandwidth (>200 Hz) and compact size, including all control hardware (e.g. the size of one Euro-board).
Figure 2.10: Block diagram of a polarisation diversity receiver. PBS are identical polarisation beam splitters. This method can be used in combination with all modulation formats. For a polarisation independent baseband signal the IF detector outputs should be quadratically proportional to the IF amplitudes.

Depending upon the SOP of the received signal the IF currents of the two branches are then given by

\[ i_\parallel = \sqrt{2}E_{10}E_s \cos \phi \cos (\omega_{IF}t + \theta_{IF}) \]  \hspace{1cm} (2-34)

\[ i_\perp = \sqrt{2}E_{10}E_s \sin \phi \cos (\omega_{IF}t + \theta_{IF}) \]  \hspace{1cm} (2-35)

From (2-34)-(2-35) it is clear that the sum of the time-averaged IF powers in both branches is polarisation independent. The easiest way of obtaining a polarisation-independent baseband signal is therefore IF power detection, with a perfectly quadratic IF-detection characteristic. However, the latter requirement can be severe for practical broadband circuits, so only a few systems using this solution have been reported [109, 110]. When IF detectors with a non-ideal (e.g. linear) characteristic are used,
the sum of the detector outputs can not be polarisation independent \((\cos \phi + \sin \phi)\). These baseband outputs should then be corrected for the actual IF power ratios by multiplying them with \(\cos \phi\) and \(\sin \phi\), respectively. Such a feed-forward combining circuit is called a ratio combiner. This is a commonly used method since it offers the possibility to correct non-ideal IF-detection characteristics.

A general drawback of polarisation diversity is the reduced receiver sensitivity compared to a single branch polarisation control system. In principle there are three additional sensitivity penalties, the first of which is caused by the excess loss of the optical diversity circuit.\(^{18}\) Since the LO signal is attenuated as well, the receiver noise penalty (1-23) increases, which is aggravated by the fact that the LO power is split over four instead of two photodiodes. Finally the receiver has one additional IF branch, giving increased IF-detection penalties which will be evaluated in Chapter 6. Generally this means that practical polarisation diversity receivers are at least 3 dB less sensitive compared to polarisation-controlled receivers. However, due to the absence of good controllers diversity is in most cases the only feasible solution.

### 2.6 System qualification

In this final section of Chapter 2 a system comparison is given from the point of view of IF-detection. The three essential categories are coherent, differentially coherent and non-coherent IF-detection, each of which will be summarised.

#### 2.6.1 Coherent IF-detection

Optimum detection can be achieved when the received signal can be compared with an exact replica of the transmitted carrier. For that purpose an LO-signal should be generated in the receiver with both the proper frequency and phase. This means that there will be phase and frequency coherence between the received signal and LO, giving coherent or synchronous IF-detection. Since this is a one-step detection process no real IF-detection is required; the desired baseband signal is obtained in the first step.

Coherent IF-detection requires extremely good phase coherence between received signal and LO. When using semiconductor lasers this can only be achieved with an (optical) PLL and very narrow linewidth (e.g. external cavity) lasers. It is above all the complexity of these small-linewidth lasers that makes receivers with coherent IF-detection unattractive at present.

The reason why I shall not analyse this type of IF-detection is however of a more fundamental nature. In the following chapters it will be shown that the combination of

\(^{18}\)The sensitivity of a polarisation diversity receiver is defined at the input of the signal PBS. In general the sensitivity should be defined at the location where the received signal enters the complete receiver unit. Sensitivities measured on the photodiodes or as detected photo-electrons are therefore highly misleading and are often a disguise for a high-loss diversity unit.
lasers with a non-negligible linewidth and non-coherent IF-detection introduces receiver sensitivity penalties, which are only a function of the relative IF bandwidth at a given SNR. Systems with coherent IF-detection, on the other hand, are ideal from this point of view. In the phase detector the received signal (which contains both signal and noise) should ideally be mixed with a noise-free LO signal (see also Figure 2-11). After post-detection low-pass filtering, the baseband signal will always contain the same amount of noise, independent of the original (IF) noise bandwidth.

The IF characteristics thus only have a limited influence on the receiver performance. However, the phase instability of the LO will now introduce sensitivity penalties of a completely different nature.

2.6.2 Differentially coherent IF-detection

In differentially coherent systems the phase coherence between received signal and LO for one decision has only to be maintained over a period in the order of one bit-time. Receivers use delay-line discriminators as phase or frequency IF detectors. Due to the fact that the discriminators exploit both phase and frequency information of the detected IF, the sensitivity of DPSK and CPFSK systems is the highest of the coherent optical systems without coherent IF-detection. However, compared with the ‘ideal’ coherent IF-detection in Figure 2-11, the noisy received signal will no longer be mixed with a ‘clean’ LO but with an equally noisy delayed IF signal. This introduces additional system penalties. Also, both for DPSK and CPFSK the required IF bandwidth must be much larger than the bitrate in order to reduce distortion.
DPSK requires a delay equal to one bit period, CPFSK receivers have a delay determined by the frequency deviation or modulation index. This means that lasers can be used with more relaxed linewidth requirements compared to systems with coherent IF-detection. CPFSK systems are therefore becoming more important, although presently the laser requirements can only be met for bitrates above 1 Gbit/s.

2.6.3 Non-coherent IF-detection

This is the oldest generation of coherent optical systems, due to the modest linewidth requirements of receivers with filter detection. The combination of direct (FSK) modulation of the laser and simple filter detection in the receiver made that most of the first generation coherent optical systems had this IF-detection mechanism. Although high-bitrate systems increasingly make use of CPFSK, at lower bitrates non-coherent IF-detection will in many cases remain the most 'economical' solution with regard to laser requirements and receiver complexity. However, in all cases the acceptable receiver sensitivity penalty is low, making it essential to have exact knowledge of the penalties associated with the IF-detection process.
Chapter 3

Non-coherent IF-detection

Introduction

This chapter introduces most of the tools that are to be used in the analysis of different types of coherent optical systems. As summarised in Chapter 2 one can distinguish three different types of coherent optical system, depending upon the IF-detection mechanism used. The first category to be analysed here is non-coherent IF-detection.

The aim of the analysis is to obtain a complete understanding of the IF-detection processes, and thus exact knowledge of the associated system penalties. The results should be fully analytical, without the need for complicated computations or simulations. However, at the same time the required accuracy is high - given the constraints of the model - since practical penalties of tenths of a dB should be determined. This excludes, for example, the extensive use of signal-to-noise ratios as they only give a limited insight into the average signal characteristics.

The best way of modeling the complex signals involved in IF-detection is by using accurate probability density functions (pdf), as this is the only method that fully utilises the stochastical properties of the signals. Although the evaluation of the pdf's can be fully accomplished by analytical means, some assumptions have to be made. Most of these are related to the idealisation of the circuitry and the signals. In order to separate the effects of optical and IF-detection all evaluations start with the same electrical IF signal, defined by its SNR. Consequently only IF-detection penalties will be analysed, since the optical detection penalties can be treated separately. When using idealised circuit models (e.g. rectangular bandpass filters) the validity and limitations of these assumptions will be indicated whenever possible.

One important general aspect of IF-detection is the influence of the laser linewidth upon the receiver performance. This is one of the main differences between the analysis presented here and most others published so far. Section 3.6 summarises some of these
results, as well as the procedure I will use to eliminate the linewidth from the analysis. This makes it possible to obtain accurate analytical results without the need for the complicated numerical computations that are normally associated with the analysis of coherent optical systems. However, for most types of receiver a summary of the linewidth effects on the BER will be given. This will concentrate on the minimum required IF-configuration not causing linewidth effects.

3.1 Receiver model

3.1.1 Circuits

IF filtering

The receiver configuration to be used is illustrated in Figure 3-1. Since we adopted the AWGN model for the transmission channel, it can be shown that this configuration is optimal [35, 98, 106]. Relevant aspects of this model are detailed in the sections to follow. For reasons given in the introduction the analysis starts with IF signals, details of which will be given in the next section. After bandpass filtering, an IF signal $s(t)$ is obtained that is detected by the IF-detector. The bandpass filter (BPF) is modeled as a rectangular filter of width $2B$ around a central frequency $f_c$. In combination with white IF noise before the filter - another common assumption - this yields a normal pdf of the BPF output noise. However, for this reason the rectangular filter is not required, since any linear BPF gives Gaussian noise output statistics with white input noise. The advantage of the rectangular BPF is to be found in the simplified mathematics, especially in the calculation of variances and Fourier spectra. Obviously, truly rectangular filters can not be realised and would give unacceptably high phase distortion near the filter edges. In practice filters will therefore not be rectangular, but feature third to fifth order roll-off at the most. The noise bandwidth is then defined by

$$B_n = \frac{\int_0^\infty H(f)H^*(f)df}{2H(f_c)H^*(f_c)}$$  \(3-1\)

which is the weighted noise bandwidth. This is not necessarily identical to the bandwidth between the −3 dB frequencies of the filter. In practice, when filters with flat frequency characteristics are used, the approximation made by using these −3 dB frequencies is accurate enough\(^1\).

\(^1\)In many practical receivers the BPF will not be a single BPF in the classical filter sense, but two cascaded lowpass and highpass characteristics. Often the lowpass is due to the roll-off of the receiver front end. The highpass is usually chosen to have a −3 dB frequency at least higher than the bitrate in order to avoid IF-to-baseband crosstalk. Since these filters are independent, and the total IF bandwidth is in the order of 1 to 5 GHz, the effective IF bandpass characteristic is in many cases nearly rectangular.
Figure 3.1: The elementary configuration of a single-filter receiver with non-coherent IF-detection and idealised filter characteristics.

IF-detection means that the information imbedded in the IF-carrier is converted to the baseband frequency range. This requires some sort of non-linear operation, usually performed by a diode or transistor. Section 3.3 will discuss the IF-detection in more detail. The detector output will be the voltage $z(t)$.

Post-detection filtering

So far the definition of the receiver is similar to most ‘classical’ digital communication systems, e.g. satellite telemetry links or digital radio relays. In radio however many systems use the same unguided medium (‘ether’), forcing them to use the available bandwidth as economically as possible. Channel spacings are therefore usually very small, requiring narrow IF filters. This is the most important difference between ‘classical’ and coherent optical receivers, and is in fact the reason for the analysis presented here. As will be shown in section 3.6 the IF bandwidth of coherent optical receivers is often many times larger than the bitrate. The IF bandwidth relative to the bitrate $1/T$ will be defined by the IF bandwidth expansion factor $m$ with\(^2\)

$$m \triangleq \frac{B_{IF}}{1/T} = 2BT$$ (3-2)

Compared to digital radio receivers where Nyquist filtering takes place in the IF-section, the SNR will be much lower due to the increased noise bandwidth. A better

\(^2\)In practice $m$ can attain any value, but for simplicity the theoretical analysis will be restricted to integer values.
SNR is obtained with post-detection Nyquist filtering, an essential aspect of this type of receivers\textsuperscript{3}. Nyquist filtering is normally defined using Nyquist's first (sampling) criterion [3, 84, 85]:

**Theorem 3.1 (Nyquist)** If pulses are applied to a rectangular lowpass filter with bandwidth $B$, the responses can be observed independently at instants separated by $1/2B$.

In other words: for zero Inter-Symbol Interference (ISI) the maximum transmission rate through a rectangular LPF of with $B$ is $2B$. And for transmitted pulses of duration $T$ the minimum filter bandwidth for ISI-free reception is $1/2T$. A digital baseband signal that has been filtered in accordance with this theorem will be called Nyquist filtered. This filtering does not have to take place in the post-detection filter alone, as long as the overall filter characteristic is according to Nyquist. Filtering by the intermediate frequency BPF's, when symmetrical around the carrier frequency, may contribute to the Nyquist filter characteristic as well; in the limiting case all filtering takes place in the IF section. Nyquist's first criterion can thus be generalised for filtering in the IF as well:

**Theorem 3.2** For transmission of a binary signal sampled every $T$ seconds and modulated on a carrier with frequency $f_{IF}$, the minimum required IF bandwidth in order to avoid ISI is $1/T$ Hz from $f_{IF} - 1/2T$ to $f_{IF} + 1/2T$.

This will be called Nyquist filtering at IF\textsuperscript{4}.

The simplest post-detection filter is thus a rectangular lowpass with bandwidth $1/2T$, but from Nyquist's criteria other possible characteristics can be deduced as well (see section 3.4). Designing good Nyquist filters is extremely important for obtaining high-sensitivity receivers, since apart from the noise bandwidth also the phase and amplitude characteristics are critical [3, 112]. The filter should ideally compensate all signal distortion accumulated in the preceding system sections, to yield an output signal with maximum eye opening.

In the case of optimum post-detection filtering the filter output $u(t)$ has the highest SNR obtainable with the given IF configuration. At this point a decision circuit therefore decides whether the originally transmitted bit was a Mark or a Space. In order to obtain a time-synchronised output, decisions are made at regular time intervals $T$ using the regenerated clock signal with frequency $1/T$. This is the third non-linear operation in the receiver, after optical- and IF-detection; it is often referred to as signal regeneration. The output of the decision circuit is the originally transmitted

\textsuperscript{3}In classical digital receivers post-detection filters are used as well, but these are less critical and used mainly for the purpose of rejecting second and higher order harmonics resulting from the IF-detection process.

\textsuperscript{4}It should be stressed that this theorem is valid only if the superposition principle applies. This means that the spectrum of the IF-signal and the spectrum of the demodulated baseband signal are not allowed to overlap.
3.1. RECEIVER MODEL

digital data but degraded by a certain amount of random errors (a Mark instead of a Space or vice versa) characterised by the Bit-Error-Rate (BER). The decision process is determined by the pdf’s of the post-detection filter output signals. Relevant aspects of decision theory are summarised in section 3.5.

3.1.2 IF signals

Signal

The IF signal after bandpass filtering will comprise both the wanted signal containing the transmitted information, and noise. In order to distinguish between the information-carrying component and signals in general, the first will be referred to as Signal, i.e. capitalised. In its most general form it can be written as the phasor

\[ S(t) = a(t) \exp(j\omega_{IF}(t) + \theta_{IF}(t)) \]  

(3-3)

\( S(t) \) is a complex voltage with time-varying amplitude, frequency or phase due to either the ASK, FSK or PSK modulation formats. The modulation is assumed ideal, which means that changes of state due to a bit transition are instantaneous. During each bit period the amplitude and frequency or phase (depending upon the modulation method) are thus considered to be constant. The amplitude noise will be incorporated in the IF noise, while the phase/frequency noise due to the laser linewidth will be treated in detail in section 3.6.

Using these first assumptions the complex IF Signal voltage at any given moment \( t \) can effectively be written as a carrier with constant amplitude \( A \), frequency \( \omega_{IF} \) and phase \( \theta_{IF} \).

\[ S(t) = A \exp(j\omega_{IF}t + \theta_{IF}) \]  

(3-4)

It is assumed throughout the analysis that the receiver operates without Inter Symbol Interference (ISI). This means that at the sampling instances \( k \cdot T \) the amplitude of the IF Signal is either \( A \) or 0. Time responses of any filter characteristics within the transmitter-receiver path are thus not important as far as the Signal is concerned. Noise filtering by the IF and baseband filters remains the only factor that influences the SNR at the decision moments. It should be stressed that analyses taking into account real time responses of filters (and thus ISI) are of a completely different nature, since they are directly related to the actual system realisation. For the unified analysis introduced here the carrier representation is the best solution.

On many occasions it will be useful to have the one-sided representation of \( S(t) \), containing positive frequencies only. This will be called the waveform \( S(t) \), which is defined as

\[ S(t) = \text{Re} \{S(t)\} = A \cos \omega_{IF}t \]  

(3-5)
Figure 3-2: The phasor representation (left) and the one-sided power density spectrum (right) of the IF carrier Signal.

(For convenience $\theta_{IF}$ is set to 0 since the phase usually does not play a role in non-coherent detection). The one-sided power density spectrum of this signal is then given by

$$2G_S(f) = \frac{1}{2} A^2 \cdot \delta(f - f_{IF})$$

(3-6)

where $\delta(\cdot)$ is the Dirac delta-function. Using the spectral definition of the variance of a signal

$$\text{var}(S) = \int_0^\infty 2G_S(f) df$$

(3-7)

the variance is readily calculated to be $\frac{1}{2} \cdot A^2$, which is of course correct for a carrier with amplitude $A$. Both the phasor representation and power density spectrum are illustrated in Figure 3-2.

Noise

The IF noise considered here is assumed to be white and normally distributed over the frequency range of interest. This means that after IF bandpass filtering the resulting noise will be bandlimited between $f_{IF} - B$ and $f_{IF} + B$, with a uniform one-sided power spectral density $n$ (see Figure 3-3). The variance or total noise power can be calculated with (3-7) and will be equal to

$$\sigma^2 = \int_{f_{IF} - B}^{f_{IF} + B} 2G_n(f) df = 2B \cdot n$$

(3-8)
3.1. RECEIVER MODEL

Let the noise power within the Nyquist bandwidth be \( N \). With (3-8) \( N \) is related to \( n \) through the expression

\[
N \triangleq n \cdot 1/T
\]  
(3-9)

Combining this relation with (3-2) and (3-8) finally yields the total IF noise variance expressed in \( N \):

\[
\sigma^2 = 2BT \cdot \frac{n}{T} = m \cdot N \triangleq N_m
\]  
(3-10)

--

Figure 3.3: The phasor representation (left) and the one-sided power density spectrum (right) of the bandpass-filtered IF noise.

At the same time it can be proven (e.g. Schwartz et al. [98]) that the bandlimited noise of the BPF output will be Gaussian- or normally distributed along two orthogonal phasor axes:

\[
n(t) = \{ x(t) + j y(t) \} \exp(j \omega_{IF} t)
\]  
(3-11)

\[
n(t) = x(t) \cos(\omega_{IF} t) - y(t) \sin(\omega_{IF} t)
\]  
(3-12)

The two quadrature noise voltages \( x(t) \) and \( y(t) \) are identically distributed but orthogonal and thus independent. The pdf's of the two variables \( x \) and \( y \) can be written as

\[
p(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{x^2}{2\sigma^2} \right) = \frac{1}{\sqrt{2\pi N_m}} \exp \left( -\frac{x^2}{2N_m} \right)
\]  
(3-13)

\[
p(y) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{y^2}{2\sigma^2} \right) = \frac{1}{\sqrt{2\pi N_m}} \exp \left( -\frac{y^2}{2N_m} \right)
\]  
(3-14)

The mean values \( < x > \) and \( < y > \) of \( x \) and \( y \) will be 0, while the variances are equal to \( < x^2 > = < y^2 > = \sigma^2 = N_m \). The variance of \( n(t) \) can now be computed by\(^5\)

\(^5\)The factors \( 1/2 \) are due to the average power of the carrier cosine and sine terms.
\[ \frac{1}{2} \cdot \langle x^2 \rangle + \frac{1}{2} \cdot \langle y^2 \rangle = N_m, \text{ identical to the variance defined by the spectral analysis (3-10)}. \]

### 3.1.3 Signal-to-Noise Ratio

The SNR can be defined in a straightforward way. Combining (3-5) and (3-12) yields the total IF waveform \( s(t) \):

\[
s(t) = \{ A + x(t) \} \cos \omega_{IF} t - y(t) \sin \omega_{IF} t \tag{3-15}
\]

(This notation will be used on most occasions. However, sometimes it will prove to be more convenient to use a Signal phasor rotated over 45°:

\[
s(t) = \left( \frac{A}{\sqrt{2}} + x(t) \right) \cos \omega_{IF} t - \left( \frac{A}{\sqrt{2}} + y(t) \right) \sin \omega_{IF} t \tag{3-16}
\]

This does not influence the validity of the analysis.)

---

Figure 3-4: The phasor representation (left) and the one-sided power density spectrum (right) of the bandpass filtered total IF signal.

The SNR \( \xi \) that follows from \( s(t) \) or from the spectral representation of \( 2G_s(f) \) in Figure 3-4 is, in both cases, equal to

\[
\xi \triangleq \frac{\text{var}(S(t))}{\text{var}(n(t))} = \frac{1}{2} \frac{A^2}{\sigma^2} = \frac{A^2}{2N_m} \tag{3-17}
\]

Since \( \xi \) depends upon the actual width of the IF BPF (and thus upon \( m \)) it cannot be used as an independent variable. I will therefore define the IF SNR within a bandwidth equal to the bitrate as \( \rho \) and use this as the independent variable in terms of which all results will be expressed.

\[
\rho \triangleq \frac{A^2}{2N} = m\xi \tag{3-18}
\]
3.2 Probability density functions

3.2.1 PDF's of the IF signal

In section 3.1.3 the IF waveform \( s(t) \) has been defined as a two-dimensional phasor, with an amplitude of the in-phase component \( A + x(t) \) and an amplitude of the quadrature component \( y(t) \). As previously mentioned, the phase of the IF signal will in most cases not play a role of importance in the IF-detection process. It is therefore more convenient to use a transformation from rectangular to polar co-ordinates;

\[
\begin{align*}
    r(t) &= \sqrt{s_I^2(t) + s_Q^2(t)} \\
    \theta &= \arctan\left(\frac{s_Q(t)}{s_I(t)}\right)
\end{align*}
\]  

(3-19)  

(3-20)

In the case of ASK, the amplitude \( r \) at a sampling moment can now be written as

\[
\begin{align*}
    r &= \sqrt{(A + x)^2 + y^2} & \text{(Mark)} \\
    r &= \sqrt{x^2 + y^2} & \text{(Space)}
\end{align*}
\]  

(3-21)  

(3-22)

In both cases the variable \( r \) is formed by the root of a sum of two squared normally distributed variables; non-central variables in the case of a Mark and central variables in the case of a Space.\(^6\) Both \( x \) and \( y \) are defined as central normal variables in (3-13) and (3-14). From the definition given in Appendix A.2 it is clear that \( r \) has either a non-central chi distribution (for a transmitted Mark) or a central chi distribution (for a transmitted Space) [61, 62]. In each case the distribution has two degrees of freedom; each degree due to one of the two orthogonal noise components. The non-central distribution has a noncentrality parameter \( \lambda \) defined by (A.3-5) and equal to \( A^2/N_m = 2\xi \). The two probability density functions of the distributions can be obtained from (A.4-2) and (A.4-16) with \( \nu = 2 \) and \( \lambda \) as given above:

\[
\begin{align*}
    p_M(r) &= p_X\left(r \bigg| 2, \frac{A^2}{N_m}\right) \\
    &= \frac{r}{N_m} \exp\left(-\frac{r^2 + A^2}{2N_m}\right) I_0\left(\frac{rA}{N_m}\right) & \text{(Mark)} \\
    p_S(r) &= p_X(r \bigg| 2) \\
    &= \frac{r}{N_m} \exp\left(-\frac{r^2}{2N_m}\right) & \text{(Space)}
\end{align*}
\]  

(3-23)  

(3-24)

The distribution in the case of a transmitted Space is Rayleigh, while the distribution of the Mark is often called generalised Rayleigh or Rice [93, 94]. Of course the two

\(^6\)A central normal variable has a mean value equal to 0, see also appendix A.2.
pdf’s can also be obtained from the pdf’s of \( x \) and \( y \) through

\[
p(r, \theta)dr \, d\theta = p(x, y)dx \, dy = p(x)p(y)dx \, dy \tag{3-25}
\]

\[
dx \, dy = r \, dr \, d\theta
\]

The pdf’s of \( r \) and \( \theta \) can then be obtained by partial integration over the unused variable, leading to (3-23) and (3-24) for \( p(r) \) and for the pdf of \( \theta \) to [106]:

\[
p_M(\theta) = \frac{1}{2\pi} \exp \left( -\frac{A^2}{2N_m} \right) + \frac{A \cos \theta}{2\sqrt{2\pi}N_m} \left\{ 1 + \text{erf} \left( \frac{A \cos \theta}{\sqrt{2N_m}} \right) \right\} \exp \left( -\frac{A^2 \sin^2 \theta}{2N_m} \right) \tag{3-26}
\]

\[
p_s(\theta) = \frac{1}{2\pi} \tag{3-27}
\]

It is important to note that neither pdf of the amplitude is Gaussian, being always limited to non-negative values of the variables. Only in the case of very large \( A \) the non-central chi distribution approaches a Gaussian distribution. In particular for moderate signal-to-noise ratios Gaussian approximations are therefore quite inaccurate.

![Probability density functions](image)

**Figure 3-5**: Probability density functions of the IF amplitude \( r \) (left) and phase \( \theta \) (left) as a function of increasing carrier amplitude \( A \).

### 3.2.2 Moments of the distributions

The first two moments around zero of \( p(r) \) can be obtained using (A.4-6),(A.4-13) and (A.4-14), the latter two for large \( A \) or \( \xi \) only. For the second moment a simpler
method can also be used:

\[
E\{r^2\} = E\{(A + x)^2 + y^2\} = x^2 + 2xA + A^2 + y^2 = A^2 + <x^2> + <y^2> = A^2 + 2N_m \tag{3-28}
\]

\[
E\{r\} = E\left\{\sqrt{(A + x)^2 + y^2}\right\} = <A\sqrt{(1 + \frac{x}{A})^2 + (\frac{y}{A})^2}> \tag{3-29}
\]

which can be approximated by

\[
E\{r\} \simeq A < 1 + \frac{1}{2} \left(\frac{2x}{A} + \frac{x^2}{A^2} + \frac{y^2}{A^2}\right) - \frac{1}{8} \left(\frac{2x}{A} + \frac{x^2}{A^2} + \frac{y^2}{A^2}\right)^2 \tag{3-30}
\]

Using only the quadratic terms of the series expansion (3-30), the expression can be reduced to:

\[
E\{r\} \simeq A < 1 + \frac{x}{A} + \frac{y^2}{2A^2} > = A + \frac{N_m}{2A} = A \left(1 + \frac{1}{8\xi}\right) \tag{3-31}
\]

\[
\text{var}(r) = N_m - \frac{N_m^2}{4A^2} = N_m \left(1 - \frac{1}{8\xi}\right) \tag{3-32}
\]

These results are identical to the ones derived using the formulae of Appendix A.4.

Two interesting aspects of the moments derived here are the following. Firstly, the expected value \(E\{r\}\) is always larger than \(A\). This is due to the quadrature component...
of the signal phasor, as illustrated in Figure 3-4 and presented again in Figure 3-6. For small \( A \) the quadrature component is of the same order as the in-phase component, leading to a wide distribution of \( \theta \) and values of \( r \) larger than \( A \).

Secondly, for large \( A \) and SNR \( \xi \), \( \theta \) will be limited to small values around 0 (see also Figure 3-5) so the contribution of the quadrature component diminishes. The distribution then resembles a one-dimensional Gaussian distribution around \( A \) with the variance \( N_m \) of the single noise component \( x \). However, this approximation is only valid around the maximum of the pdf! For small values of \( r \) the quadrature component can not be neglected and the Gaussian approximation is generally not valid.

### 3.3 IF-detection

#### 3.3.1 Diode detectors

Throughout the analysis diode detectors will be used as non-linear elements for the IF-detection. In the cases of non-coherent ASK and FSK IF-detection, the diodes perform envelope detection of the bandpass-filtered IF signals (section 3.1.1). In most classical systems [98, 106] these detectors are of the linear type, using simple capacitor charging via a series diode. It will be shown, however, that in coherent optical receivers the exact modelling of the type of IF-detection is extremely important for optimum reception. Two essential types of detection to be considered are linear and quadratic IF-detection.

When a sine wave \( r = B \cos \omega t \) is applied to an unbiased diode detector, the desired signal is the dc output variation \( z_{dc} \). All output signals at multiples of the input frequency are effectively suppressed by a post detection filter. The diode characteristic is determined by the physics of the device, and is given by \( z = \exp r \) (omitting constants). The normalised output response to an input signal \( B \cos t \) can be calculated from

\[
    z_{dc} = \frac{\int_0^{2\pi} B \cos t \, dt}{\int_0^{2\pi} \cos t \, dt} = \frac{1}{2\pi} \int_0^{2\pi} \exp(B \cos t) + \exp(-B \cos t) \, dt = I_0(B) \tag{3-33}
\]

(Using (Gr.8.431-3) and with \( \omega \) set to 1)\(^7\). With the series expansion of the modified

\[
    z_{dc} = \sum_{k=0}^{\infty} \frac{B^{2k}}{(2k)!} \cos^{2k} t
\]

With (Gr.1.320-5) the dc-component of the cosine-power can be found to be \( (2k)!/(k!)^2 \), which

---

\(^7\)A second possibility is to use the series expansion of \( \exp(B \cos t) \), knowing that only even harmonics of the cosine contribute. This gives:

\[
    z_{dc} = \sum_{k=0}^{\infty} \frac{B^{2k}}{(2k)!} \cos^{2k} t
\]

With (Gr.1.320-5) the dc-component of the cosine-power can be found to be \( (2k)!/(k!)^2 \), which
3.3. IF-DETECTION

Bessel function (Gr.8.447-1) the output of the detector can be written as:

\[ z_{dc} = 1 + \frac{B^2}{4} + \frac{B^4}{64} + \ldots \]  \hspace{1cm} (3-34)

For small input signals this shows that a diode detector will be approximately quadratic or square-law\(^8\).

For high input levels the diode characteristic changes from purely exponential into \( \sqrt{\text{exp} \, r} \) [107]. This means that the dc output increases more slowly with \( B \), approaching a linear relation. However, and this is important, the relation can not simply be taken as \( z = r \) since this would imply that no non-linear operation is performed. In the case of linear detection the relation between the input and output signals should thus be given by

\[ z = \sqrt{r^2} = |r| \]  \hspace{1cm} (3-35)

More general, the characteristic of a \( n^{th} \)-order detector should be written as

\[ z = (r^2)^{n/2} \]  \hspace{1cm} (3-36)

From now on only the two limiting cases of \( n=1 \) (linear detector) and \( n=2 \) (quadratic detector) will be treated.

Pdf of the detector output

According to the last formula, the detector output is always obtained through a power of a sum of squared signal components (since \( r^2 \) is equal to the squared in-phase and quadrature signal components). In Appendices A.3 and A.4 it is shown that pdf's of the detector output therefore belong to the families of chi-square and chi distributions. In the case of linear IF-detection the detector order \( n \) is 1 and the pdf will be (non-central) chi distributed with two degrees of freedom and - in the case of a transmitted Mark - noncentrality parameter \( \lambda = A^2/N_m \):

\[ p_M(z) = p_{\chi^2} \left( z \mid 2, \frac{A^2}{N_m} \right) \]  \hspace{1cm} (3-37)

\[ \text{var}(z) = N_m - \frac{N_m^2}{4A^2} = mN \cdot \left( 1 - \frac{1}{8\xi} \right) \]  \hspace{1cm} (SNR \gg 1) \hspace{1cm} (3-38)

\(^8\)For proper quadratic detection the input signals of the detector must be kept sufficiently small. This means however that the noise of the detector, usually the shot noise due to the diode current, becomes important and can no longer be neglected [109, 110, 114]. Effectively there will be a range of permitted input signals, limited at the lower side by the detector noise and at the higher side by the allowed deviation from the quadratic characteristic. Practical implications are discussed in Chapter 8.
CHAPTER 3. NON-COHERENT IF-DETECTION

For a transmitted Space:

\[ p_S(z) = p_X(z \mid 2) \]  \hspace{1cm} (3-39)
\[ \text{var}(z) = \left( 2 - \frac{\pi}{2} \right) \cdot mN \]  \hspace{1cm} (3-40)

For quadratic IF-detection the detector order is 2, so the statistics are chi-square with two degrees of freedom and - again only for a transmitted Mark - the same noncentrality parameter \( \lambda = A^2/N_m \):

\[ p_M(z) = p \left( z \mid 2, \frac{A^2}{N_m} \right) \]  \hspace{1cm} (3-41)
\[ \text{var}(z) = 4mA^2N + 4m^2N^2 \]  \hspace{1cm} (3-42)

and for a transmitted Space

\[ p_S(z) = p(z \mid 2) \]  \hspace{1cm} (3-43)
\[ \text{var}(z) = 4m^2N^2 \]  \hspace{1cm} (3-44)

More details about the chi and chi-square distributions can be found in the Appendices, while the exact characteristics of the detector outputs are treated in Chapter 4.

Output components

If an IF signal given by (3-15) is input to a quadratic detector, the output can be written as

\[ z = (A^2) + (2Ax) + (x^2 + y^2) \]
\[ \Delta z_{SxS} + z_{SN} + z_{NzN} \]  \hspace{1cm} (3-45)

The output signal consists of three components, which are identified according to their respective origins as Signal-times-Signal (SxS), Signal-times-Noise (SxN) and Noise-times-Noise (NzN). These three components are correlated, which means that

\[ p(z) \neq p_{SxS}(z) \ast p_{SxN}(z) \ast p_{NzN}(z) \]
\[ \neq \delta(A^2) \ast p(8A^2N,0) \ast N \cdot p(\chi^2 \mid 2) \]  \hspace{1cm} (3-46)

(The asterisk denotes convolution). The problem is that the Gaussian component in the right-hand side of the expression yields a pdf from \(-\infty\) to \(+\infty\). The proper pdf of \( z \) however, a non-central chi-square distribution, is limited to positive values, which shows that \( p(z) \) may not be calculated in this way.

The signal-to-noise ratio of the detector output can now be defined in the same way as done by Schwartz et al. [98, p.105]:

\[ SNR_o \Delta \frac{E\{z^2 \mid N = 0\}}{E\{z^2\} - E\{z^2 \mid N = 0\}} = \frac{E\{z_{SxS}^2\}}{E\{z_{SxN}^2\} + E\{z_{NzN}^2\}} \]  \hspace{1cm} (3-47)

Note that in this definition the noise is not equal to the variance only, since dc noise components are also included in the total noise.
3.3. IF-DETECTION

3.3.2 Spectral effects

Only a quadratic detector has an output spectrum that can be calculated in a simple way. Each component of the output (3-45) can be obtained by convolution of two of the input components Signal and noise, the spectra of which are illustrated in Figure 3-4. Since this is a straightforward Fourier analysis, which has already been done once by Schwartz, Bennett and Stein [98, pp.108–114], only the results will be presented here⁹.

![Diagram of power density spectrum](image)

Figure 3-7: The one-sided power density spectrum of the output of a quadratic IF-detector with a carrier plus bandlimited noise of bandwidth 2B at its input. The spectra around 2f_IF will be filtered out by a post-detection filter.

\[
2G_{SS} = A^4 \delta(f) \quad (3-48)
\]
\[
2G_{SN} = 8A^2n \quad 0 \leq f \leq B \quad (3-49)
\]
\[
2G_{NN} = 4N_m^2 \delta(f) + 4n^2 \Delta'(2B) \quad (3-50)
\]

where \(\Delta'(2B)\) is the triangle function from -2B to 2B with top value 2B and \(\Delta'(2B)\) the one-sided form from 0 to 2B with peak amplitude 4B. These results are also illustrated in Figure 3-7. Since the quadratic detector performs power detection, the

⁹In this analysis the carrier components of (3-15) have again been omitted. Also, only the spectral components around \(f=0\) are given; results of the spectral components around 2f_IF are illustrated in Figure 3-7 and are simply scaled by a factor 1/2 due to the carrier cosine term.
time-averaged (non-zero) NxN noise power also contributes to the dc detector output. The combined dc component due to SxS and NxN is given by

\[ 2G_{dc} = (A^2 + 2N_m)^2 \delta(f) \]  

(3-51)

With (3-47) the detector output SNR can now be calculated from these spectra as the ratio of the signal dc-component and all noise components below 2B (including those at dc). This yields, using the relation \( 2Bn = N_m \) from (3-10):

\[ SNR_o = \frac{A^4}{8A^2N_m + 8N_m^2} = \frac{A^4}{8mA^2N + 8m^2N^2} \]  

(3-52)

From this formula, which can be rewritten as

\[ SNR_o = \frac{\xi^2}{2} \cdot \frac{1}{1 + 2\xi} \]  

(3-53)

some interesting conclusions can be drawn. For large input SNR \( \xi \) the output SNR \( SNR_o \) will be approximately \( \xi/4 \). For small input SNR on the other hand the output SNR approaches \( \xi^2/2 \), which means that there is a change of order in the input-output relation, see Figure 3-8. Due to the increasing influence of the NxN noise component for decreasing input SNR, \( SNR_o \) will drop more rapidly. This is called the suppression characteristic of envelope detection [98, p.104], which is only one of the many effects of the NxN noise.

Autocorrelation of the noise

The autocorrelation of a noise voltage \( n(t) \) is defined as

\[ R_n(\tau) = E\{n(t)n(t + \tau)\} \]  

(3-54)

when the noise is a stationary process, i.e. the statistics are independent of time. In that case the Wiener-Kinchine theorem [119] gives the following relations between the (two-sided) power density spectrum and the autocorrelation function:

\[ G_n(f) = \int_{-\infty}^{\infty} R_n(\tau) \exp(-j\omega \tau) \, d\tau \]  

(3-55)

\[ R_n(\tau) = \int_{-\infty}^{\infty} G_n(f) \exp(j2\pi f \tau) df \]  

(3-56)

\( G_n(f) \) and \( R_n(\tau) \) form a Fourier-transform pair, so the autocorrelation of the detector output components can easily be calculated from the results in the previous paragraph.

\[ ^{10} \text{These figures are only valid for quadratic detectors. Schwartz et al. [98, pp.104-105] show that for large \( \xi \) the relation between \( \xi \) and \( SNR_o \) is given by \( SNR_o = \xi/n^2 \); with \( n \) the order of the detector. For the linear detector it can be shown using a pdf-analysis that indeed \( SNR_o = \xi \) for large \( \xi \).} \]
3.3. IF-DETECTION

The autocorrelation function of the SxN noise component can thus be found from the Fourier transform of (3-49) as

\[
R_{S\times N}(\tau) = \mathcal{F}\{4nA^2 \cdot \text{rect}(2B)\} \\
= 2A^2 N_m \cdot \frac{\sin(2\pi B\tau)}{2\pi B\tau}
\]  

(3-57)

The NxN autocorrelation function is derived from (3-50) as follows:

\[
R_{N\times N}(\tau) = \mathcal{F}\{4n^2 \Delta(2B)\} = \mathcal{F}\{4n^2 \cdot \text{rect}(2B) \ast \text{rect}(2B)\} \\
= N_m^2 \left(\frac{\sin(2\pi B\tau)}{2\pi B\tau}\right)^2
\]  

(3-58)

(The asterisk as usual denotes convolution).

As expected the autocorrelation function of the bandlimited and rectangular SxN noise spectrum is a sinc function with zeros at multiples of 1/2B, see Figure 3-9. This means that for independent samples of the SxN noise the sample frequency should be 2B or T/m Hz, exactly in accordance with Nyquist’s first sampling theorem (page 50). Although the NxN noise has spectral components up to 2B Hz, the required sampling rate is identical to the one for SxN noise since the zeros of its autocorrelation function are also multiples of 1/2B. The reason for this seeming discrepancy is Nyquist’s first
Figure 3-9: The two-sided spectral noise components at the output of a quadrature detector (left) and the associated autocorrelation functions (right).

criterion on equally spaced axis crossings in the impulse response, also called Nyquist's vestigial-symmetry theorem [84, 85] [3, Ch.5]. It can be stated as follows:

Theorem 3.3 (Nyquist) The impulse response of a LPF of width $f_1$ can be modified by extending the maximum frequency to $2f_1$ as long as vestigial symmetry around $f_1$ is maintained. This leaves the original axis crossings of the impulse response intact, while introducing additional ones.

The equivalent mathematical notation is [3, 5-35]:

$$G(f) = \begin{cases} 
1 - G_1(f) & 0 < f \leq B \\
G_1(f) & B < f \leq 2B 
\end{cases} \quad (3-59)$$

for $G_1(B - f) = -G_1(B + f)$

and $G_1(B) = \frac{1}{2}$

Obviously, the triangular NxN noise spectrum satisfies this criterion, which explains why the same sampling requirements hold as for the SxN noise\(^\text{11}\). This is a very important result which has its impact on the modelling of the post-detection filter in section 3.4.

\(^{11}\)An important group of filter characteristics satisfying this criterion is the raised cosine or RACOS with sinusoidal roll-off. Especially in practical systems these characteristics are very important for obtaining ISI-free eye-patterns. See e.g. Tomesen and Hooijmans [114].
3.4 Post-detection filtering

3.4.1 Integrate-and-dump filters

As already mentioned in 3.1.1, a post-detection filter has the dual task of noise optimisation and ISI minimisation. It is well-known [98, 106] that for maximum SNR a matched filter is required. If the IF detector output signal is defined by the Fourier-transform pair \( z(t) \leftrightarrow Z(f) \), the required impulse response \( h(t) \) and frequency transfer characteristic \( H(f) \) of the post-detection filter are:

\[
\begin{align*}
  h(t) &= z^*(T_s - t) \\
  H(f) &= Z^*(f) \exp(-j2\pi f T_s)
\end{align*}
\]  

(3-60)

(3-61)

where \( T_s \) is the sampling instant and \( * \) denotes the complex conjugate. Note that when the input noise of the matched filter is not white, but 'coloured' with spectral density function \( G_N(f) \), the transfer function should be modified as

\[
H(f) = \frac{Z^*(f)}{G_N(f)} \exp(-j2\pi f T_s)
\]  

(3-62)

When the pulses \( z(t) \) are rectangular with duration \( T \) and the input noise is white, it is easy to show that \( h(t) \) should also be rectangular:

\[
\begin{align*}
  h(t) &= \begin{cases} 
  1/T & 0 < t < T \\
  0 & \text{elsewhere}
\end{cases} \\
  &= \frac{1}{T} \text{rect}(T) * \delta \left( \frac{T}{2} \right)
\end{align*}
\]  

(3-63)

(3-64)

The frequency transfer function of this filter is of course a sinc with zeros at multiples of \( 1/T \) and unity transfer at dc. The filter output \( u(t) \) is formed by the convolution of \( z(t) \) and \( h(t) \). Writing out this operation shows that the post-detection filter can be realised as an integrator with time constant \( T \), also known as an Integrate-and-Dump (I&D) filter:

\[
u(t) = h(t) * z(t) = \frac{1}{T} \int_0^T z(\tau) d\tau
\]  

(3-65)

It is important to note that the post-detection filter defined here is not the proper matched filter for optimal reception. With non-white input noise a matched filter should compensate for the non-flat noise power spectral density by weighting as in (3-62). Since the detector output has a far from flat noise response - as illustrated in Figure 3-7 - the matched post-detection filter should compensate for this. However, as can be seen from (3-63), the I&D-filter does not perform any weighting and therefore is not the optimum solution. The receiver performance will thus not be optimal,
resulting in some penalties. Nevertheless, I&D-filters will be used as post-detection filters throughout the analysis, both because inclusion of the noise-weighting renders exact theoretical analysis impossible, and because practical post-detection filters do not perform weighting.

In the previous section it was derived that the IF-detector output, both Signal and noise, can be described by \( m \) samples, taken \( 1/2B \) or \( T/m \) seconds apart. These \( m \) independent samples \( z_i \) will now be used as the basis for the post-detection filtering analysis. First, from these samples we have to reconstruct the signal, since the input signal of the post-detection filter is a continuous time signal and not sampled. The normal way for reconstructing a signal from a series of Dirac-pulses is with a rectangular LPF of single-sided width \( m/T \), see Figure 3-11. The associated impulse response of such a filter, also illustrated in the same figure, is a sinc-function. This means that the response of one single Dirac-impulse stretches out from \(-\infty\) to \(+\infty\). With respect to the waveform reconstruction this is no problem, since at the other sampling instances \( k \cdot T/m \) the response of the sinc is zero. However, when the waveform reconstructed using sinc's is passed through the I&D-filter, a continuous time integrator, this implies that every Dirac-impulse from \(-\infty\) to \(+\infty\) contributes to the filter output. Figure 3-11 illustrates that the post-detection filter response of a single group of \( m=3 \) samples can be non-zero at the other sampling moments 0 and 2\( T \). This is only for one group of \( m \) samples. For a signal of multiple periods \( T \) the responses of all groups should be added.

Apparently, the combination of a signal reconstructed using a rectangular LPF (and thus sinc impulse response) and I&D post-detection filtering is not good. If we want to stick to the I&D-filter, which is mathematically the only solution, we have to select a different method for waveform reconstruction.
Figure 3.11: Illustration (for \( m = 3 \)) of two possibilities for the reconstruction of a continuous time signal from a series of samples, and its influence on the integrate-and-dump post-detection filter output. The filter output for a signal reconstructed with a rectangular LPF can not be given since all (unknown) samples from \(-\infty\) to \(+\infty\) contribute.
The impulse response of the new reconstruction LPF should be limited to as small a time interval as possible, preferably to \( T/m \). This can be realised using a LPF with impulse response \( \text{rect}(T/m) \), see Figure 3-11. When the samples of the IF-detector output signal are located in the center of their respective timeslots, the waveform reconstructed with the alternative LPF can be written as

\[
    z'(t) = \sum_{i=0}^{m-1} z_i \cdot \delta \left( \frac{T}{2m} + i \cdot \frac{T}{m} \right) \ast \text{rect} \left( \frac{T}{m} \right) \tag{3-66}
\]

The reconstructed waveform now consists of a series of blocks, which means that the high-frequency content of the signal is higher than with the sinc reconstruction. This is in accordance with the fact that the frequency characteristic of the new filter is a sinc-function, and stretches out to \(+\infty\). Apparently, by exchanging low frequency spectral components of the series of samples for high frequency components, a better and ISI-free waveform reconstruction can be obtained. This means at the same time that we are not using a pure I&D post-detection filter, but an I&D filter in combination with a pre-filter with impulse response \( \text{rect}(T/m) \).

The post-detection filter output is given by:

\[
    u(t) = z'(t) \ast h(t) = \sum_{i=0}^{m-1} z_i \cdot \delta \left( \frac{T}{2m} + i \cdot \frac{T}{m} \right) \ast \text{rect} \left( \frac{T}{m} \right) \ast \frac{1}{T} \text{rect}(T) \ast \delta \left( \frac{T}{2} \right) \tag{3-67}
\]

The convolution of the two rectangular functions yields a trapezoidal function defined by \( \text{trap}(\text{topwidth}/\text{bottomwidth}) \) (of height 1) and an amplitude factor \( T/m \).

\[
    u(t) = \frac{1}{T} \cdot \frac{T}{m} \sum_{i=0}^{m-1} z_i \cdot \delta \left( \frac{T}{2m} + i \cdot \frac{T}{m} \right) \ast \text{trap} \left( \frac{T - T/m}{T + T/m} \right) \tag{3-68}
\]

From the example for \( m=3 \) in Figure 3-11 it can be seen that at the decision moment \( T \) all individual trapezoidal functions give exactly a maximum contribution to the total filter output. The integrate-and-dump post-detection filter output at decision moments \( k \cdot T \), for a series of \( m \) input samples between 0 and \( T \), can now be written as:

\[
    u(T) = \frac{1}{m} \sum_{i=0}^{m-1} z_i \tag{3-69}
\]

\[
    u(k \cdot T) = 0 \quad (k \neq 1) \tag{3-70}
\]

Figure 3-11 also clearly illustrates the ISI-free response of this combination of waveform reconstruction and post-detection filtering. Since we have adopted the IF carrier representation (with a constant IF Signal amplitude), we are only interested in the pdf
3.4. POST-DETECTION FILTERING

at decision moments. When omitting the references to time, the post-detection filter output expression can be written as:

\[ u = \frac{1}{m} \sum_{i=1}^{m} z_i \]  

(3-71)

Each \( z_i \) can furthermore be written as the sum of two independent quadrature samples - for once using the notation of (3-16) -

\[ z_i = \left( \frac{A}{\sqrt{2}} + x_i \right)^2 + \left( \frac{A}{\sqrt{2}} + y_i \right)^2 \]  

(3-72)

where each sample \( x_i \) and \( y_i \) have variance \( N_m \). Since \( x(t) \) and \( y(t) \) are uncorrelated, but identically normally distributed as (3-13) and (3-14), \( z \) is also given by

\[ z_i = \sum_{j=1}^{2} \left( \frac{A}{\sqrt{2}} + x_j \right)^2 \]  

(3-73)

This yields the filter output

\[ u = \frac{1}{m} \sum_{i=1}^{2m} \left( \frac{A}{\sqrt{2}} + x_i \right)^2 \]  

\[ = \frac{1}{2m} \sum_{i=1}^{2m} (A + x_i')^2 \]  

(3-74)

where \( x' \) is a normally distributed variable with variance \( 2N_m \).

Again, the limitations of the post-detection filter model should be stressed. Apparently, a single integrate-and-dump post-detection filter is not suited for ISI-free filtering of normal waveforms. The reason for using nevertheless this filter model is the mathematical convenience that arises from the conversion of a continuous time integration into a summation of samples. If we still want to use this filter model, in combination with ISI-free reception, some pre-filtering is required. In this case we have selected a LPF with rectangular impulse response and sinc frequency transfer characteristic. This leaves more high frequency components in the reconstructed waveform. In fact the total post-detection LPF frequency characteristic is the product of the two individual characteristics, and is given by

\[ H_{tot}(f) = \frac{\sin \pi fT}{\pi fT} \cdot \frac{\sin \pi fT/m}{\pi fT/m} \]  

(3-75)

Whenever the term 'I&D' post-detection filtering will be used in the following chapters, I&D plus pre-filtering is meant.
PDF of the filter output

According to (3-74) the post-detection filter output is a sum of $2^m$ variates, which means that the pdf of $u$ can be found according to (A.1-11) by consecutive convolution of the pdf’s of each variate:

$$p(u) = p(z_1) * p(z_2) * \ldots * p(z_{2^m-1}) * p(z_{2^m})$$  \hspace{2cm} (3-76)

Since each sample $z_i$ is a squared non-central gaussian variate, all $z$ have the same non-central chi-square distribution - defined by (A.3-1) - with 1 degree of freedom and (in the case of a transmitted Mark) noncentrality parameter $A^2/2N_m$. In short: $p(z | 1, A^2/2N_m)$. An extremely important characteristic of the chi-square distribution is its reproducibility under convolution, as detailed in Appendix A.3. When convolving two chi-square pdf’s the resulting pdf is again chi-square, with the degrees of freedom and noncentrality parameters summed. $m$-Fold convolution of a chi-square pdf thus gives another chi-square pdf with $m$ times higher degrees of freedom and noncentrality parameter. The factor $1/2m$ reduces the variance of a single sample in (3-74) from $2N_m$ to $2N$. This finally yields for the pdf of $u$:

$$p(u) = p_1(z | 1, A^2/2N_m) * \ldots * p_{2m}(z | 1, A^2/2N_m)$$

$$= p(u | 2m, A^2/N)$$

$$= p(u | 2m, 2\rho)$$  \hspace{2cm} (3-77)

The importance of this result is the fact that the evaluation of the receiver performance in terms of pdf's can be continued, including the essential post-detection filter.\(^{12}\)

3.4.2 Inter-symbol interference

The analysis presented so far does not include penalties associated with Inter-Symbol Interference (ISI). The use of a carrier of infinite duration in the case of a Mark - or the absence in the case of a Space - precludes any form of time-dependency, so the statistics belong to a stationary process. A common method for inclusion of possible ISI penalties is to analyse the system response to an isolated Mark or Space pulse, surrounded by increasing amounts of opposite pulses (see e.g. Kazovsky [67]).

ISI is strongly related to the actual over-all system performance, taking into account all transfer characteristics from transmitter to post-detection filter. However, as explained in the previous sections of this chapter, most circuits in the analysis are idealised. The impulse response of the rectangular IF BPF is a function of the form $\text{sinc}(mt/T)$, with zeros at the moments $T/m$ and at least at every sample moment $kT$ not equal to 0. Whatever happens in the IF detector and post-detection filter, the zeros at the sample moments remain unchanged, so zero ISI is nevertheless guaranteed.

\(^{12}\)Whalen [119, ch.4] derives the same pdf for (radar-) receivers with post-detection summation. However, this is a completely different concept, which does not include post detection filtering. It shows on the other hand that different receiver concepts may yield the same pdf's.
In conclusion it is thus permitted to state that, however abstract the signals used, the complete receiver model does not violate ISI requirements. This means that the ISI can be eliminated from the analysis.

3.5 Decision theory

The final operation after post-detection filtering is the analog-to-digital conversion in the decision circuit. From the noisy, filtered and IF-detected signal the receiver has to decide which bit was originally transmitted. Decision theory is well-documented, so only the essential aspects will be treated here: determining the decision threshold.

A baseband signal with two probability density functions for a transmitted Mark and Space \( p_M(u) \) and \( p_S(u) \) is given. Both pdf's are known and related to the receiver configuration, which was assumed to be optimal. Due to the IF-detection - either linear or quadratic - both pdf's are limited to positive values of \( u \), and - apart from limiting cases - not identical. There will thus be at least one point \( u_b \) where the two pdf's intersect (apart from the trivial point at \( u=0 \)).

![Figure 3-12: Two arbitrary probability density functions and the threshold setting for optimum detection.](image)

The decision circuit is a bistable clocked element (especially at high bitrates usually an edge-triggered Master-Slave D-type flip-flop), which takes either of its two output states depending upon the value of the input signal relative to a threshold \( b \). The total symbol error probability can now be defined as

\[
P_e = P(0) \int_b^\infty p_S(u)du + P(1) \int_0^b p_M(u)du
\]  

(3-78)

where \( P(0) \) and \( P(1) \) are the a priori symbol probabilities for a Space and Mark
respectively. Bayes decision theory\textsuperscript{13} gives a criterion for determining the optimum decision rule [35, 119]. It defines the likelihood ratio \(l(u)\) as

\[
l(u) \triangleq \frac{p_M(u)}{p_S(u)}
\]  

(3-79)

The likelihood ratio is a stochastic variable itself, that can attain values between 0 and \(\infty\). Omitting the costs, the threshold value of the likelihood ratio is given by the constant

\[
l_0 \triangleq \frac{P(0)}{P(1)}
\]  

(3-80)

The final Bayes decision rule can now be formulated as follows: “decide that a Mark has been sent if \(l(u) > l_0\), decide that a Space has been sent if \(l(u) < l_0\)”. In formula:

\[
l(u) = \frac{p_M(u)}{p_S(u)} \quad \frac{M}{S} = \frac{P(0)}{P(1)} = l_0
\]  

(3-81)

This result can also be obtained with the method of maximum a posteriori probability, or minimum error probability (e.g. [119]).

In the case of binary communication the \(a\ priori\) probabilities \(P(0)\) and \(P(1)\) are usually identical and equal to \(\frac{1}{2}\), so the likelihood ratio reduces to 1. The value of \(u\) for which \(l(u) = l_0\) is the decision threshold \(b\) of the decision circuit. Since \(l_0 = 1\) the decision threshold is also determined by the intersection of the two pdf’s:

\[
p_S(b) = p_M(b)
\]  

(3-82)

So, if \(u\) falls within the region \(U_S\) (see Figure 3-12) it should be decided that a Space has been sent, while \(u\) in region \(U_M\) leads to a transmitted Mark. Since both \(p_M(u)\) and \(p_S(u)\) are known, it will thus be possible to derive an analytical expression for the threshold. Using this threshold, the average symbol error probability will be minimal and defined equal to the Bit-Error Rate (BER). The latter assumption is valid when, averaged over long bit sequences, the average total error rate is equal to the probability of individual symbol errors. Finally, it is important to note that, despite the fact that the \(a\ priori\) symbol probabilities are the same, even at optimum threshold the two a posteriori symbol error probabilities are not necessarily identical.

Especially in binary communication the exact threshold setting is of extreme importance for the detection performance. In the next chapter it will be shown that the BER features only a sharp minimum as a function of the threshold. In contrast to most analyses\textsuperscript{14} I have put much effort in calculating the exact threshold settings as a function of the SNR and \(m\).

\textsuperscript{13}In its most general form Bayes theory also contains ‘cost factors’ for all four types of a posteriori decisions. Since in our case Mark and Space costs are identical for wrong decisions, while the costs for good decisions are zero, they can be eliminated from the analysis.

\textsuperscript{14}There are two distinct ways in which the threshold is treated in most published analyses. The
3.6 Linewidth effects

3.6.1 Theory

The effect of the large laser linewidth upon the receiver performance is one of the main differences between 'classical' systems and coherent optical systems. It has therefore been analysed extensively by many different authors [21, 25, 26, 57, 64, 65, 103, 104], consequently only a short summary will be presented here.

![Diagram showing IF BPF and 2G(f)](image)

\[ 2G_{\text{tot}}(f) = 2G_{\text{signal}}(f) \times 2G_{\Delta \nu}(f) \]  

(3-83)

Figure 3-13: The effect of the IF laser linewidth on single-filter detection. Signal falling outside the BPF gives a detection error, so the normalised shaded areas yield the BER due to the linewidth.

The Lorentzian power density spectrum of the combined transmitter and local oscillator linewidth - from now on referred to as the IF linewidth \( \Delta \nu_{IF} \) - as given by (2-8) and Figure 2-2 will broaden the IF spectrum of the received carrier.

The first one, mostly used in communication oriented papers, leaves the parameter \( b \) in the results, without determining the exact value. Sometimes the limiting value for infinite SNR is used as a constant throughout the analysis. Personally I think that such results have little practical value. On some rare occasions people effectively compute the threshold, but always with the help of a computer and for a very limited amount of settings. This does not give a general insight in the behaviour of the threshold either.

The second approach is the one commonly used in radar evaluation, where generally the a priori probabilities of targets are unknown. There, first the probability of false alarm is determined by the integral from a fixed threshold to infinity of \( p_F(u) \). The SNR and/or the number of summed post-detection samples (see the footnote on page 70) then have to be chosen in such a way that the probability of detection - being the integral from the threshold to infinity of \( p_M(u) \) - is as close to 1 as required. Although the threshold remains essential, it is a less critical parameter compared to a communication receiver. This is essentially the way Marcum treats detection in his famous paper [76].
\[ 2G_{tot}(f) = \frac{1}{2} A^2 \cdot \delta(f - f_{IF}) \ast \frac{2}{\pi \Delta\nu_{IF} \left[ 1 + \left( \frac{2f}{\Delta\nu_{IF}} \right)^2 \right]} \]  

(3-84)

The simplest, but nevertheless effective analysis is made by Kazovsky [64]. Assuming perfect rectangular binary ASK modulation, the required bandwidth \( B_M \) of the IF bandpass filter in order to accommodate 95% of the signal power is \( 3/T \) (with \( 1/T \) the bitrate as usual). Kazovsky calculates the additional bandwidth \( \Delta f_L \) required to accommodate 95% of the phase noise power of the broadened carrier. This can be derived from the Lorentzian spectrum to be equal to \( 12.7\Delta\nu_{IF} \). The ratio of the two bandwidths with and without phase noise is then given by

\[ \frac{\sqrt{B_M^2 + (\Delta f_L)^2}}{B_M} = \sqrt{1 + \left( \frac{1}{3} \Delta f_L T \right)^2} \]  

(3-85)

and is defined as the penalty. For a penalty of 1 dB he finally derives the allowable linewidth-bitrate ratio for ASK single-filter receivers

\[ \Delta\nu_{IF} T \geq 0.18 \]  

(3-86)

A signal, in principle a carrier within the range of the BPF, falling outside the BPF due to the phase noise generates a detection error. The normalised tails of the linewidth spectrum outside the BPF are thus a measure for the bit-error ratio due to the linewidth (see Figure 3-13). Since the normalised area of these tails is independent of the SNR, the linewidth will introduce an irreducible BER; a BER-floor. A measurable floor in the BER characteristic should under all circumstances be avoided, since this means that the receiver performance can not be improved further by increasing the SNR. It should be stressed that - as long as the linewidth is not zero - there will always be a BER-floor, although it may become unobservable. Practical floors below \( 10^{-18} - 10^{-20} \) are considered to be acceptable, as will be seen from Figure 3-14 in the next section.

### 3.6.2 Linewidth floor

For practical system design it is required to know at which level the BER-floor is situated, since this is directly related to a possible sensitivity penalty. Although over the past years many authors have analysed the influence of a non-negligible laser linewidth on the receiver performance few results are fit for practical applications. Further, most papers use extensive computer analysis and can therefore not be generally applied.

Garrett and Jacobson [25] have derived an expression for the BER-floor of ASK receivers without post-detection filtering, essentially by simple integration of the (sup-
Figure 3.14: Bit-error rate curves of a single-filter receiver without post-detection filtering for increasing linewidth-to-bitrate ratios. $m=1$. The values of $\Delta \nu_{IF} T$ and associated levels of the BER-floor are indicated.

Posedly Gaussian) detector output pdf:

\[
P_{\text{floor}} = \frac{1}{2} \text{erfc} \left( \frac{1}{2} \sqrt{\frac{\pi B_{IF}}{\Delta \nu_{IF}}} \right) = \frac{1}{2} \text{erfc} \left( \sqrt{\frac{\pi m}{4 \Delta \nu_{IF} T}} \right)
\]  

(3-87)

\[
P_{\text{floor}} = \frac{1}{2} \text{erfc} \left( \frac{\pi}{\sqrt{2\gamma}} \right)
\]  

(3-88)
The parameter $\gamma$ - the IF linewidth-to-bandwidth ratio - is defined by

$$\gamma = 2\pi \Delta \nu_{IF} \frac{1}{B_{1F}} = \frac{2\pi \Delta \nu_{IF} T}{m} \quad (3-89)$$

As can be seen from (3-87) the effective floor only depends upon the IF linewidth-to-bandwidth ratio. However, the effect of post-detection filtering is still not included. A more fundamental mathematical analysis is presented in [22] by Foschini and Vannucci. They include the laser phase noise, by defining the parameter $X$ as the accumulated phase change within one bit interval $T$ at the output of a finite-time integrator IF BPF (modelled in exactly the same way as our post-detection filter):

$$X = \left| \frac{1}{T} \int_0^T \exp(j \theta(t)) dt \right| \quad 0 \leq X \leq 1 \quad (3-90)$$

$$= \ r(T) \exp(j \theta(T))$$

Figure 3-15 shows the effect of $X$: the accumulated and detected phase variations

Figure 3-15: Illustration of the effect of laser phase noise on the output of an integrating post-detection filter. $A^2T$ is the output level at the decision moment without phase noise, $(XA)^2T$ the same output with phase noise. At the right the pdf of the filter output in the absence of amplitude noise.

(which are always non-negative due to the IF-detector) within one bit period lowers the IF-detector output at the decision moment. The pdf of the filter output, as illustrated in Figure 3-15, has finite values for $u \geq 0$ (note the logarithmic scale of $p(u)$!). Even a threshold set to the absolute minimum value 0 gives a finite probability, which is equivalent to the previously defined area of the normalised tails in Figure 3-13. In order to find the pdf (and distribution) of the post-detection filter output $u(T)$, Foschini and Vannucci take the Laplace transform of the pdf - or Moment Generating Function
3.6. LINEWIDTH EFFECTS

\( (\text{MGF}) \) - of \( z(T) \) after quadratic IF-detection:

\[
M_z(s) = \exp(sT^2) \left( \frac{\sinh \sqrt{2\gamma sT}}{\sqrt{2\gamma sT}} \right)^{-1/2}
\]

\[= \exp(sT^2) \prod_{k=1}^{\infty} \left( 1 + \frac{2\gamma sT^2}{(k\pi)^2} \right)^{-1/2}
\]

They obtain from this MGF the pdf \( p_x(z) \) in the situation without post-detection filtering. The post-detection filter output can now be found after \( m \)-fold convolution of the pdf \( p(z) \), which is equivalent to \( m \)-fold multiplication in the frequency domain according to the Laplace convolution (or Faltung) theorem:

\[
M_u(s) = [M_z(s)]^m
\]

\[= \exp(msT^2) \prod_{k=1}^{\infty} \left( 1 + \frac{2\gamma sT^2}{(k\pi)^2} \right)^{-m/2}
\]

Garrett and Jacobson in [26] approximate this expression by using only \( k=1 \) and setting all other product-terms to 1, which is allowed for ‘moderate’ values of \( \gamma \). This yields the final MGF (after time normalisation, so using \( s \) instead of \( sT^2 \))

\[
M_u(s) = \exp(ms) \left( 1 + 2\gamma s/\pi^2 \right)^{-m/2}
\]

The pdf of \( u, p(u) \), is obtained by the inverse Laplace transform of (3-94), yielding [26]:

\[
p_x(u) = \frac{1}{\Gamma(m/2)} \left( \frac{m\pi^2}{2\gamma} \right)^{\frac{m}{2}} (1-u)^{\frac{m-2}{2}} \exp \left( -\frac{m\pi^2}{2\gamma} (1-u) \right)
\]

\( (u \leq 1) \)

Finally, the error rate floor is determined by the tail of \( p_x(u) \) extending below the normalised decision threshold \( b' \):

\[
P_{\text{floor}} = \int_0^{b'} \left( \frac{m\pi^2}{2\gamma} \right)^{\frac{m}{2}} \left( 1-u \right)^{\frac{m-2}{2}} \frac{\exp \left( -\frac{m\pi^2}{2\gamma} (1-u) \right)}{\Gamma(m/2)} \, du
\]

which can be solved straightforwardly with \((Gr.3.381-3)\) to be

\[
P_{\text{floor}} = \frac{\Gamma \left( \frac{m}{2}, \frac{m\pi^2}{2\gamma} (1-b') \right)}{\Gamma(m/2)} = \frac{\Gamma(m/2, Z)}{\Gamma(m/2)}
\]

where

\[
Z = \frac{m\pi^2}{2\gamma} (1-b') \quad (b' \leq 1)
\]

\[= \frac{m^2\pi}{4\Delta v_{IF} T} (1-b') \quad (b' \leq 1)
\]
Figure 3.16: The bit-error rate floor of a single-filter receiver with quadratic IF-detection and integrate-and-dump post-detection filtering as a function of the relative IF linewidth $\Delta \nu_{IF} T$ and the IF bandwidth expansion factor $m$. The threshold $b'$ is set to 0.25. The two solid dots indicate the floor of the two settings extensively evaluated by Foschini, Greenstein and Vannucci in [21]; $m=5 \Leftrightarrow \Delta \nu_{IF} T=0.27$ and $m=8 \Leftrightarrow \Delta \nu_{IF} T=0.57$

$\Gamma\left(\frac{m}{2}, Z\right)$ is the incomplete gamma function defined in (A.3-25), for which a very accurate and useful approximation is given by (C.2-8). So, with equal a priori Mark and Space probabilities, and knowing that the Space signals outside the BPF do not contribute to the floor, the approximative but simple expression for the BER floor is given by

$$P_{floor} = \frac{e}{2} \sqrt{\frac{m}{2\pi}} \frac{(Z/m)^m \exp[-(Z - m + 1)]}{(Z - m + 1)}$$

(3-100)
3.6. LINEWIDTH EFFECTS

or

\[ P_{\text{floor}} = \frac{e}{2} \sqrt{\frac{m}{2\pi}} \left( \frac{m\pi^2}{2\gamma} (1 - b') \right)^m \exp \left[ -\left( \frac{m\pi^2}{2\gamma} (1 - b') - m + 1 \right) \right] \frac{1}{\left( \frac{m\pi^2}{2\gamma} (1 - b') - m + 1 \right)} \]  

(3-101)

It can be checked that for \( m=1 \) and \( b'=0 \), with \( \Gamma(\frac{1}{2}, Z) = \sqrt{\pi} \text{erfc}(\sqrt{Z}) \) and \( \Gamma(\frac{1}{2}) = \sqrt{\pi} \), (3-97) reduces to (3-87). For increasing \( m \) the system thus becomes more tolerant to the IF linewidth, as illustrated in Figure 3-16.

The above results show the basic advantage of the coherent optical receiver design presented so far. Receivers employing single-filter IF-detection may tolerate very large IF linewidth values, as high as the bitrate, by using large bandwidth IF filters. The integrating effect of the post-detection filter both reduces the influence of the linewidth and the sensitivity penalty due to the excess IF noise bandwidth.

3.6.3 IF bandwidth

It is essential to be able to calculate the approximate level of the BER-floor due to the phase noise in the way as shown above. However, in practical system design the most important question is:

How should the coherent optical receiver - and in particular the IF bandpass filter - be designed in order to combine maximum sensitivity with minimum phase noise dependence?

Amazingly enough, only one article has dealt with this question in a general way. Foschini, Greenstein and Vannucci [21] in principle use exactly the same system model as I do, but after the definition of the pdf's the analysis is taken over by computer simulation. In the pdf \( p_M(z) \) the constant carrier amplitude \( A \) is replaced by the phase noise dependent amplitude \( XA \), with \( X \) defined by (3-90). The effective pdf in the case of a transmitted Mark - in the case of a transmitted Space and single-filter detection the phase noise has no influence - can then be calculated from

\[ p_M(z) = \int_0^1 p_M(z \mid X) p_X(X) dX \]  

(3-102)

Unfortunately the evaluation of this integral can not be done in an analytical way, so Foschini et al. resort to computer simulation. However, their results are extremely interesting. For several different linewidth-bitrate ratios \( \Delta \nu_{IF}T \) they calculate the bandwidth expansion factor \( m \) that gives the highest sensitivity at a BER of \( 10^{-9} \). In this analysis both additive noise due to the receiver and phase noise due to the laser linewidth have thus been included.
Figure 3-17: Optimum IF bandwidth versus linewidth-bitrate ratio $\Delta \nu_{IF}T$ with integrate-and-dump post-detection filtering and at a BER of $10^{-9}$. Each closed circle represents an optimal bandwidth, expressed as an integer multiple of $1/T$, and each vertical bar spans a range within $\pm 1/T$ of that bandwidth. The regression curve provides a simple function that is highly consistent with the optimal-bandwidth results. (From Foschini, Greenstein and Vannucci, IEEE Transactions on Communications, Vol. 36, No. 3, 1988, pp.306–314, ©1988 IEEE).

The regression line in Figure 3-17 gives the optimum bandwidth expansion factor $m$ as a function of the linewidth-bitrate product. The mathematical representation can be derived as

$$m_{opt} = 12(\Delta \nu_{IF}T)^{0.65}$$  \hspace{1cm} (3-103)

For these values of $m$ no effects of a floor are observed for BER values above $10^{-10}$. Furthermore they find that, especially at higher linewidth-bitrate ratios, the optimum for $m$ is very broad (although high), which means that the actual value of $m$ is not critical. It is interesting to note that for the two special cases they present in detail - $m=5 \Leftrightarrow \Delta \nu_{IF}T=0.27$ and $m=8 \Leftrightarrow \Delta \nu_{IF}T=0.57$ - the theoretical optimum values of $m$ with (3-103) are 5.1 and 8.3 respectively. The values of $m$ and the linewidth inserted in the BER-floor expression (3-100) yield values for the floor of $4.2 \times 10^{-19}$ and $1.2 \times 10^{-20}$ respectively. This is in close agreement with the previously posed assumption that acceptable error rate floors should be around $10^{-20}$ or lower.

Now that we have a confirmation for the assumption that a BER-floor at $10^{-20}$ is optimum for a combined unobservable floor and minimum IF bandwidth, we can extract
3.6. LINEWIDTH EFFECTS

the same data from Figure 3-16. Using the points where the curves $P_{floor}(\Delta \nu_{IF} | m)$ cross the line $P_{floor} = 10^{-20}$, and fitting a curve of the form $a(\Delta \nu_{IF} T)^b$, we obtain the following relation for optimum IF bandwidth expansion factor $m$:

$$m_{opt} = 11(\Delta \nu_{IF} T)^{0.565}$$  \hspace{1cm} (3-104)

Note the remarkable resemblance with the result of Foschini et al. (3-103).

![Graph showing comparison of bandwidth expansion factor with linewidth-to-bitrate ratio.](image)

**Figure 3-18:** Comparison of the optimum IF bandwidth versus linewidth-bitrate ratio $\Delta \nu_{IF} T$ with integrate-and-dump post-detection filtering and at a BER of $10^{-9}$, computed by three different methods. The solid line represents the analytical results derived in this thesis, the dashed line the results from Foschini, Greenstein and Vannucci [21] and the dotted line the theoretical results of Kazovsky [64].

Finally, the expression for $m$ may heuristically be improved by inclusion of the fixed IF bandwidth that is required for proper reception. Theoretically a bandwidth of $1/T$ should be sufficient, but $2/T$ is more realistic (remember that Kazovsky [64] used $3/T$). This yields either:

$$m = \sqrt{4 + 144(\Delta \nu_{IF} T)^{1.3}}$$ \hspace{1cm} (Foschini et al.)  \hspace{1cm} (3-105)

or

$$m = \sqrt{9 + 161(\Delta \nu_{IF} T)^2}$$ \hspace{1cm} (Kazovsky)  \hspace{1cm} (3-106)

or

$$m = \sqrt{4 + 121(\Delta \nu_{IF} T)^{1.13}}$$ \hspace{1cm} (This work)  \hspace{1cm} (3-107)
In the region of interest, with the linewidth-to-bitrate ratio between 0.01 and 1, the difference in $m_{opt}$ from these three equations is less than 1. This is also illustrated in Figure 3-18.

Finally it should be noted that a bandwidth expansion factor $m$ of 1, as used in all papers that attempt to analyse receiver performance using pdf’s, is not a realistic value since it assumes matched filtering in the IF section. It can only be used as a reference for an analysis with larger values of $m$.

### 3.6.4 Procedure

The procedure that will be followed in order to eliminate the linewidth from the analysis has now become very simple. In all cases I will assume that with (3-103) the optimum IF bandwidth expansion factor $m_{opt}$ can be calculated. Using this value of $m$ will assure an unnoticeable BER-floor, while at the same time the sensitivity will be maximal. Also, for every IF bandwidth the maximum permitted linewidth at a given bitrate can be calculated with (3-103). The relative IF bandwidth $m$ can thus be an independent parameter, effectively taking into account the laser linewidth which can be eliminated explicitly from the analysis.

### 3.7 Summary

A single-filter receiver with non-coherent IF-detection and integrate-and-dump post-detection filtering can be analysed using the probability density functions of the signals. The IF linewidth due to the laser phase noise can be eliminated from the analysis by properly adjusting the width of the IF bandpass filter. The width of this BPF, normalised to the bitrate, is the important bandwidth expansion factor $m$. In order to compensate for the low SNR caused by the large IF bandwidth, post-detection filtering is essential. The integrate-and-dump matched filter was modelled as a discrete finite-time integrator, taking $m$ independent samples of the IF detector output. However, for ISI-free reception this I&D-filter should be used in combination with pre-filtering having rectangular impulse response. For the IF-detector the two limiting cases of linear and quadratic detection have been treated. Finally, for obtaining a minimum overall error probability, the threshold of the decision circuit should be properly set at the crossing point of the two pdf’s.

The goal of the analysis was to determine all IF penalties associated with the detection scheme. This means that $P_e$ has to be found, with the SNR $\rho$ and the bandwidth expansion factor $m$ as independent parameters. This requires that the threshold setting is determined first, as a function of the same parameters.
Chapter 4

Single-filter detection

Introduction

Single-branch single-filter receivers with non-coherent IF-detection are the simplest type of coherent optical receiver. They can be used for the detection of both Amplitude Shift Keying (ASK) and Frequency Shift Keying (FSK) signals, which made them initially the most widely used receiver type. In particular at either lower bitrates (e.g. less than 500 Mbit/s) or for a large laser linewidth-bitrate ratio, the required filter bandwidth can become high through (3-103). It will then be difficult to accomodate two IF filters within the often limited receiver bandwidth, making it impossible to use dual-filter FSK detection. In many cases single-filter detection is therefore used for FSK demodulation, with the same performance as ASK.

The main difference between single-filter IF-detection and the other IF-detection schemes to be discussed in Chapter 5–7 is the fact that the detection process is not symmetrical for Mark and Space bits. For a transmitted Mark the pdf of the IF detector output is therefore different from the pdf in case of a transmitted Space. This in turn leads to a non-zero decision threshold setting in the decision process. Accurate setting of this decision process will prove to be very important for obtaining minimum BER performance of the receiver. In order not to confuse the effects on the receiver sensitivity due to the decision threshold setting and the IF bandwidth expansion, the two analyses are separated. First the receiver performance in the case of minimum IF bandwidth \( m=1 \) will be evaluated as a function of the decision threshold and the IF SNR \( \rho \). This gives a clear insight in the importance of optimum threshold settings, which can then be extended for the general case \( m > 1 \). The second analysis in section 4.2 concentrates on the effect of the increasing IF bandwidth, assuming optimum decision threshold setting.

A different problem, that will only be addressed briefly, is linear IF-detection. Sec-
4.4 summarises the few analytical results that can be obtained for this IF-detection scheme. However, general analysis of receivers with linear IF-detection is not possible and is therefore excluded.

4.1 Reference case \((m=1)\)

4.1.1 IF-detection

As a reference, the receiver performance in the case of minimum IF bandwidth \((m=1)\) will be evaluated first. This means that the effective Nyquist and Matched filtering takes place in the IF section, so a post-detection filter is not required. The pdf's of the BPF output amplitude \(r\) are then given by (3-23) and (3-24), with \(N_m = mN\) replaced by \(N\). This yields non-central and central chi pdf's with two degrees of freedom, since \(r\) is the root of a sum of two squares:

\[
\begin{align*}
    r_M &= \sqrt{(A + z)^2 + y^2} \quad (4-1) \\
n_M(r) &= p_{\chi'}(r | 2, \frac{A^2}{N}) \\
      &= \frac{r}{N} \exp\left(-\frac{r^2 + A^2}{2N}\right) I_0\left(\frac{rA}{N}\right) \quad (4-2) \\
r_s &= \sqrt{x^2 + y^2} \\
p_s(r) &= p_{\chi}(r | 2) \\
      &= \frac{r}{N} \exp\left(-\frac{r^2}{2N}\right) \quad (4-4)
\end{align*}
\]

Since \(m = 1\) the SNR in the IF passband is equal to \(\rho\). As explained in section 3.2 the phase is of no importance in the case of non-coherent IF-detection. The IF detector will then be used for peak detection of the IF signal, yielding a baseband signal \(z\) proportional to \(|r|\) in the case of a linear detector, and \(r^2\) in the case of a quadratic detector.

Linear IF-detection

Since \(z = |r|\) the pdf's of \(z\) will be identical to the two pdf's of \(r\) given above. For a transmitted Mark the pdf and the first two moments are then given by:

\[
\begin{align*}
p_M(z) &= p_{\chi'}(z | 2, \frac{A^2}{N}) \\
     &= \frac{z}{N} \exp\left(-\frac{z^2 + A^2}{2N}\right) I_0\left(\frac{zA}{N}\right) \quad (4-5)
\end{align*}
\]
4.1 REFERENCE CASE (m=1)

\[ E\{z\} = A + \frac{N}{2A} \quad (SNR \gg 1) \quad (4-6) \]
\[ E\{z^2\} = A^2 + 2N \quad (4-7) \]
\[ \text{var}(z) = N - \left( \frac{N}{2A} \right)^2 \quad (SNR \gg 1) \quad (4-8) \]

For a transmitted Space:
\[ p_S(z) = p_X(z \mid 2) \]
\[ = \frac{z}{N} \exp \left( -\frac{z^2}{2N} \right) \quad (4-9) \]
\[ E\{z\} = \sqrt{\frac{\pi}{2}} N \quad (4-10) \]
\[ E\{z^2\} = 2N \quad (4-11) \]
\[ \text{var}(z) = \left( 2 - \frac{\pi}{2} \right) N \quad (4-12) \]

Quadratic IF-detection

Now \( z = r^2 \), so \( z \) can be written as:
\[ z_M = (A + x)^2 + y^2 \quad (4-13) \]
\[ z_S = x^2 + y^2 \quad (4-14) \]

Both \( z_M \) and \( z_S \) are sums of squares, and will thus have a (non-central) chi-square distribution, with - in this case - two degrees of freedom. The pdf’s can be derived from the general expressions derived in Appendix A.3 (A.3-4) and (A.3-18). The noncentrality parameter \( \lambda \) is found with (A.3-5) to be equal to \( A^2 / N \). This yields the pdf’s\(^1\)

\[ p_M(z) = p \left( z \mid 2, \frac{A^2}{N} \right) \]
\[ = \frac{1}{2N} \exp \left( -\frac{z + A^2}{2N} \right) I_0 \left( \frac{\sqrt{2}A}{N} \right) \quad (4-15) \]
\[ E\{z\} = A^2 + 2N \quad (4-16) \]
\[ E\{z^2\} = A^4 + 8NA^2 + 8N^2 \quad (4-17) \]
\[ \text{var}(z) = 4NA^2 + 4N^2 \quad (4-18) \]

\(^1\)The pdf’s of the quadratic detector output can also be obtained from (4-5) and (4-9) using the transformation \( z = r^2 \). The pdf then follows from

\[ p(z) = p_r(\sqrt{z}) \frac{dr}{dz} \quad \frac{dr}{dz} = \frac{1}{2\sqrt{z}} \]
\[ p_S(z) = p(z \mid 2) = \frac{1}{2N} \exp \left( -\frac{z}{2N} \right) \]  
\[ E\{z\} = 2N \]  
\[ E\{z^2\} = 8N^2 \]  
\[ \text{var}(z) = 4N^2 \]  

Note that the pdf in the case of a Space, which is a negative exponential function, has a non-zero probability for \( z = 0 \). The mean values of the quadratic detector output are furthermore identical to the second moments - or total power - of the linear detector output. This is of course due to the transformation \( z = r^2 \), but it illustrates clearly that the quadratic detector is indeed a power detector. In addition, the variance is equal to half the non-signal components of \( E\{z^2\} \). This is a consequence of a half of the noise power being part of the dc output, as illustrated in section 3.3, whereas the variance only accounts for the non-dc noise powers.

### 4.1.2 Decision threshold

The likelihood ratios of the detector outputs can be found with (3-81) from section 3.5, by dividing \( p_M(z) \) and \( p_S(z) \). With \( b \) the threshold value of \( u \), this yields for a linear and quadratic detector, respectively, the equations:

\[ I_0 \left( \frac{b \frac{A}{N}}{2N} \right) \exp \left( -\frac{A^2}{2N} \right) = 1 \quad \text{(linear)} \]  
\[ I_0 \left( \frac{\sqrt{b} \frac{A}{N}}{2N} \right) \exp \left( -\frac{A^2}{2N} \right) = 1 \quad \text{(quadratic)} \]  

Unfortunately these transcendental equations cannot be solved analytically. It is possible to compute the threshold using the series expansion of the modified Bessel function. The numerical procedure for this calculation is explained in Appendix B.1. However, for system studies it is more useful to have some sort of analytical expression for the threshold, so a convenient approximation method should be derived.

In order to limit the amount of work, it is interesting to note that the two threshold equations above are identical apart from the threshold variables \( b \) and \( \sqrt{\hat{b}} \). They are thus identical when \( b_{\text{lin}} = \sqrt{b_{\text{qdr}}} \).

First, several new variables will be defined, the first of which is the normalised general threshold \( Y \). In the case of quadratic detection \( Y = b_{\text{qdr}}/2N \), while for linear detection the relation is \( Y = b_{\text{lin}}^2/2N \). By using \( Y \) the distinction between linear and quadratic IF-detection has disappeared, giving one general equation to be solved. Using \( Y \) and the IF SNR \( \rho \), (4-24) can be written as

\[ I_0(2\sqrt{Y \rho}) = \exp(\rho) \]  

(4-25)
4.1 REFERENCE CASE \((m=1)\)

where \(Y\) is the independent variable that will be used in the expressions for the BER. In the present case however, it is more convenient to write the threshold equation as

\[
I_0(\zeta \rho) = \exp(\rho) \quad \text{with} \quad \zeta = 2 \sqrt{\frac{Y}{\rho}} \quad (4-26)
\]

Using the definition of \(Y\), the relation between \(b_{qdr}\) and \(\zeta\) is given by

\[
b_{qdr} = \zeta^2 \cdot \frac{A^2}{4} \quad (4-28)
\]

\(\zeta^2\) can thus be regarded as the threshold deviation factor, a measure for the threshold deviation from the limiting value \(A^2/4\) in the case of quadratic detection. For linear detectors \(\zeta\) gives the deviation from \(A/2\):

\[
b_{lin} = \zeta \cdot \frac{A}{2} \quad (4-29)
\]

Large SNR.

Equation (4-26) can not be solved analytically, so the asymptotic expansion (Ab.9.7.1) of the modified Bessel function, valid for large arguments \(z\), will be used.

\[
I_\nu(z) = \frac{\exp(z)}{\sqrt{2\pi z}} \left\{ 1 - \frac{4\nu^2 - 1}{8z} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)}{2! (8z)^2} - \ldots \right\} \quad (4-30)
\]

When the SNR \(\rho\) has operational values, \(z\) is in the order of 60 or more; \(\nu\) is equal to 0. It is then allowed to use only the expression outside the brackets\(^2\), yielding

\[
\frac{\exp(\zeta \rho)}{\sqrt{2\pi \zeta \rho}} = \exp(\rho) \quad (4-31)
\]

which can be rewritten into

\[
\zeta - 1 = \frac{\ln 2\pi \zeta \rho}{2\rho} \quad (4-32)
\]

From here two different methods can be used. The first is a rapidly converging iteration, described in Appendix B.2, which gives an acceptable approximation limited only by the accuracy of (4-30). A second method makes use of the fact that \(\zeta\) will be very close to 1 in all cases of practical interest. In a first-order approximation \(\ln \zeta\) in (4-32) may then be replaced by \(\zeta - 1\), leading to:

\[
\zeta - 1 = \frac{\ln 2\pi \rho + \ln \zeta}{2\rho} \simeq \frac{\ln 2\pi \rho + (\zeta - 1)}{2\rho}
\]

\[
\zeta = 1 + \frac{\ln 2\pi \rho}{2\rho - 1} \quad (4-33)
\]

\(^2\)For \(\rho=60\) the relative error is \(0.21\%\), for \(\rho=80\) \(0.16\%).
Using (4-28) and (4-27) the normalised and real thresholds of the detector output become

\[ Y = (1 + \frac{\ln 2\pi \rho}{2\rho - 1})^2 \frac{\rho}{4} \]  
(4-34)

\[ b_{qdr} = (1 + \frac{\ln 2\pi \rho}{2\rho - 1})^2 \frac{A^2}{4} \]  
(4-35)

\[ b_{tin} = (1 + \frac{\ln 2\pi \rho}{2\rho - 1}) \frac{A}{2} \]  
(4-36)

For \( \rho \) tending to infinity the threshold deviation factor \( \zeta \) approaches 1, so the thresholds \( b_{qdr} \) and \( b_{tin} \) become equal to \( A^2/4 \) and \( A/2 \), respectively. For moderate values of the SNR \( \zeta \) will always be larger than 1, which can be explained with the help of figure 4-1.

![Diagram](image)

Figure 4-1: Illustration of the a posteriori probabilities of the output of a linear IF detector with a fixed detection threshold at \( A/2 \).

The only requirement for optimum detection remains the condition (3-82):

\[ p_S(z) = p_M(z) \]  
(4-37)

Both \( p_S(z) \) and \( p_M(z) \) can be calculated from the partial error integrals \( P_S(z) \) and \( P_M(z) \) in accordance with (3-78)

\[ p_S(z) = \frac{dP_S(z)}{dz} \]  
(4-38)

\[ p_M(z) = \frac{dP_M(z)}{dz} \]  
(4-39)
After IF-detection the receiver will decide that a Space was sent when \( |z| = \sqrt{r^2 < b} \). With \( b \) midway between 0 and \( A \) at \( A/2 \), \( P_S \) is thus equivalent to the area 1 within the circle with radius \( A/2 \), centered at \((0,0)\). Consequently it will be decided that a Mark was transmitted when \( |z| > b \), so \( P_M \) is equivalent to the integrated probabilities of the areas 2 and 3. When the a priori Mark and Space probabilities are identical, circles 1 and 2 yield the same probability. The a posteriori Mark probability of the linear detector output is then larger by the probability density integrated over area 3. For larger SNR the probability densities in area 3 will be very small, but not negligible. Since the partial error rates are very steep and nearly antisymmetrical, the pdf’s for \( b = A/2 \) are thus not identical.

In order to obtain equal pdf’s (or differential partial error rates) area 1 should be increased. Increasing the radius of area 1 yields a uniform increase of \( P_S \) since \( p_S \) has rotational symmetry around \((0,0)\). The Mark partial error rate \( P_M \) on the other hand decreases much more slowly, which can be explained as follows. The pdf \( p_M \) has rotational symmetry around \((A,0)\). The decrease of \( P_M \) due to the reduction of areas 2 and 3 (being the area between the straight and dashed circles in Figure 4-1) is therefore much smaller since the main contribution comes from the area around the l-axis between the straight and dashed circles around \((0,0)\). In order to obtain equal differential partial error rates \( dP_s(z)/dz \) and \( dP_M(z)/dz \) (and thus \( p_S(z) = p_M(z) \)) the increase in radius of area 1 must be relatively large, yielding a decision threshold that can be substantially larger than \( A/2 \).

Finally it should be stressed that in the above analysis the actual values of \( P_S \) and \( P_M \) are not important. The only criterion remains equal pdf’s and differential partial error rates. However, under operational conditions the two partial error rates will feature almost identical but antisymmetrical behaviour around the optimum decision threshold. This means that in case of optimum detection also the two partial error rates \( P_S \) and \( P_M \) will be nearly identical\(^3\)

\[
P_S \approx P_M \quad (4-40)
\]

Small SNR.

For small SNR the modified Bessel function in (4-26) can be approximated by taking the first two terms of the series expansion *(Gr. 8.445)*:

\[
I_\nu(z) = \sum_{k=0}^{\infty} \frac{1}{k!(\nu + k)!} \left( \frac{z}{2} \right)^{\nu+2k} \quad (4-41)
\]

\(^3\)Numerical analysis in sections 4.1.4 and 4.3 reveals that the difference between \( P_S \) and \( P_M \) at optimum decision threshold is in the order of a few percent maximum, depending upon the SNR and \( n \).
which yields

\[ I_0(z) \simeq 1 + \frac{z^2}{4} \simeq \exp \left( \frac{z^2}{4} \right) \]  \hspace{1cm} (4-42)

The threshold equation then reduces to

\[ \exp \left( \frac{\zeta^2 \rho^2}{4} \right) = \exp(\rho) \]  \hspace{1cm} (4-43)

which yields

\[ \zeta = \sqrt{\frac{4}{\rho}} \quad Y = 1 \]

\[ b_{qdr} = \frac{A^2}{\rho} = 2N \]  \hspace{1cm} (4-44)

\[ b_{lin} = \frac{A}{\sqrt{\rho}} = \sqrt{2N} \]  \hspace{1cm} (4-45)

Using normalisation to the amplitude, the threshold can be written as \( b_{qdr}/A^2 = 1/\rho \). For very low SNR the two pdf's become nearly identical, so the threshold should be set to a value much lower than the signal amplitude \( A \) or \( A^2 \).

Comparison with other results

The only other analytical expression for the threshold setting with an acceptable accuracy can be found in Schwartz, Bennett and Stein [98, p.291]. They give the approximation for the threshold of a linear detector (using my own notation) as

\[ b_{lin} = \sqrt{2N} \sqrt{1 + \frac{\rho}{4}} = \frac{A}{2} \sqrt{1 + \frac{4}{\rho}} \]  \hspace{1cm} (4-46)

For a quadratic IF detector the threshold then becomes

\[ b_{qdr} = 2N \left( 1 + \frac{\rho}{4} \right) = \frac{A^2}{4} \left( 1 + \frac{4}{\rho} \right) \]  \hspace{1cm} (4-47)

For large SNR the thresholds for linear and quadratic IF-detection become equal to \( A/2 \) and \( A^2/4 \), respectively, and equivalently for small SNR to \( \sqrt{2N} \) and \( 2N \). These results are indeed identical to the ones derived in the previous sections. However, for intermediate values of \( \rho \) the expression of Schwartz et al. is less accurate.

One other threshold approximation is derived by Salz [97]. He uses relatively tight exponential upper bounds, leading to the expression for the threshold of an ASK-receiver with quadratic IF-detection (again adapted to my own notation):

\[ Y_{qdr} = \frac{\rho}{4} \left( 1 + \frac{1}{\rho} \right)^2 \]  \hspace{1cm} (4-48)
\[ b_{qdr} = \frac{A^2}{4} \left( 1 + \frac{1}{\rho} \right)^2 = \frac{A^2}{4} \left( 1 + \frac{2}{\rho} + \frac{1}{\rho^2} \right) \] (4-49)

Neglecting the quadratic component it is easy to see that the deviation from the 'infinite' threshold is always half as large as with the result of Schwartz. Since Schwartz is already too close to the 'infinite' threshold compared to the accurate numerical result, the method of Salz is even more inaccurate.

Table 4-1: Comparison of the results of different methods for approximating the threshold of a single-filter receiver with quadratic IF-detection and \( m=1 \). The relative error of the results is also given.

<table>
<thead>
<tr>
<th>SNR ( \rho )</th>
<th>exact solution (4-24)</th>
<th>analytical approximation (4-35)</th>
<th>Schwartz' error formula (4-47)</th>
<th>Salz' error formula (4-49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0.2861</td>
<td>0.2862</td>
<td>0.2750</td>
<td>4.0</td>
</tr>
<tr>
<td>60</td>
<td>0.2755</td>
<td>0.2755</td>
<td>0.2667</td>
<td>3.3</td>
</tr>
<tr>
<td>80</td>
<td>0.2699</td>
<td>0.2699</td>
<td>0.2625</td>
<td>2.8</td>
</tr>
<tr>
<td>100</td>
<td>0.2664</td>
<td>0.2665</td>
<td>0.2600</td>
<td>2.5</td>
</tr>
<tr>
<td>120</td>
<td>0.2640</td>
<td>0.2641</td>
<td>0.2583</td>
<td>2.2</td>
</tr>
<tr>
<td>140</td>
<td>0.2623</td>
<td>0.2623</td>
<td>0.2571</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 4-1 gives a comparison between the exact threshold calculated using the algorithm described in Appendix B.1, the analytical approximation of equation (4-35) and the equations of Schwartz et al. (4-47) and Salz. This table shows the high accuracy of the analytical approximation - deviations in the fourth digit are only due to rounding of the result. The formulae of Schwartz and Salz, on the other hand, introduce inaccuracies of several percent. Since there is no literature reference with respect to the equation used by Schwartz, I am not able to discuss it in more detail. The result of Salz is clearly not to be used in an accurate analysis and good system design.

4.1.3 The Bit-Error Rate

Knowing the decision threshold, the BER can now be calculated with (3-78)

\[ P_e = P(0) \cdot P_S + P(1) \cdot P_M \] (4-50)
or, when Mark and Space have equal a priori probabilities.

\[
P_e = \frac{1}{2} \left\{ \int_{b_{\text{lin}}}^{\infty} p_S(z)dz + \int_{0}^{b} p_M(z)dz \right\}
\]  

(4-51)

In order to calculate the total error probability both integrals must be solved, denoted by \(P_S\) and \(P_M\). For linear IF-detection \(P_S\) is defined by the definite integral of the pdf \(p_x(z|2)\) (4-9)

\[
P_S = \int_{b_{\text{lin}}}^{\infty} \frac{1}{N} \exp \left( -\frac{z^2}{2N} \right) dz = \int_{b_{\text{lin}}/2N}^{\infty} \exp(-y)dy
\]

(4-52)

For quadratic detection the pdf \(p(z|2)\) (4-19) should be used:

\[
P_S = \int_{b_{\text{qdr}}}^{\infty} \frac{1}{2N} \exp \left( -\frac{z}{2N} \right) dz = \int_{b_{\text{qdr}}/2N}^{\infty} \exp(-y)dy
\]

(4-53)

Since \(b_{\text{qdr}}\) is equal to \(b_{\text{lin}}^2\), both equations are identical when using the normalised general threshold \(Y\). This implies that the BER characteristics of receivers with either linear or quadratic IF-detection are identical, as long as the threshold is adapted to the detector-order\(^4\). The integral can now readily be solved as

\[
P_S = \int_{Y}^{\infty} \exp(-y)dy = \exp(-Y)
\]

(4-54)

Calculating \(P_M\) is of course more difficult, due to the modified Bessel functions in the pdf’s. The two integrals for linear and quadratic IF detection can be found using (4-5) and (4-15) respectively:

\[
P_M = \int_{0}^{b_{\text{lin}}} \frac{z}{N} \exp \left( -\frac{z^2 + A^2}{2N} \right) I_0 \left( \frac{zA}{N} \right) dz
\]

(4-55)

\[
= \int_{0}^{b_{\text{qdr}}} \frac{1}{2N} \exp \left( -\frac{z + A^2}{2N} \right) I_0 \left( \frac{\sqrt{2}A}{N} \right) dz
\]

(4-56)

The ‘classical’ way of treating these integrals is the one introduced by Marcum [75], using the \(Q\)-function, defined by

\[
Q(\alpha, \beta) \triangleq \int_{\beta}^{\infty} y \exp \left( -\frac{y^2 + \alpha^2}{2} \right) I_0(\alpha y)dy
\]

(4-57)

\(^4\)Since a change of detector order is equivalent to a transformation of the parameter in the pdf’s, it is easy to show that the threshold setting must be changed using the same transformation. This is valid for all detector characteristics.
4.1 REFERENCE CASE (m=1)

Using the same generalised threshold $Y$ as for $P_S$, $P_M$ can now be expressed in terms of Marcum’s Q-function:

$$
P_M = \int_0^{\sqrt{2Y}} y \exp\left(-\frac{y^2 + 2\rho}{2}\right) I_0\left(\sqrt{2\rho}y\right) dy = 1 - Q\left(\sqrt{2\rho}, \sqrt{2Y}\right)$$

(4-58)

The disadvantage of the Q-function is the fact that it does not give any insight in the behaviour of $P_M$. It is on the other hand well documented for numerical evaluation. Shnidman [102] gives a particularly good overview of the different forms of the Q-function, as well as a fast numerical method for solving it. Schwartz et al. [98, App.A] give a comprehensive description of the properties of the Q-function.

A different method is given by Panter\(^5\) [87, p.708], using successive integration by parts (see Appendix C.1). With this method $P_M$ can be written as an infinite sum of Bessel functions (C.1.7):

$$
P_M = \exp(-\rho - Y) \sum_{k=1}^{\infty} \left(\frac{Y}{\rho}\right)^k \frac{I_k(2\sqrt{\rho Y})}{k!}$$

(4-59)

Using the series expansion of the Bessel function ($Ab.9.6.10$) or ($Gr.8.445$) this can be rewritten into

$$
P_M = \exp(-\rho - Y) \sum_{k=0}^{\infty} Y^{k+1} \sum_{j=0}^{\infty} \frac{(\rho Y)^j}{j!(j + k + 1)!}$$

$$
= 1 - \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{\infty} \frac{Y^k}{k!}$$

(4-60)

Finally - by properly interchanging the summation order, regrouping of the terms and use of the incomplete gamma function - $P_M$ can also be expressed as a sum of gamma functions (C.2.23):

$$
P_M = 1 - \exp(-\rho) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \frac{\Gamma(1 + j, Y)}{\Gamma(1 + j)}$$

(4-61)

It can be checked that all three expressions of $P_M$ behave properly for the three limiting cases: for $\rho = 0$ $P_M = 1 - P_S$; for $m$ going to infinity $P_M$ becomes 0; for the threshold $Y = 0$ $P_M$ becomes 0.

---

\(^5\)Although Panter gives the correct method, the final result contains a (typing-)error: the argument of the modified Bessel function must be $Ab/N$ instead of $A^2b/N$. 
Combining (4-54) and (4-60) the total BER of a single filter receiver is now given by

\[
P_e = \frac{1}{2} \left\{ 1 + \exp(-Y) - \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{j} \frac{Y^k}{k!} \right\}
\]

\[
= \frac{1}{2} \left\{ 1 - \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{j-1} \frac{Y^{1+k}}{(1+k)!} \right\}
\]

(4-62)

A convincing and powerful check on this result is obtained by differentiating \(P_e\) to \(Y\). This yields

\[
\frac{dP_e}{dY} = \frac{1}{2} \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \left\{ \sum_{k=0}^{\infty} \frac{y^k}{k!} - \sum_{k=0}^{\infty} \frac{y^{k+1}}{(k+1)!} \right\}
\]

\[
= \frac{1}{2} \exp(-\rho - Y) \left\{ \sum_{j=0}^{\infty} \frac{\rho^j}{j!} - \sum_{j=0}^{\infty} \frac{(\rho Y)^j}{(j!)^2} - 1 + 1 \right\}
\]

(4-63)

A minimumin \(P_e\) is attained for \(dP_e/dY=0\), in which case the above equation can be reduced - as shown in Appendix C.3 - to

\[
\exp(\rho) = I_0(2\sqrt{\rho Y})
\]

(4-64)

which is exactly the threshold equation (4-24). This confirms that the BER indeed features an extreme for the optimum threshold setting. With the above expression it is now possible to evaluate the BER as a function of both the threshold setting \(Y\) - where the optimum setting is known through (4-35) - and the IF SNR \(\rho\).

### 4.1.4 Sensitivity

#### Optimum decision threshold

With the optimum threshold setting the sensitivity of the receiver will be maximal or, equivalently, the required SNR \(\rho\) minimal\(^6\). Figure 4-2 shows the BER-curve in the case of optimum threshold setting compared to the BER-curve for (ideal) synchronous detection given by (2-15). For a BER of \(10^{-9}\) the required SNR \(\rho\) is equal to 76.63 (18.84 dB), 0.26 dB higher than the 72.05 (18.58 dB) for coherent IF-detection. This is identical to the results obtained by Noé [82] and Enning et al. [20].

Figure 4-3 shows the BER as a function of the decision threshold for different signal-to-noise ratio's of a single-filter receiver with quadratic IF-detection. (For linear

\(^6\)The associated optical input power \(P_s\), which is the real sensitivity of the receiver, can be obtained from \(\rho\) using equation (1-20).
Figure 4-2: Bit-error-rate curves of a single-filter ASK receiver as a function of the SNR $\rho$. $m=1$, $Y$ is the generalised threshold. For comparison the curve in case of ideal synchronous IF detection is also indicated.

IF-detection one should simply take the root of each value of $b_{qdr}$. For increasing $\rho$, the line $b_{opt}(\rho)$ clearly approaches the limiting value of $b = A^2/4$, in accordance with (4-35). For $\rho=80$ the two individual contributions to $P_e$, $P_S/2$ and $P_M/2$, have been plotted. This shows that the behaviour of $P_S$ and $P_M$ around the optimum threshold setting is nearly identical, apart from the sign of the derivatives. There is no type of error that dominates, and at $b_{opt}$ the two errors $P_S$ and $P_M$ are almost identical. This is in accordance with (4-40) on page 89, giving almost equal a posteriori partial error
Figure 4-3: The overall BER of a single-filter ASK receiver with quadratic IF detection, as a function of the decision threshold and the SNR $\rho$. $b_{opt}(\rho)$ is the line interconnecting the points of optimum threshold—minimum BER, while $b_{\infty}$ is the often used fixed threshold for infinite SNR. $m=1$.

rates$^7$.

Fixed decision threshold

In principle the optimum threshold setting should always be selected in order to obtain maximum sensitivity. This has however one big disadvantage: the threshold setting is SNR-dependent. In practical systems it is extremely difficult to measure the SNR; measuring the signal amplitude is easy, but measuring the signal power normalised to the noise level is much more difficult. A fixed threshold is therefore commonly used. This can be explained by the fact that most (coherent optical) receivers use some sort

---

$^7$This result can be used to advantage when numerically calculating the overall BER. In general the expression for $P_M$ is much more complex than the expression for $P_S$. So, by using the latter in combination with the proper threshold setting, the overall BER can be found without solving $P_M$. This can save a considerable amount of computing time.
of Automatic Gain Control (AGC) or amplitude limiting, yielding a baseband output signal of constant average amplitude. At the input of the decision circuit the threshold can then be set to a fixed value, usually the limit for large SNR of $Y = \rho / 4$, since a receiver is normally operated at low bit-error rates\(^8\).

![Graph showing the influence of a fixed non-optimum threshold at the value for infinite SNR.]  

**Figure 4-4:** The influence of a fixed non-optimum threshold at the value for infinite SNR, for a single-filter ASK receiver with quadratic IF-detection. The partial error rates $P_S$ and $P_M$ are indicated by different shades, showing that $P_S$ is considerably larger.

The easiest way to see the effect of a fixed threshold $Y = \rho / 4$ on the overall BER is by using the expression for $P_M$ using Marcum’s Q-function (4-58). For large $\alpha$ the Q-function $Q(\alpha, \beta)$ can be approximated by $1 - \frac{1}{2} \text{erfc}(\frac{\alpha}{2} - \frac{\beta}{2})$ [98, App.A], yielding

$$
P_M = 1 - Q(\sqrt{2\rho}, \sqrt{2Y}) \\
\simeq \frac{1}{2} \text{erfc}(\sqrt{\rho} - \sqrt{Y}) = \frac{1}{2} \text{erfc} \left( \frac{\sqrt{\rho}}{2} \right)
$$

(4-65)

Note that this exactly coincides with the result for the overall BER of ASK receivers with coherent IF-detection (2-15). This expression can again be approximated for large

---

\(^8\)At high bit-error rates this method introduces sensitivity penalties. Assuming a linear IF peak detector, the gain control input signal can be written using (4-6) as

$$
\hat{u}_{\text{peak}} = A \left( 1 + \frac{1}{4\rho} \right)
$$

For low SNR $\rho$ the AGC peak detector output increases relative to the infinite-SNR value $A$. This will be interpreted by the AGC as too high an input signal amplitude, whereupon the IF gain will be decreased. This results in a decision threshold setting that is relatively too high, yielding an increased BER.
\[ P_e = \frac{1}{2} \left\{ \exp \left( -\frac{\rho}{4} \right) + \frac{2}{\sqrt{\pi \rho}} \exp \left( -\frac{\rho}{4} \right) \right\} \]  
\[ \approx \frac{1}{2} \exp \left( -\frac{\rho}{4} \right) \]

This is the expression given by most references (e.g. [98, 106, 3, 119, 35]) as the theoretical performance of ASK receivers with non-coherent IF-detection in the case of large SNR. The required SNR \( \rho \) for a BER of \( 10^{-9} \) is 80.12 (19.04 dB), 0.46 dB higher compared to the receiver with coherent IF detection and 0.19 dB higher than the 76.63 of the receiver with optimum threshold. The BER-curve is included in Figure 4-2.

Figure 4-5: Illustration of the required SNR-improvement relative to the optimum threshold situation, in order to obtain the same BER at a fixed threshold setting.

From (4-67) and Figure 4-4 it is clear that, in the case of a fixed threshold at the value for infinite SNR \( (b_{\infty}) \), the errors of \( P_S \) ('false alarms') dominate. Figure 4-5 shows that \( P_M \) then contributes less than 2% to the overall BER, thus justifying the
4.2. NON-OPTIMUM IF BANDWIDTH ($m > 1$)

4.2.1 Post-detection filtering

When $m$ is larger than 1 the IF SNR $\xi$ reduces to $\rho/m$, so additional post-detection filtering must eliminate the extra amount of noise. In section 3.4 it has been shown that it is possible to model the post-detection filter as a finite-time integrator, taking $m$ independent samples of the IF-detector output signal (3-72) or (3-73). In the case of quadratic IF-detection and a transmitted Mark $u$ is given by (3-74):

$$u = \frac{1}{2m} \sum_{i=1}^{2m} (A + x_i')^2$$

(4-68)

The noise samples $x_i'$ are normally distributed with variance $2N_m = 2mN$. In the case of linear IF-detection $u$ is found by

$$u = \frac{1}{m} \sum_{i=1}^{m} \sqrt{(A + x_i)^2 + y_i^2}$$

(4-69)

In both cases $u$ is the sum of $m$ samples of the detector output $z$. The pdf of $u$ can thus be obtained by $m$-fold convolution of the pdf of $z$.

However, there is an important difference between the two different detector types. As will be shown in the next section, analysis of the quadratic detector with post-detection filtering can be performed in an analytical way without any simplification or approximation. This is not the case for the linear detector, due to the root in (4-69). Linear IF-detection will therefore be treated separately in section 4.4.

**Quadratic IF-detection**

For quadratic IF-detection the pdf's (4-15) and (4-19) of the reference case $m=1$ should be convolved $m$ times in order to obtain the pdf of $u$. This leads to central and non-central chi-square distributions with $2m$ degrees of freedom and noncentrality parameter $A^2/N$. Details on these pdf's can be found in Appendix A.3. For a transmitted Mark the pdf and moments are:
\[ p_M(u) = p\left( u \mid 2m, \frac{A^2}{N} \right) = \frac{1}{2N} \left( \frac{u}{A^2} \right)^{m-1} \exp \left( -\frac{u + A^2}{2N} \right) I_{m-1} \left( \frac{A}{N} \sqrt{u} \right) \] (4-70)

\[ E\{u\} = A^2 + 2mN \] (4-71)

\[ E\{u^2\} = 4N(m+1)(mN + A^2) + A^4 \] (4-72)

\[ \text{var}(u) = 4NA^2 + 4mN^2 \] (4-73)

And finally for a transmitted Space:

\[ p_S(u) = p(u \mid 2m) = \frac{1}{2N \Gamma(m)} \left( \frac{u}{2N} \right)^{m-1} \exp \left( -\frac{u}{2N} \right) \] (4-74)

\[ E\{u\} = 2mN \] (4-75)

\[ E\{u^2\} = 4m(m+1)N^2 \] (4-76)

\[ \text{var}(u) = 4mN^2 \] (4-77)

Some simple checks can be performed on these results. All expressions for both pdf’s and moments reduce to the ones of the previous section by setting \( m=1 \). The results for a transmitted Mark reduce to those of a Space when \( A=0 \).

Compared to the pdf’s of the IF detector output in section 3.3 the degrees of freedom have increased by a factor \( m \), while the noncentrality parameters have been decreased by the same factor. Relative to the variances of the detector outputs (3-38) and (3-42), the variances of the post-detection filter outputs are a factor \( m \) lower. This shows the ultimate reason for using post-detection filtering when the IF bandwidth is large: the IF bandwidth expansion factor \( m \) is compensated by the post-detection improvement factor \( m \) when proper I&D filtering is used. However, a look at the variance - especially of the quadratic detector with post-detection filtering - reveals that the Noise-times-Noise component still depends on \( m \). This justifies analysis of the effective bandwidth-dependent sensitivity penalty in the next sections.

### 4.2.2 The decision threshold for quadratic detection

The decision threshold can be obtained by setting \( p_S(u) = p_M(u) \) according to (3-81). This yields:

\[ I_{m-1} \left( \frac{\sqrt{b_{\text{quad}}A}}{N} \right) = \frac{1}{\Gamma(m)} \left( \frac{\sqrt{b_{\text{quad}}A}}{2N} \right)^{m-1} \exp \left( \frac{A^2}{2N} \right) \] (4-78)

\[ I_{m-1}(2\sqrt{Y\rho}) = \frac{1}{\Gamma(m)} (Y\rho)^{m-1} \exp \rho \] (4-79)
where $Y$ is the normalised general threshold $b_{qdr}/2N$. This equation can be solved in the same way as already presented for the reference situation $m=1$, by using either an exact numerical computation (detailed in Appendix B.1) or the less accurate approximate iteration from Appendix B.2.

Large SNR

It is however more interesting to try to obtain an analytical expression for the threshold setting. Again the threshold deviation factor $\zeta$ is used, and the threshold equation (4-79) is simplified using the approximation for the modified Bessel function with large argument (4-30):

$$\frac{\exp(\zeta \rho)}{\sqrt{2\pi \zeta \rho}} = \frac{\exp \rho}{\Gamma(m)} \left( \frac{\zeta \rho}{2} \right)^{m-1}$$  \hspace{1cm} (4-80)

By taking the natural log at both sides, and after approximating $\ln(\zeta - 1)$ as $\zeta - 1$, the resulting expression for $\zeta$ is

$$\zeta - 1 = \frac{\ln \pi + (3 + 2m) \ln 2 + (2m - 1) \ln \rho - 2 \ln \Gamma(m)}{2\rho - 2m + 1}$$  \hspace{1cm} (4-81)

This expression reduces to (4-33) for $m=1$. A further simplification can be obtained through substitution of $\ln \Gamma(m)$ by the approximation given by (Gr.8.343), although this degrades the accuracy slightly. The final result is

$$\zeta - 1 = \frac{(2m - 1) \ln(\rho/2m) + \ln 2 + 2m}{2\rho - 2m + 1} \hspace{1cm} (\rho \gg 1)$$  \hspace{1cm} (4-82)

The optimum threshold settings can be obtained in the same way as for $m=1$ through the relations

$$Y = \frac{\zeta^2 P}{4}$$  \hspace{1cm} (4-83)

$$b_{qdr} = \frac{\zeta^2 A^2}{4}$$  \hspace{1cm} (4-84)

Independent of $m$ the threshold deviation factor $\zeta$ approaches 1 for large SNR, so the threshold $b_{qdr}$ approaches $A^2/4$. At a given SNR the decision threshold shifts to higher values for increasing $m$, which can intuitively be explained with the help of Figure 4-1. With an increase of the IF bandwidth more noise will appear at the detector output, giving higher tails of the pdf's. There will consequently be a higher probability density in the area 3 of Figure 4-1, which must be compensated by a larger radius of circle 1. This is equivalent to a shift to a higher threshold.
Figure 4-6: Optimum decision threshold settings, as a function of the IF SNR \( \rho \) and the IF bandwidth expansion factor \( m \), of a single-filter ASK receiver with quadratic IF-detection. Dashed lines interconnect points \((\rho, m)\) of constant BER.

Small SNR

For small values of the SNR \( \rho \) the modified Bessel function can again be approximated using (4-41), which yields:

\[
I_{m-1}(z) \simeq \frac{1}{\Gamma(m)} \left( \frac{z}{2} \right)^{m-1} \left( 1 + \frac{z^2}{4m} \right)
\]

(4-85)

\[
\simeq \frac{1}{\Gamma(m)} \left( \frac{z}{2} \right)^{m-1} \exp \left( \frac{x^2}{4m} \right)
\]

(4-86)

Substituting this into the threshold equation gives

\[
\frac{1}{\Gamma(m)} \left( \frac{\zeta \rho}{2} \right)^{m-1} \exp \left( \frac{\zeta^2 \rho^2}{4m} \right) = \frac{\exp(\rho)}{\Gamma(m)} \left( \frac{\zeta \rho}{2} \right)^{m-1}
\]

(4-87)

which finally yields

\[
\zeta = \sqrt{\frac{4m}{\rho}} \quad \text{Y = m}
\]

\[
b_{qdr} = \frac{A^2}{m \rho} = \frac{2N}{m}
\]

(4-88)
4.2. NON-OPTIMUM IF BANDWIDTH \((m \geq 1)\)

Compared to the situation for \(m=1\) the threshold setting becomes even lower relative to the 'signal amplitude' (in fact there is not much Signal left for low SNR), which is caused by the broadening pdf's for increasing \(m\).

Accuracy of the approximation

The accuracy of the approximation for large SNR \((4.82)\) has been investigated. Figure 4-7 shows the absolute value of the relative error of the approximated normalised general threshold \(Y\) in \((4.83)\) compared to the exact solution of the threshold equation \((4.79)\). The exact solution has been computed by numerical iteration as in Appendix B.1. From this figure we see that the relative error in the range \((\rho, m)\) of interest is equal to \(10^{-3}\) or lower. It should be noted that the error becomes 0 at the bottom of the ridge and attains negative values in the upper right corner. It can thus be concluded that the approximation has a very high accuracy, and can be used for all practical values of the SNR \(\rho\) and bandwidth expansion factor \(m\).

Comparison with other results

To the best of my knowledge the analytical expressions presented here have not been derived before. In general, expressions for the threshold setting when \(m\) is not equal to 1 are extremely rare. The single expression I was able to find is by Salz [97, p.2203], who gives (with his notation converted into mine) the threshold for ASK detection with quadratic IF-detection:

\[
b = \left(1 + \frac{m}{\rho}\right)^2 \cdot \frac{A^2}{4}\]

(4.89)

Since this expression is derived using approximations and upper bounds, the accuracy can be expected to be rather low. One check that can be performed is the behaviour for large SNR, which yields \(A^2/4\). However, for operational values of \(\rho\) and \(m\) the deviations can be as high as tens of percents.

Two more papers have presented numerical results regarding the threshold setting as function of \(m\). The first is by Foschini, Greenstein and Vannucci [21], who give the threshold for the two situations \(m=5 \Leftrightarrow \rho=100\) and \(m=8 \Leftrightarrow \rho = 112.2\) (20.5 dB). They found these combinations as the nearest integers to the optimum solutions\(^9\) for

\(^9\)In an earlier stage of the investigations leading to this thesis I derived an expression for the threshold which closely matched the results obtained by Foschini, Greenstein and Vannucci. However, this expression, published in [41], was based on an approximation using \((C.2.8)\) and proved to be less accurate. It read:

\[
\frac{b}{2N} = \frac{\rho}{4} + m + (m-1)\ln\left(\frac{e}{2\pi}\right) - \frac{8}{\rho^2}(m-1)e
\]

The inaccuracy relative to the exact solution is only a few percent.
Figure 4-7: Contour plot, indicating the accuracy of the threshold settings computed with the approximation (4-82). The contours present the absolute error of the approximated $Y$ (4-83) relative to the exact solution of the threshold equation (4-79). Dots refer to the three special cases that are analysed in detail in Appendix B4: $m=1 \leftrightarrow \rho=80$, $m=5 \leftrightarrow \rho=100$ and $m=8 \leftrightarrow \rho=112$.

$m$ when the relative linewidth $\Delta \nu_{IF} T$ was 0.27 and 0.57 respectively. (Indeed (3-103) gives 5.12 and 8.33 for these values of the linewidth).

A second paper is by Humblet and Azizoğlu [50]. Although this paper deals with optical amplification where $m$ defines the relative bandwidth of an optical bandpass filter, the pdf’s and statistics are identical. They give threshold values for the two cases $m=1 \leftrightarrow \rho=76$ and $m=30 \leftrightarrow \rho=130$ respectively. The results of all of these simulations, as well as the exact results derived with (4-82) are listed in table 4-2.

In general the thresholds that follow from the simulations of Foschini et al. and Humblet et al. are lower than the exact thresholds computed by the program described in Appendix B.1.
Table 4-2: Comparison of the results of several decision threshold calculations for ASK single-filter receivers with quadratic IF-detection.

<table>
<thead>
<tr>
<th>Authors</th>
<th>$m$</th>
<th>$\rho$</th>
<th>$b_{\text{exact}}/A^2$</th>
<th>$b/A^2$</th>
<th>error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwartz, Bennett &amp; Stein</td>
<td>1</td>
<td>76</td>
<td>0.271</td>
<td>0.263</td>
<td>-3.0</td>
</tr>
<tr>
<td>Humblett &amp; Azizoğu</td>
<td>1</td>
<td>76</td>
<td>0.271</td>
<td>0.258</td>
<td>-4.8</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>130</td>
<td>0.584</td>
<td>0.346</td>
<td>-4.1</td>
</tr>
<tr>
<td>Foschini, Greenstein &amp; Vannucci</td>
<td>5</td>
<td>100</td>
<td>0.339</td>
<td>0.320</td>
<td>-5.6</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>112</td>
<td>0.372</td>
<td>0.347</td>
<td>-6.6</td>
</tr>
</tbody>
</table>

4.2.3 The Bit-Error Rate, quadratic detection

Again the two integrals in (3-78) must be solved (assuming equal a priori Mark and Space probabilities):

$$P_e = \frac{1}{2}P_s + \frac{1}{2}P_M$$

$$= \frac{1}{2} \left\{ \int_b^\infty p_s(u)du + \int_0^b p_M(u)du \right\}$$  \hspace{1cm} (4-90)

First, $P_S$ can easily be obtained from $p(u|2m)$ (4-74):

$$P_S = \int_{b_{\text{std}}}^\infty \frac{1}{2N\Gamma(m)} \left( \frac{u}{2N} \right)^{m-1} \exp \left( -\frac{u}{2N} \right) du$$  \hspace{1cm} (4-91)

$$= \int_Y \frac{1}{\Gamma(m)} y^{m-1} \exp(-y)dy$$  \hspace{1cm} (4-92)

With the standard integral (Gr.3.351-2)

$$\int_a^\infty y^n \exp(-by)dy = \exp(-ab) \sum_{k=0}^n \frac{n!}{k!} \frac{a^k}{b^{n-k+1}}$$  \hspace{1cm} (4-93)

the expression for $P_S$ can easily be found as:

$$P_S = \exp(-Y) \sum_{k=0}^{m-1} \frac{Y^k}{k!}$$  \hspace{1cm} (4-94)
or

$$P_S = \frac{\Gamma(m, Y)}{\Gamma(m)}$$  \hspace{1cm} (4-95)

The latter form is obtained directly from (4-94) since this is the definition of the incomplete gamma function $\Gamma(m, Y)$, see also Appendix A.3 (A.3-20) and (A.3-25). Useful approximations without the gamma functions are derived in Appendix C.2 (C.2-8):

$$P_S = e^{\sqrt{\frac{m}{2\pi}}} \left(\frac{Y}{m}\right)^m \frac{\exp[-(Y - m + 1)]}{(Y - m + 1)}$$  \hspace{1cm} (4-96)

The inaccuracy of this approximation is 8% for $m=1$ and lower for increasing $m$ and $Y$.

$P_M$ is obtained from the pdf $p(u|2m, \lambda)$ (4-70) with $\lambda$ equal to $A^2/N$:

$$P_M = \int_0^{b_{\alpha+2}} \frac{1}{2N} \left(\frac{u}{A}\right)^{\frac{m-1}{2}} \exp\left(-\frac{u + A^2}{2N}\right) I_{m-1} \left(\frac{A}{N} \sqrt{u}\right) du$$  \hspace{1cm} (4-97)

$$= \int_0^Y \left(\frac{y}{\rho}\right)^{\frac{m-1}{2}} \exp(-y - \rho) I_{m-1}(2\sqrt{\rho y}) dy$$  \hspace{1cm} (4-98)

The latter expression is identical to the definition of the generalised Marcum Q-function:

$$Q_m(\alpha, \beta) \triangleq \int_\beta^\infty z \left(\frac{z}{\alpha}\right)^{m-1} \exp\left(-\frac{z^2 + \alpha^2}{2}\right) I_{m-1}(\alpha z) dz$$  \hspace{1cm} (4-99)

$$Q_m(\sqrt{2\alpha}, \sqrt{2\beta}) \triangleq \int_\beta^\infty \left(\frac{z}{\alpha}\right)^{m-1} \exp(-z - \alpha) I_{m-1}(2\sqrt{\alpha z}) dz$$  \hspace{1cm} (4-100)

This leads to the compact expression

$$P_M = 1 - Q_m(\sqrt{2\rho}, \sqrt{2Y})$$  \hspace{1cm} (4-101)

Again, as for $m=1$, this expression can not be solved analytically and does not give too much insight. With the method of successive integration by parts, detailed in Appendix C.2, $P_M$ can be written as a sum of Bessel functions (C.2-19):

$$P_M = \exp(-\rho - Y) \sum_{k=0}^\infty \left(\frac{Y}{\rho}\right)^{\frac{k+m}{2}} I_{k+m}(2\sqrt{\rho Y})$$  \hspace{1cm} (4-102)

For $m=1$ this expression reduces to (4-59). While Shnidman [102, form(12)] gives this result as one of the elementary forms of the generalised Marcum Q-function. By
4.2. NON-OPTIMUM IF BANDWIDTH \((m>1)\)

writing out the Bessel functions in their series expansion one obtains

\[ P_M = \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{\infty} \frac{Y^{m+k+j}}{(m+k+j)!} \]  

(4-103)

\[ = 1 - \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{\infty} \frac{Y^k}{k!} \]  

(4-104)

The last sum in the latter expression is the definition of the incomplete gamma function \(\Gamma(1+j,Y)\) (A.3-25), so the expression for \(P_M\) can be written as:

\[ P_M = 1 - \exp(-\rho) \sum_{j=0}^{\infty} \frac{\rho^j \Gamma(m+j,Y)}{j! \Gamma(m+j)} \]  

(4-105)

Note that this is in accordance with the general rule given in Appendix A.3 (A.3-33) that \(P_M\) is a mixture of \(P_S\) distributions weighted by the Poisson distribution [62, ch.28] (Ab.26.4.25). With \(\rho = \lambda/2, \theta = 2m\) and \(P_S = \Gamma(2m,Y)\) (A.3-33) reduces to (4-105).

Starting with (4-104) some mathematical manipulations - explained in Appendix C.2 - yield an expression that is very useful for analysis of the overall BER, namely

\[ P_M = 1 - P_S - \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k+m-1)!} \]  

(4-106)

With this expression (and also with all other results for \(P_M\) given above) it is easy to check the proper behaviour for some of the limiting cases: For \(\rho\) equal to 0 all sums will be zero and \(P_M\) reduces to 1–\(P_S\). For \(m\) to infinity or the threshold \(Y\) equal to 0, \(P_M\) becomes 0.

Combining the results for \(P_S\) and \(P_M\) (4-106) the overall BER of the single-filter receiver can now be written as

\[ P_e = \frac{1}{2} \left\{ 1 - \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k+m-1)!} \right\} \]  

(4-107)

In Appendix C.3 it is verified that after differentiation to \(Y\) the minimum value for \(P_e\) is obtained when \(Y\) satisfies the threshold equation (4-79). With this result the BER can thus be calculated for all values of the IF bandwidth expansion factor \(m\), the normalised threshold setting \(Y\) and the IF SNR \(\rho\).
Figure 4-8: Bit-error-rate curves of a single-filter ASK receiver with quadratic IF-detection as a function of the SNR $\rho$ and the IF bandwidth expansion factor $m$.

4.3 Sensitivity and penalties ($m>1$)

4.3.1 Numerical evaluation

The expression for the BER (4-107) can be used for numerical evaluation of the receiver behaviour$^{10}$. With the optimum threshold setting (4-82) the BER-curves for different

---

$^{10}$The numerical evaluation is performed on an IBM mainframe computer, using simple Fortran programs. High accuracy double-precision variables are used throughout these programs. Infinite
Table 4-3: The required SNR $\rho$ for a BER of $10^{-9}$ of a single-filter ASK receiver with quadratic IF-detection and optimum decision threshold setting, as a function of the IF bandwidth expansion factor $m$. $\Delta SNR$ is the sensitivity penalty in dB relative to $m=1$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\rho$</th>
<th>$\Delta SNR$ dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76.63</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>81.40</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>85.12</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>88.30</td>
<td>0.62</td>
</tr>
<tr>
<td>5</td>
<td>91.12</td>
<td>0.75</td>
</tr>
<tr>
<td>6</td>
<td>93.70</td>
<td>0.87</td>
</tr>
<tr>
<td>7</td>
<td>96.07</td>
<td>0.98</td>
</tr>
<tr>
<td>8</td>
<td>98.28</td>
<td>1.08</td>
</tr>
<tr>
<td>9</td>
<td>100.37</td>
<td>1.17</td>
</tr>
<tr>
<td>10</td>
<td>102.35</td>
<td>1.26</td>
</tr>
</tbody>
</table>

$m$ can be calculated first. Figure 4-8 shows a gradual shift of these curves to lower sensitivity values with increasing $m$, which means that there is an increasing penalty as a function of $m$. The reason is the combination of non-white (and non-Gaussian) IF-detector output noise and integrate-and-dump post-detection filtering. An ideal post-detection filter with proper matched filter weighting of the noise characteristic would thus not give an IF-bandwidth dependent sensitivity penalty. With the I&D filter, the penalty was predicted in section 4.2, since the NxN variance of the output of this filter depends on $m$. Since the spectral shape of the NxN noise is triangular (see Figure 3-7) this penalty does not increase too fast with $m$. Figure 4-9 and table 4-3 both illustrate the optimum sensitivity and the sensitivity penalty $\Delta SNR$ as a function of $m$.

One can easily fit a curve through the values of $\Delta SNR$ for discrete values of $m$, as illustrated in Figure 4-9. This yields the compact relation

$$\Delta SNR = 0.708 \log m + 0.541 \log^2 m \quad [dB]$$

(4-108)
Due to the nature of the pdf’s and BER expressions, analytical results can only be obtained for integer values of $m$. However, the above expression is equally valid for non-integer and intermediate values. This is our final result, useful in system design.

summations like in (4-107) are terminated when the relative increment of each term is reduced to a factor $10^{-12}$. 
Figure 4-9: The sensitivity penalty $\Delta SNR$ for a BER of $10^{-9}$ of a single-filter ASK receiver with quadratic IF-detection and optimum decision threshold, as a function of the IF bandwidth expansion factor $m$.

and analysis. The complex and lengthy analysis thus produced a practical result that can be compressed into the simple form (4-108). The accuracy of the expression is better than $\pm 0.01$ dB up to $m=20$. Of course one can derive more complex curve fittings, but that is hardly required for practical use.

4.3.2 Comparison with other results

There is not much literature with which to compare our results. Noé [82] has performed the same analysis using numerical fast convolution of the pdf's. My results for the quadratic detector are identical to his. Foschini, Greenstein and Vannucci [21] and Cimini and Foschini [11] do not give figures of the same form as Figure 4-9, but from their BER curves as a function of the relative linewidth one can derive the sensitivity penalties. The trend in $\Delta SNR$ is the same as I have, but since they do not use the proper threshold setting (see section 4.2.2) they obtain sensitivity penalties that are consistently high by a few tenths of a dB. The important implication of their numerical results, however, is the fact that it proves that there is an optimum IF bandwidth expansion factor $m$ for every combination of linewidth and SNR. For this optimum of $m$, the effect of the linewidth can be eliminated and the receiver behaves as if the linewidth were zero. This provides a validation of the procedure I follow for eliminating the linewidth from the analysis, as explained in section 3.6.
4.4 Linear IF-detection

4.4.1 Post-detection filtering

Apart from the definition of the post-detection filter output (4-69), nothing has been said about linear IF-detection. The reason for this is expressed by the following conjecture:

Conjecture 4.1 It is not possible to analyse (coherent optical) receivers with non-coherent linear IF-detection and post-detection filtering in an analytical way.

I shall now try to indicate what is possible and what not. There are good reasons for the lasting 'interest' in linear IF-detection, despite the fact that nobody has actually solved the problem mathematically. First of all, as has been shown in section 4.1, the problem does not exist in the absence of post-detection filtering. In most 'classical' system analyses related to this topic, linear IF-detectors with Rician (chi) statistics can be used if post-detection filtering is not included. This method has also been used in section 4.1. Another reason is the fact that the commonly used diode peak detectors are mostly operated at high bias. This gives a characteristic that is certainly not quadratic, but better approximated by a linear transfer\[11^\].

Secondly, studying linear IF-detection analytically in the same way as done for quadratic IF-detection is impossible. Post-detection filtering, which is essentially summation of samples and thus convolution of the pdf's, can not be studied analytically since the convolution integrals are too complex. The only exception to this is the convolution of two central chi pdf's (4-9), which yields (C.4.11):

\[
p_{t}(u|4) = \frac{u}{N} \exp \left( -\frac{u^2}{N} \right) \\
+ \sqrt{\frac{\pi}{2N}} \exp \left( -\frac{u^2}{2N} \right) \left\{ \frac{u^2}{N} - 1 \right\} \Phi \left( \frac{u}{\sqrt{2N}} \right)
\]

(4-109)

This is the same result as Marcum's [75, 105]), but including the noise variance \(N\) and the post-detection filter normalisation of \(\frac{1}{2}\). This result immediately shows one of the problems associated with convolving chi distributions: the results contain multiple components that are only partly related to the original pdf's and are much more complicated. For example, convolving two pdf's (4-109) in order to obtain \(p_{t}(u|8)\) is already impossible. The pdf for \(m=2\) and a transmitted Mark, which should be obtained by convolution of two non-central chi pdf's (4-5), can not be found analytically either.

\[11^\]The main difference between quadratic and linear IF-detection is the fact that quadratic detection can be realised in practice [109, 110], whereas a perfectly linear transfer function can never be obtained. However, realising detectors with the proper quadratic behaviour and a large enough dynamic range is also very difficult. Linear detectors, or at least detectors with quasi-linear characteristics resembling that, therefore remain popular.
The partial error rate for transmitted Spaces can be calculated from (4-109) - using some approximations (!), see (C.4-18) - to be

$$P_S = \sqrt{2\pi Y} \exp(-Y) + \frac{1}{2} \exp(-2Y)$$

(4-110)

where $Y$ is again the normalised threshold $b_{th}^2/2N$. Under normal conditions the second term is negligible. Compared to the result for quadratic IF-detection and $m=2$ - yielding with (4-94) $(1 + Y) \exp(-Y)$ - the partial error rate is of the same order, but slightly higher at the same value of $Y$. However, and this is the problem, this does not mean that linear IF-detection yields a higher BER compared to quadratic IF-detection. For a complete BER analysis it is a prerequisite to know both $P_S$ and $P_M$, or otherwise $P_S$ and the exact threshold. However, the latter must be determined from the - unknown - pdf’s.

### 4.4.2 Characteristic functions method

Pdf for transmitted Space

A completely different approach has been proposed by Marcum [75], at least in case of a transmitted Space. Marcum extends the validity of the pdf $p_x(u|2)$ to negative values of $u$, and then states that this does not influence the convolution of these anti-symmetric functions for large values of $u$. When this is true, the characteristic function of the pdf - or the Fourier transform - can be obtained from pair 710.1 of Campbell and Foster [6]:

$$p \exp(ap^2) \overset{x \to -x}{\leftrightarrow} \frac{-x}{4\sqrt{\pi a}} \exp\left(-\frac{x^2}{4a}\right)$$

(4-111)

This yields the characteristic function

$$\mathcal{F}(p) = (-1)^\sqrt{2\pi Np} \exp\left(\frac{Np^2}{2}\right)$$

(4-112)

The first action of the post-detection filter - summation of $m$ samples - is equivalent to $m$-fold convolution of the pdf and thus also to $m$-fold multiplication of the characteristic function. This gives the new Fourier function:

$$\mathcal{F}(p) = (-1)^m (2\pi N)^{m/2} p^m \exp\left(\frac{mNp^2}{2}\right)$$

(4-113)

The inverse Fourier transform of this function is given by pair 709.0 of Campbell and Foster:

$$p^m \exp(ap^2) \overset{x \to -x}{\leftrightarrow} \frac{(-1)^m}{2^m/2 - 1} \sqrt{\pi a} m+1 \exp\left(-\frac{x^2}{4a}\right) \text{He}_m\left(\frac{x}{\sqrt{2a}}\right)$$

(4-114)
4.4. LINEAR IF-DETECTION

Figure 4-10: Illustration of the generalisation made by Marcum, and its effect on the cross-correlation of two central chi pdf's.

where $He_m(\cdot)$ is the Hermite polynomial defined by\textsuperscript{12}

$$He_m(x) = x^m - \frac{m(m - 1)}{2} x^{m - 2} + \frac{m(m - 1)(m - 2)(m - 3)}{2 \cdot 4} x^{m - 4} \ldots (4-115)$$

This leads to the pdf of the $m$-fold convolved chi pdf:

$$p(u) = \frac{(2\pi)^{\frac{m-1}{2}}}{\sqrt{N^m (m + 2)}} \exp \left( -\frac{u^2}{2mN} \right) He_m \left( \frac{u}{\sqrt{mN}} \right)$$  \hspace{1cm} (4-116)

Finally the averaging action of the post-detection filter - i.e. dividing by $m$ - should be accounted for. Again the SNR will be divided, so (with $u^2 := u^2/m$) the final pdf of the post-detection filter output is given by:

$$p_t(u|2m) = \frac{1}{\sqrt{2\pi N}} \left( \frac{2\pi}{m} \right)^{m/2} \exp \left( -\frac{u^2}{2N} \right) He_m \left( \frac{u}{\sqrt{N}} \right)$$  \hspace{1cm} (4-117)

(\(\rho \gg 1\))

For $m=1$ - with $He_1(x) = x$ - and $m=2$ - with $He_2(x) = x^2 - 1$ - the pdf reduces to

$$p_t(u|2) = \frac{u}{N} \exp \left( -\frac{u^2}{2N} \right)$$  \hspace{1cm} (4-118)

$$p_t(u|4) = \sqrt{\frac{\pi}{2N}} \exp \left( -\frac{u^2}{2N} \right) \left( \frac{u^2}{N} - 1 \right)$$  \hspace{1cm} (4-119)

\textsuperscript{12}The notation $He_m(\cdot)$ denotes the less commonly used definition of the Hermite polynomial, see for the alternative definition e.g. (Gr.8.950-2) or (Ab.22.2.14).
The first expression is identical to the original pdf $p_X(u|2)$, so for $m=1$ the post-detection filter has indeed no effect. For $m=2$ the approximated pdf is identical to (4-109), when in the latter expression the first exponential is neglected and the error-function approximated by 1. This shows that the approximation is very reliable. As an illustration the pdf’s for $m=3$ and 4 are given below:

$$p_t(u|6) = \frac{2\pi}{3\sqrt{3}N} u \exp\left(-\frac{u^2}{2N}\right) \left(\frac{u^2}{N} - 3\right)$$  \hspace{1cm} (4-120)

$$p_t(u|8) = \frac{\pi}{8} \sqrt{\frac{2\pi}{N}} \exp\left(-\frac{u^2}{2N}\right) \left(\frac{u^4}{N^2} - 6\frac{u^2}{N} + 3\right)$$  \hspace{1cm} (4-121)

Partial error rate, Space

The partial error rate $P_S$ can be computed from the above pdf using $(Gr.7,373)$ with modified integration boundaries and the different definition of the Hermite polynomial:

$$\int_b^\infty \exp\left(-\frac{x^2}{2}\right) H_m(x) dx = \exp\left(-\frac{b^2}{2}\right) H_{m-1}(b)$$  \hspace{1cm} (4-122)

This yields

$$P_S = \frac{1}{\sqrt{2\pi}} \left(\frac{2\pi}{m}\right)^{m/2} \exp(-Y) H_{m-1}(\sqrt{2Y})$$  \hspace{1cm} (4-123)

with $Y = b_{lin}^2/2N$ the normalised threshold. For large SNR the Hermite polynomial can be approximated as $H_m(z) \approx z^m$, giving the simplified expression for the partial error rate:

$$P_S = \frac{1}{2\sqrt{\pi Y}} \left(\frac{4\pi Y}{m}\right)^{m/2} \exp(-Y)$$  \hspace{1cm} (4-124)

For $m=1$ to 4 the expressions for $P_S$ using the exact result (4-123) are

$$(m = 1) \quad P_S = \exp(-Y)$$  \hspace{1cm} (4-125)

$$(m = 2) \quad P_S = \sqrt{\pi Y} \exp(-Y)$$  \hspace{1cm} (4-126)

$$(m = 3) \quad P_S = \frac{2\pi}{3\sqrt{3}} \exp(-Y)(2Y - 1)$$  \hspace{1cm} (4-127)

$$(m = 4) \quad P_S = \frac{\pi}{4} \sqrt{\pi Y} \exp(-Y)(2Y - 3)$$  \hspace{1cm} (4-128)

For $m=1$ the result is identical to (4-54), while for $m=2$ the error rate is a factor $\sqrt{2}$ lower compared to the more exact result (4-110).
4.5 SUMMARY

Pdf for transmitted Mark

This analysis is much more complicated, since the simplification used for the Space-pdf - the extension of the pdf to negative values of z - is not valid in this case. The modified Bessel function makes that the pdf-function does not belong to a standard Fourier-transform pair. The Bessel function should therefore be approximated in some other way, but in general this yields inaccurate results that will not be presented here.

4.5 Summary

In this chapter, an extensive analysis of ASK single-filter receivers with (quadratic) IF-detection has been given. Using a complete set of pdf's it is possible to model the IF signals, the detector output and the post-detection output. From these the overall BER can be calculated as a function of the SNR and the threshold setting, of which the latter is extremely important and critical. For minimum BER at a given SNR the optimum threshold setting should always be used; it is located at the cross-over point of the Mark and Space pdf's. A unique and compact expression has been derived for the threshold setting as a function of the SNR and $m$, which makes it possible to analyse the effects of non-optimum settings on the BER-performance of the receiver. Finally, simple expressions and figures were derived for the sensitivity penalty of the receiver as a function of $m$. These expressions can be used for system design and analysis.

Analysis of single-filter receivers with linear IF-detection is extremely difficult. Mathematically this is caused by the convolution of (non-central) chi pdf's. Using some approximations, expressions for the Space partial error rate have been obtained but analysis of the Mark situation is impossible thus far. This implies that this type of receiver can not be fully analysed, since both partial error rates would then be required.

Although ASK receivers are the least advanced of all coherent optical receiver types, they are very fundamental in the sense that all important aspects of system analysis are involved. The performance of ASK receivers is therefore often a reference for more complicated receiver concepts. It will be shown in the following that many other types of receiver - such as FSK receivers - can be seen as combinations of single-filter receivers. The tools that have been introduced in this chapter can therefore be used to advantage in the analysis of FSK and diversity receivers in the next two chapters.
Chapter 5

Dual-filter detection

Introduction

Single-filter IF-detection, as treated in the previous chapter, has two disadvantages. First, only one half of the received signal is detected, the Mark component. Theoretically ASK means that the transmitter should be switched off during a Space, so no signal is received at all. However, in practice pure ASK is not possible using semiconductor lasers, since a change in modulation current always introduces a shift in wavelength of the laser. Electrical ASK modulation of the laser thus also results in very large deviation FSK of the optical output, with typical deviations of tens of GHz. The resulting effect is a combined ASK-FSK signal. This means that the step from ASK to FSK is much smaller than it would appear theoretically, requiring only that the amplitude of the modulating current is decreased in order to maintain an acceptable frequency deviation\(^1\). This opens the way to IF-detection of all the received signal, resulting in a higher sensitivity.

Secondly, ASK requires extremely accurate setting of the decision threshold, depending upon the IF signal-to-noise ratio (formula (4-82)). In practice this means that either complicated control loops are needed to readjust the threshold setting to the optimum, or that a fixed non-optimum setting is used, resulting in a reduced receiver sensitivity. An IF-detection mechanism not requiring a complicated threshold setting would thus be favourable, which is the case for dual-filter FSK detection.

\(^1\)This works also the other way round: in case of direct FSK modulation of a laser diode ASK will be introduced as well. For low modulation index the ASK can usually be neglected, otherwise one should correct for the unbalance in the IF filters. In the following it is assumed that the residual ASK is negligible.
Figure 5-1: Block diagram of a dual-filter FSK receiver with non-coherent IF-detection. The theoretical filter characteristics are also illustrated.

5.1 Receiver model

5.1.1 A second IF branch

In order to detect the complete FSK IF signal the ASK receiver model of Figure 3-1 needs to be extended with a second IF branch. This second branch must contain an IF bandpass filter (BPF) centered at the second FSK IF peak, as well as a second IF detector. Apart from the centre frequency of the second BPF, the two IF branches are assumed to be identical as far as bandwidth, noise and detection order are concerned. Their centre frequencies are defined as $f_0$ and $f_1$ respectively, while the two BPF have the same bandwidth $B_{IF} = m \cdot 1/T$. The separation of the two BPF centre frequencies is set equal to the peak-to-peak FSK frequency deviation $\Delta f$. This means that when $m$ is larger than $\Delta f T$ the two IF filters will overlap in a region around the average IF. Although this may happen in practice\(^2\) it will be assumed that $m$ is equal to or smaller than $\Delta f T$. Like in Chapter 3 it is assumed that the IF noise before the (two) BPF’s is white. The two BPF’s then pass two different portions of this white IF noise, for which case Whalen has proven [119, p.50] that the noise components at the filter outputs are completely decorrelated. Finally the assumption already used for ASK -

\(^2\)See for example the Philips dual-filter receiver [39, 42, 44, 109, 110], which uses two IF BPF’s with cosine and sine frequency characteristics. Chapter 8 gives a detailed analysis of this system.
5.1. **RECEIVER MODEL**

i.e. infinite modulation speed - is still valid, which means that the transitions between $f_0$ and $f_1$ and back are instantaneous. Consequently there will be Signal in only one BPF at a time and the system can be analysed using the stationary situation with carriers.

### 5.1.2 FSK decision theory

In the description of the receiver configuration given above, it was assumed without proof that the two-filter approach is the best choice for optimum non-coherent FSK detection. For understanding the detection mechanisms involved with FSK detection it may be interesting to have a closer theoretical look at the decision theory involved\(^3\). For equal *a priori* Mark and Space probabilities, the likelihood ratio (3-81) in section 3.5 is given by

\[
l(s) = \frac{p_M(s)}{p_S(s)} \begin{cases} M > S & 1 \\ S > M & \end{cases}
\]  

(5-1)

The total IF signal $s(t)$ is the sum of the Signal component $S(t)$ and the bandlimited noise $n(t)$ with spectral density $n$ (defined in (3-8)). The signals $s(t)$, $S(t)$ and $n(t)$ are all one-dimensional, since the phase has been eliminated in the way described in section 3.2. Only the signal amplitudes are thus considered here. The noise has a normal distribution with variance $N_m$ determined by the IF bandwidth, while the variable $s(t)$ has a non-central normal distribution with the same variance:

\[
p_i(s) = \frac{1}{\sqrt{2\pi N_m}} \exp \left(-\frac{(s - S_i)^2}{2N_m}\right)
\]  

(5-2)

where $S_i$ are the IF Signal amplitudes for a transmitted Mark ($i = M$) or Space ($i = S$). The likelihood ratio can then be written as

\[
l(s) = \frac{\exp\left(-\frac{(s - M)^2}{2N_m}\right)}{\exp\left(-\frac{(s - S)^2}{2N_m}\right)} \begin{cases} M > S & 1 \\ S > M & \end{cases}
\]  

(5-3)

Taking the natural logarithm at both sides yields

\[
(s - S)^2 - (s - M)^2 \begin{cases} M > S & 2N_m \cdot \ln 1 = 0 \\ S > M & \end{cases}
\]

\[
2(S_M - S_S) \cdot s \begin{cases} M > S & S_M^2 - S_S^2 \\ S > M & \end{cases}
\]  

(5-4)

---

*This part is adapted from Schwartz, Bennett and Stein [98, p.70-72] and Stein and Jones [106, p.264-271].*
In principle the two IF Signal powers $S_M^2$ and $S_S^2$ are identical, so after integration over $T$ seconds (which does not change the basic relation) the resulting likelihood expression reads

$$\int_0^T (S_M - S_S) \cdot s(t) \, dt \begin{cases} > & M \\ < & S \end{cases} 0$$ (5-5)

First of all this result shows that - for an optimum likelihood ratio and thus a minimum overall BER - the IF signal $s(t)$ should be cross-correlated with both $S_M$ and $S_S$. The outputs of the two cross-correlators should be detected and subtracted for comparison with the threshold. For this cross-correlation operation several possibilities are available, the simplest one being a single IF matched filter with the response

$$h(t) = S_M(T - t) - S_S(T - t)$$ (5-6)

(see also Schwartz [98] formula (2-5-20)). A second possibility is of course two IF branches, with matched filters $h(t) = S_M(T - t)$ and $h(t) = S_S(T - t)$, respectively, followed by two separate IF-detectors and a baseband subtraction. The third form, more related to practical systems, is two IF bandpass filters optimised for $S_M$ and $S_S$, after IF-detection and subtraction followed by one post-detection filter for effective matched filtering. It is clear that this form is exactly the configuration presented above for FSK receivers with non-coherent IF-detection. In principle the post-detection filter and the two IF BPF’s (apart from their respective centre frequencies) are identical to the filters in the ASK receiver.

It is said that for proper FSK transmission the two signalling states must be orthogonal. In general it can be shown [98, p.77] that for systems not using synchronous IF-detection, maximum SNR and sensitivity are obtained when signals are orthogonal. For maximum discrimination between the Mark and Space states of an orthogonal signal the response of the Mark-filter to a Space-signal must be zero, and vice versa [106, p.266]. In combination with (5-6) this leads to

$$\int_0^T u_M(t) \cdot u_S^*(t) \, dt = 0$$ (5-7)

This can be achieved with $90^\circ$ out of phase signals of the same frequency, or by using different frequencies with the frequency deviation $\Delta f$ set to multiples\(^4\) of $1/T$. The first method yields DPSK, the second is equivalent to FSK.

For IF matched filtering the separation required for orthogonality is identical to the separation needed for zero filter overlap, in which case it has already been shown that

\(^4\)Schwartz, Bennett and Stein [98, p.81], using a slightly more detailed analysis, derive the optimum modulation index for orthogonality as 0.7. This is for synchronous FSK detection, and not a practical value for two-filter FSK detection, but in CPFSK it is indeed the most widely used modulation index, see section 7.2.
the signals are decorrelated [119]. As a consequence of the wider IF filters - which are required for a good linewidth tolerance - the practical dual-filter frequency deviation is usually several times $1/T$.

A last result of the FSK likelihood expression (5-5) is the optimum threshold setting, which is equal to 0. The receiver will thus simply compare the amplitudes of the detected filter outputs, and decide that a Mark has been sent if the Mark-filter gives the larger output, and vice versa. As long as the signals and the receiver are completely symmetrical the threshold thus remains equal to 0 under all conditions, giving a much simpler decision mechanism in the receiver than for ASK.

### 5.1.3 IF signals

Assuming a transmitted Mark, with all Signal in the BPF centered at $f_1$, the two BPF outputs can be written as:

\[
\begin{align*}
    s_0(t) &= v(t) \cos(\omega_0 t) - w(t) \sin(\omega_0 t) \\
    s_1(t) &= \{A + x(t)\} \cos(\omega_1 t) - y(t) \sin(\omega_1 t)
\end{align*}
\]  

(5-8)

The voltages $v(t)$ and $w(t)$ are the in-phase and quadrature components of the noise voltage in the lower BFP, while $x(t)$ and $y(t)$ are the same for the upper filter. The two orthogonal components of one filter output are uncorrelated, while the two pairs are also uncorrelated when the filters do not overlap. This means that all four noise voltages are identically distributed but uncorrelated.

The signal amplitudes of the two BPF outputs are given - in analogy to (3-21), and omitting the time indication - by

\[
\begin{align*}
    r_0 &= \sqrt{v^2 + w^2} \\
    r_1 &= \sqrt{(A + x)^2 + y^2} \quad \text{or} \quad \sqrt{\left(\frac{A}{\sqrt{2}} + x\right)^2 + \left(\frac{A}{\sqrt{2}} + y\right)^2}
\end{align*}
\]  

(5-10)

(5-11)

Within each Signal-carrying filter the SNR is identical to $\xi$ (3-17) as defined for ASK, while the SNR within a bandwidth $1/T$ is also equal to $\rho = A^2/2N$ (3-18).

### 5.1.4 Post-detection filtering

The IF-detectors in the FSK receiver can again be of any order $n$, as long as the two detectors are identical. Denoting the two detector outputs by $z_1 = r^n_1$ and $z_0 = r^n_0$, the resulting detected signal after subtraction of the detector outputs is given by

\[
    z = z_1 - z_0 = (r^n_1)^{n/2} - (r^n_0)^{n/2}
\]  

(5-12)

As mentioned above, the filters in the FSK receiver will be identical to the filters in the ASK receiver, so the post-detection integrate-and-dump filter can be modelled as
the finite-time integrator (3-71).

\[ u = \frac{1}{m} \sum_{i=1}^{m} (z_{1,i} - z_{0,i}) \]
\[ = \frac{1}{m} \sum_{i=1}^{m} z_{1,i} - \frac{1}{m} \sum_{i=1}^{m} z_{0,i} \]  
\[ (5-13) \]

As was the case for single-filter detection, the performance of receivers with linear and quadratic IF-detection will be different. For quadratic IF-detection the two detector outputs are \( p_1(z|2) \) and \( p_0(z|2) \), the pdf of \( u \) becomes

\[ p(u) = p_1(u|2m) * p_0(-u|2m) \]  
\[ (5-14) \]

For linear IF-detection the pdf's of the detector outputs are \( p_{\chi,1}(z|2) \) and \( p_{\chi,0}(z|2) \) respectively. Using the same notation \( p_1(u|2m) \) for a \( m \)-fold convolved chi pdf as in Chapter 4, the post-detection filter output is defined by

\[ p(u) = p_{1,1}(u|2m) * p_{1,0}(-u|2m) \]  
\[ (5-15) \]

The pdf of \( u \) can thus be obtained by \( m \)-fold convolution of the pdf of the Mark detector output, followed by a second convolution with the other detector output; \( m \)-fold convolved and mirrored in \( u=0 \). In Appendix A.1 it is shown that this is equivalent to determining the cross-correlation between the two effective matched filter outputs of the two receiver branches. Note that this is in accordance with the decision theory of section 5.1.2, which states that the received signal should be cross-correlated in the matched filters of the two branches.

### 5.1.5 Linewidth effects

The analysis of the linewidth effects in dual-filter FSK receiver is much the same as performed for ASK single-filter receivers in section 3.6. However, in contrast to single-filter reception an error will not be generated when the Signal falls outside its filter, but rather when it falls inside the other filter. Depending upon the separation between the two IF filters, there is a finite probability that the 'wrong' filter accumulates signal during the integration time \( T \). However, as illustrated in Figure 5-2, the pdf \( p_0(u) \) of this filter decreases very rapidly. The resulting pdf after subtraction of the two detector outputs extends from \(-1\) to \(+1\). For \( 0 \leq u \leq 1 \) the form of \( p(u) \) is almost identical to \( p_1(u) \), for \(-1 \leq u < 0 \) it resembles \( p_0(u) \). The error rate floor probability can then be calculated by integrating the resulting \( p(u) \) from \(-1\) to \( 0 \).

It is not within the scope of this section to give a complete analysis of the linewidth effects in dual-filter receivers. However, from Figure 5-2 - where \( \log(p(u)) \) is plotted - it can be expected that for equal \( \Delta \nu_{IF} \) and \( m \) values, the probability of the error rate floor for FSK will be orders of magnitude lower compared to ASK. This is confirmed
5.1. RECEIVER MODEL

![Graphs showing log[p(u)] vs u/A^2T for FSK and ASK reception.]

**Figure 5-2:** Left: Typical pdf's of the two FSK branch outputs, in the case of non-zero IF linewidth. Right: The resulting pdf's after subtraction of the two branches. The probabilities of the error rate floors for ASK and FSK reception are both indicated.

By the results of Foschini, Greenstein and Vannucci [21, Fig.12]. A regression line through their optimal values of \( m \) for different linewidth-to-bitrate ratios yields an expression that closely resembles (3-103):

\[
m_{opt} = 8(\Delta \nu_{IF}T)^{0.65}
\]

(5-16)

In the case of a fixed IF filter bandwidth - constant \( m \) - FSK is more tolerant to the linewidth:

\[
\Delta \nu_{IF,FSK} = \left( \frac{12}{8} \right)^{1/0.65} \Delta \nu_{IF,ASK} = 1.86 \cdot \Delta \nu_{IF,ASK}
\]

(5-17)

In view of the limited accuracy of the above analysis one can thus state more generally that FSK is nearly twice as tolerant to the linewidth compared to ASK. Or, from a different viewpoint, for a certain linewidth, FSK gives the same performance with a 1.5 times lower filter bandwidth. This is another advantage over ASK, and one that is not commonly known.
5.2 IF-detection

5.2.1 Probability density functions

The pdf's $p_1(u|2m)$ and $p_0(u|2m)$ in (5-14) are the pdf's of the detector outputs as derived in section 4.2. It has been shown that the pdf's for a transmitted Mark and Space are symmetrical when mirrored in the vertical $u=0$ axis. Therefore, it is sufficient to derive the pdf of $u$ for a transmitted Mark since the pdf for a transmitted Space can be obtained by substituting $-u$ for $u$.

When the cross-correlation of the two signals is calculated through the convolution (5-14) one complication arises. In a standard convolution of two pdf's limited to non-negative values of the variables, the resulting pdf will have non-negative variables as well. In the case of a cross-correlation convolution, with one of the pdf's mirrored, two different convolution integrals must be calculated, see Appendix A.1 (A.1-16) and (A.1-17).

\begin{align}
 p(u) &= \int_{0}^{\infty} p_1(y) p_0(y-u) \, dy \quad (u \leq 0) \\
 p(u) &= \int_{u}^{\infty} p_1(y) p_0(y-u) \, dy \quad (u > 0)
\end{align}

(5-18) 
(5-19)

The resulting pdf thus stretches out from $-\infty$ to $+\infty$, as could be expected. The second integral, with dependent integration boundaries, can in most cases not be solved analytically when the pdf's to be convolved are chi or chi-square. However, as illustrated in Figure 5-3, for calculating the BER only the negative tail of the pdf is required, given by the first integral. I will consequently not calculate the second integral, although this means that the complete pdf of $u$ is not obtained.

Quadratic IF-detection

In the case of quadratic IF-detection $p_1(u|2m)$ and $p_0(u|2m)$ are identical to the pdf's for a transmitted Mark and Space of the ASK receiver with quadratic IF-detection; (4-70) and (4-74). The expression for $p(u)$ after subtraction of $u_0$ from $u_1$ now becomes

\begin{align}
 p(u) &= p_1\left(u \bigg| 2m, \frac{A^2}{N}\right) * p_0\left(-u \bigg| 2m\right) \quad \text{Mark} \\
 p(u) &= p_1(u|2m) * p_0\left(-u \bigg| 2m, \frac{A^2}{N}\right) \quad \text{Space}
\end{align}

(5-20) 
(5-21)

Since only the tail $u \leq 0$ is of importance this gives the convolution integral:
5.2. IF-DETECTION

![Probability Density Function](image)

Figure 5.3: The probability density function of the post-detection filter output of a dual-filter FSK receiver with IF detector order \( n \). \( \chi^n \) is the single-filter pdf for a transmitted Mark, while \( M \) is the resulting dual-filter pdf after convolution. \( S \) is the dual-filter pdf for a transmitted Space. Only the shaded area is of importance for calculating the BER.

\[
\begin{align*}
    p(u) &= \left( \frac{1}{2N} \right)^2 \left( \frac{1}{2AN} \right)^m \frac{m^{-1}}{\Gamma(m)} \exp \left( -\frac{A^2}{2N} + \frac{u}{2N} \right) \\
    &\quad \cdot \int_0^\infty y^{m^{-1}} (y - u)^{m-1} \exp \left( -\frac{y}{N} \right) I_{m-1} \left( \frac{A}{N} \sqrt{y} \right) dy \quad (u \leq 0)
\end{align*}
\]

(5.22)

The analysis of this expression is detailed in Appendix D.1 and yields

\[
\begin{align*}
    p(u) &= \frac{1}{4N} \left( -\frac{u}{4N} \right)^{m^{-1}} \exp \left( -\frac{A^2}{4N} + \frac{u}{2N} \right) \\
    &\quad \cdot \sum_{k=0}^{m-1} \frac{1}{\Gamma(m-k)} \left( \frac{N}{u} \right)^k L_k^{m-1} \left( -\frac{A^2}{4N} \right) \quad (u \leq 0, \text{ Mark})
\end{align*}
\]

(5.23)

\( L_k^{m-1}(\cdot) \) is a generalised Laguerre polynomial of degree \( k \), which is a member of the family of orthogonal polynomials. Concerning the orthogonality one special remark needs to be made, since this is only defined when the argument of \( L_k^{m-1}(\cdot) \) is positive. However, in the expression for \( p(u) \) given above the argument is always equal to \(-\rho/2\) and thus negative. This means that the polynomial loses its orthogonality, although
the definition (A.3-9) remains valid:

\[
I_k^{m-1}(-x) = \sum_{i=0}^{k} \binom{k + m - 1}{k - i} \frac{x^i}{i!}
\]  

(5-24)

For a transmitted Space \( u \) should be replaced by \(-u\). For \( m=1 \) (5-23) reduces to

\[
p(u) = \frac{1}{4N} \exp\left(-\frac{A^2}{2N} + \frac{u}{2N}\right) \quad (u \leq 0, \text{ Mark})
\]  

(5-25)

For other \( m \) a more compact approximation can be obtained from (5-23), although it should be stressed that it is not allowed to use this expression for BER-analyses. It is only fit for showing the general form of the pdf-tail, and reads (D.1-10):

\[
p(u) = \left(\frac{1}{4N}\right)^{m-1} \exp\left(-\frac{A^2}{4N} + \frac{u}{2N}\right) \left(\frac{A^2}{4} - u\right)^{m-1}
\]  

(5-26)

\[(u \leq 0, \text{ SNR} \gg 1, \text{ Mark})\]

For \( m=1 \) this expression is exact, but for increasing \( m \) it becomes less accurate. It shows nevertheless that the pdf-tail for negative \( u \) is an exponential function multiplied with a second form-function \((A^2/4 - u)^{m-1}\). The fact that (5-23) may only be used for negative \( u \) is clearly illustrated by the fact that this form-function introduces a zero in the pdf at \( u = +A^2/4 \), whereas the real pdf for positive \( u \) resembles the original chi-square pdf of the detector output (see also Figure 5-3).

NxN noise

Since only the tail of \( p(u) \) can be written in an analytical form it is not possible to calculate the variance, since this requires knowledge of the complete pdf. However, by setting \( A \) to 0 - being equivalent to the situation that neither filter contains Signal - the variance of the NxN noise can be determined. With \( A=0 \), and after elimination of the sum, \( p(u) \) can be written as

\[
p(u) = \frac{1}{\sqrt{\pi N}} \frac{1}{\Gamma(m)} \left(\frac{1}{4N}\right)^m (-u)^{m-1/2} K_{m-1/2} \left(-\frac{u}{2N}\right) \quad (u \leq 0)
\]  

(5-27)

For reasons of symmetry the complete expression for \( p(u) \) in this case becomes (D.1-15):

\[
p(u) = \frac{1}{\sqrt{\pi N}} \frac{1}{\Gamma(m)} \left(\frac{1}{4N}\right)^m |u|^{m-1/2} K_{m-1/2} \left(\frac{|u|}{2N}\right)
\]  

(5-28)

\( K_{\nu}(-) \) is a modified spherical Bessel function of the third kind and order \( \nu \). For \( m=1 \) this Bessel function reduces to \( \sqrt{\pi/2u} \exp(-u) \) (Gr.8.469-2) and \( p(u) \) can be
written as
\[
p(u) = \frac{1}{4N} \exp \left( -\frac{|u|}{2N} \right)
\]  
(5-29)

It is easy to show that this result can also be obtained by autocorrelation of (4-19).

The variance of \( p(u) \) is calculated in Appendix D.1 and is equal to \( 8mN^2 \). This is identical to either twice the variance of one quadratic IF detector (3-44) divided by the post-detection improvement factor \( m \), or correspondingly twice the \( N \times N \) variance of the single-filter receiver (4-77). This proves that in the limit of no Signal the pdf’s derived for the dual-filter receiver are exact. It also shows that the \( N \times N \) noise variance has been doubled due to the second filter, giving a symmetrical pdf around \( u=0 \). I will come back to this increased \( N \times N \) noise at the end of the next section, since it plays an important role in the degradation of the sensitivity of the receiver.

5.2.2 The Bit-Error Rate

The BER can be determined by integration of the two pdf tails
\[
P_e = P(0) \int_0^\infty p_S(u) du + P(1) \int_{-\infty}^0 p_M(u) du
\]  
(5-30)

Assuming equal a priori Mark and Space probabilities \( P(0) \) and \( P(1) \), and knowing that the two pdf’s are identical when mirrored in \( u=0 \), this can be reduced to
\[
P_e = \int_{-\infty}^0 p_M(u) du = \int_0^\infty p_S(u) du
\]  
(5-31)

The resulting BER has been derived in Appendix D for quadratic IF-detection:
\[
P_e = \left( \frac{1}{2} \right)^m \exp \left( -\frac{A^2}{4N} \right) \sum_{k=0}^{m-1} \left( \frac{1}{2} \right)^k L_k^{m-1} \left( -\frac{A^2}{4N} \right)
\]  
(5-32)

or, with \( A^2/2N = \rho \),
\[
P_e = \left( \frac{1}{2} \right)^m \exp \left( -\frac{\rho}{2} \right) \sum_{k=0}^{m-1} \left( \frac{1}{2} \right)^k L_k^{m-1} \left( -\frac{\rho}{2} \right)
\]  
(5-33)

By writing out the Laguerre polynomial as a separate sum, an expression is obtained that is well-suited for fast and accurate numerical evaluation:
\[
P_e = \left( \frac{1}{2} \right)^m \exp \left( -\frac{\rho}{2} \right) \sum_{k=0}^{m-1} \left( \frac{1}{2} \right)^k \sum_{j=0}^{k} \binom{m+k-1}{k-j} \frac{1}{j!} \left( \frac{\rho}{2} \right)^j
\]  
(5-34)
Figure 5-4: Bit-error-rate curves of a dual-filter FSK receiver with quadratic IF-detection as a function of the SNR $\rho$ and the IF bandwidth expansion factor $m$.

For $m=1$ the BER-expression reduces to the well-known formula (2-21) [98, p.298] [106, p.240]

$$P_e = \frac{1}{2} \exp \left( -\frac{A^2}{4N} \right) = \frac{1}{2} \exp \left( -\frac{\rho}{2} \right)$$  \hspace{1cm} (5-35)

In general the BER can be written as the product of a fixed part equal to (5-35) and a BER deterioration factor $C(m)$. For $m=2$ to 5 $C(m)$ is given by:

$$P_e = \frac{1}{2} C(m) \exp \left( -\frac{\rho}{2} \right)$$  \hspace{1cm} (5-36)
5.3. SENSITIVITY AND PENALTIES

\[ C(2) = 1 + \frac{1}{4} \left( \frac{\rho}{2} \right) \]  
\[ C(3) = 1 + \frac{3}{8} \left( \frac{\rho}{2} \right) + \frac{1}{32} \left( \frac{\rho}{2} \right)^2 \]  
\[ C(4) = 1 + \frac{29}{64} \left( \frac{\rho}{2} \right) + \frac{1}{16} \left( \frac{\rho}{2} \right)^2 + \frac{1}{384} \left( \frac{\rho}{2} \right)^3 \]  
\[ C(5) = 1 + \frac{65}{128} \left( \frac{\rho}{2} \right) + \frac{23}{256} \left( \frac{\rho}{2} \right)^2 + \frac{5}{768} \left( \frac{\rho}{2} \right)^3 + \frac{1}{6144} \left( \frac{\rho}{2} \right)^4 \]  
(5-37)  
(5-38)  
(5-39)  
(5-40)

The order of \( C(m) \) thus increases with \( m \), giving a much larger multiplier for the baseline \( m=1 \) BER. Therefore it is to be expected that increasing \( m \) introduces an additional sensitivity penalty.

5.3 Sensitivity and penalties

Calculating the BER with (5-32) is simpler than was the case for ASK, since the (normalised) threshold is not a variable here. The BER only depends upon the SNR \( \rho \) and the bandwidth expansion factor \( m \). As would be expected, the BER-curves in Figure 5-4 shift to lower sensitivities for increasing \( m \), which is again due to the increasing N\(x\)N noise falling within the post-detection filter bandwidth.

For \( m=1 \) the sensitivity is equal to 40.06 or 16.03 dB. This is 0.46 dB away from the sensitivity of 36.02 dB for synchronous IF-detection (2-19), which is confirmed by the computations of Noé [82]. Fitting a curve through the resulting penalties for discrete values of \( m \) in Figure 5-5 yields a very compact expression:

\[ \Delta SNR = 1.085 \log m + 0.625 \log^2 m \text{ [dB]} \]  
(5-41)

This result may be extended to non-integer values of \( m \), giving another design tool for practical coherent systems. The accuracy of the above expression is in the order of 0.01 dB up to \( m=10 \). This is more than enough for practical system design.

It is interesting to note that the coefficients of (5-41) for FSK are not identical to the coefficients of (4-108) for ASK. This means that the sensitivity penalties of these two modulation schemes are different growing functions of \( m \).

5.4 FSK-ASK sensitivity improvement

The sensitivity improvement associated with FSK dual-filter demodulation, relative to ASK single-filter demodulation, will be defined as \( \Delta FSK \). Taking into account the difference of 2.82 dB between 40.06 (FSK) and 76.63 (ASK) for \( m=1 \), \( \Delta FSK \) can be found by subtraction of (5-41) and (4-108), yielding

\[ \Delta FSK = 2.82 - \Delta SNR_{FSK} + \Delta SNR_{ASK} \]  
\[ = 2.82 - 0.377 \log m - 0.084 \log^2 m \text{ [dB]} \]  
(5-42)
Table 5-1: The required SNR $\rho$ for a BER of $10^{-9}$ of a dual-filter FSK receiver with quadratic IF-detection, as a function of the IF bandwidth expansion factor $m$. $\Delta SNR$ is the sensitivity penalty in dB's relative to $m=1$, $\Delta FSK$ is the sensitivity improvement of FSK over ASK. For comparison the values for ASK have been copied from Table 4-3.

<table>
<thead>
<tr>
<th>$m$</th>
<th>FSK</th>
<th>ASK</th>
<th>$\Delta FSK$ dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40.06</td>
<td>76.63</td>
<td>2.82</td>
</tr>
<tr>
<td>2</td>
<td>43.79</td>
<td>81.40</td>
<td>2.69</td>
</tr>
<tr>
<td>3</td>
<td>46.63</td>
<td>85.12</td>
<td>2.61</td>
</tr>
<tr>
<td>4</td>
<td>49.61</td>
<td>88.30</td>
<td>2.56</td>
</tr>
<tr>
<td>5</td>
<td>51.10</td>
<td>91.12</td>
<td>2.51</td>
</tr>
<tr>
<td>6</td>
<td>53.00</td>
<td>93.70</td>
<td>2.47</td>
</tr>
<tr>
<td>7</td>
<td>54.73</td>
<td>96.07</td>
<td>2.44</td>
</tr>
<tr>
<td>8</td>
<td>56.55</td>
<td>98.28</td>
<td>2.42</td>
</tr>
<tr>
<td>9</td>
<td>57.86</td>
<td>100.37</td>
<td>2.39</td>
</tr>
<tr>
<td>10</td>
<td>59.29</td>
<td>102.35</td>
<td>2.37</td>
</tr>
</tbody>
</table>

Since this formula is obtained by subtracting two expressions each having an accuracy of 0.01 dB, the resulting accuracy is in the order of 0.02 dB. The numerical results for $m=1$ to 10 are included in table 5-1.

This is a result not commonly observed, since it is often stated (e.g. [3, 98, 106]) that the FSK-ASK improvement is 3 dB, which follows from (4-67) and (5-35), $\frac{1}{2} \exp(-\rho/4)$ and $\frac{1}{2} \exp(-\rho/2)$ respectively. However, it should be remembered that the approximation (4-67) is valid only in the limit of large SNR $\rho$. In all other cases the exact expression for the ASK BER (4-62) must be used, resulting in $\Delta FSK$ as derived here.

A simple explanation of the supposed 3 dB sensitivity improvement is that with dual-filter detection twice as much signal is detected than with single-filter detection; both for a received Mark and Space. However, this neglects the additional noise effects of the second filter branch. Referring to (3-47) the SNR of the detector output of a single-filter receiver can be written as:

$$SNR_o(ASK) = \frac{S^2}{\text{var}(SxN) + \text{var}(NxN)}$$  \hspace{1cm} (5-43)

In the case of dual-filter detection, with all signal in one branch and after subtraction
of the two detector outputs, $SNR_o$ becomes

$$SNR_o(FSK) = \frac{S^2}{\text{var}(SxN) + 2\text{var}(NxN)}$$ (5-44)

It is the doubled amount of NxN noise, due to the second FSK receiver branch, that causes less than 3 dB improvement of FSK. We have thus identified two effects that may increase the total sensitivity penalty of a receiver. The first is the IF bandwidth necessary to accommodate the IF laser linewidth, which, through the expansion factor $m$, reduces the sensitivity when the linewidth increases. This has been shown for single-filter and dual-filter detection, but will also prove to be valid for differentially coherent IF-detection.

The second penalty is associated with the demodulation scheme. Adding additional filter branches may increase the receiver sensitivity considerably, but there is always an associated penalty due to the additional NxN noise contributed by those branches. This is valid not only for the dual filter case presented here, but in general for M-ary FSK. Analysis of M-ary signalling systems can be performed using a straightforward extension of the method presented here.
5.5 Summary

Frequency Shift Keying with dual-filter IF-detection is a much more advanced detection scheme than ASK with single-filter detection, as has been analysed in Chapter 4. This is caused by the fact that binary FSK is an orthogonal signalling scheme, which gives the possibility to detect the Mark and Space signals separately. When the frequency separation of Mark and Space is large enough, a perfectly symmetrical dual-filter receiver can be used, yielding symmetrical detection statistics. The decision threshold can then be set to 0 under all circumstances, a considerable improvement compared to ASK single-filter detection where the threshold setting is critical for optimal detection. A further advantage of dual-filter FSK detection is the higher tolerance to the IF linewidth, 1.86 times that of ASK single-filter reception.

The mentioned theoretical advantage of FSK over ASK has of course its effect on the realisation of practical coherent optical systems using FSK detection. In Chapters 3 and 4 it has been shown that coherent optical systems with non-zero IF linewidth may require a considerable IF bandpass filter bandwidth, depending upon the IF linewidth-to-bitrate ratio. The combined requirements of non-overlapping BPF's - for orthogonal signalling - and large BPF bandwidth - in order to avoid a BER-floor due to the linewidth - may lead to a excessive overall IF receiver bandwidth. This can be expressed by a low receiver bandwidth efficiency. In practice dual-filter IF-detection is therefore not often used. Reduced bandwidth BPF's can be used when the laser linewidth is decreased, but once smaller linewidths are available it is preferable to use CPFSK modulation (Chapter 7). Nevertheless, the theoretical analysis of FSK dual-filter receivers is the basis on which the analysis of systems with differentially coherent detection will be built. FSK dual-filter reception may prove to be a very robust detection scheme, giving reliable and high performance detection. In Chapter 8 such a practical FSK receiver will be analysed in detail.

Although the statistics of the FSK receiver are symmetrical for Mark and Space signals, this does not mean that the analysis is simpler. In fact, only the tails of the probability density functions can be derived analytically, which fortunately suffices for evaluation of the BER. The resulting expression for the bit error rate as a function of the SNR $\rho$ and the IF bandwidth expansion factor $m$ can be used for an accurate analysis of the BER-curves, the sensitivity and the penalties. Again, as for ASK, a very simple and compact expression for the sensitivity penalty (at a BER of $10^{-9}$) as a function of $m$ has been derived.

Finally, the effective gain of FSK over ASK has been investigated. Although generally a sensitivity improvement of 3 dB is associated with FSK, it has been shown that the actual gain is always less. This is caused by the additional NxN noise of the second FSK branch, which contributes noise even when all signal is in the other branch. Although this penalty remains between 0.2 and 0.7 dB up to $m=10$, it is an essential characteristic of multi-branch IF-detection. It leads to much higher penalties with, for example, diversity configurations that are treated in the next chapter.
Chapter 6

Diversity reception

Introduction

Although wide IF bandwidth is uncommon in the more classical radio-oriented transmission systems, both the ASK and the FSK analyses presented thus far are not specific to coherent optical transmission. It was only implicitly assumed that the receivers were equipped with an optical polarisation control in order to obtain maximum IF amplitudes. Two receiver techniques presented in Chapter 2 that are more specific to coherent optical detection are polarisation and phase diversity. Diversity reception is a well known technique in the world of radio for the suppression of fading or unwanted reflections. Commonly employed methods are space (or antenna) diversity, frequency diversity, angle diversity, time (signal repetition) diversity and the polarisation diversity already mentioned [98].

However, there is one major difference between fading mechanisms associated with radio transmission, and polarisation fading in an optical fibre. A propagating radio signal may fade completely, even when e.g. multiple receiver antennas (space diversity) or both polarisation states (polarisation diversity) are used. This means that the signal in all diversity branches of the receiver reduces to below a certain threshold, giving rise to what is known as a deep fade [98]. Fades can partly be compensated by a properly chosen combining method, but when the fade is really 'deep' this will lead to an outage of the receiver. Increasing the number of diversity branches reduces the outage rate, at the same time yielding a considerable sensitivity gain since on average more signal will be received [98, sect.10-4]. However, the outage rate will never reduce to zero.

The fading phenomena in the case of radio transmission are caused by the fact that the radio 'ether' is a far from static transmission medium that is shared by many different (and thus interfering) systems. Moreover, atmospheric propagation may give
rise to statistical fluctuations of the individual signals, due to e.g. rain attenuation and atmospheric scintillations. This is precisely the difference in coherent optical fibre transmission, where each link has in principle its own, unique and relatively static transmission environment in the form of a single-mode fibre. Under normal conditions the properties of a SMF will barely vary, with the exception of the effective optical path length. Since the wavelength of the light is around 1.5 \mu m, the integrated temperature effect over tens of kilometers of fibre may result in a considerable shift of the optical signal phase. Since practical fibres have slightly elliptical cores the resulting phase shift will be randomly distributed along the two orthogonal output polarisation axes. Mechanical stress and microphonics may introduce relatively fast variations of the fibre birefringence. The output State of Polarisation (SOP) is thus random and fluctuating, with typical bandwidth values in the order of 20–100 Hz. However, and this is the main difference with radio transmission, the total amount of optical power at the fibre output will not change, since the attenuation remains constant. Therefore, with good receivers outages should not occur, since at all times the received amount of optical power is enough for proper reception. In practice the requirements of a polarisation diversity receiver are very stringent, allowing only a minimal sensitivity variation as function of the incoming SOP.

Polarisation diversity is the only practical alternative to polarisation control, and is commonly used in most experimental coherent optical systems\(^2\). Diversity has the main advantage that no mechanical or high power control mechanisms, with all their inherent problems, are required. However, the essence of diversity is the use of multiple receiver branches in order to increase the probability of reception. As was the case with the second branch of an FSK receiver, it is again assumed that all diversity branches have identical characteristics. Any imbalance between the branches will cause polarisation dependent sensitivity penalties. These will be calculated in sections 6.1.3 and 6.1.5.

Phase diversity reception, already introduced in section 2.4.3, has a different purpose. It can be shown that the IF amplitude in a homodyne ASK od DPSK receiver depends on the optical phases of received signal and local oscillator. Due to the non-zero linewidth of both laser sources the IF amplitude will thus fluctuate, with a finite probability of dropping to zero. By using multiple receiver branches, with different

---

\(^1\)The ellipticity may be very low, in the order of 10\(^{-3}\) or 10\(^{-4}\), but the integrated effect over tens or hundreds of kilometers of fibre can give a non-negligible effect. Furthermore, the orientation of two orthogonal polarisation axes (say horizontal and vertical) defined at the beginning of a fibre does not remain fixed in space, since the fibre may be rotated, twisted and spliced. This means that the orientation of the main polarisation axes at the output of the fibre can not be known in advance.

\(^2\)Using Polarisation Maintaining Fibre (PMF) throughout the optical network is not considered a realistic option. Apart from the costs, the finite extinction ratio will always give crosstalk between the two orthogonal polarisation states. After tens of kilometers this will again result in a random SOP.
(orthogonal) IF phases, this problem can be eluded, although again at the expense of multiple branches.

This chapter will investigate the influence of these additional IF branches, first for polarisation diversity and then for ASK phase diversity.

6.1 Polarisation diversity

6.1.1 Receiver model

The block diagram of a polarisation diversity receiver has been given in Figure 2-10. Depending upon the effective polarisation split angle $\phi$ - which has a uniform distribution between 0 and $2\pi$ and is determined by the actual SOP - the received signal power is split between two receiver branches with the use of a Polarisation Beam Splitter (PBS). The local oscillator light usually has a linear polarisation, since it passes through an optical isolator$^3$. Using Polarisation Maintaining Fibre (PMF) 50:50-couplers, the LO-power can be split evenly over both optical receiver branches in order to obtain the same theoretical sensitivity for both branches. The LO and signal components having the same polarisation are then combined in another PMF coupler, resulting in (amongst others) the intermediate frequency in the way described in Chapter 1. In accordance with (2-34) and (2-35) the IF photo currents in both receiver branches - denoted by $\parallel$ for the parallel, and $\perp$ for the perpendicular branch - are given by

$$i_{\parallel} = A \cos \phi \cos(\omega_{IF}t + \theta_{IF,\parallel})$$

$$i_{\perp} = A \sin \phi \cos(\omega_{IF}t + \theta_{IF,\perp})$$

This shows that the sum of the time-averaged IF powers is indeed identical to the IF power of a polarisation controlled single branch receiver, since $\cos^2 \phi + \sin^2 \phi = 1$. Also, irrespective of the value of $\phi$, both branches will never attain zero amplitude at the same instant.

Combining

Given these boundary conditions, an optimal way of combining the branches should be found. First of all it is clear that IF-combining is a non-starter, since the two IF phases $\theta_{IF,\parallel}$ and $\theta_{IF,\perp}$ are random. What remains are methods of post-IF-detection combining, notably selection combining, maximal-ratio combining or equal-gain combining [98, ch.10]. The principles of these three methods are illustrated in Figure 6-1.

$^3$Optical isolators have (45$^\circ$ rotated) polarisers at both ends. Between these polarisers an Yttrium-Iron-Garnet (YIG) $\lambda/4$ waveplate surrounded by a strong permanent magnet rotates the linear input SOP spatially over 45$^\circ$ according to the Faraday-effect. The outgoing light is thus always linearly polarised.
It is clear that selection combining, i.e. selecting the branch with the largest SNR without using the signal in the other branch, is not acceptable. This will only yield the maximum SNR at those rare moments when all the Signal power is in one branch. Selection combining thus introduces a polarisation dependent sensitivity penalty which fluctuates between 0 and 3 dB.

Maximal-ratio combining and equal-gain combining are originally concepts from

Figure 6-1: Three different methods of diversity combining:
From left to right counterclockwise: selection combining, maximal-ratio combining and equal-gain combining.
DET = any type of IF detector, LIN = linear IF detector, QDR = quadratic IF detector, SEL = selector.
6.1. POLARISATION DIVERSITY

radio, often employed as a form of pre-detection combining [98]. One can develop optical counterparts according to the same principles. The first, maximal-ratio combining, is often used in combination with linear IF-detection. Defining the amplitudes of the two IF Signals as $A \cos \phi$ and $A \sin \phi$, the outputs of two linear detectors are found with (3-31) to be

$$u_\parallel = A \cos \phi + \frac{mN}{2A \cos \phi}$$  \hspace{1cm} (6-3)

$$u_\perp = A \sin \phi + \frac{mN}{2A \sin \phi}$$  \hspace{1cm} (6-4)

$$(SNR = A^2/2N \gg 1)$$

Simply summing these two signals would still result in a polarisation dependent baseband Signal (omitting the detected noise components) of

$$u_{sum} = \sqrt{2}A \sin \left( \phi + \frac{\pi}{4} \right)$$  \hspace{1cm} (6-5)

On the other hand, if the detected signals (or alternatively the IF signals\(^4\)) are first multiplied by $\cos \phi$ and $\sin \phi$, respectively, a polarisation independent output signal is obtained. It is this measurement of the SNR in both IF branches that complicates the combiner, since the IF peak detectors that measure the IF amplitude give exactly the same outputs - but time-averaged - as (6-3) and (6-4). In the case of low SNR values the peak detectors detects more noise than Signal. This can be compensated for, but the whole set-up becomes extremely complicated. At the same time the advantage of maximal-ratio combiners is that with all Signal in one branch the gain of the second branch is reduced to zero, effectively cancelling the influence of that branch. The theoretical sensitivity degradation of maximal-ratio combining for perfectly parallel or perpendicular polarisation is therefore zero (see also [20]). Although coherent optical systems using maximal-ratio combining have been reported, it is certainly not the ideal solution. For example Imai [53] shows that in the case of non-ideal gain stages - which means not having a perfectly linear SNR-gain relation - additional sensitivity penalties arise.

The best results in terms of polarisation independence and electronics complexity are obtained with quadratic IF-detection followed by direct post-detection summation. Since there is no difference in the gain control of the two diversity branches this method can be considered as a form of equal-gain combining. Using perfectly quadratic IF detectors the two detector outputs are given - in accordance with (4-71) - as

$$u_\parallel = A^2 \cos^2 \phi + 2mN$$  \hspace{1cm} (6-6)

$$u_\perp = A^2 \sin^2 \phi + 2mN$$  \hspace{1cm} (6-7)

\(^4\)In principle control of the post detector baseband signals will be much simpler, compared to gain control of the large-bandwidth IF amplifiers. Maintaining a flat frequency transfer characteristic and balance between the branches becomes nearly impossible for bandwidths in excess of 2 GHz.
Direct summation of these two outputs yields $A^2 + 4mN$, which is polarisation independent at the expense of a twofold increase in the noise. This will obviously introduce a sensitivity penalty, which will be analysed in the next section. The only practical problem associated with this combining method is the required quadraticity of the IF detector transfer function. Depending upon the type of non-linear element, this is limited on the high side by the generation of higher harmonics and at the lower side by the noise floor of the post-detection receiver [109, 110]. However, it shows that this method gives the best performance when properly executed, which is the reason why it will be analysed as the ideal form of polarisation diversity combining.

6.1.2 BER for ASK polarisation diversity

The block diagram of the ASK polarisation diversity receiver with quadratic IF-detection and equal-gain post-detection summation is shown in Figure 6-2. With a fraction $\cos^2 \phi$ of the Signal IF power in the parallel branch and a fraction $\sin^2 \phi$ in the perpendicular branch, the pdf's of the two quadratic IF detector outputs (in the case of a transmitted Mark) are given as

$$p_{\parallel}(z) = p \left( z \mid 2, \frac{A^2 \cos^2 \phi}{mN} \right)$$

$$p_{\perp}(z) = p \left( z \mid 2, \frac{A^2 \sin^2 \phi}{mN} \right)$$

The combiner is a simple summation circuit, which means that the resulting signal
6.1. POLARISATION DIVERSITY

Figure 6-3: Bit-error-rate curves of a polarisation diversity single-filter ASK receiver with quadratic IF-detection, as a function of the SNR $\rho$ and the IF bandwidth expansion factor $m$. For comparison the single-branch BER-curve is also given (dashed).

and pdf before post-detection filtering become

$$z_{sum} = z_{\parallel} + z_{\perp}$$  \hspace{1cm} (6-10)

$$p_{sum}(z) = p_{\parallel}(z) \ast p_{\perp}(z)$$  \hspace{1cm} (6-11)

Since both $p_{\parallel}(z)$ and $p_{\perp}(z)$ belong to chi-square distributions, the resulting pdf after convolution is again chi-square with the number of degrees of freedom and non-
Table 6-1: The required SNR $\rho$ in dB for a BER of $10^{-9}$ of single-branch and polarisation diversity single-filter ASK receivers as a function of the IF bandwidth expansion factor $m$. $\Delta SNR$ is the sensitivity penalty in dB relative to $m=1$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>Single branch $\rho$</th>
<th>$\Delta SNR$ dB</th>
<th>Pol. diversity $\rho$</th>
<th>$\Delta SNR$ dB</th>
<th>$\Delta Div_{ASK}$ dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76.63</td>
<td>-</td>
<td>81.40</td>
<td>-</td>
<td>0.26</td>
</tr>
<tr>
<td>2</td>
<td>81.40</td>
<td>0.26</td>
<td>88.30</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>3</td>
<td>85.12</td>
<td>0.46</td>
<td>93.68</td>
<td>0.61</td>
<td>0.42</td>
</tr>
<tr>
<td>4</td>
<td>88.30</td>
<td>0.62</td>
<td>98.28</td>
<td>0.82</td>
<td>0.47</td>
</tr>
<tr>
<td>5</td>
<td>91.12</td>
<td>0.75</td>
<td>102.35</td>
<td>0.99</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>93.68</td>
<td>0.87</td>
<td>106.05</td>
<td>1.15</td>
<td>0.54</td>
</tr>
<tr>
<td>7</td>
<td>96.07</td>
<td>0.98</td>
<td>109.45</td>
<td>1.29</td>
<td>0.57</td>
</tr>
<tr>
<td>8</td>
<td>98.28</td>
<td>1.08</td>
<td>112.60</td>
<td>1.41</td>
<td>0.59</td>
</tr>
<tr>
<td>9</td>
<td>100.38</td>
<td>1.17</td>
<td>115.60</td>
<td>1.52</td>
<td>0.61</td>
</tr>
<tr>
<td>10</td>
<td>102.35</td>
<td>1.26</td>
<td>118.40</td>
<td>1.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Centrality parameters summed:

$$p_{sum}(z) = p(z \mid 4, \frac{A^2}{mN})$$  \hspace{1cm} (6-12)

After post-detection filtering by an integrate-and-dump filter the final pdfs of the polarisation diversity receiver for a transmitted Mark and Space become

$$p_{sum,M}(u) = p(u \mid 4m, \frac{A^2}{N})$$  \hspace{1cm} (6-13)

$$p_{sum,S}(u) = p(u \mid 4m)$$  \hspace{1cm} (6-14)

From this result it is clear that the baseband voltage $u$ has indeed become polarisation independent at the expense of a doubled number of degrees of freedom. We know from section 4.3 that this introduces a sensitivity penalty. Compared to the single-branch results of Chapter 4, exactly the same formulae for calculating the threshold (4-82), the BER (4-107) and the sensitivity penalty (4-108) can be used if one replaces every $m$ by $2m$. Figure 6-3 shows the resulting BER-curves, while table 6.1.2 gives the sensitivities that are obtained at a BER of $10^{-9}$. The polarisation diversity sensitivity penalty $\Delta SNR$ relative to $m=1$ can be found by fitting a curve to the discrete values of Table 6.1.2 and Figure 6-4:

$$\Delta SNR = 0.963 \log m + 0.662 \log^2 m \hspace{1cm} [dB]$$  \hspace{1cm} (6-15)
6.1. POLARISATION DIVERSITY

Figure 6-4: The sensitivity penalty $\Delta SNR$ in dB for a BER of $10^{-9}$ of single branch and polarisation diversity single-filter ASK receivers with quadratic IF-detection, as a function of the IF bandwidth expansion factor $m$.

The difference between (4-108) and (6-15) now gives the ASK diversity penalty $\Delta Div_{ASK}$, as illustrated in Figure 6-4:

$$\Delta Div_{ASK} = 0.262 + 0.255 \log m + 0.121 \log^2 m \ [dB]$$ (6-16)

Up to $m=10$ the accuracy of these results is better than 0.02 dB, which is more than enough for practical system design and evaluation. More important, with these results the effect on receiver sensitivity of using the polarisation diversity scheme can be evaluated quantitatively. This shows that the minimum penalty due to this scheme is 0.26 dB, which is caused by the NxN noise from the diversity branch. The value of 0.26 dB has been confirmed from analyses by Noé [82] and Enning [20]. The first author also gives the correct penalties for higher $m$, obtained by numerical convolution of the pdf's.

Even with the total Signal in one branch, the second diversity branch still contributes NxN noise. For increasing $m$ the NxN noise both from the original branch and the second diversity branch increase, making that the sensitivity penalty $\Delta SNR$ of a diversity receiver rises more rapidly than that of a single branch receiver. Therefore the diversity penalty $\Delta Div_{ASK}$ also increases with $m$, although at a very slow rate. For practical values of $m$ between 2 and 8 the diversity penalty remains almost constant between 0.4 and 0.6 dB, although one should not forget that the total sensitivity penalty due to $m$ can become as high as 1.5 dB.
6.1.3 SNR imbalance

A very practical problem associated with polarisation diversity is obtaining two identical IF branches, which has been a basic assumption in the analysis so far. Particularly for high IF bandwidths it is difficult to fabricate, for example, front ends with identical noise performance, or IF AGC amplifiers with equal bandwidth characteristics. Other factors that may upset the balance between the two diversity branches are the losses in the components of the optical polarisation diversity unit, photodiode responsivities and in general all electrical parasitics. This means that the peak IF SNR's at the inputs of the two quadratic IF detectors will in practice be different, influencing the detection process.

For simplicity it is assumed, as before, that the noise levels in both branches remain identical and that only the Signal amplitudes are influenced by the imbalance. An IF Signal voltage-imbalance factor $\kappa$ will be used. Arbitrarily taking the parallel branch as the reference, the pdf's of the IF detector outputs of both diversity branches (6-8) and (6-9) can be modified into:

$$p_\parallel(z) = p\left(z \left| \beta, \frac{A^2 \cos^2 \phi}{mN}\right.\right)$$

$$p_\perp(z) = p\left(z \left| \beta, \frac{\kappa^2 A^2 \sin^2 \phi}{mN}\right.\right)$$

After post-detection summation and filtering, the final pdf's of the decision circuit are derived as

$$p_M(u) = p\left(u \left| 4m, \frac{A^2}{N}(\cos^2 \phi + \kappa^2 \sin^2 \phi)\right.\right)$$

$$p_S(u) = p(u|4m)$$

As expected, for $\kappa$ not equal to 1 the expression for $p_M(u)$ has become polarisation dependent. For determining the resulting BER two effects must now be taken into account. First of all the effective IF SNR, which follows from $\text{SNR} = \lambda/2$, is now polarisation dependent and varies between $A^2/2N$ and $\kappa^2 A^2/2N$. In the absence of other effects this would introduce a sensitivity penalty equal to $20\log \kappa$ dB, since the SNR should be increased by this amount in order to maintain the same BER. However, a changing SNR also means that the optimum decision threshold should be adapted through (4-82):

$$\zeta - 1 = \frac{(2m - 1) \ln(\kappa^2 \rho/2m) + \ln 2 + 2m}{2\rho - 2m + 1} \quad (\rho \gg 1)$$

---

5Practical systems will be optimized for proper performance (e.g. a BER of $10^{-9}$) of the poorest branch. The normal branch then shows a lower BER. However, for the analysis it is more illustrative to define the performance of the normal branch at a BER of $10^{-5}$ and then evaluate the additional sensitivity penalty in the poor branch. This means that in the analysis $\kappa$ is always lower than 1.
6.1. POLARISATION DIVERSITY

Figure 6-5: Example for $\kappa=0.8$ and $m=4$ of the influence of IF SNR imbalance on the BER, depending upon the type of threshold used.

- circles: fixed threshold $b_A$.
- square: optimum threshold $b_{opt}$.
- crosses: fixed threshold $b_{av}$ for minimum average BER.

There are therefore two possibilities. If the threshold is optimised using (6-21) the sensitivity penalty will indeed be limited to $20\log\kappa$ dB as above, when all of the Signal is in the 'poor' branch. If the threshold is left in the setting $b_A$ that would be optimal for maximum sensitivity with all Signal in the 'good' branch, the BER becomes higher, see Figure 6-5. This can be interpreted as a lower SNR, giving the same BER with optimum threshold. With a fixed threshold the sensitivity penalty will thus be higher than the original $20\log\kappa$ dB. In particular for $m$ larger than 1 and high imbalance the additional penalty becomes in the order of 2-3 dB or more, see Figure 6-6. In practice it will be difficult to use optimum threshold settings, since one must find a way to measure the effective polarisation split angle $\phi$ and then design the proper control algorithm for
6.1.4 BER for FSK polarisation diversity

The analysis of FSK polarisation diversity is to a large extent similar to that of ASK polarisation diversity. Again two identical IF branches are assumed (see Figure 6-7), with quadratic IF-detection followed by equal-gain post-detection combining. The pdf's of the two IF detector outputs in the case of a transmitted Mark can be written with (5-20) as
6.1. POLARISATION DIVERSITY

Figure 6-7: Block diagram of an FSK polarisation diversity receiver with quadratic IF-detection and equal-gain combining.

\[ p_\parallel(z) = p_1 \left( z, \frac{\cos^2 \phi A^2}{mN} \right) \ast p_0 (-z|2) \]  \hspace{1cm} (6-22)

\[ p_\perp(z) = p_1 \left( z, \frac{\sin^2 \phi A^2}{mN} \right) \ast p_0 (-z|2) \]  \hspace{1cm} (6-23)

The combiner is again a summation circuit, which means that the resulting signal and pdf before post-detection filtering become

\[ z_{\text{sum}} = z_\parallel + z_\perp \]  \hspace{1cm} (6-24)

\[ p_{\text{sum}}(z) = p_\parallel(z) \ast p_\perp(z) \]  \hspace{1cm} (6-25)

The two components of \( p_\parallel(z) \) and \( p_\perp(z) \) both belong to chi-square distributions, so the resulting pdf after convolution is obtained by combining\(^6\) the identical chi-square

---

\(^6\)Notice that the resulting pdf is obtained analytically by first summing the Mark-filter outputs and Space-filter outputs, respectively. The resulting two signals are then subtracted. In practice this can be an alternative way of obtaining the sum-signal, although rarely (if at all) used.
Figure 6-8: Bit-error-rate curves of a polarisation diversity dual-filter FSK receiver with quadratic IF-detection, as a function of the SNR $\rho$ and the IF bandwidth expansion factor $m$. For comparison the single-branch BER-curve for $m=1$ is given (dashed).

pdf parts:

$$p_{\text{sum}}(z) = p\left(z \bigg| 4, \frac{A^2}{mN}\right) * p(-z|4)$$

(6-26)

After post-detection filtering by an integrate-and-dump filter the final pdf's of the
6.1. POLARISATION DIVERSITY

Table 6-2: The required SNR $\rho$ in dB for a BER of $10^{-9}$ of single-branch and polarisation diversity dual-filter FSK receivers as a function of the IF bandwidth expansion factor $m$. $\Delta SNR$ is the sensitivity penalty in dB relative to $m=1$ and $\Delta FSK$ the FSK over ASK sensitivity improvement.

<table>
<thead>
<tr>
<th>$m$</th>
<th>Single branch $\rho$</th>
<th>$\Delta SNR_{dB}$</th>
<th>Pol. diversity $\rho$</th>
<th>$\Delta SNR_{dB}$</th>
<th>$\Delta Div_{FSK}$</th>
<th>$\Delta FSK_{Div}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40.06</td>
<td>-</td>
<td>43.79</td>
<td>-</td>
<td>0.38</td>
<td>2.70</td>
</tr>
<tr>
<td>2</td>
<td>43.79</td>
<td>0.38</td>
<td>49.01</td>
<td>0.49</td>
<td>0.49</td>
<td>2.56</td>
</tr>
<tr>
<td>3</td>
<td>46.63</td>
<td>0.66</td>
<td>53.00</td>
<td>0.83</td>
<td>0.56</td>
<td>2.47</td>
</tr>
<tr>
<td>4</td>
<td>49.01</td>
<td>0.88</td>
<td>56.35</td>
<td>1.10</td>
<td>0.61</td>
<td>2.42</td>
</tr>
<tr>
<td>5</td>
<td>51.10</td>
<td>1.06</td>
<td>59.29</td>
<td>1.32</td>
<td>0.65</td>
<td>2.37</td>
</tr>
<tr>
<td>6</td>
<td>53.00</td>
<td>1.22</td>
<td>61.95</td>
<td>1.51</td>
<td>0.68</td>
<td>2.33</td>
</tr>
<tr>
<td>7</td>
<td>54.73</td>
<td>1.36</td>
<td>64.38</td>
<td>1.67</td>
<td>0.71</td>
<td>2.30</td>
</tr>
<tr>
<td>8</td>
<td>56.35</td>
<td>1.48</td>
<td>66.65</td>
<td>1.82</td>
<td>0.73</td>
<td>2.28</td>
</tr>
<tr>
<td>9</td>
<td>57.86</td>
<td>1.60</td>
<td>68.77</td>
<td>1.96</td>
<td>0.75</td>
<td>2.26</td>
</tr>
<tr>
<td>10</td>
<td>59.29</td>
<td>1.70</td>
<td>70.78</td>
<td>2.09</td>
<td>0.77</td>
<td>2.23</td>
</tr>
</tbody>
</table>

Polarisation diversity receiver for a transmitted Mark and Space become

\[
p_M(u) = p\left(u \left| 4m, \frac{A^2}{N}\right.\right) * p\left(-u \left| 4m\right.\right) \tag{6-27}
\]

\[
p_S(u) = p\left(u \left| 4m\right.\right) * p\left(u \left| 4m, \frac{A^2}{N}\right.\right) \tag{6-28}
\]

Like it was the case for ASK polarisation diversity, exactly the same pdf's as for the single-branch receiver are obtained, but with $2m$ instead of $m$. Compared to the single-branch results of Chapter 5, the same formulae for calculating the BER (5-32) and the sensitivity penalty (5-41) can be used if every $m$ is replaced by $2m$. The resulting BER-curves are illustrated in Figure 6-8. The polarisation diversity sensitivity penalty $\Delta SNR$ relative to $m=1$ can again be found by fitting a curve through the discrete values of Table 6-2 and Figure 6-9:

\[
\Delta SNR = 1.461 \log m + 0.625 \log^2 m \tag{6-29}
\]

The difference between (5-41) and (6-29) now gives the FSK diversity penalty $\Delta Div_{FSK}$:

\[
\Delta Div_{FSK} = 0.383 + 0.376 \log m \tag{6-30}
\]
Figure 6-9: The sensitivity penalty $\Delta SNR$ for a BER of $10^{-9}$ of single branch and polarisation diversity single-filter ASK and dual-filter FSK receivers with quadratic IF-detection, as a function of the IF bandwidth expansion factor $m$. The reference SNR is the value of 40.06 for single-branch dual-filter FSK detection.

Finally, using (6-15) and (6-29), the FSK-over-ASK sensitivity improvement $\Delta FSK$ for polarisation diversity receivers can be derived:

$$\Delta FSK_{div} = 2.70 - \Delta SNR_{FSK, div} + \Delta SNR_{ASK, div}$$
$$= 2.70 - 0.498 \log m + 0.037 \log^2 m \quad [dB] \quad (6-31)$$

The numerical accuracy of these results is again in the order of 0.02 dB, which makes them well suited for practical systems design and evaluation. Compared to ASK single-filter detection, the polarisation diversity penalty for FSK dual-filter detection $\Delta Div_{FSK}$ is slightly higher, with a minimum value for $m=1$ of 0.38 dB. In accordance with the theory presented in sections 5.4 and 6.1.2 this increased penalty is caused by the additional filters, which contribute extra NxN-noise.

6.1.5 SNR imbalance

The arguments for the existence of IF SNR imbalance in ASK receivers (section 6.1.3) apply equally well for FSK receivers. In fact FSK dual-filter receivers may be considered as systems having four different IF-branches (Figure 6-7), thereby increasing the practical problem of maintaining perfect IF balance. However, here it will be assumed that the imbalance arises before the branch point to the two Mark and Space
Figure 6-10: The polarization dependent sensitivity penalty in dB of a polarization diversity dual-filter FSK receiver with IF SNR imbalance. Penalties are relative to the single-branch $m=1$ sensitivity of 40.06.

filters, so both feature the same SNR reduction $\kappa^2$. This yields for the two branch outputs in case of a transmitted Mark, and again assuming that the ‘parallel’-branch has optimum SNR:

$$p_\parallel(z) = p_1 \left( z \left| 2, \frac{\cos^2 \phi A^2}{mN} \right. \right) * p_0 (-z|2) \quad (6-32)$$

$$p_\perp(z) = p_1 \left( z \left| 2, \frac{\kappa^2 \sin^2 \phi A^2}{mN} \right. \right) * p_0 (-z|2) \quad (6-33)$$

After post-detection summation and filtering this results in the following polarization-dependent pdf:

$$p_M(u) = p \left( u \left| 4m, \frac{A^2}{N}(\cos^2 \phi + \kappa^2 \sin^2 \phi) \right. \right) * p(-u|4m) \quad (6-34)$$

$$p_S(u) = p \left( u|4m \right) * p \left( -u \left| 4m, \frac{A^2}{N}(\cos^2 \phi + \kappa^2 \sin^2 \phi) \right. \right) \quad (6-35)$$

In contrast to (6-19) and (6-20), the Mark and Space-pdf’s for FSK are still mirrored around the optimum threshold setting $u=0$, so no threshold adjustment is required. This means that the polarization dependent penalty fluctuates between 0 and
20\log \kappa \text{ dB}, as illustrated in Figure 6-10. In practice FSK dual-filter receivers are thus slightly less sensitive to IF SNR imbalance than the ASK single-filter systems. Eventually this is due to the fact that FSK is an orthogonal signalling scheme, offering more 'symmetrical' system behaviour.

6.2 ASK phase diversity

6.2.1 Receiver phase model

ASK phase diversity has been considered in section 2.4.3 as a means of overcoming linewidth-induced IF signal fading in ASK homodyne receivers. In principle, several types of phase diversity receivers can be used, the simplest of which uses two branches. This is the 2x2 phase diversity receiver, requiring 90° out-of-phase (I&Q) IF-signals. The 3x3 phase diversity has been described in section 2.4.3 and will again be analysed as an illustrative example. However, with a proper optical 8-port 4x4 phase diversity systems, with four identical IF branches, are also possible [63, 108].

The phase-problem associated with ASK\textsuperscript{7} homodyne reception becomes apparent when equations (3-4) and (3-11) are modified for zero IF:

\begin{align}
S(t) &= A \exp(j\theta_{IF}) \\
n(t) &= \{x(t) + jy(t)\}\exp(j\theta_{IF}) \\
      &= v(t) + jw(t)
\end{align}

where \(x, y, v\) and \(w\) are identically Gaussian distributed noise voltages with variance \(2N_m = 2mN\). The resulting IF-voltage\textsuperscript{8} then follows as

\begin{align}
s(t) &= A \cos \theta_{IF} + v(t) \quad \text{(Mark)} \\
s(t) &= v(t) \quad \text{(Space)}
\end{align}

With respect to the Signal-component we note that it depends on the instantaneous IF-phase \(\theta_{IF}\), which is determined by the random optical Signal and LO-phases. Furthermore both optical phases vary due to the laser linewidth, yielding an unstable IF-phase. \(\theta_{IF}\) can therefore attain any value, reducing the IF Signal amplitude to zero for \(\theta_{IF} = \frac{\pi}{2} + k \cdot \pi\). This problem can be overcome by generating additional (orthogonal) IF currents: in a dual-branch (or 2x2) phase-diversity receiver a second IF-branch with Signal-component \(A \sin \theta_{IF}\). According to the diversity principle there will then always be Signal in at least one branch, while the total amount of power in the two branches remains constant.

\textsuperscript{7}And also DPSK, but this will be postponed to the next chapter. However, once the analysis for ASK has been performed, the results for DPSK can be obtained in a very straightforward manner.

\textsuperscript{8}Theoretically there is no IF in the case of homodyne reception. I will nevertheless employ the term 'IF' for those signals between the opto-electronic conversion and the squaring (IF) non-linear circuits.
6.2. ASK PHASE DIVERSITY

Figure 6-11: Power density spectra of the IF signals in ASK heterodyne (left) and homodyne receivers (right).

The noise in (6-38) has been reduced to one single noise voltage $v(t)$. In accordance with the theory presented in 3.2 this means that the degrees of freedom have also been reduced to one. The variance of this noise voltage is nevertheless equal to the total IF-variance of the heterodyne receiver, which can be proven in two ways. The first one is simply the mathematical equality for the IF-variance:

$$\text{var}(n_{het}) = \frac{1}{2} (\langle x^2 \rangle + \langle y^2 \rangle) = \frac{1}{2}(N_m + N_m) = N_m$$

(6-40)

which is exactly the variance of $v(t)$. The second method uses Figure 6-11, which shows that the advantage of the reduced noise bandwidth due to the homodyne reception is cancelled by the increased noise spectral density as a result of the spectral folding around 0 Hz. From the same figure it is also clear that the IF SNR does not change either, since the absence of the $\cos \omega_{IF} t$ carrier also increases the 'carrier' amplitude at 0 Hz.

6.2.2 2x2 phase diversity receivers

The IF pdf’s of the two I and Q-branches in a 2x2 ASK phase diversity receiver are chi distributed, with one degree of freedom. After quadratic IF-detection the following pdf’s are obtained for the I-branch:

$$p_{I,M}(z) = p(z \mid 1, \frac{A^2 \cos^2 \theta}{mN})$$

(6-41)

$$p_{I,S}(z) = p(z \mid 1)$$

(6-42)

And in the same way for the Q-branch:
Figure 6-12: Block diagram of a 2x2 2-branch ASK homodyne phase diversity receiver with quadratic IF-detection and I&D post-detection filtering.

\[
p_{Q,M}(z) = p\left(z \left| 1, \frac{A^2 \sin^2 \theta}{mN} \right. \right) \quad (6-43)
\]

\[
p_{Q,S}(z) = p(z|1) \quad (6-44)
\]

Post-detection summation and I&D-filtering yield the final pdf in \( u \) through summation of the degrees of freedom and noncentrality parameters, followed by \( m \)-fold multiplication:

\[
p_M(u) = p\left(u \left| 2m, \frac{A^2}{N} \right. \right) \quad (6-45)
\]

\[
p_S(u) = p(u|2m) \quad (6-46)
\]

Indeed this result is independent of the random IF-phase \( \theta \), confirming the principle of multiport phase diversity reception. More important, the above result is identical to the 'standard' expressions for the pdf's of a single-branch ASK heterodyne receiver (4-70) and (4-74). With the proper decision threshold setting, the results of a two-branch 2x2 ASK phase diversity receiver are thus identical to those of a single-branch ASK heterodyne receiver. Obviously, the total amount of IF-bandwidth in a receiver remains the critical parameter for a sensitivity analysis.
6.2. ASK PHASE DIVERSITY

6.2.3 2x2 phase diversity with IF imbalance

It was shown for polarisation diversity, that systems relying on some form of \( \cos^2 \theta + \sin^2 \theta \) (or \( \phi \)) summation are sensitive to imbalance in the IF branches. In the case of phase diversity this imbalance can manifest itself as amplitude or phase mismatch. Using the definitions of the imbalance of Figure 6-13, with a voltage amplitude imbalance \( \kappa \) and a phase mismatch \( \Delta \theta \), the final pdf \( p_M(u) \) is given by

\[
p_M(u) = p \left( u \left| 2m, \frac{A^2}{N} \{\cos^2 \theta + \kappa^2 \sin^2(\theta + \Delta \theta)\} \right. \right) \quad (6-47)
\]

\[
= p \left( u \left| 2m, \frac{A^2}{N} K_{2x2}(\theta, \kappa, \Delta \theta) \right. \right) \quad (6-48)
\]

with

\[
K_{2x2}(\theta, \kappa, \Delta \theta) = \cos^2 \theta + \kappa^2 \sin^2(\theta + \Delta \theta) \quad (6-49)
\]

The factor \( K_{2x2}(\theta, \kappa, \Delta \theta) \) is the phase diversity amplitude factor. It is easily verified that \( K_{2x2}(\theta, 1, 0) \), or in short \( K(\theta) \), is equal to 1. Next, the sensitivity penalties for the two limiting cases of \( \kappa=1 \) and \( \Delta \theta=0 \) are investigated. First, with the phase mismatch \( \Delta \theta \) equal to 0, the pdf reduces to

\[
p_M(u) = p \left( u \left| 2m, \frac{A^2}{N} \{\cos^2 \theta + \kappa^2 \sin^2 \theta\} \right. \right) \quad (6-50)
\]

\[
= p \left( u \left| 2m, \frac{A^2}{N} K_{2x2}(\theta, \kappa, 0) \right. \right) \quad (6-51)
\]

![Figure 6-13: Definition of the imbalance \( \kappa \) and phase mismatch \( \Delta \theta \) of a 2x2 ASK phase diversity receiver.](image)
The phase angle $\theta$ giving maximum amplitude variation (and thus SNR-variation) is found by differentiation to $\theta$ of $K_{2x2}(\theta, \kappa, 0)$ and yields

$$ \frac{dK_{2x2}(\theta, \kappa, 0)}{d\theta} = -(1 - \kappa^2) \sin 2\theta $$

(6-52)

This function has minima for $\theta = \frac{\pi}{2} + k \cdot \pi$. For these values of $\theta$ it is easily verified that $K_{2x2}(\theta, \kappa, 0)$ is equal to $\kappa^2$. This finally yields a sensitivity penalty of $20 \log \kappa$ dB.

Table 6-3: The maximum acceptable amplitude imbalance $\kappa$ and phase mismatch $\Delta \theta$ for maintaining a given sensitivity penalty $\Delta SNR$ of a 2x2 ASK phase diversity receiver.

<table>
<thead>
<tr>
<th>$\Delta SNR$ dB</th>
<th>$\Delta \theta$=0</th>
<th>$\Delta \theta$=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>$-20 \log \kappa$</td>
<td>$\Delta \theta$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.988</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.977</td>
<td>0.2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.944</td>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
<td>0.891</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The other limit is the situation of no amplitude imbalance ($\kappa=1$), with the variable part of the noncentrality parameter equal to

$$ K_{2x2}(\theta, 1, \Delta \theta) = \cos^2 \theta + \sin^2(\theta + \Delta \theta) $$

(6-53)

Differentiation to $\theta$ first yields

$$ \tan 2\theta = \cotan \Delta \theta $$
$$ \sin 2\theta \sin \Delta \theta = \cos 2\theta \cos \Delta \theta $$
$$ \cos(2\theta + \Delta \theta) = 0 $$

(6-54)

This expression has minima for $\theta = \frac{3}{4} \pi - \Delta \theta/2 + k \cdot \pi$, with

$$ K_{2x2}(\theta, 1, \Delta \theta) = 1 - \sin \Delta \theta $$

(6-55)

The associated sensitivity penalty is thus equal to $10 \log(1 - \sin \Delta \theta)$. Some numerical results for the two limiting cases of $\kappa=1$ and $\Delta \theta=0$ are given in Table 6-3.

As it was in the case of FSK polarisation diversity, an amplitude imbalance of $20 \log \kappa$ dB yields a maximum (phase-dependent) sensitivity penalty of the same amount. The tolerable phase mismatch is rather high, which can be explained by
the fact that the amplitude decrease of the phasor along its principle axis goes with $\sin^2 \Delta \theta$. In a dual-branch 2x2 ASK phase diversity system it is thus important to keep the amplitude balance as close to 1 as possible, whereas phase mismatch is less problematic. When both the amplitude and the phase feature deviations from their optimum values the tolerable variations of each for obtaining a certain sensitivity penalty become lower.

### 6.2.4 3x3 phase diversity receivers

The block diagram of a 3x3 phase diversity receiver was given in Figure 2-8. In line with the previous analysis of 2x2 receivers, the pdf’s of the three detector outputs can be derived as

$$
p_1(z) = p \left( z \left| 1, \frac{2A^2 \cos^2 \theta}{3mN} \right. \right) \quad (6-56)
$$

$$
p_2(z) = p \left( z \left| 1, \frac{2A^2 \cos^2 (\theta + 2\pi/3)}{3mN} \right. \right) \quad (6-57)
$$

$$
p_3(z) = p \left( z \left| 1, \frac{2A^2 \cos^2 (\theta - 2\pi/3)}{3mN} \right. \right) \quad (6-58)
$$

The usual post-detection summation and filtering yield the final pdf’s

$$
p_M(u) = p \left( u \left| 3m, \frac{2A^2}{3N} K_{3x3}(\theta) \right. \right) \quad (6-59)
$$

$$
p_S(u) = p(u|3m) \quad (6-60)
$$

The value of the phase diversity amplitude factor $K_{3x3}(\theta)$ is obtained by the summation of the squared cosine components and ideally equals $\frac{3}{2}$ (see (2-33)). In that case the noncentrality parameter of $p_M(u)$ reduces again to $A^2/N$ like for the 2x2 receiver. However, for this 3x3 phase diversity receiver the degrees of freedom have been increased by a factor 1.5. In general it can then be concluded that the degrees of freedom of a $k \times k$-branch ASK phase diversity receiver are equal to $km$.

Using the results of section 4.3 with $m=1.5$ gives a sensitivity of 79.05, 0.14 dB from the result for the 2x2 phase diversity receiver. Equation (4-108) can be modified by replacing $m$ by $1.5m$, which yields the numerical expression of the sensitivity penalty of a 3x3 ASK phase diversity receiver:

$$
\Delta SNR = 0.899 \log m + 0.541 \log^2 m \quad [dB] \quad (6-61)
$$

### 6.2.5 3x3 phase diversity with IF imbalance

For the analysis of IF imbalance in 3x3 phase diversity receivers it will first be assumed that only one branch features amplitude imbalance and phase mismatch. The analysis
can then easily be extended to the more general case of phase and/or amplitude imbalance in two branches (one branch is always used as an ideal reference). Using the definition of $\kappa$ and $\Delta \theta$ in Figure 6-14 yields the following pdf:

$$p_M(u) = p \left( u \left| 3m, \frac{2A^2}{3N} \left( \kappa^2 \cos^2(\theta + \Delta \theta) + \cos^2(\theta - \frac{2\pi}{3}) + \cos^2(\theta + \frac{2\pi}{3}) \right) \right. \right)$$  \hspace{1cm} (6-62)

$$\quad = p \left( u \left| 3m, \frac{2A^2}{3N} K_{3x3}(\theta, \kappa, \Delta \theta) \right. \right)$$  \hspace{1cm} (6-63)

with

$$K_{3x3}(\theta, \kappa, \Delta \theta) = \kappa^2 \cos^2(\theta + \Delta \theta) + \cos^2(\theta - \frac{2\pi}{3}) + \cos^2(\theta + \frac{2\pi}{3})$$  \hspace{1cm} (6-64)

![Figure 6-14: Definition of the imbalance $\kappa$ and phase mismatch $\Delta \theta$ of a 3x3 ASK phase diversity receiver.](image)

In the case of zero phase mismatch ($\Delta \theta = 0$) differentiation to $\theta$ of the amplitude factor $K_{3x3}(\theta, \kappa, 0)$ of the noncentrality parameter yields

$$\frac{dK_{3x3}(\theta, \kappa, 0)}{d\theta} = (1 - \kappa^2) \sin 2\theta$$  \hspace{1cm} (6-65)

Minima occur for $\theta = 0 + k \cdot \pi$, with the value of $K_{3x3}(\theta, \kappa, 0)$ equal to $\frac{1}{2} + \kappa^2$. This yields a sensitivity penalty relative to the optimum value $\frac{3}{2}$ of $10\log[(1 + 2\kappa^2)/3]$. Results are included in Table 6-4.
6.2. ASK PHASE DIVERSITY

Table 6-4: The maximum acceptable amplitude imbalance $\kappa$ and phase mismatch $\Delta \theta$ for maintaining a given sensitivity penalty $\Delta SNR$ of a 3x3 ASK phase diversity receiver. Tolerable phase mismatches are given for the case of a single mismatched phasor ($\Delta \theta$) and the case of two equally mismatched phasors ($\Delta \theta_1 = \Delta \theta_2$).

<table>
<thead>
<tr>
<th>$\Delta SNR$ dB</th>
<th>$\Delta \theta = 0$</th>
<th>$\kappa = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log \kappa$ dB</td>
<td>$\Delta \theta^o$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.983</td>
<td>0.15</td>
</tr>
<tr>
<td>0.2</td>
<td>0.966</td>
<td>0.30</td>
</tr>
<tr>
<td>0.5</td>
<td>0.915</td>
<td>0.77</td>
</tr>
<tr>
<td>1.0</td>
<td>0.832</td>
<td>1.60</td>
</tr>
</tbody>
</table>

The case of perfectly matched amplitudes and a phase error $\Delta \theta$ in the 3x3 phase diversity receiver requires slightly more goniometrical manipulations. Differentiation of the phase diversity amplitude factor gives

$$\frac{dK_{3x3}(\theta, 1, \Delta \theta)}{d\theta} = -[\sin 2\theta (\cos 2\Delta \theta - 1) + \cos 2\theta \sin 2\Delta \theta]$$

(6-66)

and setting this equation to zero yields

$$\tan 2\theta = \cotan \Delta \theta = \tan \left(\frac{\pi}{2} + \Delta \theta\right)$$

(6-67)

Minima are then obtained for $\theta = \frac{\pi}{4} - \Delta \theta/2 + k \cdot \pi$. Filling in the value $\frac{\pi}{4} - \Delta \theta/2$ for $\theta$ in $K_{3x3}(\theta, 1, \Delta \theta)$ leads, after splitting of all the cosines, to

$$K_{3x3}(\theta, 1, \Delta \theta) = \left(\cos^2 \frac{\pi}{4} + \cos^2 \frac{5\pi}{12} + \cos^2 \frac{11\pi}{12}\right) \cos^2 \frac{\Delta \theta}{2}$$

$$+ \left(\sin^2 \frac{\pi}{2} + \sin^2 \frac{5\pi}{12} + \sin^2 \frac{11\pi}{12}\right) \sin^2 \frac{\Delta \theta}{2}$$

$$- \frac{1}{2} \left(1 - \sin \frac{10\pi}{12} - \sin \frac{22\pi}{12}\right) \sin \Delta \theta$$

(6-68)

Surprisingly enough this expression reduces to

$$K_{3x3}(\theta, 1, \Delta \theta) = \frac{3}{2} - \sin \Delta \theta$$

(6-69)

Relative to the maximum constant amplitude factor of $\frac{3}{2}$ the sensitivity penalty is thus equal to $10\log[1 - \frac{3}{2} \sin \Delta \theta]$. Results are again included in Table 6-4. However,
since it is unlikely that in practice only one branch features a phase mismatch, a first order approximation of the situation with two (small) phase mismatches is also given. Assuming that both mismatches are of the same order this yields

$$K(\theta, 1, \Delta \theta_1, \Delta \theta_2) = \frac{3}{2} - \sin \Delta \theta_1 - \sin \Delta \theta_2$$  \hspace{1cm} (6-70)

This results in a nearly two times lower phase mismatch tolerance for maintaining the same sensitivity penalty. From the table it follows that as a good rule of thumb the tolerable phase mismatch in degrees of each diversity branch is equal to 10 times the sensitivity penalty in dB.

Comparing the results from Table 6-3 and 6-4 shows that the 3x3 phase diversity receiver is more tolerant to amplitude variations and approximately equally tolerant to phase mismatches. The first effect can be explained as follows. In a 2x2 receiver, reducing the amplitude of a branch to zero also yields a total amplitude (or K-factor) of zero for certain phase angles $\theta$. In a 3x3 receiver on the other hand we have 3 phasors that contribute to the total amplitude. Due to their 120° phase difference, reduction to zero of one amplitude does not lead to zero total amplitude, since the combined contributions of the other two are always larger than $\frac{1}{2}$. (This yields the result $K_{3x3}(\theta, \kappa, 0) = \frac{1}{2} + \kappa^2$). Since the ideal amplitude factor in a 3x3 phase diversity receiver is $\frac{3}{2}$, variations of $\pm 1$ due to one phasor amplitude are relatively less important. With respect to the phase mismatch the same holds, since a phase mismatch of one phasor can never reduce the total amplitude to zero. For one single mismatched phasor this yields a 1.5 times larger tolerance. For two equally mismatched phasors the phase tolerance becomes of the same order for 2x2 and 3x3 phase diversity reception. These results compare relatively well with the figures obtained through numerical simulation by de Krom [69], although his results are generally 20-50% more tolerant to phase mismatch. In general the balance of the optical 3x3 hybrid in phase diversity (ASK) receivers is extremely important. Fused 3x3 optical couplers have been measured by Pietzsch and Gottwald [29, 91] and Dumont [16], which shows that the general performance of these components can be good, although in practice imbalance and phase mismatch can amount to several %.

6.3 Conclusions

Diversity receivers use additional IF-branches to overcome fading problems associated with the optical polarisation or laser linewidth. It has been shown that using the proper detection and combining methods indeed results in receivers that are independent of the aforementioned parameters. In practice this means that polarisation diversity receivers with quadratic IF-detection and equal-gain combining should be used. For multi-branch ASK phase diversity receivers quadratic detectors are also required. In both cases it is essential for proper operation that all IF-branches of a receiver are
6.3. CONCLUSIONS

exactly identical, which may be one of the major problems in realising practical high-bitrate diversity receivers. In case of imbalances between the IF-branches, additional penalties arise that can amount to several dB's.

These imbalance-penalties would add to the intrinsic penalties associated with diversity reception. As was the case with FSK in the previous chapter, adding filters and/or branches always introduces sensitivity penalties due to the increased NxN post-IF-detection noise. From the theoretical analysis it follows that this NxN noise is caused by the increased degrees of freedom of a diversity configuration: since the (additive) noise in an additional diversity branch may be considered as independent, one essentially adds a complete second receiver. Mathematically this means that in all results of the single-branch receivers from Chapter 4 and 5 the degrees of freedom $m$ may be replaced by $2m$ for polarisation diversity reception. This introduces sensitivity penalties of several tenth's of a dB, with a minimum for $m=1$ of 0.26 dB for ASK and 0.38 dB for FSK. Up to $m=10$ these values increase slowly by no more than 0.4 dB.

It has been shown for all types of diversity reception that the balance between the IF-branches is of extreme importance. In general the principle of diversity reception, be it polarisation or phase, is based on the post-IF-detection summation $\cos^2 \phi + \sin^2 \phi = 1$ or some similar relation. The variable $\phi$ then depends on either the polarisation state or the instantaneous IF-phase, and in the case of ideal quadratic IF-detection and identical branches can be eliminated in this way. Any deviation from this ideal situation immediately introduces polarisation- or phase-dependent relations and thus fading of the receiver sensitivity.

Polarisation diversity is up to now the most widely - if not the only - used polarisation handling scheme in what may be called ‘practical systems’, i.e. systems that can operate outside laboratory conditions [10, 44, 52, 54, 96, 100]. However, this is more due to the absence of any alternative control scheme with a comparable reliability than to the advantages of polarisation diversity. For increasing bitrates and the associated IF-bandwidths it becomes exceedingly difficult to obtain identical IF-branches. One of the few remaining possibilities is then (silicon bipolar) integration of the IF electronics, thus reducing the parasitics for obtaining well-determined characteristics.

Phase diversity has the advantage of homodyne (zero-IF) optical detection, facilitating the use of ‘low-frequency’ electronics. Originally this scheme was thought to be ideal for overcoming the problems associated with the large IF laser linewidth. However, the developments in reducing the laser linewidth have seen so considerable that this raison d'être of phase diversity systems has disappeared. In my opinion the only

---

9 Polarisation control is, in my opinion, from a systems point of view the ideal solution, requiring a minimum of IF-branches. However, up to now all published polarisation control methods feature at least one of the following three drawbacks:
- they are too slow, mostly requiring use of some mechanical control method.
- they have too high an insertion loss, especially in the case of integrated optics or other exotic solutions.
- they are not reliable, particularly when SOP-variations are induced by mechanical operations (fibre squeezing, pulling, rotating, etc.).
possible application of this type of system will be at extremely high bitrates - 5 Gbit/s or more - where electrical receiver bandwidth becomes very expensive [18]. With respect to the number of phase diversity branches to be used, which can range from 2 to 4, one should again bear in mind the general rule of thumb for all diversity systems; additional branches \(\implies\) additional noise bandwidth \(\implies\) sensitivity penalties.

Although diversity reception can often be a good solution, one should never forget the associated penalties and consider the required work in order to obtain - and maintain - really identical IF-branches.
Chapter 7

Differentially coherent IF-detection

Introduction

In Chapter 2 it was shown that systems using differentially coherent IF detection, like DPSK and CPFSK, are second best when maximum receiver sensitivity is considered. Ideally, optimum sensitivity is obtained by synchronous (IF) detection, where the phase of the received signal is compared with an absolute reference signal generated in an (optical) PLL. Differentially coherent IF detection is an intermediate solution between non-coherent IF-detection on the one hand (which does not use phase information at all) and pure homodyne PSK detection with an OPLL. In the case of DPSK and CPFSK only the phases at the end of two consecutive bits are compared, they mainly differ in their phase-behaviour over a bit period.

To this end the delayed signal phase of a certain bit-period is used as a phase reference for the next bit-period. This is advantageous as no optical PLL is required; the demodulator consists of a simple delay line and a comparator/phase detector. However, it is clear that the phase reference of a receiver with differentially coherent IF-detection is of much poorer quality compared to that of a receiver with an OPLL. Especially in the case of a small SNR due to additive noise the receiver performance will deteriorate rapidly. On the other hand, differentially coherent IF-detection is much more tolerant towards the laser linewidth as opposed to OPLL detection, since only the bit-to-bit phase is compared. In particular the latter condition, coupled with the difficulty in obtaining extremely narrow linewidth lasers (for PSK linewidth values in the order of 10 kHz are required), makes that differentially coherent IF-detection is becoming the most widely used demodulation scheme in coherent optical systems. Of the two possible modulation formats that can be used, CPFSK and DPSK, the first
one is generally employed, since it has the highest linewidth tolerance of the two. A second important reason for choosing CPFSK is that direct laser modulation can be used, without the need for external phase modulators.

Compared to non-coherent single- or dual-filter IF-detection, CPFSK gives two additional system parameters. The first is the frequency deviation $\Delta f = 2f_d$, which is essential for the system performance analysis since the complete IF spectrum is contained within the single IF bandpass filter\(^1\). Practical values of the CPFSK modulation index $M$ - the peak-to-peak frequency deviation relative to the bitrate - are between 0.5 and 1.5, and have a large impact upon the receiver performance. The required IF bandwidth, the linewidth tolerance, the required delay and the ISI are all a function of the modulation index. A second parameter that will prove to be very critical is the detection efficiency $\delta$, i.e. the ratio of the frequency discriminator period to the peak-to-peak frequency deviation. Both additional parameters are a cause of difficulty in the analysis of CPFSK systems.

Since the frequency deviation in DPSK is set to zero and replaced by an instantaneous $180^\circ$ phase shift, DPSK is in theory simpler than CPFSK. It will on the other hand be shown that this tightens the requirements on the settings of the remaining parameters, such as modulation depth, IF center frequency and delay time. In section 7.5 it is shown that in the limit of ideal detection, DPSK and CPFSK obtain the same performance and sensitivity, i.e. 3 dB better than non-coherent dual-filter FSK IF-detection.

### 7.1 Delay-line discrimination

#### 7.1.1 Receiver model

The block diagram of a receiver with differential IF-detection is identical to the diagram of the ASK receiver (Figure 3-1) with the exception of the IF detector. After an IF bandpass filter (BPF) with again a width $m \cdot 1/T$, the signal is split into two branches, one of which is delayed by a time $\tau$. The direct and delayed IF signals at the input of the IF phase detector, $s(t)$ and $d(t)$ respectively, can now be written as

$$s(t) = x(t)\cos \omega_{IF}t - y(t)\sin \omega_{IF}t$$

$$d(t) = v(t)\cos \omega_{IF}t - w(t)\sin \omega_{IF}t$$

The definitions of the SNR (3-17) and (3-18) remain the same. In contrast to filter detection, the IF spectrum falls within one single BPF, which requires that the frequency modulation index $M$ be defined (in the same way as Iwashita [55]):

$$M \triangleq \frac{\Delta f}{1/T} = 2f_dT$$

---

\(^1\)In the case of dual-filter FSK detection the actual value of the modulation index was of no importance, as long as it was so large that the two IF filters did not overlap.
7.1. DELAY-LINE DISCRIMINATION

Figure 7-1: Block diagram of a receiver with differentially coherent IF-detection. The characteristics of the BPF and the delay line discriminator in the case of CPFSK are illustrated below.

where

\[ \Delta f = 2f_d \quad (7-4) \]

Here, \( \Delta f \) and \( f_d \) are the peak-to-peak and the carrier-to-peak frequency deviation, respectively. The theoretical minimum value of \( M \) is 0.5 (Minimum Shift Keying, MSK), but practical values are often between 0.7 and 1.0. This has to do with the fact that the MSK spectrum, which is the most compact spectrum for frequency modulation, becomes more sensitive to asymmetric filtering of its sidelobes, introducing ISI. According to simulations by Kazovsky and Jacobson [66, 67] a relatively large IF bandwidth is required for \( M=0.5 \), with \( m \) at least 3.7 in order to avoid large ISI penalties. For modulation indexes \( M \) between 0.7 and 1.0 the required IF bandwidth expansion factor is around 2, for higher \( M \) it increases due to the broadening of the spectrum (see Figure 7-2). For DPSK Park, Jr. [89] has also shown that \( m=1 \) leads to
Figure 7-2: The optimum IF bandwidth expansion factor \( m \) of CPFSK systems, as function of the modulation index \( M \). For comparison, Carson’s rule is also plotted (dashed). For perfect decorrelation of the direct and delayed IF-signals, \( m \) should be above the dash-dotted line. Uncertainty margins of the simulations are also indicated. (Data are from Kazovsky and Jacobson, "Multichannel CPFSK coherent optical communications systems", J. of Lightwave Techn., Vol.7, No.6, 1989, pp.972–982 [67], ©1989 IEEE).

Unacceptable levels of ISI. He suggests to use wider IF filters instead, without mentioning the exact width. In contrast to non-coherent IF-detection, \( m=1 \) obviously has no meaning in the case of CPFSK and DPSK, since \( m \) must be larger than 2. Also note that the required IF bandwidths are not always in agreement with the more ‘classical’ Carson rule [7, 8], which is valid for wide deviation FSK and gives the required relative IF bandwidth as\(^2\) \( m = 2 + M \).

The average intermediate frequency should be positioned at a zero of the frequency discriminator characteristic, which is given by

\[
H(f) = \text{Re}\{\exp(j\omega t)\exp(-j\omega(t-\tau))\} = \cos(\omega \tau)
\]

\[ (7-5) \]

\(^2\)The classical Carson rule reads [7, 8]:

\[
B_{IF} = 2(B + f_d)
\]

with \( B \) the baseband spectral width and \( f_d \) the largest single-sided frequency deviation. With \( B_{IF}/B = m \) and \( f_d = MB/2 \) we obtain the ‘digital’ version of Carson’s rule.
7.1. DELAY-LINE DISCRIMINATION

The frequency detection curve is illustrated in Figure 7-1 and has zeros at frequencies $f_c = (1 + 2n)/4\tau$ with $n$ an integer. The half period of the frequency detection curve - i.e. the distance between a maximum and a minimum - is defined as $f_m = 1/2\tau$. In practice the second ($n=1$) or third ($n=2$) zero-crossing are used in order to avoid overlapping of the IF and baseband spectra. Additional design flexibility can be obtained by using a 90° electrical hybrid in one of the discriminator branches. This yields a sine-like frequency discriminator characteristic with zeros at $f_c = n/2\tau$. The detection efficiency $\delta$ is defined as the ratio of the peak-to-peak signal frequency deviation $\Delta f$ and the half period of the detection curve $f_m$. This yields

$$\delta \equiv \frac{\Delta f}{f_m} = \frac{2M\tau}{T} \quad (7-6)$$

The IF bandwidth expansion factor $m$ and the detection efficiency $\delta$ are the two independent variables to be used in the performance analysis of CPFSK receivers. The modulation index $M$ is now no longer a variable, since it is incorporated in $\delta$.

In the presence of phase noise $\theta(t)$ the discriminator output signal can finally be written as

$$z(t) = \cos(2\pi(f_c + f_d)\tau + \theta(t) - \theta(t - \tau))$$

$$= \cos[(2n + 1)\frac{\pi}{2} \pm \delta\frac{\pi}{2} + \Delta\theta(\tau)] \quad (7-7)$$

$\Delta\theta(\tau)$ is defined as $\theta(t) - \theta(t - \tau)$. For $\delta=1$ the expression reduces to

$$z(t) = \pm \cos \Delta\theta(\tau) \quad (7-8)$$

where a Mark gives + and a Space -.

Influence of phase noise

It should be stressed that, even in the absence of laser phase noise, the performance of receivers using delay-line discriminators can deteriorate due to phase noise. This is caused by the fact that the delayed signal itself is used as a phase reference for the

---

3 Due to the high conversion losses in most mixers, IF-to-baseband crosstalk can be a serious problem. By separating the IF and baseband signals in the frequency domain many problems can be avoided. Careful balancing of the mixer electronics [109, 110] can also reduce the crosstalk, allowing for more spectral overlapping. In practical heterodyne CPFSK systems the third zero-crossing is often located at two times the bitrate [10]. This yields $f_c=2/T$ and $\tau=5T/8$, which means that the delay is in the order of half a bitperiod. For a detection efficiency of 1 this requires a modulation index of 0.8.

4 With the 90°-hybrid one can obtain a smaller half period of the frequency discriminator curve at the same central IF, and consequently use a smaller optimum frequency modulation index. Using the same central IF of 5 GHz as in [10] and the $n=5$ zero-crossing, we obtain a discriminator half period of $\frac{5}{8}$ or 1.7 GHz. This yields an optimum frequency modulation index of $M=0.68$, and a much more compact spectrum than at $M=0.8$. 

next bit. However, the SNR of the signal can be low, especially for received powers close to the sensitivity limit. The bandlimited white amplitude noise of the signal carrier causes variations of the instantaneous signal phase (see e.g. Figure 3-6), which degrades its function as a stable phase reference. Jain and Blachman [4, 58, 59] have derived an expression for the BER of a receiver with hard limiting in the IF, followed by delay-line frequency or phase demodulation. Although coherent optical systems never use hard limiting\(^5\) and post-detection filtering is not included in the analysis, it is a good illustration of the effect of reduced SNR on this type of demodulation. The BER then reads (with \(p_1 = p_2 = \sqrt{\rho}\))

\[
P_e = \frac{1}{2} - \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(k + 1/2) \Gamma(k + 3/2)}{(2k + 1)^2} \cdot \rho^{2k+1} \left[ \Phi(k + \frac{1}{2}, 2k + 2, -\rho) \right]^2
\]

\[
= \frac{1}{2} - \frac{\rho \exp(-\rho)}{2} \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k + 1} \left[ I_k \left( \frac{\rho}{2} \right) + I_{k+1} \left( \frac{\rho}{2} \right) \right]^2
\]

(7-9)

This equation is the general expression for the BER of a modulation scheme where the detector output can be written as [59]

\[
z = A \cos(\theta_1 \pm \theta_2)
\]

(7-10)

and the phase variations \(\theta\) are purely the result of the additive Gaussian noise. It can be approximated [4, 55] for large SNR \(\rho\) to

\[
P_e = \frac{1}{2} \exp(-\rho)
\]

(7-11)

which is indeed identical to the asymptotic performance of CPFSK systems (2-22).

### 7.1.2 Linewidth effects

Blachman [4] and Iwashita et al. [55] have modified (7-9) in order to include linewidth and detection efficiency. This gives\(^6\)

\[
P_e = \frac{1}{2} - \frac{\rho \exp(-\rho)}{2} \sum_{k=0}^{\infty} (-1)^k \frac{1}{2k + 1} \left[ I_k \left( \frac{\rho}{2} \right) + I_{k+1} \left( \frac{\rho}{2} \right) \right]^2 \cdot \exp\left[-(2k + 1)^2 \pi \Delta v F \tau \right] \cdot \cos \left\{ (2k + 1)(1 - \delta) \frac{\pi}{2} \right\}
\]

(7-12)

---

\(^5\)The combination of hard limiting in the IF and the phase noise of the lasers would result in an unacceptably high floor in the BER characteristic.

\(^6\)My parameter \(\delta\) is identical to the parameter \(\beta\) in [55].
7.1. *DELAY-LINE DISCRIMINATION*

![Diagram](image)

**Figure 7-3:**
Left: The phasor diagrams of the two discriminator input signals where \( d(t) \) is taken as the reference. In the case of a transmitted Mark both phasors must be in the right half plane for a proper decision.

Right: The pdf of the phase difference \( \Delta \theta \). Shaded regions of the pdf yield decision errors, \( \delta = 1 \).

Using this equation, Iwashita has determined the sensitivity penalty as a function of \( \Delta \nu_{IF}T \) and \( \delta \). From [55, Fig.4] one can obtain the following expression for the receiver sensitivity penalty as a function of the IF linewidth \( \Delta \nu_{IF} \):

\[
\Delta SNR = 4.4 \times 10^4 (\Delta \nu_{IF}T)^{1.88} \quad [dB] \quad (\delta = 1)
\]  
(7-13)

For \( \delta \) not equal to one it is more difficult to derive a modified expression, but the following result is accurate to within about 0.2 dB:

\[
\Delta SNR \approx 4.4 \times 10^4 \left( \frac{\Delta \nu_{IF}T}{\delta^2} \right)^{1.88} \quad [dB] \quad (0.5 < \delta < 1)
\]  
(7-14)

Note that this analysis does not incorporate post-detection filtering.

A different type of analysis is performed by Garrett and Jacobson [56] and Franz [23] for delay-line demodulators in general, and DPSK receivers in particular. They model the pdf of the phase variations of the IF signal as having a normal distribution (see also section 3.6), a result of the random phase walk. They also assume that the IF filter bandwidth is much larger than the IF linewidth, which means that the phase noise is passed undisturbed. The pdf of the discriminator output, which is essentially the difference of the input phases (7-8), therefore has a variance determined by the IF.
linewidth and the delay between the two discriminator branches:

\[ p(\Delta \theta) = \frac{1}{2\pi \sqrt{\Delta \nu_{IF} \tau}} \exp \left( -\frac{\Delta \theta^2}{4\pi \Delta \nu_{IF} \tau} \right) \] (7-15)

where the variance of \( \Delta \theta \) is equal to \( 2\pi \Delta \nu_{IF} \tau \). This variance will be denoted by \( \gamma' \),

\[ \gamma' = 2\pi \Delta \nu_{IF} \tau \] (7-16)

which is a modification of the IF-linewidth-to-bandwidth ratio \( \gamma = 2\pi \Delta \nu_{IF} T/m \) as used for ASK.

A bit error floor appears when, even in the situation of infinite SNR, wrong decisions are made due to, in this case, the phase noise. With infinite SNR the discriminator acts as a perfect phase comparator. In the case of a transmitted Mark a proper decision is made when both the phasor of \( s(t) \) and \( d(t) \) are in the right half plane (see Figure 7-3). The BER-floor then follows from

\[ P_{\text{floor}} = 2Pr \left\{ (1 + 4k) \frac{\pi}{2} \leq \Delta \theta \leq (3 + 4k) \frac{\pi}{2} \right\} \quad k = 0, 1, 2, \ldots \] (7-17)

\[ = 2 \sum_{k=0}^{\infty} Q \left( (1 + 4k) \frac{\pi}{2} \right) - Q \left( (3 + 4k) \frac{\pi}{2} \right) \]

\[ = \sum_{k=0}^{\infty} \text{erfc} \left[ \frac{(1 + 4k) \frac{\pi}{4}}{\sqrt{\Delta \nu_{IF} \tau}} \right] - \text{erfc} \left[ \frac{(3 + 4k) \frac{\pi}{4}}{\sqrt{\Delta \nu_{IF} \tau}} \right] \] (7-18)

where \( Q(x) \) and \( \text{erfc}(x) \) are the complementary normal distribution function and the complementary error function, respectively. Both are defined in Appendix A.2. It is easy to show that the first term for \( k=0 \) dominates for all practical values of interest, so the expression for \( P_{\text{floor}} \) can be reduced to

\[ P_{\text{floor}} = \text{erfc} \left[ \frac{1}{4} \sqrt{\frac{\pi}{\Delta \nu_{IF} \tau}} \right] \approx \frac{4}{\pi} \sqrt{\Delta \nu_{IF} \tau} \exp \left( -\frac{\pi}{16 \Delta \nu_{IF} \tau} \right) \] (7-19)

\[ \approx \frac{1}{\sqrt{\pi}} \sqrt{\frac{8\gamma'}{\pi^2}} \exp \left( -\frac{\pi^2}{8\gamma'} \right) \] (7-20)

The analysis presented above does not incorporate the influence of the IF bandwidth, since this is a basic assumption. However, the IF bandwidth expansion factor \( m \) appears in the result when post-detection filtering is included. For this, we'll suppose that the assumption of undistorted linewidth characteristics after IF filtering remains valid. The IF linewidth-to-bandwidth ratio \( \Delta \nu_{IF} T/m \) is thus small. The IF detector output \( z(t) \) can be written, in the case of optimum detection with \( \delta=1 \) and a transmitted Mark, as (7-8):\[ z = \cos \Delta \theta \] (7-21)
where $\Delta \theta$ is equal to the phase difference $\theta(t) - \theta(t-\tau)$ and has variance $\gamma'$. The pdf of $\Delta \theta$ is still given by (7-15), so the pdf of $z$ can be obtained from $p(z)dz = p(\Delta \theta)d\Delta \theta$. With $\Delta \theta = \arccos z$ and $-\pi \leq \Delta \theta \leq 0$ this yields

$$p(z) = \frac{1}{\sqrt{2\pi \gamma'}} \exp \left( -\frac{\arccos^2 z}{2 \gamma'} \right) \frac{1}{\sqrt{1 - z^2}} \quad (-1 \leq z \leq 1) \tag{7-22}$$

In analogy to the single-filter case, the post-detection filter output is obtained by averaging $m$ samples of the IF detector output according to (3-71). This is equivalent to $m$-fold convolution of $p(z)$, which leads to the final expression for $p(u)$:

$$p(u) = (2\pi \gamma' (1 - u^2))^{-m/2} \exp \left( -\frac{m \arccos^2 u}{2 \gamma'} \right) \quad (-1 \leq u \leq 1) \tag{7-23}$$

Note that this is the product of an Gaussian-like pdf centered at $u=1$ (!) and a function with two asymptotes at $u = \pm 1$. The resulting pdf is thus skewed in the direction of $u=1$.

The BER-floor can now be calculated by integration from $-1$ to $0$ (the decision threshold) of $p(u)$:

$$P_{floor} = (2\pi \gamma')^{-m/2} \int_{-1}^{0} \exp \left( -\frac{m \arccos^2 u}{2 \gamma'} \right) (1 - u^2)^{-m/2} du \tag{7-24}$$

Using the inverse substitution from the one used previously, i.e. $\Delta \theta = \arccos u$, the integral can be simplified. After insertion of a factor 2 for the symmetrical behaviour of $\Delta \theta$ as in Figure 7-3, we obtain

$$P_{floor} = 2(2\pi \gamma')^{-m/2} \int_{-\pi/2}^{\pi/2} \exp \left( -\frac{m \Delta \theta^2}{2 \gamma'} \right) (1 - \cos^2 \Delta \theta)^{-m/2} d\Delta \theta \tag{7-25}$$

$$= -2(2\pi \gamma')(1-m)/2 \int_{-\pi/2}^{\pi/2} \exp \left( -\frac{m \Delta \theta^2}{2 \gamma'} \right) \sin^{-m} \Delta \theta \ d\Delta \theta \tag{7-26}$$

It can be checked that for $m=1$ this expression reduces to (7-20). The general solution of the integral is more difficult due to the sine-term. A lower bound of the BER-floor can be obtained by setting $\sin \Delta \theta = 1$, which yields:

$$P_{floor} = -2(2\pi \gamma')^{1-m}/2 \int_{-\pi/2}^{\pi/2} \exp \left( -\frac{m \Delta \theta^2}{2 \gamma'} \right) d\Delta \theta \tag{7-27}$$

The integral can be solved using (Gr.3.321-2) and gives

$$I = \frac{2\pi \gamma'}{2} \left\{ \text{erfc} \left( \frac{\pi}{\sqrt{2}\gamma'} \right) - \text{erfc} \left( \frac{\pi}{2\sqrt{2}\gamma'} \right) \right\}$$

$$\approx -\frac{\sqrt{2\pi \gamma'}}{2} \text{erfc} \left( \sqrt{\frac{\pi^2}{8\gamma'}} \right) \tag{7-28}$$
Figure 7-4: The bit error rate floor of a CPFSK or DPSK receiver with delay-line frequency discrimination/IF-detection, as a function of the relative IF linewidth and the IF bandwidth expansion factor m.

The final expression for the lower bound of the BER-floor thus becomes:

\[ P_{floor} = (2 \pi \gamma')^{(1-m)/2} \text{erfc} \left( \sqrt{\frac{\pi^2}{8 \gamma'}} \right) \]

\[ \approx \sqrt{\frac{8 \gamma'}{m \pi^3 (2 \pi \gamma')^{m-1}}} \exp \left( -\frac{\pi^2 m}{8 \gamma'} \right) \]

(7-29) \hspace{1cm} (7-30)

with \( \gamma' = 2 \pi \Delta \nu_{IF} \tau \).

These linewidth results are valid for both DPSK and CPFSK. In the case of DPSK \( \tau = T \), while for CPSK it follows from (7-6) that \( \tau \) is equal to \( \delta T / 2M \). With \( \delta = 1 \) and \( M = 0.5 \) the required delay time of CPFSK is thus identical to the DPSK delay. The theoretical performance of MSK (essentially a form of phase modulation with ramp functions instead of step functions) is thus equal to DPSK. For \( M = 1 \) the delay time reduces to \( \tau = T/2 \), which means that the linewidth may double in order to obtain
the same BER floor and sensitivity penalty. Due to the steep erfc-function, the BER floor, for the same linewidth, decreases rapidly with increasing modulation index.

7.1.3 Comparison of CPFSK and FSK linewidth requirements

Now that we have obtained expressions for the linewidth-induced BER-floor of receivers with either single-filter, dual-filter or delay-line IF-detection, it is possible to make a comparison between the different detection schemes. To this end we must find a common independent parameter, since for non-coherent IF-detection the floor is expressed as a function of \( m \), while for differentially coherent IF-detection the parameter is \( \tau \). A better, and independent, parameter is the frequency deviation \( M \). For FSK with dual-filter detection we'll assume that the two IF filters just touch in the center. The frequency modulation index is then equal to \( m \). For ASK with single-filter detection the same \( M \) as for FSK will be used.

For CPFSK/DPSK the following relation between \( \Delta \nu_{IF} \tau \) and \( m \) can be derived from (7-30) and Figure 7-4 for obtaining a BER-floor at \( 10^{-20} \):

\[
\Delta \nu_{IF} \tau = 0.0045m \quad (P_{floor} = 10^{-20}) \tag{7-31}
\]

By doubling the IF bandwidth, the receiver thus becomes twice as tolerant to the IF linewidth. Substituting\(^7\)

\[
\frac{\tau}{T} = \frac{\delta}{2M} \tag{7-32}
\]
yields (for optimum detection efficiency \( \delta = 1 \))

\[
\Delta \nu_{IF} T = 0.009Mm \tag{7-33}
\]

Inverting (5-16) gives the same type of relation for FSK dual-filter detection:

\[
\Delta \nu_{IF} T = 0.04M^{1.54} \tag{7-34}
\]

In the same way we find with (3-103) for ASK

\[
\Delta \nu_{IF} T = 0.02M^{1.54} \tag{7-35}
\]

These three relations are illustrated in Figure 7-5. All three IF-detection schemes feature an increasing linewidth tolerance for increasing frequency modulation index and IF bandwidth. In contrast to the theoretical analyses published so far [23, 24, 55, 56], which only evaluated receivers without post-detection filtering, we find that the tolerance of the different IF-detection methods is of the same order. This is a very

\(^7\)The next part applies only for CPFSK, since the delay time \( \tau \) of DPSK remains equal to \( T \) under all circumstances.
important result, since it implies that for CPFSK-DPSK extremely narrow linewidth lasers are not always required. By taking wider IF filters, which may be another problem by itself, one can increase the linewidth tolerance of all schemes. In Figure 7-5 it is assumed that for CPFSK the IF bandwidth expansion $m$ must, if possible, be at least 2 higher than $M$. It also shows that e.g. $m=2$ can be used for a very limited range of modulation indexes only, from 0.5-1. At the same time an IF linewidth of 1-2% of the bitrate is required.
7.2. CPFSK

How can it be explained that non-coherent and differentially coherent IF-detection yield approximately the same linewidth sensitivity? To this end we must take a look at the parameters that we are comparing; the frequency modulation index $M$ and the frequency variation of the IF $\Delta \nu_{IF}$. In the absence of additive Gaussian noise, the noise after frequency detection is determined by the phase noise which is converted from a frequency variation into a baseband amplitude variation. The baseband signal amplitude is determined by the frequency deviation. Assuming for the moment a linear frequency discrimination characteristic one can derive the following relation:

$$SNR_o = \frac{\bar{u}_{signal, baseband}}{\sqrt{\langle u \rangle_{noise, baseband}}} \propto \frac{\Delta f}{\Delta \nu_{IF}}$$  \hspace{1cm} (7-36)

This shows that the minimum SNR, and thus the BER-floor, is determined by the ratio of the frequency deviation and the frequency noise or linewidth. Both FSK and CPFSK therefore have the same mechanism for lowering the BER-floor by increasing the frequency deviation. The only difference that remains between the two schemes is the form of the frequency discriminator curve. For the delay line discriminator this is a smooth cosine, while for FSK non-coherent dual-filter detection it is a rectangular characteristic. The latter can be regarded as a clipped version of the former. Indeed, if we compare the required linewidth for CPFSK at $m = 2M$ with the requirement at the same $M$ for FSK we find nearly the same values. This proves that, when optimally performed, both IF-detection schemes can extract the same information from the IF signal. The remaining differences are due to the fact that the two analyses are based on different assumptions.

7.2 CPFSK

7.2.1 IF-detection

After IF bandpass filtering by a filter with bandwidth $m/T$, the filtered IF signal $s(t)$ is obtained (7-1), which can be written more conveniently as

$$s(t) = x(t)\cos \omega_c t - y(t)\sin \omega_c t$$  \hspace{1cm} (7-37)

The amplitude signals $x(t)$ and $y(t)$ contain both Signal and noise, with a total SNR

$$\xi = \frac{A^2}{2mN} = \frac{\rho}{m}$$  \hspace{1cm} (7-38)

The delayed signal $d(t)$ now becomes

$$d(t) = x(t - \tau)\cos \omega_c (t - \tau) - y(t - \tau)\sin \omega_c (t - \tau)$$
$$= v(t)\cos \omega_c (t - \tau) - w(t)\sin \omega_c (t - \tau)$$  \hspace{1cm} (7-39)

\footnote{Remember that the total IF bandwidth of the CPFSK receiver has the expansion factor $m$, while each filter of an FSK receiver has the same expansion. For proper comparison we must thus compare the situation of FSK-$m$ with CPFSK-$2m$ at the same $M$.}
The noise phasors $x(t), y(t)$ and $v(t), w(t)$ are decorrelated when $\tau > 1/B_{IF}$ or $\tau > T/m$, whereas the real delay time is equal to $\delta T/2M$. Thus, for decorrelated signals $m$ should be larger than $2M$. This is illustrated in Figure 7-2. For the practical, ISL-dictated IF bandwidths as presented in section 7.1.1 the real delay is always larger than $T/m$. The only exception is $M=2$, where $\tau = T/4$ and $T/m$ becomes equal to $T/3$. The latter is most probably due to the accuracy of the simulation, and is not considered as a violation of the general rule given above. In general it is therefore valid to assume that the delayed signal is decorrelated.

The output $z(t)$ of the frequency discriminator is now given by the four cross-products of the components of $s(t)$ and $d(t)$:

$$
z(t) = x(t)v(t)\cos \omega_c t \cos \omega_c (t - \tau)
+ y(t)w(t)\sin \omega_c t \sin \omega_c (t - \tau)
- x(t)w(t)\cos \omega_c t \sin \omega_c (t - \tau)
- y(t)v(t)\sin \omega_c t \cos \omega_c (t - \tau)
= \frac{1}{2} x(t)v(t) \{ \cos \omega_c \tau + \cos(2\omega_c t - \omega_c \tau) \}
+ \frac{1}{2} y(t)w(t) \{ \cos \omega_c \tau - \cos(2\omega_c t - \omega_c \tau) \}
+ \frac{1}{2} x(t)w(t) \{ \sin \omega_c \tau - \sin(2\omega_c t - \omega_c \tau) \}
- \frac{1}{2} y(t)v(t) \{ \sin \omega_c \tau - \sin(2\omega_c t - \omega_c \tau) \}
$$

(7-40)

(7-41)

First of all the second harmonics around $2\omega_c$ will be filtered out by the post-detection filter. Furthermore $\omega_c$ must be located in a zero-crossing of the detection characteristic, so

$$
\omega_c = \frac{(2n + 1)}{4} \cdot \frac{2\pi}{\tau}
$$

(7-42)

This implies that $\cos \omega_c \tau$ is equal to 0, while $\sin \omega_c \tau$ is equal to $(-1)^n$. This yields

$$
z(t) = \frac{1}{2} \{ x(t)v(t) - y(t)w(t) \} \cdot (-1)^n
$$

(7-43)

As previously mentioned, the second zero-crossing at $n=1$ is often used, although the analysis yields identical results for all $n$.

The four components $v(t), w(t), x(t)$ and $y(t)$ are all (non)-central normally distributed, but the product of normal variates is not easily calculated. The usual way of treating these forms is by writing them as sums of squares [68, 98, 106], which yields (for $n=1$ and omitting all references to time):

$$
z(t) = \frac{1}{2} \{ yv - xw \}
= \frac{1}{8} \{ (x - w)^2 + (y + v)^2 - (x + w)^2 - (y - v)^2 \}
= \frac{1}{8} \{ z_1 - z_2 \}
$$

(7-44)

(7-45)
with
\[ z_1 = (x - w)^2 + (y + v)^2 \]  \hspace{1cm} (7-46)\]
\[ z_2 = (x + w)^2 + (y - v)^2 \]  \hspace{1cm} (7-47)\]

Since \( v, w, x, y \) are normally distributed, the forms \((x \pm w)\) and \((y \pm v)\) are also normally distributed. The variates \( z_{1,2} \) are then sums of squared normal variates and have (non)-central chi-square distributions.

![Figure 7-6: The phasors of the two discriminator input signals \( d(t) \) and \( s(t) \), with the former used as reference.](image)

**Left**: The total phase rotation, including the average IF.
**Right**: The IF phase rotation relative to \( \omega_c \).

Using \( d(t) \) at \( t=0 \) as a reference, and with the help of Figure 7-6, the four phasor components can now be written as:
\[ d(t) : v := A + v \]  \hspace{1cm} (7-48)\]
\[ w := w \]  \hspace{1cm} (7-49)\]
\[ s(t) : x := A \cos \omega_d \tau + x \]  \hspace{1cm} (7-50)\]
\[ y := \Xi(t) A \sin \omega_d \tau + y \]  \hspace{1cm} (7-51)\]

where \( \omega_d = \omega_c / 2\pi \) is the carrier-to-peak frequency deviation, and thus equal to \( \Delta f / 2 \), see also (7-3). \( \Xi(t) \) is the data content, with possible values of 1 for a Mark and -1 for a Space. The four (new) noise signals \( v, w, x, y \) have identical (central) normal distributions with variances \( mN \). The output component \( z_1 \) is then
\[ z_1 = (A \cos \omega_d \tau + x - w)^2 + (A + \Xi A \sin \omega_d \tau + y + v)^2 \]  \hspace{1cm} (7-52)\]
and obviously has a chi-square distribution with two degrees of freedom. The variance of \( x - w \) and \( y + v \) is \( 2mN \), so the non-centrality parameter of the pdf of \( z_1 \) is found by

\[
\lambda_1 = \frac{1}{\sigma^2} \sum_i A_i^2 = \frac{A^2 \cos^2 \omega_d \tau + (1 + \Xi(t) \sin \omega_d \tau)^2}{2mN} = \frac{A^2}{mN} (1 + \Xi(t) \sin \omega_d \tau)
\]

(7-53)

In the same way the \( \lambda \) of \( z_2 \) can be derived, resulting in the following two pdf's:

\[
p(z_1) = p \left( z \left| 2, \frac{A^2 \alpha}{mN} \right. \right) = p \left( z \left| 2, \frac{A^2 \alpha}{mN} \right. \right)
\]

(7-54)

\[
p(z_2) = p \left( z \left| 2, \frac{A^2 \beta}{mN} \right. \right) = p \left( z \left| 2, \frac{A^2 \beta}{mN} \right. \right)
\]

(7-55)

The two parameters

\[
\alpha = 1 + \Xi(t) \sin \omega_d \tau
\]

(7-56)

\[
\beta = 1 - \Xi(t) \sin \omega_d \tau
\]

(7-57)

can be seen as detection improvement factors, since the detection efficiency \( \delta \) influences \( \alpha \) and \( \beta \) through the relation \( \omega_d \tau = \delta \pi / 2 \). The sum of \( \alpha \) and \( \beta \) is always equal to 2, but \( \delta \) determines if either can attain the values of 0 and 2, depending upon the transmitted data. In other words, the rotation of the phasor in Figure 7-6 is exactly \( 90^\circ \) (necessary for maximum signal power) only if \( \delta \) is equal to 1.

The pdf of the frequency discriminator output is obtained by subtraction of \( z_1 \) and \( z_2 \) according to (7-45). Using the definition of the cross-correlation from Appendix A this gives:

\[
p(z) = p \left( z \left| 2, \frac{A^2 \alpha}{mN} \right. \right) * p \left( -z \left| 2, \frac{A^2 \beta}{mN} \right. \right)
\]

(7-58)

With an integrate-and-dump post-detection filter acting as a finite-time integrator, the final pdf of the filter output is obtained through \( m \)-fold convolution of (7-58), yielding

\[
p(u) = p_1 \left( u \left| 2m, \frac{A^2 \alpha}{N} \right. \right) * p_2 \left( -u \left| 2m, \frac{A^2 \beta}{N} \right. \right)
\]

Mark

(7-59)

\[
p(u) = p_1 \left( u \left| 2m, \frac{A^2 \beta}{N} \right. \right) * p_2 \left( -u \left| 2m, \frac{A^2 \alpha}{N} \right. \right)
\]

Space

(7-60)

(7-61)

In the case of optimum detection \( \delta = 1 \), (and e.g. a Mark) \( \alpha = 2 \) and \( \beta = 0 \), so \( p(u) \) reduces to

\[
p(u) = p \left( u \left| 2m, \frac{2A^2}{N} \right. \right) * p \left( -u | 2m \right)
\]

Mark

(7-62)
7.2. CPFSK

This is the same expression as (5-20) for dual-filter FSK detection, but with a two 
times higher noncentrality parameter - and thus SNR. This confirms that CPFSK has 
indeed a 3 dB better performance than non-coherent FSK IF-detection.

7.2.2 The Bit-Error Rate

The difference between pdf's $p(u)$ - (5-20) for FSK and (7-59) for CPFSK - is that for 
CPFSK in general two non-central chi-square pdf's must be cross-correlated. For FSK, 
with one central and one non-central pdf to be cross-correlated, the convolution could 
still be done, but with two non-central pdf's this becomes very difficult. Therefore the 
BER will not be derived by computing the pdf of $u$ first, but through integration over 
all combinations of $u_1$ and $u_2$ that yield a negative $u$ for a transmitted Mark. Again 
we assume equal a priori Mark and Space probabilities, so the BER can be written as:

$$P_e = Pr(|u_1 - u_2| \leq 0) = Pr(|u_2| \geq |u_1|)$$

$$= \int_{u_1=0}^{\infty} p_1(u_1) \int_{u_2=0}^{\infty} p_2(u_2) \, du_2 \, du_1$$

(7-63)

First the inner integral - denoted by $I$ - is solved, which is equivalent to calculating the 
conditional probability $Pr(u_2 \geq u_1 | u_1)$. When this integral is written in the following form

$$I = \int_{-\infty}^{u} p_2(y) \, dy = 1 - \int_{0}^{u} p_2(y) \, dy$$

(7-64)

it becomes clear that, given the fact that $p_2(y)$ is a non-central chi-square pdf, the 
integral in the right-hand side of the expression has the same form as the partial error 
rate expression $P_M$ of single-filter IF-detection (4-98). With respect to the single-filter 
situation in section 4.2.3, $\rho$ is equal to $A^2 \beta / 2N$ and $u / 2N$ should be substituted for $Y$. The integral $I$ can then be written as a modified form of (4-104) and (4-105):

$$I = \exp \left( -\frac{u + A^2 \beta}{2N} \right) \sum_{j=0}^{\infty} \frac{1}{j!} \left( \frac{A^2 \beta}{2N} \right)^j \sum_{k=0}^{m+j-1} \frac{1}{k!} \left( \frac{u}{2N} \right)^k$$

(7-65)

$$= \exp \left( -\frac{A^2 \beta}{2N} \right) \sum_{j=0}^{\infty} \frac{1}{j!} \left( \frac{A^2 \beta}{2N} \right)^j \frac{\Gamma(m+j,u/2N)}{\Gamma(m+j)}$$

(7-66)

It can easily be checked with the help of the definition of the incomplete gamma 
function (Gr.8.352) (A.3-25) that for both $m \to \infty$ and $u=0$ the integral becomes 
equal to 1.

The error rate $P_e$ is now calculated from

$$P_e = \int_{0}^{\infty} p_1(u) \, I \, du$$
\[ P_e = \frac{1}{2^N} \left( \frac{1}{A^2 \alpha} \right)^{m+1} \exp \left( -\frac{A^2(\alpha + \beta)}{2N} \right) \cdot \sum_{j=0}^{\infty} \frac{1}{j!} \left( \frac{A^2 \beta}{2N} \right)^j \sum_{k=0}^{m+j-1} \frac{1}{k!} \left( \frac{1}{2N} \right)^k \cdot \int_0^\infty u^{m+1} \exp \left( -\frac{u}{N} \right) I_{m-1} \left( \frac{A}{N \sqrt{\alpha u}} \right) du \]  

(7-67)

The integral, which will be denoted by \( II \), is of the same form as the standard integral (A.3-8). It can therefore be solved as

\[ II = k! \left( \frac{A \sqrt{\alpha}}{2} \right)^{m+1} N^{k+1} \exp \left( \frac{A^2 \alpha}{4N} \right) L_k^{m-1} \left( -\frac{A^2 \alpha}{4N} \right) \]  

(7-68)

Combining these two expressions, and taking into account that \( \alpha + \beta \) is always equal to 2, yields the final expression for the BER of a CPFSK receiver with delay-line discrimination [46]:

\[ P_e = \left( \frac{1}{2} \right)^m \exp \left( -\frac{A^2}{N} + \frac{A^2 \alpha}{4N} \right) \cdot \sum_{j=0}^{\infty} \frac{1}{j!} \left( \frac{A^2 \beta}{2N} \right)^j \sum_{k=0}^{m+j-1} \left( \frac{1}{2} \right)^k L_k^{m-1} \left( -\frac{A^2 \alpha}{4N} \right) \]  

(7-69)

Using the definition of the Laguerre polynomial (Gr.8.970-1) (A.3-9) and substituting \( \rho \) for \( A^2/2N \) yields

\[ P_e = \left( \frac{1}{2} \right)^m \exp \left( -2 \rho + \frac{\rho \alpha}{2} \right) \sum_{j=0}^{\infty} \frac{(\rho \beta)^j}{j!} \cdot \sum_{k=0}^{m+j-1} \left( \frac{1}{2} \right)^k \frac{\Gamma(k + m)}{\Gamma(k - l + 1) \Gamma(m + l)} \frac{1}{l!} \left( \frac{\rho \alpha}{2} \right)^l \]  

(7-70)

In the case of a transmitted Mark and optimum detection (\( \delta=1 \)) the two detection improvement factors \( \alpha \) and \( \beta \) become equal to 2 and 0 respectively. This means that in (7-70) only the term \( j=0 \) gives a non-zero contribution to \( P_e \), in which case this reduces to:

\[ P_e = \left( \frac{1}{2} \right)^m \exp(-\rho) \sum_{k=0}^{m-1} \left( \frac{1}{2} \right)^k L_k^{m-1}(-\rho) \]  

(7-71)

\[ = \left( \frac{1}{2} \right)^m \exp(-\rho) \sum_{k=0}^{m-1} \left( \frac{1}{2} \right)^k \sum_{l=0}^{k} \binom{m + k - 1}{k - l} \frac{\rho^l}{l!} \]  

(7-72)
Figure 7-7: BER curves of a heterodyne CPFSK receiver with delay-line frequency detection, as a function of the IF bandwidth expansion factor $m$ and the detection efficiency $\delta$.

This is exactly the same expression as for the BER of a dual-filter receiver with quadratic IF-detection (5-34), but with $\rho$ instead of $\rho/2$. It may thus be concluded that the ultimate sensitivity of CPFSK is indeed a factor of 2 (or 3 dB) better compared to dual-filter FSK with non-coherent IF-detection. As in FSK, the BER may be written as the combination of a fixed part $\frac{1}{2} \exp(-\rho)$ and a BER deterioration factor
Table 7-1: The sensitivity penalty $\Delta SNR$ in dB of a heterodyne CPFSK receiver with delay-line frequency detection as a function of the IF bandwidth expansion factor $m$ and the detection efficiency $\delta$. Penalties are relative to the configuration $m = \delta = 1$, with a sensitivity of 20.03.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\Delta SNR$ [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.39</td>
</tr>
<tr>
<td>3</td>
<td>0.66</td>
</tr>
<tr>
<td>4</td>
<td>0.88</td>
</tr>
<tr>
<td>5</td>
<td>1.06</td>
</tr>
<tr>
<td>6</td>
<td>1.22</td>
</tr>
<tr>
<td>7</td>
<td>1.35</td>
</tr>
<tr>
<td>8</td>
<td>1.48</td>
</tr>
<tr>
<td>9</td>
<td>1.60</td>
</tr>
<tr>
<td>10</td>
<td>1.70</td>
</tr>
</tbody>
</table>

$C(m|\delta)$:

$$P_e = \frac{1}{2} C(m|\delta) \exp(-\rho) \quad (7-73)$$

$$C(1|1) = 1 \quad (7-74)$$

$$C(2|1) = 1 + \frac{1}{4} \rho \quad (7-75)$$

$$C(3|1) = 1 + \frac{3}{8} \rho + \frac{1}{32} \rho^2 \quad (7-76)$$

$$C(4|1) = 1 + \frac{29}{64} \rho + \frac{1}{16} \rho^2 + \frac{1}{384} \rho^3 \quad (7-77)$$

$$C(5|1) = 1 + \frac{65}{128} \rho + \frac{23}{256} \rho^2 + \frac{5}{768} \rho^3 + \frac{1}{6144} \rho^4 \quad (7-78)$$

7.2.3 Sensitivity and penalties

The sensitivity of the optical heterodyne CPFSK receiver with delay-line frequency detection can be evaluated using (7-70), with both the IF bandwidth expansion factor $m$ and the detection efficiency $\delta$ as parameters. Figure 7-7 shows the BER-curves for some values of $m$ between 1 and 6, and $\delta$-values of 1.0, 0.7 and 0.5. The ultimate
7.2. CPFSK

Figure 7-8: The sensitivity penalty $\Delta SNR$ of a heterodyne CPFSK receiver, as a function of the IF bandwidth expansion factor $m$ and the detection efficiency $\delta$. The dashed line gives the results of the dual-filter receiver with non-coherent IF as a reference.

sensitivity of CPFSK for $m=1$ and $\delta=1$ is 20.03 or 13.01 dB, exactly 3 dB better than dual-filter FSK.

Figure 7-8 shows the sensitivity penalty $\Delta SNR$ for increasing $m$ and $\delta$ (remember that it has already been indicated that $m=1$ is no realistic value for CPFSK and that $m$ will in general be larger than 2). For every value of $\delta$ a compact formula of the form (5-41) can be derived, e.g.:

\[
\Delta SNR = 0.625 \log^2 m + 1.085 \log m \quad [dB] \quad (\delta = 1) \quad (7-79)
\]
\[
\Delta SNR = 0.538 \log^2 m + 0.204 \log m + 2.139 \quad [dB] \quad (\delta = 0.7) \quad (7-80)
\]

For $m=1$ the sensitivity penalties of CPFSK and dual-filter FSK are identical, as would be expected.

Non-optimum detection efficiency

The most interesting aspect of Figure 7-8 and Table 7-1 is the development of the sensitivity penalty when both $m$ and $\delta$ are non-optimum. At lower values of the detection efficiency the sensitivity penalty $\Delta SNR$ depends much less upon $m$. In fact the penalty becomes unacceptably high for $\delta$ lower than 0.8-0.7, while for $\delta=0.6$ the performance of CPFSK is even lower than that of dual-filter FSK with non-coherent IF-detection. In contrast to the IF bandwidth-dependent sensitivity penalty, which
Figure 7-9: The sensitivity penalty $\Delta SNR$ in dB of a heterodyne CPFSK receiver as a function of the detection efficiency $\delta$. Curves are given for $m=2$, 4, 6, 8 and 10, each one referenced to its sensitivity for $\delta=1$.

typically increases by about 1-1.5 dB up to $m=10$, the efficiency-dependent penalty easily becomes 5 dB.

The large influence of the detection efficiency can be explained by looking at the detection process in the delay-line discriminator. The cosine frequency discriminator characteristic as depicted in Figure 7-1 is valid only for sinusoidal input Signals. The phase rotation in the delayed branch ensures that the two mixer inputs are periodically orthogonal, resulting in a zero in the characteristic. However, the phase of the Signal and noise voltages is of no importance for the $SxN$ and $NxN$ output components of the mixer. The total output can thus be presented as

$$u_o = S_s x S_d + S_s x N_d + S_d x N_s + N_s x N_d$$
$$= S^2 \cos(\omega_d \tau) + 2SxN + NxN$$

(7-81)

The factor $\omega \tau$ is equal to $\frac{\pi}{2} + n \cdot \pi + \frac{\pi}{2} \cdot \delta$. The output SNR of the mixer is then:

$$SNR_o = \frac{S^2 \sin(\frac{\pi \delta}{2})}{2SxN + NxN}$$

(7-82)

This clearly shows that non-optimum detection with $\delta$ lower than 1 gives a reduced Signal component, while the amount of output noise remains the same. The output SNR thus critically depends upon an optimum detection efficiency. It should at the
same time be stressed that the 'intuitive' explanation presented here is not exact since
the sensitivity penalty $\Delta SNR$ in Figure 7-8 is larger than $20 \log \sin(\frac{\pi}{2} \delta)$. This can be
explained in the following way. With respect to the pdf $p(u)$ (7-59), $\delta$ determines the
skewness. For optimum detection efficiency, $p(u)$ results from the cross-correlation of
a non-central and a central chi-square pdf. Apart from the negative tail, this yields
a resulting pdf that differs only marginally from the original non-central chi-square
pdf. For decreasing $\delta$ however, both pdf’s become non-central until, for $\delta = 0$, they are
identical. But cross-correlation of identical pdf’s yields a symmetrical pdf around 0,
with equal Mark and Space error probabilities of exactly $\frac{1}{2}$. The average value $<u>$
of $p(u)$ thus shifts more rapidly to 0 than that of a single chi-square pdf, giving a
higher sensitivity penalty.

CPFSK-FSK sensitivity improvement

From the analysis above one can also derive a reason for the times two higher sensitivi-
ty of CPFSK relative to FSK. Comparing a quadratic FSK detector - which essentially
multiplies the bandpass filtered IF signal with itself - and a delay-line frequency dis-
riminator yields one important difference. In the FSK quadratic IF-detector two
components $S_x N_2 + S_2 x N_1$ (where the subscripts 1 and 2 denote
the numbers of the inputs$^9$). However, since both inputs have the same signal these
two components are fully correlated and add coherently.

On the other hand, in the delay-line discriminator the two noise signals at the
direct and delayed inputs are decorrelated. The two SxN-components are then also
decorrelated and add incoherently. This yields, while the SxS components of FSK and
CPFSK are identical, a times two lower SxN noise output power for CPFSK. It is easy
to show that the same holds for the NxN noise. The noncentrality parameter, which
is directly linked to the SNR, will thus be a factor of 2 higher in the case of CPFSK,
resulting in a two times or 3 dB higher sensitivity.

On the other hand it should be remembered that the IF bandwidth expansion
factors are defined differently. For CPFSK $m$ covers the complete IF band, while for
FSK dual-filter detection each filter has an index $m$. In the case of wide deviation
FSK with delay-line discrimination one must thus compare CPFSK with index $2m$ with
FSK and index $m$. The sensitivity penalty for CPFSK relative to $\rho = 20.03$ ($m = 1$) is
the same as (5-41):

$$\Delta SNR_{CPFSK} = 1.085 \log m + 0.625 \log^2 m \quad [dB] \quad (7-83)$$

In the case of very wide deviation FSK modulation with delay-line (CPFSK) demodu-
lization this expression can be modified as

$$\Delta SNR_{CPFSK} = 1.085 \log(M + m) + 0.625 \log^2(M + m) \quad [dB] \quad (7-84)$$

$^9$An IF detector has in principle only one input. But the output of a quadratic detector can be
modelled as the product of two 'virtual' inputs, both having the same input signal. The output
components are then the result of pairwise multiplication, with one component from each input.
For the evaluation of dual-filter versus delay-line IF-detection, this sensitivity penalty must be compared with the dual-filter $\Delta SNR$ relative to the same SNR of 20.03 (1):

$$\Delta SNR_{FSK} = 3 + 1.085 \log m + 0.625 \log^2 m \quad [dB] \quad (7-85)$$

It is easily verified that for very large frequency modulation indexes $M$ the performance of the delay-line discriminator becomes inferior to that of the dual-filter IF-detector. However, for moderate modulation indexes (smaller than 20) CPFSK delay-line demodulation remains more sensitive than dual-filter detection.

### 7.3 Polarisation diversity CPFSK

#### 7.3.1 The Bit-Error Rate

Polarisation diversity CPFSK is the most commonly used coherent optical detection scheme for systems operating at bitrates above 1 Gbit/s. Like for ASK and FSK systems with non-coherent IF-detection, the influence of adding a second diversity branch can be translated into a doubling of the degrees of freedom. Since this is a straightforward extension of the analyses presented in Chapter 6 and section 7.2, it will not be repeated here. In the case of polarisation diversity heterodyne CPFSK with delay-line differentially coherent IF-detection, the pdf of the post-detection filter output for a transmitted Mark is given by (7-59) with all factors $m$ replaced by $2m$:

$$p(u) = p_1 \left( u \left| 4m, \frac{A^2\alpha}{N} \right. \right) \ast p_2 \left( -u \left| 4m, \frac{A^2\beta}{N} \right. \right) \quad (7-86)$$

The BER then becomes identical to (7-69) with, again, all $m$ replaced by $2m$:

$$P_e = \left( \frac{1}{2} \right)^{2m} \exp \left( -\frac{\rho\alpha}{2} \right) \sum_{j=0}^{\infty} \frac{(\rho\beta)^j}{j!} \cdot \sum_{k=0}^{2m+j-1} \left( \frac{1}{2} \right)^k L_k^{2m-1} \left( -\frac{\rho\alpha}{2} \right)$$

(7-87)

With (7-87) the sensitivity penalties $\Delta SNR$ and the diversity penalties $\Delta Div$ can be computed. These are listed in Table 7-2 and are illustrated in Figure 7-10. The compact expression for $\Delta SNR$ at optimum detection efficiency is identical to that of FSK:

$$\Delta SNR = 1.461 \log m + 0.625 \log^2 m \quad [dB] \quad (\delta = 1) \quad (7-88)$$

#### 7.3.2 SNR imbalance

As already mentioned, polarisation diversity CPFSK is the most commonly used modulation and optical detection scheme at high bitrates. However, this requires IF fre-
Table 7-2: Sensitivity penalties $\Delta SNR$ and polarisation diversity penalties $\Delta Div$ (both in dB) of CPFSK polarisation diversity receivers. All $\Delta SNR$ are relative to the ultimate CPFSK SNR at a BER of $10^{-9}$ of 20.03. The penalties $\Delta Div$ are relative to the sensitivity of the single branch CPFSK receiver with the same $m$ and $\delta$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\delta = 1.0$</th>
<th>$\delta = 0.7$</th>
<th>$\delta = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta SNR$ dB</td>
<td>$\Delta Div$ dB</td>
<td>$\Delta SNR$ dB</td>
</tr>
<tr>
<td>1</td>
<td>0.39</td>
<td>0.39</td>
<td>2.25</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>0.49</td>
<td>2.45</td>
</tr>
<tr>
<td>3</td>
<td>1.22</td>
<td>0.56</td>
<td>2.62</td>
</tr>
<tr>
<td>4</td>
<td>1.48</td>
<td>0.60</td>
<td>2.78</td>
</tr>
<tr>
<td>5</td>
<td>1.70</td>
<td>0.64</td>
<td>2.92</td>
</tr>
<tr>
<td>6</td>
<td>1.89</td>
<td>0.67</td>
<td>3.05</td>
</tr>
<tr>
<td>7</td>
<td>2.06</td>
<td>0.71</td>
<td>3.16</td>
</tr>
<tr>
<td>8</td>
<td>2.21</td>
<td>0.73</td>
<td>3.27</td>
</tr>
<tr>
<td>9</td>
<td>2.35</td>
<td>0.75</td>
<td>3.38</td>
</tr>
<tr>
<td>10</td>
<td>2.47</td>
<td>0.77</td>
<td>3.47</td>
</tr>
</tbody>
</table>

frequencies spanning the range from e.g. 3 to 8 GHz (for typical 2.5 Gbit/s systems), which makes it exceedingly difficult to obtain perfectly symmetrical IF branches. The same analysis as made for non-coherent FSK IF-detection with unequal SNR’s in the two diversity branches (section 6.1.5) can be performed. Using the CPFSK results obtained so far, equation (6-34) can thus be modified into the pdf of a polarisation diversity heterodyne CPFSK receiver with IF SNR imbalance $\kappa^2$:

$$p(u) = p_1( u \left| 4m, \frac{A^2 \alpha (\cos^2 \phi + \kappa^2 \sin^2 \phi)}{N} \right) *$$

$$* p_2 (-u \left| 4m, \frac{A^2 \beta (\cos^2 \phi + \kappa^2 \sin^2 \phi)}{N} \right)$$

(7-89)

As expected, for $\kappa$ not equal to 1 the final pdf after post-detection filtering has become polarisation dependent. With all signal in the ‘good’ branch ($\phi = 0 + n \cdot \pi$) expression (7-89) reduces to (7-86) and the maximum sensitivity is obtained. The other extreme occurs when all signal is in the ‘poor’ branch ($\phi = \frac{\pi}{2} + n \cdot \pi$), in which case the pdf reduces to

$$p(u) = p_1( u \left| 4m, \frac{A^2 \alpha \kappa^2}{N} \right) * p_2 (-u \left| 4m, \frac{A^2 \beta \kappa^2}{N} \right)$$

(7-90)
In order to obtain the same noncentrality parameter as in the optimum situation, a SNR equal to $\rho_{\text{opt}}/\kappa^2$ is required. Depending upon the actual SOP the sensitivity will thus fade between $\rho_{\text{opt}}$ and $\rho_{\text{opt}}/\kappa^2$, yielding an additional (polarisation dependent) penalty of $20\log\kappa$ dB. This yields exactly the same results as presented in Figure 6-10.

### 7.4 CPFSK phase diversity

#### 7.4.1 Receiver model

CPFSK phase diversity has been introduced in Chapter 2 as a way of optimising the IF bandwidth efficiency. By shifting the average IF to 0 Hz one obtains 'quasi-homodyne' detection, requiring less high-frequency electronics for a certain bitrate than with heterodyne reception. With the average IF equal to 0, the *Space* and *Mark* frequencies have the same value at opposite sides of 0 Hz, so at $\pm \Delta f/2 = \pm f_d$. Phase diversity CPFSK receivers therefore require a delay-and-cross-multiply frequency discriminator that can distinguish between positive and negative intermediate frequencies. In fact the frequency is not important, from a detection point of view the phase determines
the bit-value\textsuperscript{10}. For proper reception, phase diversity receivers require an electrical bandwidth equal to \( m/2T \), which is considerably less than the value \( f_c + m/2T \) for CPFSK heterodyning\textsuperscript{11}. The minimum required \( m \), given the modulation index \( M \), can again be obtained from Figure 7-2. These relaxed bandwidth requirements can in practice be used for increasing the frequency deviation, resulting in less stringent linewidth requirements, or simply for reduction of IF signal distortion by increasing the bandwidth.

For obtaining an asymmetric sine-like frequency discriminator characteristic around 0 Hz, one can use to advantage the 90\(^\circ\) out-of-phase IF signals from an optical 90\(^\circ\)-hybrid\textsuperscript{12}, e.g. a modified polarisation diversity unit [2, 27, 43, 81], a 3x3 fused coupler [16, 47] or a proper 4x4 fused coupler [63]. When using a symmetrical eight-port optical hybrid, the two IF Signal components are given by

\[
\begin{align*}
\hat{s}_I(t) &= \text{Re} \left\{ \frac{A}{\sqrt{2}} \exp(j\Xi(t)\omega_d t + \Delta\theta_{IF}) \right\} \\
&= \frac{A}{\sqrt{2}} \cos(\omega_d t + \Delta\theta_{IF}) \\
\hat{s}_Q(t) &= \text{Re} \left\{ \frac{A}{\sqrt{2}} \exp(j\Xi(t)\omega_d t + \Delta\theta_{IF} + \pi/2) \right\} \\
&= \frac{A}{\sqrt{2}} \Xi(t) \sin(\omega_d t + \Delta\theta_{IF})
\end{align*}
\]  

(7-91)

(7-92)

where \( \Xi(t) \) is the data content \( \pm 1 \). This result shows that the in-phase signal \( \hat{s}_I(t) \) can be regarded as a stationary carrier at a frequency \( f_d \), with the quadrature signal leading or lagging 90\(^\circ\) depending upon the actual bit sent. In contrast to ASK and DPSK phase diversity receivers there is no real diversity with signal fading between different branches. In fact, this type of detection requires that the two IF SNR's are identical, so an optical polarisation controller is needed\textsuperscript{13}.

\textsuperscript{10}For automatic frequency control (AFC) the real positive or negative frequencies remain an essential requirement, necessitating frequency discrimination.

\textsuperscript{11}Some typical examples for comparison:

- 'standard' 2.5 Gbit/s CPFSK heterodyne systems (e.g. [10]) have a central IF of 5 GHz, with
an IF between 3 and 8 GHz at a modulation index of 0.7–0.8. This gives an electrical bandwidth
efficiency of 0.3125.

- CPFSK phase diversity systems have been realised with a bitrate of 1 Gbit/s and an IF band-
width of 2.3 GHz [16, 27, 43, 47] at \( M=2.5 \), as well as a bitrate of 2.5 Gbit/s and a bandwidth of 4.0 GHz at \( M=1 \). This yields bandwidth efficiencies of 0.43 and 0.625 respectively.

\textsuperscript{12}Realising an electrical 90\(^\circ\) phase shifter from 0 to several GHz is (nearly) impossible, since
multi-decade electrical delay elements do not exist. The optical method on the other hand makes
use of the phases and/or the polarisation properties of the incoming optical signals, yielding a
frequency-independent 90\(^\circ\) phase rotation.

\textsuperscript{13}Theoretically one can combine CPFSK phase diversity and polarisation diversity in one
receiver, yielding a four-branch IF section. This is however not a practical solution, since the
polarisation diversity power splitting and the subsequent number of photodiodes (6 or 8) reduce
7.4.2 IF-detection

Assuming that indeed two identical and 90° out-of-phase IF signals are obtained by any of the different methods, the discriminator output \( z \) is given by:

\[
    z = z_1 - z_2 = \text{Re}\{s_I d_Q\} - \text{Re}\{s_Q d_I\}
\]  

\[ (7-93) \]

\((s(t)\) and \(d(t)\) are now simply the ‘baseband’ components, since the ‘carrier’ is equal to 0). Like for heterodyne CPFSK, the direct and delayed signal can be considered decorrelated when \(\tau > 1/B_{IF} = T/m\). For practical values of the frequency detection period \(f_m\) and the IF bandwidth expansion factor \(m\) this is true. Note that the IF bandwidth \(B_{IF}\) and \(f_m\) have been defined in exactly the same way as for heterodyne CPFSK, so in this case from negative to positive frequencies. The first component \(z_1\)

the LO-power on the photodiodes dramatically, leading to high receiver noise sensitivity penalties. Obtaining proper balance between the IF branches is an additional problem.
Figure 7-12: The phasors of the signals \( s_I \) and \( d_Q \) at the first frequency discriminator mixer.

of \( z \) can be written as

\[
z_1 = \text{Re}\{(x + jy)(v + jw)\}
\]

\[= xv - yw\]

\[= \frac{1}{4} \left\{ \left[ (x + v)^2 + (y - w)^2 \right] - \left[ (x - v)^2 + (y + w)^2 \right] \right\}\]  \hspace{1cm} (7-94)

\[= \frac{1}{4} \{z_A - z_B\}\]  \hspace{1cm} (7-95)

Omitting the scaling factor \( \frac{1}{4} \), \( z_1 \) is obtained by subtracting two non-central chi-square distributed variables \( z_A \) and \( z_B \). Using \( s_I(0) \) as reference, the different components of \( z_A \) and \( z_B \) can be obtained from Figure 7-12 as

\[
s_I: x := \frac{A}{\sqrt{2}} + x \hspace{1cm} (7-96)
\]

\[y := y\]  \hspace{1cm} (7-97)

\[
d_Q: v := \Xi(t) \frac{A}{\sqrt{2}} \sin \omega_d \tau + v \hspace{1cm} (7-98)
\]

\[w := -\Xi(t) \frac{A}{\sqrt{2}} \cos \omega_d \tau + w\]  \hspace{1cm} (7-99)

Note that, since \( \omega_d \tau = \delta \frac{\pi}{2} \), a non-optimum detection efficiency \( \delta \) reduces the phase rotation to less than the required 90°. Knowing that the variance of each pair out of the normal variables \( v, w, x, y \), is equal to \( 2N_m = 2mN \), the noncentrality parameters
are given by

\[
\lambda_A = \frac{A^2 \cos^2 \omega_d \tau + A^2 (1 + \Xi \sin \omega_d \tau)^2}{2 \cdot 2mN} = \frac{A^2 \alpha}{2mN}
\]

\[
\lambda_B = \frac{A^2 \cos^2 \omega_d \tau - A^2 (1 + \Xi \sin \omega_d \tau)^2}{2 \cdot 2mN} = \frac{A^2 \beta}{2mN}
\]

(7-100)

(7-101)

where \(\alpha\) and \(\beta\) are defined as in (7-56) and (7-57). The first pdf \(p_1(z)\) is now given by

\[
p_1(z) = p(z_A) * p(-z_B)
\]

\[
= p_A \left( z \left| 2, \frac{A^2 \alpha}{2mN} \right. \right) * p_B \left( -z \left| 2, \frac{A^2 \beta}{2mN} \right. \right)
\]

(7-102)

The second half of \(p(z)\), the pdf \(p_2(z)\), can be calculated in exactly the same way. This time taking the delayed in-phase signal \(d_i(t)\) as a reference, the following signal components are derived:

\[
d_I : v := \frac{A}{\sqrt{2}} + v
\]

(7-103)

\[
w := w
\]

(7-104)

\[
s_Q : x := -\Xi(t) \frac{A}{\sqrt{2}} \sin \omega_d \tau + x
\]

(7-105)

\[
y := -\Xi(t) \frac{A}{\sqrt{2}} \cos \omega_d \tau + y
\]

(7-106)

which yields the pdf

\[
p_2(z) = p(z_C) * p(-z_D)
\]

\[
= p_C \left( z \left| 2, \frac{A^2 \beta}{2mN} \right. \right) * p_D \left( -z \left| 2, \frac{A^2 \alpha}{2mN} \right. \right)
\]

(7-107)

Finally, \(z\) can be obtained by subtraction of \(z_1\) and \(z_2\), which gives for the pdf's:

\[
p(z) = p(z_1 - z_2) = p_1(z) * p_2(-z)
\]

\[
= \left[ p_A(z) * p_B(-z) \right] * \left[ p_C(-z) * p_D(z) \right]
\]

(7-108)

Identical pdf's may be combined after summation of the degrees of freedom and noncentrality parameters, yielding

\[
p(z) = p \left( z \left| 4, \frac{A^2 \alpha}{mN} \right. \right) * p \left( -z \left| 4, \frac{A^2 \beta}{mN} \right. \right)
\]

(7-109)
As usual, I&D post-detection filtering is equivalent to $m$-fold convolution of $p(z)$, giving $m$ times higher degrees of freedom and noncentrality parameters.

$$p(u) = p\left(u \left| 4m, \frac{A^2\alpha}{N}\right.\right) \ast p\left(-u \left| 4m, \frac{A^2\beta}{N}\right.\right)$$  \hspace{1cm} (7-110)

This result is identical to that obtained for polarisation diversity CPFSK reception. Expressions for the BER and figures related to sensitivity penalties can therefore be found in section 7.3. Apparently the type of diversity, either polarisation or phase, does not determine the receiver performance. The number of IF branches on the other hand is directly related to the order of the receiver.

### 7.4.3 Effects of SNR imbalance

In the limit of optimum performance, i.e. equal IF SNR’s and 90° phase difference, a CPFSK polarisation diversity receiver and a CPFSK phase diversity receiver yield the same sensitivity. The situation is different for non-optimum settings. Due to the fact that, for frequency discrimination, the I- and Q-branches are cross-multiplied, a reduction of the SNR in one branch will reduce the efficiency of both multipliers. Again, introducing the amplitude imbalance factor $\kappa$, the components of $z_1$ can be written as

\[
s_I : x := \frac{A}{\sqrt{2}} + x
\]

\[
y := y
\]

\[
d_Q : v := \Xi(t)\frac{A}{\sqrt{2}} \sin \omega_d \tau + v
\]

\[
w := -\Xi(t)\frac{A}{\sqrt{2}} \cos \omega_d \tau + w
\]\n
(7-111) \hspace{1cm} (7-112) \hspace{1cm} (7-113) \hspace{1cm} (7-114)

It is easy to see that the noncentrality parameters of e.g. $z_A$ and $z_B$ are now given by

\[
\lambda_A = \frac{A^2}{2} \left(1 + \kappa^2 + 2\kappa\Xi(t) \sin \delta \frac{\pi}{2}\right)
\]

\[
\lambda_B = \frac{A^2}{2} \left(1 + \kappa^2 - 2\kappa\Xi(t) \sin \delta \frac{\pi}{2}\right)
\]\n
(7-115) \hspace{1cm} (7-116)

where $\omega_d \tau$ has been substituted by $\delta \frac{\pi}{2}$. Defining the $\kappa$-dependent detection improvement factors $\alpha_\kappa$ and $\beta_\kappa$ as

\[
\alpha_\kappa = \frac{1 + \kappa^2}{2} + \kappa\Xi(t) \sin \delta \frac{\pi}{2}
\]

\[
\beta_\kappa = \frac{1 + \kappa^2}{2} - \kappa\Xi(t) \sin \delta \frac{\pi}{2}
\]\n
(7-117) \hspace{1cm} (7-118)
the final expression for the pdf after post-detection filtering is a straightforward modification of (7-110):

\[ p(u) = p\left( u \mid 4m, \frac{A^2\alpha_k}{N} \right) \ast p\left( -u \mid 4m, \frac{A^2\beta_k}{N} \right) \]  

(7-119)

Figure 7-13: The combined sensitivity penalty \( \Delta SNR \) in dB of a CPFSK phase diversity receiver, due to the SNR imbalance \( \kappa \), the detection efficiency \( \delta \) and the IF bandwidth expansion factor \( m \). The sensitivity of 20.03 at \( m = \delta = \kappa = 1 \) is taken as reference value \( \Delta SNR = 0 \).

First of all we note that the sum of \( \alpha_k \) and \( \beta_k \) is equal to \( 1 + \kappa^2 \), compared to 2 for optimum SNR balance. For decreasing \( \kappa \) both \( \alpha_k \) and \( \beta_k \) reduce, becoming equal to \( \frac{1}{2} \) in the limit of \( \kappa = 0 \). So the influence of \( \kappa \) and \( \delta \) is a reduction of the skewness of the pdf’s, yielding larger tail areas and error probabilities. There is however one difference in the effects due to \( \delta \) - which only reduces the SxS detector output - and \( \kappa \). The output SNR of one phase diversity mixer in the case of non-optimum \( \kappa \) can be written as

\[ SNR_o = \frac{\kappa^2 SxS}{\kappa^2 SxN + SxN + N x N} \]  

(7-120)

So in contrast to the effect of \( \delta \) not only the SxS-component, but also one of the SxN-components is reduced. For example, a \( \kappa^2 \) of 0.5 or -3 dB will not yield a \( \Delta SNR \) of
7.4. CPFSK PHASE DIVERSITY

3 dB but slightly less, since the noise is also reduced. This is verified by Figure 7-13. For high imbalance the effect of the reduced SxN noise becomes negligible, so the sensitivity penalty becomes linearly dependent upon \( \kappa \).

7.4.4 Compensation of IF phase mismatch

In particular when 3x3 couplers are used as 90°-hybrids, the phase of the IF branches may deviate from ideal I&Q performance. Assuming that the I and Q phasors have equal length, and taking the I-branch as a reference, the Q-phasor may (theoretically) rotate in two different ways. Both possibilities are illustrated in Figure 7-14. The difference with SNR imbalance as treated in the previous section is, that the phase mismatches can be compensated. From Figure 7-12 we know that ideally the Space Q-phasor leads by 90° with respect to the I-phasor, and the Mark Q-phasor lags by 90°.

In case of optimum detection efficiency \( \delta = 1 \) these phasors are rotated over \( \omega_d \tau = \frac{\pi}{2} \) and become aligned with the I-axis.

![Diagram](image)

Figure 7-14: The I and Q phasors at the photodiode outputs in case of phase mismatch. The type of mismatch at the left requires a shift of the IF spectrum, the type at the right requires a smaller frequency modulation index.

The first type of mismatch, on the left side of Figure 7-14, needs two different phase rotations due to the delay in the discriminator branches. The phase rotation for a Space, \( \omega_d, S \tau \), must be equal to \( \omega_d \tau - \Psi \), while the Mark phase rotation \( \omega_d, M \tau \) must be \( \omega_d \tau + \Psi \). Since the effective phase rotation of the IF carrier is relative to the central frequency, the complete IF spectrum must be shifted. The (negative) Space IF must become lower, while the (positive) Mark IF must be higher. The complete spectrum must thus be shifted to the right on a real frequency scale, by an amount...
2Ψ/π·f_d. Obviously, for negative Ψ the spectrum must be shifted to the left. In the extreme situation of Ψ = π/2 the Space IF has shifted to 0 Hz and the Mark frequency to 2f_d.

The second type of phase mismatch, illustrated on the right side of Figure 7-14, still requires equal phase rotations for the Mark and Space phasors. In the case illustrated here, f_d (and thus ω_d) must be reduced by the amount 2Ψ/π·f_d. This is equivalent to a reduction of the modulation index M by the factor 2Ψ/π. Note that for Ψ = π/2 the frequency deviation has reduced to 0.

In principle it is thus possible, in the case of CPFSK phase diversity reception, to compensate for phase mismatch between the two IF branches. It should be stressed, however, that this is only valid for small mismatch values Ψ. For larger deviations one spectral lobe shifts to dc (causing fading due to the linewidth) or the modulation index becomes zero. In those cases other detection mechanisms will cause increasing penalties. In practical systems this can nevertheless be an important tool for performance optimisation.

### 7.4.5 Linewidth effects

CPFSK phase diversity has been shown to be very bandwidth efficient, while yielding the same sensitivity as polarisation diversity heterodyne CPFSK reception. There is a second advantage of phase diversity, not mentioned up to now, which results in a considerable advantage over heterodyne reception. This is the higher linewidth tolerance relative to heterodyne CPFSK, since it is possible with phase diversity reception to use higher modulation indexes and/or IF bandwidths. The relations derived for heterodyne CPFSK in section 7.1.2 remain valid, but due to a higher M and m the upper curves in Figure 7-5 can be used. This may easily give a four times higher linewidth tolerance.

### 7.5 DPSK

#### 7.5.1 Receiver model

The block diagram of a heterodyne DPSK receiver is identical to that of a heterodyne CPFSK receiver (Figure 7-1), with the exception of the value of τ. Compared to CPFSK modulation, DPSK is easier to analyse due to the reduced number of parameters, since in principle the modulation index has disappeared. For DPSK demodulation the delay time should be set to T, which gives a phase detector with a frequency characteristic as illustrated in Figure 7-15. Note that the periodicity of this characteristic is much higher compared to that of a CPFSK frequency discriminator. The influence of the linewidth on DPSK demodulation has been analysed in section 7.1.2, with results given by (7-19) and Figure 7-4.
7.5.2 IF-detection

Copying (7-41) without the second order terms that will be filtered out by the post-detection filter gives

\[ z(t) = \frac{1}{2} \{x(t)v(t) + y(t)w(t)\} \cos \omega_c \tau + \]
\[ + \frac{1}{2} \{x(t)w(t) - y(t)v(t)\} \sin \omega_c \tau \]  

(7-121)

where \(x\) and \(y\) are the in-phase and quadrature noise components of the direct signal, and \(v\) and \(w\) of the delayed signal. For DPSK the IF carrier is constant, so for maximum signal output the IF should be located in a maximum or minimum of the phase detector characteristic. This gives

\[ \omega_c \tau = \omega_c T = n \cdot \pi \]  

(7-122)

---

Figure 7-15: Block diagram of a heterodyne DPSK receiver (top) and the associated frequency characteristic of the phase detector (bottom).
Figure 7.16: Phasor diagram of ideal DPSK delay line detection, with \(d(t)\) and \(s(t)\) the two detector inputs. The delayed signal \(d(t)\) is taken as the reference.

\[
f_c = \frac{n}{2T} \quad (7-123)
\]

The detector output can now be written as

\[
z(t) = \frac{1}{2} \{x(t)v(t) + y(t)w(t)\} \cdot (-1)^n \quad (7-124)
\]

Notice that the in-phase components of \(s(t)\) and \(d(t)\) as well as the quadrature components are mutually multiplied, whereas for CPFSK cross-multiplication of in-phase and quadrature components is used. Taking \(n=4\) and omitting all references to time we obtain a similar expression as derived for CPFSK:

\[
z = \frac{1}{8} \{(x + v)^2 + (y + w)^2 - (x - v)^2 - (y - w)^2\} \quad (7-125)
\]

\[
= \frac{1}{8} \{z_1 - z_2\} \quad (7-126)
\]

with

\[
z_1 = (x + v)^2 + (y + w)^2 \quad (7-127)
\]

\[
z_2 = (x - v)^2 + (y - w)^2 \quad (7-128)
\]

Taking \(d(t)\) as a reference, the signal components can be written as

\[
d(t) : \quad v := A + v \quad (7-129)
\]

\[
\quad w := w \quad (7-130)
\]

\[
s(t) : \quad x := \Xi(t)A + x \quad (7-131)
\]

\[
\quad y := y \quad (7-132)
\]
7.5. **DPSK**

With the variances of \( x \pm v \) and \( y \pm w \) equal to \( 2mN \), the pdf's \( z_1 \) and \( z_2 \) can be derived in the same way as done for CPFSK.

\[
p_1(z) = p \left( \frac{A^2}{2mN} (1 + \Xi(t))^2 \right) \\
p_2(z) = p \left( \frac{A^2}{2mN} (1 - \Xi(t))^2 \right)
\]

(7-133) (7-134)

After post-detection filtering and cross-correlation of \( p_1(z) \) and \( p_2(z) \) the final pdf in \( u \) is obtained:

\[
p(u) = p_1 (u \left| 2m, \frac{A^2}{2N} (1 + \Xi(t))^2 \right) * \\
+ p_2 (-u \left| 2m, \frac{A^2}{2N} (1 - \Xi(t))^2 \right)
\]

(7-135)

For a transmitted **Mark** this reduces to

\[
p(u) = p_1 (u \left| 2m, \frac{2A^2}{N} \right) * p_2 (-u \left| 2m \right)
\]

(7-136)

which is exactly the same result as for ideal CPFSK detection (7-62). So in the limit of optimum detection heterodyne CPFSK and DPSK have indeed the same performance. The results for DPSK can therefore be found in sections 7.2.2 and 7.2.3 with \( \delta \) set to 1.

**Polarisation and phase diversity**

Additionally, for polarisation and/or phase diversity the pdf's, and thus the BER, can be found with the by now standard method of multiplying the degrees of freedom by the amount of additional branches. DPSK then obtains the same diversity sensitivity penalties as CPFSK for \( \delta = 1 \). For \( m=1 \) the polarisation diversity BER can be found with (7-87) and/or (7-72), and is given by

\[
P_e = \frac{1}{2} \left( 1 + \frac{\rho}{4} \right) \exp(-\rho)
\]

(7-137)

This the same result as derived by Okoshi [86] and Cheng [9]. For a sensitivity of \( 10^{-9} \) the diversity penalty \( \Delta Div \) is equal to 0.39 dB as for CPFSK. This value was first found by Glance [28].

7.5.3 **Non-optimum DPSK (de-)modulation**

It has been mentioned in the introduction that DPSK, compared to CPFSK, has the 'advantage' of fewer independent parameters. This can, on the other hand, also be
Figure 7-17: Phasor diagrams of the IF signals in case of a modulation depth phase error of $2\psi$ in the DPSK modulator.
Left : without central IF correction.
Right : with central IF correction for obtaining minimum BER.

seen as a disadvantage since it requires a much more accurate setting of the remaining parameters. In a practical DPSK system these are the modulation depth, the delay time $\tau$ and the IF centre frequency $f_c$.

Modulation depth phase error

The modulation depth of the DPSK phase modulator indicates the actual phase shift in case of a transmitted Space, which must be exactly$^{14}$ 180°. With a delay of $T$ seconds and the IF located at $n$ times $1/2T$, a modulation depth not equal to $\pi$ radians does not affect detection of Mark bits. Space bits on the other hand will not introduce the required $\pi$ radians phase shift, resulting in reduced detection and an increased error rate. It has already been shown in Chapter 4, section 4.1.4, that minimum BER is in principle attained when the Mark and Space partial error rates are nearly identical. In the case of non-optimum modulation depth this can be achieved by changing the average IF, displacing it from the extreme of the frequency detection curve. When the modulation depth phase error is equal to $2\psi$ (see Figure 7-17) this can be partially corrected by setting the IF $\Delta \omega = \psi/\tau$ below the nominal value.

Using the right part of Figure 7-17 yields the following phasor components of the

---

$^{14}$Remember that DPSK, which is a differential coding scheme, introduces a 180° phase shift in case of a transmitted Space, and no phase shift in case of a Mark (see section 2.3.3 and Figure 2-5).
delayed and direct IF signals:
\[
\begin{align*}
  d(t) : & \quad v := A + \nu \\
  & \quad w := w \\
  s(t) : & \quad x := \Xi(t)A \cos \psi + x \\
  & \quad y := A \sin \psi + y
\end{align*}
\]
(7-138) \hspace{1cm} (7-139) \hspace{1cm} (7-140) \hspace{1cm} (7-141)

Since these phasor components are nearly identical to the ones for CPFSK on page 175, the pdf of \( u \) can straightforwardly be derived as
\[
\begin{align*}
p(u) &= p \left( u \left| 2m, \frac{A^2 (1 + \Xi(t) \cos \psi)}{N} \right. \right) \\
& \quad \times p \left( -u \left| 2m, \frac{A^2 (1 - \Xi(t) \cos \psi)}{N} \right. \right)
\end{align*}
\]
(7-142)

This result is identical to the one for CPFSK (7-59) when we substitute
\[
\cos \psi = \sin \delta \frac{\pi}{2}
\]
(7-143)

Table 7-3: Sensitivity penalties \( \Delta SNR \) in dB for non-ideal DPSK heterodyne detection, as a function of the modulation depth phase error \( \psi \), the central IF deviation \( \Delta f_c \) or the delay time deviation \( \Delta \tau \). The corresponding value of the CPFSK detection efficiency \( \delta \), used for computing the sensitivity with (7-69), is also given.

| \( \frac{|2\psi|}{\pi} \) | \( \Delta f_c T \) | \( \Delta \tau f_c \) | \( \delta_{CPFSK} \) | \( \Delta SNR \) [dB] |
|----------------|---------------|---------------|----------------|-----------------|
|                |               |               |                | \( m \)         |
|                |               |               |                | 1   | 2   | 3   | 4   |
| 0              | 0             | 0             | 1              | 0.39 | 0.66 | 0.88 |
| 0.1            | 0.025         | 0.025         | 0.9            | 0.43 | 0.70 | 0.91 | 1.10 |
| 0.2            | 0.050         | 0.050         | 0.8            | 1.23 | 1.39 | 1.54 | 1.67 |
| 0.3            | 0.075         | 0.075         | 0.7            | 2.14 | 2.25 | 2.35 | 2.45 |
| 0.4            | 0.100         | 0.100         | 0.6            | 3.42 | 3.49 | 3.55 | 3.62 |
| 0.5            | 0.125         | 0.125         | 0.5            | 4.89 | 4.93 | 4.97 | 5.02 |

Obviously, \( \psi \) and \( \delta \) are related in the sense that they are both indicators of the ideality of the modulation and demodulation/IF-detection processes. Due to their respective definitions, \( \psi \) must be zero, while \( \delta \) must be 1. Also, for CPFSK the phasor rotation must be equal to \( \frac{\pi}{2} \), after which the cross components of the direct and
delayed signal are multiplied; \( z = xw - yv \) (7-43). For DPSK the phasor rotation must be equal to \( \pi \), resulting in multiplication of in-phase components: \( z = xv + yw \) (7-124). So by using the value

\[
\delta = 1 - \frac{|2\psi|}{\pi}
\]  

(7-144)

in the CPFSK BER-expression (7-69) the BER of the equivalent DPSK receiver can be computed. Note that in (7-144) \( 2\psi \) is the modulation depth phase error, so \( 2\psi/\pi \) is the relative modulation depth phase error.

Non-optimum delay-line detection

From Figure 7-16 it is obvious that the phase rotation due to the delay line should be \( n \cdot \pi \) for a Mark and \( n \cdot \pi - \pi \) for a Space. Since the phase rotation is given by \( f_{IF} \tau \), variations of either \( f_{IF} \) or \( \tau \) will introduce an effective phase rotation deviation. The total phase rotation is given by

\[
\theta = (\omega + \Delta \omega_c)(\tau + \Delta \tau)
\]  

(7-145)

With \( \omega \) equal to \( n\pi/T \) and \( \tau \) equal to \( T \), the optimum phase rotation is indeed \( n \cdot \pi \). The phase rotation deviation \( \Delta \theta \) is then given by (omitting the second order effect
due to the product of $\Delta \omega_c$ and $\Delta \tau$):

$$
\Delta \theta \approx \Delta \omega_c \tau + \Delta \tau \omega_c
$$

$$
= 2\pi \left( \Delta f_c T + \frac{n \Delta \tau}{2T} \right)
$$

The phasor components can be derived from Figure 7-19:

Figure 7-19: Phasor diagram of the direct and delayed IF signals of a DPSK delay line receiver, in case of central IF or delay time deviations.

$$
d(t) : v := A + v
$$

$$
w := w
$$

$$
s(t) : x := \Xi(t) A \cos \Delta \theta + x
$$

$$
y := -\Xi(t) A \sin \Delta \theta + y
$$

This yields in the usual way the pdf of the post-detection filter output

$$
p(u) = p \left( u \left| 2m, \frac{A^2(1 + \Xi(t) \cos \Delta \theta)}{N} \right. \right) \ast
$$

$$
\ast p \left( -u \left| 2m, \frac{A^2(1 - \Xi(t) \cos \Delta \theta)}{N} \right. \right)
$$

which is exactly the same result as (7-142) with $\psi$ replaced by $\Delta \theta$. Accordingly the modulation depth phase error and the non-optimum delay-line detection have the same effect. The results can thus be found in Figure 7-18 using

$$
\frac{2\psi}{\pi} = 4 \left( \Delta f_c T + \frac{n \Delta \tau}{2T} \right)
$$
This shows for example that for \( m = 2 \) (a common practical value) and an additional sensitivity penalty of 1 dB relative to the optimum situation \( \tau = T \) and e.g. \( f_c = 2/T \), the centre frequency or the delay time may deviate only 2.5% and 1.25% respectively. At 1 Gbit/s and a central IF of 2 GHz this is equivalent to 50 MHz and 12.5 ps. These exigent requirements make DPSK particularly difficult for practical implementation.

### 7.6 Orthogonal signalling

It has been shown in this chapter that CPFSK and DPSK have a theoretical performance that is 3 dB better compared to systems with non-coherent IF-detection. CPFSK is thus twice as sensitive as FSK. The question is why, since CPFSK is obviously a form of FSK.

As a first approach it can be stated that delay-line differentially coherent CPFSK IF-detection uses the phase information contained by the modulation. An argument in favour of this theory is the fact that the delay-line discriminator has a smooth curve, which detects the signal not only at its extreme positions but also in between, during transitions. The additional information from the phase should then double the amount of detected energy, increasing the sensitivity by a factor of two. However, from the receiver analysis presented in this chapter there is no strong evidence for this theory. It is true that a delay line discriminator compares the phases of two IF signals at a certain moment, but this does not support the increased sensitivity mechanism. Additionally, in a situation where the IF carriers are static (long strings of identical bits) the effect should then be absent. This is, however, not so.

A better explanation is found by looking at DPSK (and in principle CPFSK with \( M = 0.5 \) as well) and the way it is detected. For DPSK the delay of the delay-line discriminator is set equal to \( T \), so that the phase of a bit is compared with the phase of the previous bit. This means that every bit is used twice; once as the bit of which the phase is compared with the reference and once as the reference itself. In DPSK, and MSK, we should therefore not use single bits as the basic information and power carrying elements, but pairs of bits instead. Since a pair of duration \( 2T \) contains twice as much power and information as a single bit, it is easily understood that this yields a 3 dB higher sensitivity.

We now also understand the orthogonality of the signals involved in DPSK and CPFSK detection, since they should be orthogonal as it was in the case of simple FSK detection. It has been stated on page 120 that orthogonal (D)PSK signalling requires 90° out-of-phase signals. At first sight one then wonders how the DPSK signals with 0 or 180° IF phases can be orthogonal. However, considering pairs of bits it is evident that the combinations 0°-0° (1,1) and 180°-180° (-1,-1) (which belong to a Mark) are orthogonal to the pairs 0°-180° (1,-1) and 180°-0° (-1,1) that belong to a Space, see also Stein [106, ch.13] and Schwartz, Bennett and Stein [98, sec.7-6]. For CPFSK with modulation index \( M \) equal to 0.5 (MSK) the delay is also \( T \) seconds, so the same
orthogonality and sensitivity improvement apply [14].

The last question is how DPSK and CPFSK are related, or in other words: why do DPSK and CPFSK have the same theoretical performance? To answer this question it is easiest to consider DPSK as a limiting form of CPFSK. It has been assumed throughout the analysis that CPFSK features step-like frequency variations at the bit transition moments. This means that during the bit-time $T$ the phase of the IF CPFSK signal linearly integrates by an amount $\pm \omega_d T = \pm \pi M$, depending upon the bit sent. At the next decision moment only the accumulated phase change is of importance, not the way in which it is obtained. This means for example that smoother modulating frequency steps may be used, thereby reducing the width of the spectrum. Or, and this is the connection between CPFSK and DPSK, the phase change may be accumulated in a very short time. The limiting case is a Dirac delta impulse with energy content $\pm \frac{\pi}{2}$, introducing step-wise phase changes at every bit transition. This is DPSK. Indeed experiments have been reported where DFB lasers were directly DPSK modulated using impulses [12, 101]. However, on the receiver side the DPSK demodulation is independent of the way in which the signal is generated, so these systems are not treated separately. A practical drawback of this method is the fact that at high bitrates - say more than 1 Gbit/s - obtaining pulses much shorter than the bittime will be very difficult.

### 7.7 Summary

It has been shown that CPFSK and DPSK - in the limit of optimum IF-detection - have a 3 dB better performance than FSK dual-filter reception with non-coherent IF-detection. This improvement arises from the fact that delay-line differentially coherent IF-detection requires 2-bit long observation periods, doubling the amount of detected energy. From a different point of view one may also attribute the gain to the fact that not only the frequency, but also the phase-information contained by the IF signal is used by the detection.

Since CPFSK and DPSK both use one single IF bandpass filter to receive a compact IF spectrum, additional parameters become important for the IF-detection. For CPFSK these are the frequency modulation index $M$ and the detection efficiency $\delta$. In particular the latter should be as close to 1 as possible, since non-optimum detection rapidly introduces high sensitivity penalties. These can simply be explained by considering the fact that the delay-line phase rotation only affects 'coherent' - i.e. not noise-like - signals. The advantage of CPFSK is the possibility to optimise the detection efficiency in practical systems through small adjustments of the modulation index. In contrast, DPSK has fewer operational degrees of freedom, making it harder to operate. Since the required amount of phase-shift is fixed, only the delay-time and the central IF remain as parameters, which should be accurately set. Deviations of $\tau$ and $f_c$ from their ideal values introduce the same sort of sensitivity penalties as caused
by the non-optimum detection efficiency of CPFSK.

The influence of the laser linewidth on CPFSK (and DPSK) is also investigated, showing some remarkable results. For this analysis it is, however, essential to take the post-detection filtering into account. In contrast to what is generally said about these types of systems, it is shown that in principle the linewidth requirement is of the same order as for ASK and FSK with non-coherent IF-detection. In particular in the case of CPFSK with delay-line discrimination one can relax the requirements considerably by increasing the modulation index. Both CPFSK and DPSK become more linewidth-tolerant by using wider IF bandpass filters. Only for CPFSK systems with very small frequency deviations (and DPSK) in combination with relatively narrow IF filtering, the linewidth requirements become much more stringent than for non-coherent IF-detection.

Finally, both CPFSK and DPSK can be used in combination with polarisation and phase diversity. According to the procedure developed in the previous chapter, this invariably means that the degrees of freedom of the final pdf and BER-expressions double, leading to small sensitivity penalties. One system concept that has been studied in more detail is CPFSK phase diversity (quasi-homodyne) reception, which is the nearly optimal combination of CPFSK modulation, low-bandwidth homodyne optical detection and high linewidth tolerance. The theoretical performance is completely identical to that of polarisation diversity heterodyne CPFSK detection. It is one of the two types of practical systems - the other being FSK heterodyne dual-filter reception - to be investigated experimentally in the next chapter.
Chapter 8

Experimental verification

Introduction

For the systems considered in Chapter 3–7 all penalties related to the IF-detection method and realisation can be computed. In combination with the optical penalties, such as excess loss of components, the shot noise/receiver noise penalty and the known electrical penalties, a complete analysis of practical receivers can now be performed. It is obvious that a precise knowledge of all system penalties is extremely important, both in the design and evaluation phase of system engineering. Properly designed systems should have a reasonable balance between the different types of identifiable penalties, while the amount of unknown penalty, the famous and convenient “non-ideal electronics”, should certainly not account for more than 10% of the total penalty. For example, low LO output power and/or a noisy receiver can lead to a receiver noise penalty of many dB’s, which would in that case dominate all other penalties. It is then rather futile to put a lot of effort in optimising the IF-detection in order to gain a few tenth of a dB. On the other hand, with recently available high (LO) optical output powers the aforementioned receiver noise penalty can become very small, resulting in dominant IF-detection penalties. This makes the reduction of these IF-detection penalties recommendable.

A major problem associated with IF-detection penalties is the fact that they are hard to measure. Assuming ideal threshold settings, the sensitivity penalty $\Delta SNR$ is only a function of the IF bandwidth expansion factor $m$. However, it is clear that changing $m$ in a practical system is extremely difficult, requiring major changes in the multi-GHz IF-electronics. Even so, changing only one parameter in such high-frequency electronics is nearly impossible (e.g. a change of $m$ will probably affect the noise- and gain-characteristics as well), making it hard to compare the different configurations. This means that in most cases only a very limited amount of measured
data is available for evaluation of the IF-detection penalties. Mostly the measurements are related to single branch vs. diversity performance of receivers, although we and others have performed experiments with changing \( m \) by increasing the bitrate. In most cases the proper IF performance must be deduced from a complete system analysis with exact knowledge of all penalties.

In the following sections two different systems will be analysed in detail. Both are practical systems, realised at the ‘Wideband Communication Systems’ group of the Philips Research Laboratories, Eindhoven, The Netherlands. The first is a 140 Mbit/s-280 Mbaud polarisation diversity FSK heterodyne system with dual-filter quadratic IF-detection. By using the same IF electronics for 560 Mbit/s reception we could reduce \( m \) by a factor of four in multi-bitrate measurements. The other system is a second generation 1 Gbit/s polarisation control CPFSK phase diversity receiver with differentially coherent delay-and-cross-multiply IF-detection. For both systems a complete sensitivity analysis including all penalties will be presented. The FSK system is used for a detailed evaluation of the diversity penalty. Finally the few measurement results available in literature are compared to the theoretical results, although this is often much more difficult for want of a complete set of system parameters.

8.1 FSK heterodyne system description

The elementary block diagram of the Coherent Multi-Channel (CMC) FSK receiver is illustrated in Figure 8-1, showing those blocks that are essential for the analysis\(^1\). A CMC-system is used for distribution and reception of a large comb of coherent optical channels in optical frequency division multiplex (OFDM). The signals are transmitted by coherent optical transmitters, in this case directly FSK-modulated DFB single-frequency lasers. Details on these transmitters can be found in [120].

The heart of the receiver is the tunable local oscillator (LO) laser, in this case a Philips three-section Distributed Bragg Reflector (DBR) [71, 72, 77, 78, 79]. One section, the Gain-section, is biased above threshold for proper lasing of the device, while the sub-threshold currents through the phase- and Bragg-sections can be used for frequency-tuning of the laser. Since the optical channels are relatively far apart (10 GHz, which is required for avoiding inter-channel interference in combination with the relatively high IF-value\(^2\)) an optical tuning range of several hundreds of GHz is essential. Practical tuning ranges of the DBR's used are between 500 and 800 GHz. Due to the electrical control, tuning can be very fast, only limited by the settling times of the control loops [48, 78]. The linewidth of a DBR laser is not constant under tuning,

---
\(^1\)Additional circuitry not shown here are e.g. the IF bandwidth switch and its controller [78], the IF and baseband peak detectors for the AGC, buffers and bandlimiting filters in the baseband section, the phase detector, oscillator and all other components of the PLL, and all details of the interfacing to the local intelligence hardware and software.

\(^2\)Results of channel spacing measurements with the CMC-system can be found in [44].
8.1. FSK HETERODYNE SYSTEM DESCRIPTION

but is guaranteed to be below 30 MHz. Average laser linewidth-values are 15-20 MHz, which is identical to the typical linewidth of the single-section Distributed Feedback Lasers (DFB) in the CMC-transmitters [120]. The maximum specified IF linewidth in the CMC-receivers is 60 MHz, yielding the relatively large $\Delta \nu_{IF}$ value of 0.43. With (5-16) the required bandwidth expansion $m$ of each (I) FSK filter is found to be 4.6, being equivalent to about 650 MHz. In practice the real frequency deviation is set to the higher value of 1200 MHz, yielding a modulation index of 8.6. However, this means that linewidth-effects are absolutely unobservable in the receiver, thus supporting the procedure for eliminating the linewidth from the analysis as presented in Chapter 3.

It should be stressed that, as part of a European RACE-project, five identical CMC-receivers had to be built to specification\(^3\). The principle objectives of our receivers were high sensitivity and fast channel selection [40, 48, 78, 79, 115], which had to be realised in a reproducible way. To this end the technology used in the receivers

\(^3\)This is also the reason for most of the non-optimum values of some of the parameters, such as e.g. the too large frequency deviation. All values are compromises between the specifications of the different participants in the project and often had to be defined before system experiments could confirm the goodness of the assumption.
was frozen in 1989, except for the introduction of a new generation of front ends in 1991. Compared to the present state-of-the-art some values may therefore be slightly outdated, but this is largely compensated by the fact that world-wide the receivers are amongst the best-engineered [44]. The different receivers will be referred to as PHIL-01 to -05 where required, the characteristics of the important receiver sections will now be described briefly.

8.1.1 Balanced front ends

The balanced front-ends used in the receivers have an electrical bandwidth of 2.3 GHz [37, 49], which is limited by the choice of the upper resonance of the receiver noise tuning characteristic. The input stage of the electrical circuit is of the transimpedance type, with in the first generation front ends a GaAs MESFET and in the second generation a GaAs HEMT as active element. For the photo detection, PIN diodes with an active area of 80 µm, a total capacitance of 1.0 pF and a (packaged) responsivity of 0.8 A/W are used. Due to the large total input capacitance - the combined effect of photodiode, MESFET and parasitics - the intrinsic –3 dB bandwidth of a MESFET input-stage is only 700 MHz, which can be improved using serial tuning between the photodiode and the MESFET. However, this results in far from flat noise characteristics, with a times four higher noise level around 1200 MHz (see Figure 8-2). Between 1.2 and 2.5 GHz the noise level reduces considerably due to the serial input resonance at 2 GHz, obtaining an average equivalent noise level over the full bandwidth of 8 pA/√Hz per photodiode. The transfer characteristic of the front ends is flat within 1 dB from 0 to 2.1 GHz, and the –3 dB bandwidth is 2.3 GHz. The measured common-mode rejection ratio (CMRR) of a balanced front end is better than 20 dB up to at least 1 GHz. Effects of the LO-laser relative intensity noise (RIN) will, in combination with the low LO-power on the photodiodes, consequently be suppressed. See also Figure 2-3.

The 2.5 times higher transconductance of the HEMT's can be used to lower the input impedance of the input stage, thus shifting the input pole to a higher frequency (1200 MHz). The reduced impedance also lowers the quality-factor of the input series resonance, giving a broader and less distinct noise-reduction. The resulting equivalent input noise, illustrated in Figure 8-2, is therefore much flatter, yielding an average noise-level of 5 pA/√Hz. It can be expected that this noise reduction by a factor 2.5 (5²/8²=0.4) gives a considerable sensitivity improvement in the receivers.

A HEMT, for High Electron Mobility Transistor, is an improved MESFET (MEltal Semiconducto Field Effect Transistor) with a much thinner gate structure (typically 0.25 µm). This yields a considerably higher transconductance gm at about the same capacitance levels, which in turn yields a much higher transition frequency ft. For comparison: the best MESFET's used in the first generation front ends have an ft of 23 GHz and a gm of 45 mS, where the HEMT's attain values of 50 GHz and 120 mS.
8.1. FSK HETERODYNE SYSTEM DESCRIPTION

![Graph showing equivalent input noise characteristics of typical MESFET and HEMT front ends of the FSK receivers.]

Figure 8-2: Equivalent input noise characteristics of typical MESFET and HEMT front ends of the FSK receivers.

8.1.2 IF AGC-amplifiers

The IF-section in the receivers has two functions: bandwidth control and gain control. The gain is being controlled by the Automatic Gain Control (AGC) loop, maintaining the post-detection baseband signal at a constant level [115]. Under normal operation this means that the frequency discriminator inputs are also kept at a level that guarantees proper quadratic operation. However, bandwidth control is an equally important function, both for noise reduction and crosstalk minimisation. It is obvious that - given the bandwidth required for distortionless reception - the post-IF-detection noise should be minimised by reducing the IF bandwidth as much as possible. In practical systems an additional constraint is the finite IF-to-baseband crosstalk suppression of the IF-detector. Particularly in this case where the crosstalk is rather high, a strict separation of the IF and baseband signals in the frequency domain is required. This crosstalk is a consequence of the high conversion loss due to the perfectly quadratic IF-detection.

The resulting overall IF bandpass characteristic is illustrated in Figure 8-3 and has -3 dB frequencies of 400 and 2300 MHz. For optimum reception the central IF must be located in the center of the IF passband at 1350 MHz. Using the 1200 MHz frequency deviation, the Signal frequencies are then at 750 MHz (Space) and 1950 MHz (Mark), leaving about 400 MHz between these frequencies and the band edges. This is enough for suppression of excess Nyquist-filtering in the IF, so the FSK signals are
Figure 8-3: The measured IF bandpass characteristic of one of the FSK receivers. The method used for the measurement is heterodyning of two lasers, so the photodiode, the balanced front end and the IF amplifier are included in this overall characteristic. The horizontal scale is 300 MHz/div, the vertical scale 3 dB/div.

essentially passed undistorted by the IF bandpass filter\(^5\). Due to the 400 MHz highpass characteristic of the lower edge of the IF band, the IF and baseband signals are clearly separated.

### 8.1.3 Frequency discriminator

The frequency discriminator is the most critical circuit in these FSK receivers, since it determines much of the overall performance. For proper polarisation-, pattern- and temperature-independent\(^6\) reception the detection must be perfectly quadratic, with a 2.5 GHz bandwidth and reasonable IF-to-baseband crosstalk suppression. The quadratic detection is achieved with low-capacitance Schottky-diodes that are biased by a small current (0.1 mA) at their knee-voltage. In combination with reasonably

\(^5\)This will not be the case for the 560 Mbit/s FSK signals in the experiments to be described in section 8.2.4.

\(^6\)The temperature behaviour of the receiver is outside the scope of this thesis. However, temperature-independent operation of the discriminator is very critical for maintaining perfectly quadratic IF-detection. Due to the strong temperature dependence of the Schottky-diode biasing, special measures had to be taken for the suppression of any drift.
small input signals (about 70 mVpp maximal) this assures that all even harmonics other than the 'wanted' 2\textsuperscript{nd}-harmonic are suppressed. The odd harmonics are effectively suppressed by balancing the complete discriminator circuit, see Figure 8-4 [109, 110].

The dynamic range of the quadratic detection is limited on the low side by the noise floor of the baseband amplifier following the discriminator and on the high side by the fourth harmonic. The noise floor is determined by the combined noise from the amplifier and the discriminator [114]. Since the load of the discriminator on the amplifier input is highly capacitive (namely the four Schottky diodes in parallel) and a relatively high amplifier bandwidth of several hundred MHz is required, the same transimpedance design of the optical front end is used, but with a lower transimpedance to compensate for the much larger capacitance. The noise of the amplifier and the shot noise of the four diodes together yield a noise floor at the output of the baseband amplifier of -83 dBm. When a 20 dB baseband SNR is required a minimum input power level of -30 dBm is necessary in the dominant IF-branch. On the high side the
Figure 8-5: Measured levels at the output of a frequency discriminator of the different harmonics of a 150 MHz input signal. The ratio of the slopes of the first and \( n^{th} \) harmonics is equal to \( n \). The noise floor of the baseband amplifier following the discriminator is \(-83\) dBm.

The fourth harmonic must be at least 35 dB below the second harmonic for proper quadratic detection [118]. For higher input powers there is too much power-transfer from the second to the fourth harmonic, reducing the quadratic behaviour. In practice the discriminator input must therefore remain below \(-18\) dBm, yielding a 12 dB dynamic range. This is considerably better than obtained with other methods, e.g. in [20].

\(^7\)In [20] the dynamic range is only 4.8 dB, with serious deviations from quadratic behaviour both for lower and higher input powers. However, with our discriminator the quadratic dynamic range is infinite, since it stretches out from \(-18\) dBm to infinitely small input powers. The dynamic range is limited by the noise performance, not by the quadraticity. This means that the signal in the IF-branch with the lowest SNR is always perfectly quadratically detected. The dynamic range limitations thus only apply for the maximum input power in the high-SNR IF-branch.
8.1. FSK HETERODYNE SYSTEM DESCRIPTION

Figure 8-6: Measured effective frequency discrimination curves for different input powers.

It is important to consider the way in which the IF-detection is performed by this discriminator. Despite the fact that it is constructed as one single unit, functionally it is a dual-filter detector. The shorted stubs in Figure 8-4 act as highpass filters, whereas the open stubs act like lowpass filters, with respective theoretical transfer characteristics [109]

\[ H_{LPF} \propto \cos \left( \frac{f \pi}{f_r} \right) \]  \hspace{1cm} (8-1)
\[ H_{HPF} \propto \sin \left( \frac{f \pi}{f_r} \right) \]  \hspace{1cm} (8-2)

The resonance frequency \( f_r \) is determined by the length of the stubs and in theory set to 2.7 GHz, yielding a cross-over point of the two filter characteristics at exactly 1.35 GHz (-3 dB frequency of both filters). The effective discriminator curve, as given in Figure 8-6, is obtained after subtraction of the lowpass from the highpass filter characteristic. Including the quadratic IF-detection this is given by [109, 114]:

\[ i_{out} \propto i_{HPF}^2 - i_{LPF}^2 = -\cos \left( \frac{f \pi}{f_r} \right) \]  \hspace{1cm} (8-3)

In practice the theoretical value of 2.7 GHz had to be increased in order to compensate for the influence of parasitics at higher frequencies. Although the cross-over frequency
is maintained at 1.35 GHz, this is no longer at the -3 dB points of the LPF and HPF. The real -3 dB frequencies are estimated to be 100 MHz above and below the cross-over point at 1450 and 1250 MHz, respectively. In combination with the outer edges of the IF-band at 400 and 2300 MHz, this yields effective filter bandwidths of 1050 MHz and a bandwidth expansion factor $m$ of 1050/140=7.5. In contrast to the ideal situation assumed in the analysis of dual-filter IF-detection in Chapter 5, the two filters overlap in the middle. The potential effect of this on the IF-detection will be neglected when comparing the theoretical and practical system performance.

### 8.1.4 Baseband combining

As a result of all the effort put into the design of the frequency discriminator, the baseband combiner can be very simple. Due to the perfectly quadratic IF-detection no post-processing is required, and the two baseband signals can be summed directly. At high frequencies current summation is the best solution, so the output currents of two common-emitter buffer stages are fed through the same load resistor.

### 8.1.5 Biphase line-coding

One of the drawbacks of direct DFB laser modulation is the fact that these lasers feature a dip in the frequency modulation characteristic (see e.g. [39]). This dip is caused by the combined effects of thermal and injection-induced fm, which have opposite signs. At low frequencies the fm is temperature-induced, in principle due to the heating and cooling of the entire chip as a function of the changing currents. At high frequencies the fm is caused by the carrier density effects with changing currents, which result in variations of the cavity refractive index and thus the wavelength.

Due to their opposite signs, the two fm phenomena cancel each other at a certain frequency, resulting in a very low fm sensitivity; the dip. For standard DFB laser diodes this dip is somewhere in the frequency range 100 kHz-20 MHz, depending upon the actual laser parameters. However, variations in the absolute fm level become apparent one decade higher or lower. This means that any baseband signal having components around the fm dip will not be properly modulated. Passive equalisation of the fm characteristic is very difficult; due to the phase reversal a rather complicated equaliser would be required, which should furthermore be optimised for every individual laser. The only practical solution is therefore to modify the modulating signal in such a way that no signal components are near the dip. This can be obtained by linecoding.

Many different linecoding schemes can be used for reduction of the amount of low-frequency spectral content. However, most of these schemes have the drawback that they increase the bitrate that is actually present at the input of the decision circuit, thus increasing the noise bandwidth (e.g. 5B6B-coding\(^8\)), or that they use multi-level

---

\(^{8}\) 5B6B-coding, a form of mBnB-coding, means that from the original data groups of 5 bits
coding (AMI, bipolar, etc.). A higher bitrate means that the sensitivity decreases, while multi-level coding requires extremely linear transfer characteristics and (in the case of FSK) a much higher IF-bandwidth. Biphase is a special form of linecoding in the sense that the actual baudrate is indeed increased, although not necessarily at the decision switch. This linecode has the advantage that coding is very simple, using a logical exclusive or (EXOR) function on data and clock (see Figure 8-7). A logical 0 is thus converted in 1-0, a logical 1 into 0-1 (1B2B-coding). The biphase signal is therefore dc-free, and has at least a transition in the center of every biphase byte. The 'classical' way of decoding biphase - first regeneration at the biphase baudrate, thereafter digital biphase-to-baseband decoding - yields a 3 dB lower sensitivity since the Nyquist post-detection filter and the decision circuit operate at twice the baseband bitrate. However, biphase coding can also be regarded as (analogue) multiplication of the data and clock signals. This means that the data signal is subcarrier multiplexed on a carrier equal to the clock frequency. It is therefore possible to demodulate the received biphase signal in the same way, i.e. analogue subcarrier demultiplexing by multiplying with the regenerated clock frequency [36, 38, 39, 113]. The resulting baseband signal can then be Nyquist filtered at the original bitrate, yielding maximum

are transformed into groups of 6 bits. When the 5 bits are identical the 6th bit is inverted, thus limiting the amount of low-frequency power.
sensitivity. The measured penalty due to the use of biphase linecoding was found to be less than 0.3 dB for patterns up to $2^{23}-1$ PRBS (Pseudo-Random Binary Sequence), and this form of linecoding is therefore copied by many others in practical coherent optical systems [32, 33, 83].

For the system analysis only one aspect of the biphase linecoding is of importance. Since the analogue demodulation is a form of synchronous detection, the information contained in the biphase signal around the carrier frequency is preserved and only converted to its original baseband form. Effectively the complete transmit-receive chain thus operates as if the bitrate were 140 Mbit/s and not the 280 Mbaud. This is extremely important in the analysis, since it means that the effective IF bandwidth expansion and the linewidth influence can be computed using 140 Mbit/s. In fact the system can be analysed as if no biphase line-coding was used.

8.1.6 Post-detection filtering and regeneration

Like in most practical high-bitrate optical transmission systems a RAised COSine (RACOS) post-detection filter is used. This group of filters has transfer functions of the raised cosine form, yielding impulse responses with zeros at every multiple of
8.1. FSK HETERODYNE SYSTEM DESCRIPTION

$T/2$ and the sampling moments $kT$. Theoretically this avoids inter-symbol interference (ISI). The transfer function and impulse response of a RACOS-1 pulse are given by [3]:

\[
H(f) = \frac{1}{2} [1 + \cos(\pi f T)] \quad (0 \leq f \leq 1/T) \tag{8-4}
\]

\[
h(t) = \frac{\sin \pi t/T}{2\pi t} \cdot \frac{\cos \pi t/T}{1 - (2t/T)^2} \tag{8-5}
\]

The first part of the impulse response (8-5) gives the conventional sinc-function with zeros at every sample moment $kT$ except $k=0$. The second part generates additional zeros between sample moments, except at $t = \pm T/2$. When the filter has rectangular input pulses, and must have RACOS-1 pulses at its output, the frequency transfer characteristic becomes:

\[
H'(f) = \frac{\pi f}{2} \left[ \frac{1}{\sin \pi f T} + \frac{1}{\tan \pi f T} \right] \quad (0 \leq f \leq 1/T) \tag{8-6}
\]

This characteristic can be approximated by using a conventional LC lowpass filter in combination with a parallel resonant blocking filter, introducing a forced zero at $f=1/T$ in the filter transfer characteristic [3, 112]. Figure 8-8 shows the measured post-detection filter characteristic, with a bandwidth of 87.5 MHz. The part of the characteristic extending beyond the zero at 140 MHz introduces a small sensitivity penalty due to the additional noise bandwidth.

Since the output signal of the analogue biphase demodulator contains the low-frequency components associated with PCM-coded video signals\footnote{The lowest frequency component of a 140 Mbit/s 2$^{23}$−1 PRBS signal is 17 Hz. The lowest frequency component of a single TV-frame is 25 Hz. However, depending upon the picture actually transmitted, the lowest component can be nearly dc.} the Nyquist filter must be dc-coupled. It therefore contains separate dc- and hf-branches, which must be carefully adjusted in order to give a perfectly flat response from dc to several MHz. A control loop within the filter cancels the effects of drift in the dc-coupled stages.

The dc-coupled post-detection filter output is connected to one of the two inputs of a balanced decision circuit bipolar IC. The other input is connected to the dc decision threshold that can be accurately adjusted for minimum overall BER. The decision circuit itself is a master-slave D-type edge-triggered flip-flop with a 650 MHz bandwidth, triggered by the regenerated 140 MHz clock of which the phase, and thus the decision moment, can be accurately set.
Figure 8-9: Front view of a 140 Mbit/s FSK coherent multi-channel receiver. The receiver is fully remote-controlled.

Figure 8-10: Rear view of the same receiver with opened back panel. The lowest level contains the optical polarisation diversity unit, the top level all IF and baseband electronics.
Figure 8-11: Top view of the 140 Mbit/s FSK CMC receiver with the cover removed. The panel on the right contains the two IF branches in parallel, with the four fibres coming in from the board below. Above the IF section is the baseband board with the Nyquist filter and PLL. The left part of the box contains, from top to bottom, the microprocessor rack, the LO-board including temperature and current stabilisation, and power supplies.
Figure 8.12: Measurement set-up for the receiver sensitivity of the FSK receivers. The PC is for wavelength control of the transmitter. In the absence of the video-codec that would normally provide the channel identification (CID) information to the receiver, the CID-word is hard-wired to the receiver.
8.2. **FSK HETERODYNE SYSTEM MEASUREMENTS**

8.1.7 Construction of the receivers

The construction of the receivers is illustrated by the three photographs of Figures 8-9–8-11. The receivers are fully remote-controlled using a standard TV remote control unit. A display provides information about the status of the receiver, such as the channel to which it is locked or the kind of system error that has been detected. The only hardware interfaces to the receiver are the optical input, the 140 Mbit/s data and clock outputs, and a parallel bus with the channel identification (CID) word coming from the video decoder. The receiver is built in a modular way for ease of maintenance.

8.2 **FSK heterodyne system measurements**

8.2.1 140 Mbit/s sensitivity measurements

The set-up for the sensitivity measurements of the CMC FSK-receivers is illustrated in Figure 8-12. Data signals are obtained from a high-bitrate data generator, and can be varied from 1-0-1-0 to PRBS $2^N-1$, with $N$ from 7 to 23. Unless mentioned otherwise, $N=23$ was used for the measurements. The digital baseband data are converted to a biphase-coded signal in the transmitter [120] and used for direct FSK-modulation of the DFB laser. After calibration of the receiver input signal with an accurate optical power meter, the incoming signal could be attenuated in steps as small as 0.01 dB with the optical attenuator. After reception and regeneration by the receiver the data and clock signal were fed to an error detector.

There is a clear division in the measured receiver sensitivities, with one group of receivers around $-50$ dBm and another around $-52.5$ dBm. The first group (PHIL-01 and -02) is based on the first generation receivers with MESFET front-ends, the second group has the improved HEMT front-ends. PHIL-04 is an exception to this rule since it is an experimental transceiver instead of a normal receiver\(^ {10}\). This causes a higher loss of both the signal and LO optical circuits, as well as a higher polarisation dependence due to the complicated optics. PHIL-05 has a higher polarisation dependence and pattern dependence due to a slightly non-optimal polarisation diversity unit, with different fiber lengths of the two branches. In the following, PHIL-01 and -03 are therefore considered as typical examples of the first and second generation receivers.

\(^ {10}\) Receiver PHIL-04 has a HEMT front-end and nevertheless a sensitivity around $-50$ dBm, because this receiver is a so-called 'coherent dialogue transceiver' [111]. It is capable of both receiving and transmitting data, the latter by direct modulation of the LO-laser. This requires optics between the LO and the polarisation diversity unit for coupling part of the LO-power upstream for transmission, introducing additional sensitivity penalties.
Table 8-1: Measured sensitivity figures of the five CMC FSK-receivers as a function of polarisation. The optimum polarisation is defined as parallel (∥). The sensitivity variation with the pattern length between 1-0-1-0 and PRBS N=23, and the LO-power after the optical isolator and on the four photodiodes are also given. Not all combinations and characteristics could be measured before delivery of receivers.

<table>
<thead>
<tr>
<th>type</th>
<th>Receiver PHIL-</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>01</td>
<td>02</td>
</tr>
<tr>
<td>pol. ‖</td>
<td>MESFET</td>
<td>MESFET</td>
</tr>
<tr>
<td></td>
<td>-49.2</td>
<td>-50.6</td>
</tr>
<tr>
<td>pol. ⊥</td>
<td>-50.5</td>
<td>0.2</td>
</tr>
<tr>
<td>pol. dep.</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>pattern dep.</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>( P_{LO} )</td>
<td>ex isolator</td>
<td>-1.3</td>
</tr>
<tr>
<td></td>
<td>on 4 diodes</td>
<td>-2.8</td>
</tr>
</tbody>
</table>

8.2.2 Analysis of sensitivity penalties

Obtaining a good value for the measured receiver sensitivity is nice, but for proper system analysis it is equally important to know whether this measured value is indeed the best possible result. At this point the important question should thus be asked:

In how far do the measured sensitivities agree with the values predicted by our theoretical analysis?

In order to analyse the match between theory and practice all the known sensitivity penalties of each receiver have to be listed (table 8-2).

Before this, the shot-noise limit is calculated with (1-22), which yields a value of -61.5 dBm for a wavelength \( \lambda = 1.560 \mu m \), a SNR of 40.06, a bitrate \( 1/T \) of 139.264 Mbit/s and ideal quantum efficiency (100%).

For the responsivity \( R \) of the photodiodes the worst-case value of 0.8 A/W is used. The excess loss of the polarisation diversity units including the optical connector (which has an intrinsic loss of about 1 dB) are measured for each unit prior to installation in the receiver. The total LO-power on all four photodiodes (see Table 8.2.1) must first be transformed into a dc photocurrent per diode:

\[
I_{photo} = \frac{1}{4R} 10^{P_{LO, total}/10} \quad (8-7)
\]
Using the shot noise formula (2-1), or the alternative form \( \sqrt{\langle i^2 \rangle} = \sqrt{2qI_{\text{photo}}} \) in pA/\( \sqrt{\text{Hz}} \), one can then compute the equivalent shot noise due to the LO-powers on the photodiodes. This yields 5.8 and 9.6 pA/\( \sqrt{\text{Hz}} \) for PHIL-01 and -03 respectively. For calculating the receiver noise we use the individual receiver noise characteristics similar to Figure 8-2, which show that, given the central IF of 1350 MHz and the peak-to-peak frequency deviation of 1200 MHz, the Mark and Space IF-signals experience different noise levels. An average receiver noise level can be obtained from the pA/\( \sqrt{\text{Hz}} \) values.
of the front-ends at frequencies of 750 and 1950 MHz. These two values should then be squared, summed and averaged\(^\text{11}\). For PHIL-01 we obtain\(^\text{12}\) through \((8.7^2 + 6.8^2)/2\) a value of \(6.1 \cdot 10^{-24} \text{ A}^2/\text{Hz}\), and for PHIL-03 in the same way \(25 \cdot 10^{-24} \text{ A}^2/\text{Hz}\). This gives values of 7.8 and 5.1 pA/√Hz, respectively. The receiver noise penalty can then finally be computed with (1-23), yielding the penalties of 4.5 and 1.1 dB as given in the table. Note the impressive sensitivity improvement that can be obtained with HEMT input stages.

Table 8-2: The different penalties contributing to a reduced sensitivity of the 140 Mbit/s FSK receivers. Values are given for a first generation receiver with MESFET front end (PHIL-01) and a second generation HEMT receiver (PHIL-03). The error margins of the different penalties and sensitivities are indicated as well. Total error margin is calculated as the root of the sum of squared (r.s.s.) individual penalties.

<table>
<thead>
<tr>
<th></th>
<th>PHIL-01</th>
<th>PHIL-03</th>
<th>margin</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>shot noise limit</td>
<td>-61.5</td>
<td>-61.5</td>
<td>0.1</td>
<td>dBm</td>
</tr>
<tr>
<td>excess loss diversity</td>
<td>-2.3</td>
<td>-2.1</td>
<td></td>
<td>dB</td>
</tr>
<tr>
<td>responsivity (0.8 A/W)</td>
<td>-2.0</td>
<td>-2.0</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>receiver noise</td>
<td>-4.5</td>
<td>-1.1</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>RIN (140 dB/Hz)</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.1</td>
<td>dB</td>
</tr>
<tr>
<td>IF bandwidth (m=7.5)</td>
<td>-1.4</td>
<td>-1.4</td>
<td>0.05</td>
<td>dB</td>
</tr>
<tr>
<td>pol. diversity comb.</td>
<td>-0.7</td>
<td>-0.7</td>
<td>0.05</td>
<td>dB</td>
</tr>
<tr>
<td>IF linewidth (50 MHz)</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>dB</td>
</tr>
<tr>
<td>biphase coding</td>
<td>-0.3</td>
<td>-0.3</td>
<td>0.05</td>
<td>dB</td>
</tr>
<tr>
<td>decision circuit</td>
<td>-0.4</td>
<td>-0.4</td>
<td>0.1</td>
<td>dB</td>
</tr>
<tr>
<td>total known penalty</td>
<td>11.7</td>
<td>8.1</td>
<td>0.34</td>
<td>dB</td>
</tr>
<tr>
<td>theoretical sensitivity</td>
<td>-49.8</td>
<td>-53.4</td>
<td>0.34</td>
<td>dBm</td>
</tr>
<tr>
<td>remaining penalty</td>
<td>0.6</td>
<td>0.4</td>
<td></td>
<td>dB</td>
</tr>
<tr>
<td>measured sensitivity</td>
<td>-49.2</td>
<td>-53.0</td>
<td>0.1</td>
<td>dBm</td>
</tr>
</tbody>
</table>

The RIN-induced penalty is more difficult to compute, since the exact value of the RIN is not known. Based on practical experience, values between -150 and -140 dB/Hz are generally accepted as being realistic for the type of DBR used in the CMC system. When the receiver noise is first calculated with (1-23), the additional

\(^{11}\)In theory these values should be calculated through the weighted noise integrals over the Mark and Space frequency regions. In practice we use simply the values at the two center frequencies.

\(^{12}\)Note that Figure 8-2 does not belong to PHIL-01. The values of 8.7 and 6.8 pA/√Hz are the real measured data.
8.2. FSK HETERODYNE SYSTEM MEASUREMENTS

RIN penalty can be obtained through a modification of (2-14)

$$\Delta SNR = \frac{2(n_{to} + n_{rec})}{2(n_{to} + n_{rec}) + \frac{n_{RIN}}{CMRR}}$$  (8-8)

where $P_{LO,d}$ is the optical power on one photodiode. With values for $P_{LO,d}$ of 0.131 and 0.233 mW for PHIL-01 and -03, respectively, and a common-mode rejection ratio $CMRR$ of 20 dB we obtain very low RIN-induced sensitivity penalties. For a RIN of -150, -145 and -140 dB/Hz receiver PHIL-01 yields values of 0.003, 0.008 and 0.025 dB, respectively, while PHIL-03 gives 0.001, 0.003 and 0.01 dB. Conservatively, the RIN-induced penalty in Table 8-2 is therefore set to $0.1 \pm 0.1$ dB. Deviations may be due to a less than optimal CMRR and imbalances in the 2x2 couplers and/or the photodiode responsivities.

The IF penalties can be computed with the results from the theoretical analysis in Chapter 5. Using the bandwidth expansion factor $m$ of 7.5 as derived previously, the penalty of the excess IF bandwidth is easily found using (5-41) to be the mentioned value of $-1.4$ dB. The 0.7 dB penalty associated with the polarisation diversity equaliser combining of the second IF-branch is obtained in an equally simple way from (6-30). It is clear that these two penalties together make up 20-25% of the total known penalties, so that accurate knowledge of them is absolutely essential for practical system analysis.

It is easily verified that the linewidth-induced penalty is indeed very small. The IF-linewidth-to-bitrate ratio was given as 0.43. Since it was found in (5-17) that FSK has a 1.86 times higher linewidth tolerance than ASK, the BER-floor computation should be done with an effective ASK linewidth-to-bitrate ratio of 0.23. With (3-99) we find $Z = 143.3$ and with (3-100) the BER-floor at $1.7 \times 10^{-52}$. This means that the sensitivity penalty due to the BER-floor will indeed be negligible.

The 0.3 dB penalty due to the analogue biphase coding has been explained in the section on the linecoding. Finally, the 0.4 dB attributed to the decision circuit is, in fact, the combined result of improper Nyquist post-detection filtering, excess baseband noise bandwidth and noise of the decision circuit itself. Detailed analysis by Tomesen in [113] yields 0.3-0.4 dB; the worst-case value is assumed here.

The summed penalties of the two receiver types are 11.7 and 8.1 dB; the difference is caused by the improved front end performance. This leaves respectively 0.6 and 0.4 dB of unexplained penalties, which are acceptably low values. These penalties are probably caused by imperfections of the IF characteristics, generally known as the 'non-ideal electronics'. For example, looking at PHIL-01, the noise levels for the $Space$ (at 750 MHz) and the $Mark$ (at 1950 MHz) are the already mentioned 8.7 and 6.8 pA/$\sqrt{Hz}$. At the same time the signal at 1950 MHz obtains a 1.5 dB higher gain due to the input resonance (similar to Figure 8-3). The $Mark$-signal therefore has a SNR that is 1.6 (from the noise) times 1.4 (from the gain) or 2.3 times higher (3.6 dB) than the SNR of the $Space$. When the frequency discriminator curve is
not completely symmetrical around the zero at 1350 MHz, this figure can vary even more. Effectively the receiver therefore does not perform perfectly symmetrical dual-filter detection, but tends towards single-filter detection. Accordingly the effective IF-detection method falls somewhere half-way between single- and dual-filter detection. With this phenomenon a penalty of 0.4-0.6 dB can easily be explained. It shows the importance of having flat IF characteristics, all the way from photodiode and front-end to frequency discriminator.

In conclusion it can be asserted that there is an excellent agreement between the theoretical and practical performance of the two receivers.

8.2.3 Polarisation dependence

Since nearly perfect quadratic IF-detection is used in the coherent receivers, the polarisation dependence of the completed units should be very low. It has been shown in the section introducing the frequency discriminator, that only higher-order harmonics can cause polarisation-dependent penalties. For practical reasons - a higher dynamic range and discriminator output voltage - the discriminator input power level for a 50:50 polarisation split factor has been set to -13 dBm, which is 5 dB higher than the limit of -18 dBm derived previously. The 2nd-4th harmonic ratio thus decreases to 24 dB, instead of the required 35 dB, reducing to 20 dB with all signal in one branch. A small polarisation-dependent penalty due to higher-order harmonics can therefore be expected.

Figure 8-14 shows the measured results of receiver PHIL-02. First, the polarisation-dependence of the receiver is extremely low, only 0.15-0.20 dB, which means that the IF-detection must indeed be nearly perfectly quadratic. The sensitivities with all signal in one of the IF-branches ($\phi=0$ or 1) are about 0.1 dB lower compared to those of the other SOP's. This shows that there is indeed some harmonic distortion, but at an acceptably low level. Under normal operating conditions, however, the probability that the linear polarisations are attained is small, giving an average sensitivity that is close to the peak sensitivity at $\phi=0.5$. These results show clearly the high performance that can be obtained with quadratic IF-detection in combination with direct baseband equal-gain combining, giving a good practical solution for polarisation independent coherent optical reception. The practical polarisation-dependent penalty with this method is also much lower than that of systems using e.g. ratio combining [53, 70].

Figure 8-14 also gives the sensitivities of single-branch reception. These were measured with the complete polarisation diversity optics and electronics inserted for normal operation, but with one of the two frequency discriminator outputs disconnected. With all signal in the operational branch, this procedure is identical to eliminating the NxN noise of the second branch, while keeping all SNR's identical. The measured sensitivity improvements when disconnecting the second branch is in both cases 0.75 dB, in excellent agreement with the theoretical value of 0.7 dB from (6-29) as given in Table 8-2. The difference of 0.05 dB is lower than the measurement accuracy of the
8.2. FSK HETERODYNE SYSTEM MEASUREMENTS

Figure 8.14: The measured receiver sensitivity of receiver PHIL-02 for different input polarisations. The sensitivities of single-branch reception are also given, i.e. with one discriminator output disconnected.

sensitivity, estimated to be about 0.1 dB. Finally it can be shown that the absolute sensitivities of both IF-branches are within 0.1 dB of each other, an indication of the excellent reproducibility of the receiver construction.

8.2.4 560 Mbit/s sensitivity measurements

One advantage of the relatively high IF bandwidth expansion factor \( m = 7.5 \) in the 140 Mbit/s receivers is the fact that higher bitrates can readily be accommodated\(^\text{13} \). Without changing the IF electronics we have therefore modified receiver PHIL-02 for 560 Mbit/s reception. Only a new 560 Mbit/s Nyquist post-detection filter had to replace the original 140 Mbit/s filter. No form of linecoding was used for the experiments, so a certain pattern-dependence could be expected. The receiver was

\(^{13}\) Although it should be remembered that due to the biphase linecoding the IF baudrate is already 280 Mbaud, not 140 Mbit/s. This leaves less room for bitrate increase.
used for what we called coherent multi-bitrate multi-channel reception [5, 42, 45, 110]. This offered the possibility to receive signals of different bitrates with the same receiver, where only the baseband sections had to be duplicated for every individual bitrate.

Since we used the same IF sections for the 560 Mbit/s experiments, the IF bandwidth of each of the bandpass filters remained 1050 MHz. The bandwidth expansion factor thus was reduced to a value as low as 1.9. Through (5.41) it can therefore be expected that the sensitivity reduction due to the four times higher bitrate will not
be 6 dB, but lower. Using (5-41) for $m=7.5$ and 1.9 yields a difference in sensitivity penalty of 1.3 dB. Starting from the $-50.6$ dBm sensitivity at 140 Mbit/s, the predicted 560 Mbit/s receiver sensitivity is $-50.6 + 6.0 - 1.3$ or $-45.9$ dBm. The measured value of the sensitivity for a 1-0-1-0 pattern is $-45.5$ dBm, yielding a relative improvement due to the reduced $m$ of 0.9 instead of 1.3 dB. The 0.4 dB discrepancy is probably due to imperfect scaling of the Nyquist filter realisation.

![Graph](image_url)

**Figure 8-16**: The receiver sensitivity of receiver PHIL-02 for different input polarisations when used for 560 Mbit/s reception. Also the sensitivities of single-branch reception - i.e. with one discriminator output disconnected - are given, both with the IF set to the average and the optimum value.

The polarisation dependence of the 560 Mbit/s receiver was also measured. Results illustrated in Figure 8-16 show a dependence of 1.2 dB, considerably larger compared to the 140 Mbit/s operation. These figures are measured with the IF center frequency optimised for a 50:50 IF power split factor between the branches. This optimisation was required because both diversity branches had different optimum IF settings. Still using this average IF setting, the measured polarisation diversity penalties (when disconnecting one IF branch) are 0.4 and 0.5 dB, respectively. Due to the larger baseband
noise bandwidth there is an increased penalty due to the noise floor of the baseband amplifier immediately following the discriminator output of 0.08 dB. With (6-29) the theoretical diversity penalty for $m=1.9$ is equal to 0.49 dB. The differences between theory and experiment are therefore 0.17 and 0.07 dB for both branches.

It should be remembered that the experiments at 560 Mbit/s were not as well engineered as those at 140 Mbit/s. Since we used the same frequency deviation - the same laser diodes were used, so the IF linewidth remained 40-60 MHz - the spectrum was distorted at the IF band edges, and differently for both branches. This is also the reason for the different optimum IF settings of the two branches, resulting in a much higher polarisation dependence. It can therefore be concluded that 560 Mbit/s is too high a bitrate for a receiver with $m=1.9$, in accordance with the general rule that the filter bandwidth should at least be twice the bitrate, or even three times according to Kazovsky [64]. Nevertheless, the feasibility of multi-bit rate reception was clearly demonstrated by these experiments, which also presented the unique opportunity to verify the analytical results.

### 8.2.5 Analysis of results

As already mentioned, it is very difficult to separately measure the relatively small penalties that follow from the analysis. From the measurements presented so far the following verifications are possible:

- The polarisation diversity penalty $\Delta Div$ for the two different values of the IF bandwidth expansion factor $m=1.9$ and 7.5.

- The sensitivity penalty due to the IF bandwidth when $m$ increases from 1.9 to 7.5.

- The total sensitivity penalty of a polarisation diversity receiver with $m=7.5$, including the penalties related to the IF-detection.

The first measurement is accurate to the extent that $\Delta Div$ only is measured. This yields accuracies of 0.05 dB on a total penalty of 0.75 dB for $m=7.5$, and 0.10–0.15 dB on a total of 0.5 dB for $m=1.9$. Such accuracies are in accordance with the respective levels of engineering of the 140 and 560 Mbit/s experiments. Moreover, the value of 0.05 can be considered as the absolute minimum uncertainty in a sensitivity measurement, which is normally more in the order of 0.1 dB. The uncertainty in the determination of $m$, especially in the case of 560 Mbit/s, does not cause major errors, although one would expect differently. The main uncertainty is due to the bandpass filter edge frequencies located near the discriminator zero-crossing, which could not be measured directly. A 100 MHz error in the estimation of the BPF bandwidth gives a variation of 0.7 in $m$ at 140 Mbit/s, and 0.2 at 560 Mbit/s. However, due to the slow increase of $\Delta Div$ with $m$ this only gives errors in the theoretical value of $\Delta Div$ of 0.01 and 0.02 dB respectively. It can therefore be concluded that these measurements are an accurate confirmation of the theoretical results obtained in Chapter 5 and 6.
8.2. *FSK HETERODYNE SYSTEM MEASUREMENTS*

![Graph showing IF bandwidth expansion factor vs. penalty (dB) with various markers indicating different parameters and data points.](image)

**Figure 8-17:** The measured results of the 140 and 560 Mbit/s experiments using receiver PHIL-02, compared with the theoretical predictions. The sensitivity variations due to both the polarisation diversity and the bitrate increase are shown.

The two other types of measurements are less accurate for confirming the theory, since they do not measure isolated penalties. The second measurement for example, the sensitivity gain from 140 to 560 Mbit/s due to the reduced $m$, required different baseband circuitry, including the high-frequency equalisation of a transmitter laser diode. The influence of the exact Nyquist-filter frequency transfer characteristic on the measured penalty is therefore very large, making it impossible to tell which part of the remaining 0.4 dB penalty belongs to $\Delta Div$.

The third measurement, i.e. the complete receiver sensitivity and consequently the total sensitivity penalty relative to the shot noise limit, certainly does not provide any discrimination between penalties. However, the fact that the discrepancy between theory and practice is only 0.4-0.6 dB on a total amount of 8.5-12.4 dB gives confidence in the analysis. Furthermore there is a very plausible explanation for this remaining penalty, namely the Mark-Space IF imbalance. The total amount of IF-related penalties makes up 18 and 26% of the total known penalties for the two receivers analysed, which shows at least the importance of incorporating them in a total system analysis. Especially in the event of increased LO-powers\(^\text{14}\), increased photodiode responsivities

---

\(^{14}\)Typical values for LO-powers have increased from -3 dBm four years ago - i.e. the specification for the CMC LO lasers – to +3 dBm nowadays, with peak values of +10 dBm.
or lower polarisation diversity excess loss\(^{15}\) the total sensitivity penalty may reduce to 4.5-5 dB, in which case the IF-detection penalties will make up 40-50% of the total penalty.

8.3 CPFSK phase diversity measurements

8.3.1 Receiver configuration

A 1 Gbit/s CPFSK phase diversity receiver has been constructed according to the principles presented in section 2.4.3 Figure 2-9, and section 7.4 Figure 7-11. The main advantage of phase diversity reception, as explained in Chapter 2, is of course the quasi-homodyne reception, requiring electronics of much lower frequency. A second advantage is the higher linewidth-tolerance, which is related to the frequency deviation. Since the IF bandwidth requirements are less severe in case of phase diversity reception it is easier to use a frequency modulation index of about 1, requiring the lowest relative IF bandwidth according to Figure 7-2 [67]. Assuming optimum detection efficiency, a higher modulation index \(M\) gives a smaller delay time \(\tau\) with (7-6). Finally, a lower \(\tau\) in turn yields a higher linewidth-tolerance as in Figure 7-5. Another advantage is that due to the spectral folding of the IF-spectrum around 0 Hz, the noise and amplitude characteristics of Mark and Space are identical. This is much more difficult, if not impossible, in case of 5-10 GHz IF electronics for heterodyne reception. A practical disadvantage of CPFSK phase diversity systems is the need for polarisation control, while at the same time two IF-stages are used as in polarisation diversity.

For the 1 Gbit/s CPFSK system as many circuits as possible from the FSK heterodyne system described above have been used. This included the 2.3 GHz HEMT front ends and the AGC IF stages, giving an IF bandwidth expansion factor\(^{16}\) \(m\) of 4.6. The delay-and-cross-multiply frequency discriminator [13] has been built as a single silicon integrated circuit [31, 43], giving superior balance between the different branches compared to hybrid solutions. The discriminator has a zero-crossing at 0 Hz and its maxima at \(\pm 1.3\) GHz (see Figure 8-18), requiring a modulation index \(M\) of 2.6 at 1 Gbit/s for optimum detection efficiency with \(\delta=1\). The required delay of the discriminator branches is found from

\[
\tau = \frac{1}{4f_m} = \frac{T}{2M}
\]

yielding a delay time of 192 ps or \(T/5.2\). The post-detection filter is of the same design as used for 140 and 560 Mbit/s, but with all component values scaled for 1 Gbit/s. Finally, the decision circuit is a master-slave D-Flip-Flop silicon bipolar IC of

\(^{15}\)Both these values have also increased, yielding responsivities for 2.5 GHz photodiodes of 95% and polarisation diversity units with an excess loss of 0.5 dB.

\(^{16}\)Notice that the IF bandwidth expansion is determined from the positive AND negative IF bandwidth.
Figure 8-18: The measured overall frequency discriminator curve of the 1 Gbit/s CPFSK phase diversity receiver, including the 3x3 coupler, the front ends and the IF-stages. The central zero crossing is at 0 Hz, the horizontal scale is 600 MHz/div. The curve is obtained with the receiver AFC-loop locked and 'fast' (about 10 MHz) sinusoidal frequency modulation of the transmitter with a modulation index of 6000. This gives a ±3 GHz frequency sweep through the IF detection curve [16, 27].

the same technology as the frequency discriminator, with a maximum operating speed of 7.4 GHz.

Both the transmitter and the local oscillator are three-section DBR lasers at a wavelength of 1.54 μm. On the transmitter side the laser fm-characteristic had to be equalised, in order to obtain a flat response up to 1 GHz. The main problem with the LO-laser was the low optical output power, only −3 dBm after optical isolation.

90°-hybrid

The 90°-hybrid is a critical unit in a phase diversity receiver, since it has the dual task of signal-LO combining and 90° out-of-phase I&Q IF-generation. In principle there are two ways of obtaining I&Q IF-signals with an optical 90°-hybrid. The first one uses a 'standard' polarisation diversity unit as in Figure 2-10 with prescribed input polarisation states of both signal and LO-light. For example, a circularly polarised LO and a linearly polarised signal, under 45° relative to the main axis of the polarisation beam splitter (PBS), yield the required IF signals [27, 43]. However, this method is impractical due to the temperature-induced phase rotations in the optical path's
Figure 8.19: Eye pattern of the 1 Gbit/s CPFSK phase diversity receiver after Nyquist post-detection filtering, measured at a BER of $10^{-9}$. Horizontal scale is 200 ps/div, vertical scale is 50 mV/div.

between the beam splitters and the couplers. This can be overcome by using a phase-control loop as in [2], although in integrated optics only. Despite the fact that the 'polarisation diversity 90°-hybrid' theoretically gives completely balanced I&Q-signals it could not be used stably due to the slowly rotating IF phase difference, giving fading of the discriminator output.

An alternative is the use of fused fibre couplers as 90°-hybrids, be it either 3x3 [16, 47] or 4x4 [63]. Due to the poor performance of 4x4 couplers [108], 3x3 devices were selected for the experiments, although these still showed a large excess loss\(^{17}\).

The main advantage of fused fibre couplers is that the signal-LO combining and the 90° IF-generation take place at the same instant and location, avoiding any phase-delay problems as in the polarisation diversity unit.

Taking one of the three phasors as reference 1, the Q-phasor is obtained by subtracting the other two coupler outputs, see Figure 8-20. This means that the receiver configuration is not completely symmetrical, having one single and one dual front end. At the same time the SNR’s of the IF-branches are not identical either, introducing an imbalance $\kappa$. From Figure 8-20 it is clear that the Q-branch has a $\sqrt{3}$ times larger amplitude, or a 3 times larger signal power. However, using our double front end with

\(^{17}\)With presently available commercial 3x3 couplers the excess loss has gone down from 1.5 dB, as used in the experiments, to 0.2 dB.
Figure 8-20: The 3x3 fused fibre coupler used as 90°-hybrid (left) and the resulting phasors of the optical and electrical IF-signals (right).

electrical combining the receiver noise has doubled as well. This yields a 1.5 times larger Q-branch SNR compared to the I-branch, giving an imbalance factor

$$\kappa = 10^{SNR_I/20SNR_Q}$$

(8-10)

The factor 1.5 (1.8 dB) yields a $\kappa$ of 0.81, which determines through (7-119) and Figure 7-13 the sensitivity penalty as a function of $\delta$ and $m$. With $\delta=1$ and $m=4.6$ the resulting penalty due to the SNR-imbalance is equal to 1.1 dB.

Another penalty related to the 3x3 coupler is caused by the loss of signal in the balanced front end. When the two 120° out of phase phasors are subtracted for obtaining the Q-phasor, ideally their in-phase components should cancel and their quadrature components add (see figure 8-20). Since these in-phase components are not used in the detection process, but are at the same time equivalent to $1/6^{th}$ of the generated IF-power, this results in a 0.8 dB sensitivity penalty.

Finally, a drawback of the 3x3 optical coupler is that their is no proper RIN-cancellation for all inputs. The I-branch has no cancellation at all, since a single photodiode front end is used. When a balanced front end is used for the subtraction of the two other branches, the Q-branch gives RIN-suppression. At high optical LO output powers the presence of RIN may introduce an additional sensitivity penalty in the I-branch. One possible solution is the alternative triangular photodiode arrangement proposed by Langenhorst et al. [73], although this reduces the bandwidth when the same hardware is used. A better, and more general solution is the use of a single 4x4 coupler with the proper IF-phases at the outputs. These phases should be exactly
Figure 8-21: The measured BER-curves of the 1 Gbit/s CPFSK phase diversity receiver for different pattern lengths.

0-90-180-270°, but this requires a special fabrication process of these devices as shown by Mortimer and Thijssen [80, 108]. So far we have not found 4x4 couplers that fulfil all specifications.

8.3.2 Sensitivity measurement and analysis

Figure 8-21 shows the measured BER-characteristics of the phase diversity receiver. For short pattern length the sensitivity of −43.4 dBm is 12.4 dB from the shot noise limit at −55.8 dBm. Longer patterns give additional penalties, due to the absence of linecoding
8.3. CPFSK PHASE DIVERSITY MEASUREMENTS

Table 8-3: The different penalties contributing to the reduced sensitivity of the 1 Gbit/s CPFSK phase diversity receiver. Penalties are given for two different values of the local oscillator Relative Intensity Noise (RIN). Approximate error margins are indicated for the individual penalties, as well as for the theoretical and measured sensitivity values. The total error margin is calculated as the root of the sum of squared individual penalties.

<table>
<thead>
<tr>
<th>RIN = ( \frac{CMRR}{\text{penalties}} )</th>
<th>(-145 \text{ dB/Hz})</th>
<th>(-140 \text{ dB/Hz})</th>
<th>margin</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>shot noise limit</td>
<td>-55.8</td>
<td>-55.8</td>
<td>0.1</td>
<td>dBm</td>
</tr>
<tr>
<td>excess loss 3x3 coupler</td>
<td>-1.5</td>
<td>-1.5</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>responsivity (0.8 A/W)</td>
<td>-2.0</td>
<td>-2.0</td>
<td>0.05</td>
<td>dB</td>
</tr>
<tr>
<td>I&amp;Q-generation</td>
<td>-0.8</td>
<td>-0.8</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>receiver noise</td>
<td>-2.6</td>
<td>-2.6</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>RIN</td>
<td>-0.1</td>
<td>-1.0</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>IF bandwidth ((m=4.6))</td>
<td>-1.6</td>
<td>-1.6</td>
<td>0.01</td>
<td>dB</td>
</tr>
<tr>
<td>SNR imbalance</td>
<td>-1.4</td>
<td>-2.6</td>
<td>0.2</td>
<td>dB</td>
</tr>
<tr>
<td>IF linewidth (50 MHz)</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>dB</td>
</tr>
<tr>
<td>decision circuit</td>
<td>-0.4</td>
<td>-0.4</td>
<td></td>
<td>dB</td>
</tr>
<tr>
<td>total known penalty</td>
<td>10.4</td>
<td>12.5</td>
<td>0.43</td>
<td>dB</td>
</tr>
<tr>
<td>theoretical sensitivity</td>
<td>-45.4</td>
<td>-43.3</td>
<td>0.43</td>
<td>dBm</td>
</tr>
<tr>
<td>remaining penalty</td>
<td>2.0</td>
<td>-0.1</td>
<td></td>
<td>dB</td>
</tr>
<tr>
<td>measured sensitivity</td>
<td>-43.4</td>
<td>-43.4</td>
<td>0.1</td>
<td>dBm</td>
</tr>
</tbody>
</table>

and the non-flat fm-characteristics of the DBR laser after external equalisation. All known penalties contributing to this total sensitivity penalty are listed in Table 8-3.

Some of these penalties, e.g. the photodiode responsivity and the decision circuit, are obtained directly from the 140 Mbit/s Table 8-2. Others, like the receiver noise, are easily calculated from the data available. These yield a 95 \( \mu \text{A} \) LO-generated photocurrent or 5.5 pA/\( \sqrt{\text{Hz}} \), which gives in combination with the 5.1 pA/\( \sqrt{\text{Hz}} \) of the receiver noise the penalty of 2.6 dB. The penalty due to the I&Q-generation has been explained already, while the penalty due the IF bandwidth expansion is found with (7-88) and Figure 7-10.

The penalty due to the IF linewidth of 50 MHz can be found using (7-30). With \( \tau = T/5.2 \) the variance \( \gamma' \) is equal to 0.06, which gives, in combination with \( m=4.6 \), a BER-floor of \( 5.4 \times 10^{-42} \). The sensitivity penalty is therefore set to zero\(^{18} \), since so

\(^{18}\) This result is different from the one presented in the ECOC-paper [47]. Since the moment of writing of that paper we solved the problem of the linewidth in CPFSK receivers, which showed
low a floor does not introduce a penalty. This result shows the high linewidth tolerance of CPFSK reception when larger frequency deviations and IF bandwidths are used.

The RIN-induced penalty can be considerable in a receiver using a 3x3 optical coupler. The l-branch has in this configuration no RIN-suppression, which may cause problems at high LO-powers. Furthermore, the experimental 3x3 couplers are far from symmetrical, as one would require. Measurements by Dumont [16] gave the following phasor amplitudes and phases, when the l-output is taken as the reference $1 \neq 0^\circ$: $0.89 \angle 131^\circ$ for output 2 and $0.83 \angle 250^\circ$ for output 3. Ideal subtraction of output 3 from 2 yields a 'quadrature' phasor of amplitude 1.50 and phase 101°. Note that the phase angle is not exactly 90°, although this can be compensated by reducing the modulation index $M$ a fraction $\frac{11}{90}$ or 12%. Signal imbalance is also less than the theoretical value of $\sqrt{3}$ or 1.73.

In contrast to the well-engineered 140 Mbit/s system, this phase diversity set-up was much more experimental, while using older lasers. There were (low-reflectivity) optical connectors between the 3x3 coupler and the photodiodes to test different coupler-receiver configurations. However, this did not improve the balance between the different IF branches, so a low CMRR can be expected. Since older lasers are used, the RIN is also assumed to be relatively high, and is taken to be -145 or -140 dB/Hz. The RIN-induced receiver penalty will now be evaluated for one 'good' receiver with RIN=-145 dB/Hz and CMRR=10 dB, and one 'poor' receiver with RIN=-140 dB/Hz and CMRR=5 dB. The RIN-induced equivalent input noise in the photodiode can be found with

$$\sqrt{\langle i^2 \rangle} = \sqrt{R^2 P_{lo,d}^2 L_c RIN} = RP_{lo,d} L_c \sqrt{RIN}$$

(8-11)

where $L_c$ is the optical coupler excess loss and $P_{lo,d}$ the optical power per photodiode. For -145 and -140 dB/Hz this gives 5.3 and 9.4 pA/$\sqrt{\text{Hz}}$, respectively. Using (8-8) and the values of $n_{lo}=5.5$ pA/$\sqrt{\text{Hz}}$ and $n_{lo}=5.1$ pA/$\sqrt{\text{Hz}}$ yields penalties of 0.1 dB for the 'good' and 1.0 dB for the 'poor' receiver. This is a penalty due to the additional RIN in both IF branches, and must thus be included in Table 8-3.

A second effect of the RIN, and thus the deteriorated SNR in the l-branch, is an increased SNR imbalance. The total imbalance is the result of on the one hand the I&Q-generation in the 3x3 coupler, and on the other hand the RIN-cancellation in the Q-branch. This can be expressed in the following way:

$$\kappa^2 = \frac{p_Q}{p_I}$$

$$= \frac{S_Q}{S_I} \cdot \frac{n_{lo} + n_{rec} + n_{RIN}}{2(n_{lo} + n_{rec}) + \frac{n_{RIN}}{CMRR}}$$

(8-12)

that our previous estimation was wrong. A new analysis of the system then revealed that the RIN caused at least part of this penalty, not the linewidth.
This yields for the 'good' receiver a $\kappa$ of 1.28, or an imbalance of 2.2 dB. From Figure 7-13 we can now derive for $m=4.6$ and $\kappa=2.2$ dB a total IF-detection penalty of 3.0 dB. This includes the 1.6 dB due to $m$ for $\kappa=0$, so the sensitivity penalty due to the imbalance is 1.4 dB. In the same way we find for the 'poor' receiver a $\kappa$ of 1.52 or 3.65 dB, and consequently a total IF-detection penalty of 4.2 dB. This gives a worst case imbalance-induced sensitivity penalty of 2.6 dB.

The remaining penalty of the CPFSK phase diversity system varies between 0.1 and 2.0 dB, depending upon the actual value of the RIN and CMRR. (For the FSK system the values were 0.4-0.6 dB). Unfortunately it is not possible to tell which of these two is the right value. Assuming that in practice the performance was more 'poor' than 'good', the actual difference between theory and practice is of the same order as for the FSK system. When the RIN and CMRR performance of the system is 'good', some other causes of additional sensitivity penalties must be found. In the case of phase diversity, imbalance between Mark and Space characteristics can not be a source of penalties, since it has been shown already that due to the spectral folding around dc the filtering and SNR's are perfectly symmetrical. A more plausible cause is IF-to-baseband crosstalk, since for phase diversity IF and baseband spectrally overlap. Although the frequency discriminator, a fully symmetrical and balanced IC, should suppress IF-to-baseband crosstalk considerably, any remaining crosstalk would immediately distort the detected baseband signal. This might easily introduce a few tenths of a dB penalty. Other causes may be improper Nyquist post-detection filtering, ISI due to asymmetrical IF passband filtering and an increased sensitivity to optical reflections.

8.4 Accuracy of system measurements

It should be clear from the system analyses presented here that accurate receiver performance evaluation - i.e. with an accuracy of 0.1 dB - is extremely difficult. The uncertainty in the measured sensitivity values is in many practical cases already larger than the required 0.1 dB, mainly due to temperature and/or SOP variations during the measurements. The resulting BER-curves are therefore often a regression line with an accuracy of typically 0.1-0.2 dB. Some of the individual penalties contributing to the total sensitivity penalty have error margins of at least 0.2 dB, especially the receiver noise and RIN penalty, as well as the linewidth penalty. Excess losses in the optical part of the receiver typically have 0.1 dB accuracies, due to the measurement methods and, very important, variations in (dirty) optical connectors under practical system analysis conditions.

The total margins due to the combined effect of all the individual penalties in Table 8-2 and 8-3 are not obtained by direct summation of all individual margins. This would yield an unrealistically large value, not in agreement with the accuracy of the analysis. Instead, the overall margin is obtained by taking the root of the sum of the
squared (r.s.s.) of individual penalties, or summation of the respective error variances. This is permissible since most penalties are uncorrelated. The resulting overall error margins are then 0.33 and 0.43 for the FSK and CPFSK system respectively, which means that the effective discrepancies between theory and measurement are 0.25-0.95 dB (PHIL-01) and 0.05-0.75 dB (PHIL-03). With the CPFSK system we have the problem that the exact value of the RIN and CMRR is not known, so it useless to give a range here. On the low side these discrepancies are almost identical to the measurement accuracy of 0.1-0.2 dB, on the high side the value is approximately 1 dB for most cases. For each type of receiver there is a good explanation for the remaining penalties, i.e. non-flat IF characteristics and IF-to-baseband crosstalk.

The breakdown of individual penalties presented here is the best that can be obtained by practical system analysis. In order to reduce the remaining error margin even further, an extremely detailed and time-consuming analysis of the system in question should be performed, including the spectral and time-domain effects of each individual system element. Apart from the high costs in time and manpower, the result would be rather useless due to its high sensitivity to variations of the system parameters.

8.5 Results from literature

Obtaining data from literature for checking the theoretical results was much more difficult than expected. The main reason for this is the lack of knowledge about the complete system configurations and the way measurements were performed. This makes it, in most cases, impossible to determine the exact contributions of all the individual penalties. I will nevertheless consider some published systems, indicating the theoretical contributions of their IF-detection penalties and the resulting discrepancies.

In principle two groups of receiver analysis can be found, of which one deals with the influence of $m$ on the sensitivity, either by increasing the bitrate or changing the real IF bandwidth, and the other group with polarisation diversity penalties. Systems belonging to each of these two groups will be analysed briefly, with results summarised in Tables 8-4 and 8-5.

IF bandwidth penalties

One of the first coherent optical systems measured extensively was a polarisation control 34 Mbit/s dual-filter FSK system of Emura et al. from NEC [17, 19]. It used two discrete IF-filters, centered at 500 and 1000 MHz with respective bandwidths of 400 and 450 MHz. This system was used for experiments at 34 and 140 Mbit/s, yielding measured sensitivities of $-60.7$ and $-55.0$ dBm respectively and thus a sensitivity degradation of 5.7 dB.

Since the two IF-filters have different bandwidth, the average $m$ is used for the calculations. This is an acceptable simplification, since the two values are very close.
Table 8.4: Comparison of different system experiments for measuring the influence of IF bandwidth expansion and the associated sensitivity penalty $\Delta SNR$. The theoretical values for $\Delta SNR$ and the resulting discrepancies are also given.

<table>
<thead>
<tr>
<th>Reference and type</th>
<th>Bitrate or $B_{IF}$</th>
<th>Measured penalty $DB$</th>
<th>Theory $\Delta SNR$ $DB$</th>
<th>Error $DB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEC [17, 19] SF-FSK</td>
<td>34 Mb/s $m=12.5$</td>
<td>140 Mb/s $m=3.0$ total 5.7</td>
<td>bitrate 6.15 $\Delta SNR$ 0.45</td>
<td>1.30 0.85</td>
</tr>
<tr>
<td></td>
<td>140 Mb/s $m=4.6$</td>
<td>200 Mb/s $m=3.25$ total 2.2</td>
<td>bitrate 1.55 $\Delta SNR$ -0.65</td>
<td>0.25 0.9</td>
</tr>
<tr>
<td></td>
<td>140 Mb/s $m=3.0$</td>
<td>140 Mb/s $m=4.6$ $\Delta SNR$ 0.3</td>
<td>0.35 0.05</td>
<td></td>
</tr>
<tr>
<td>AT&amp;T [15] 147 Mb/s DPSK</td>
<td>300 MHz $m=2.0$</td>
<td>400 MHz $m=2.7$ $\Delta SNR$ 0.45</td>
<td>0.20 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>300 MHz $m=2.0$</td>
<td>1000 MHz $m=6.8$ $\Delta SNR$ 1.15</td>
<td>0.95 0.20</td>
<td></td>
</tr>
<tr>
<td>Philips [109, 110] pol.div. FSK</td>
<td>140 Mb/s $m=7.5$</td>
<td>560 Mb/s $m=1.9$ total 5.1</td>
<td>bitrate 6.0 $\Delta SNR$ 0.9</td>
<td>1.3 0.4</td>
</tr>
</tbody>
</table>

It yields values of $m=12.5$ and 3.0 for 34 and 140 Mbit/s, respectively. Formula (5-41) then gives total IF-penalties $\Delta SNR$ of 1.94 for $m=12.5$ and 0.66 for $m=3.0$, and thus a sensitivity improvement of 1.28 dB. Combined with the 6.15 dB reduction due to the increased bitrate this should give a total sensitivity degradation of 4.85 dB, which is 0.85 dB below the reported value of 5.7 dB.

The same receiver was also modified for higher bitrates [19] by using IF BPF's of 600 and 700 MHz, centered at 600 and 1250 MHz. At 140 Mbit/s ($m_{av}=4.6$) this gave a sensitivity of $-54.7$ dBm, at 200 Mbit/s ($m_{av}=3.25$) $-52.5$ dBm or 2.2 dB lower. The theoretical penalties $\Delta SNR$ (5-41) are 0.99 and 0.72 dB respectively, yielding a 0.25 dB sensitivity improvement. Together with the 1.55 dB degradation due to the increased bitrate this should give a 1.3 dB sensitivity reduction. The discrepancy is thus almost identical to the previous result, 0.9 dB. At this point it is interesting to note that the measured discrepancy in our own 140-560 Mbit/s system was 0.4 dB, which was much higher than the other types of discrepancy measured. This makes it valid to assume that the use of existing receivers for 4-5 times higher bitrates introduces at least 0.5-1.0 dB additional sensitivity penalty, most probably
due to increased IF spectral distortion introducing ISI.

A final check that can be performed with these measurements is the influence of the increased IF bandwidth on the 140 Mbit/s sensitivity. The first measurement, at \( m=3.0 \), gave a sensitivity of \(-55.0 \) dBm, while \( m=4.6 \) yielded \(-54.7 \) dBm, a 0.3 dB reduction. The theoretical result closely matches these data, giving a \( \Delta SNR \) reduction of 0.66-0.99 (\( =-0.33 \)) dB.

A second system that is used for measuring the effects of changing IF characteristics is by Delavaux et al. from AT&T [15], an experimental 147 Mbit/s DPSK heterodyne receiver. The IF was set at four times the bitrate or 591 MHz. The bandwidth of the IF BPF - a 5-pole 0.2 dB ripple Chebyshev filter - was changed from 300 to 400 and 1000 MHz, giving \( m=2 \), 2.7 and 6.8 respectively. Using the sensitivity at \( m=2 \) as a reference, this gives theoretical penalties \( \Delta SNR \) with (7-79) of 0.58-0.39=0.19 dB and 1.34-0.39=0.95 dB. Measured sensitivity reductions are 0.45 dB from \( m=2 \) to 2.7 and 1.15 dB from \( m=2 \) to 6.8. This yields discrepancies between theory and practice of 0.25 and 0.2 dB, which is well within the measurement accuracy claimed by the authors.

Polarisation diversity penalties

Since polarisation diversity has been investigated extensively for the past few years, some more data is available on the diversity penalty \( \Delta Div \). Results of polarisation diversity system measurements by us and four other research groups are listed in Table 8-5. This covers bitrates from 140 Mbit/s to 2.5 Gbit/s, single- and dual-filter FSK detection, DPSK and CPFSK, and finally three different measurement methods. One problem with three of these systems is the fact that no IF bandwidth is mentioned. The values given in the table with a question mark are from myself, where I assumed that IF and baseband spectra are not allowed to overlap and that the IF passband is symmetrical around the central IF. The accuracy in \( m \) is therefore around 0.2-0.5, which gives only minor deviations due to the flat slope of \( \Delta Div \) as a function of \( m \).

All systems referenced here are, given their respective modulation schemes, standard polarisation diversity receivers, with IF block diagrams identical to those given in Chapters 6 and 7. The only complication is the method used for measuring the diversity penalty. NEC and KDD have used the same method as we use, i.e. the receiver is first constructed as a complete polarisation diversity receiver. All signal light is then directed into one single IF-branch, and the signal-free branch is disconnected at the baseband combining point only. The second measurement is then made with the still signal-free IF-branch connected. This yields a direct measurement of the polarisation diversity penalty \( \Delta Div \).
Table 8-5: Results for four polarisation diversity systems reported by different groups, in order of increasing bitrate, with measured values for $\Delta Div$. The theoretical values and the resulting discrepancies are given as well. Three different measuring methods have been used, detailed in the footnote. Values with a question mark are estimates by the author.

<table>
<thead>
<tr>
<th>Reference Type</th>
<th>$f_{IF}$ (GHz)</th>
<th>$M, \tau$</th>
<th>$m$</th>
<th>Diversity Penalty method</th>
<th>$\Delta Div$ (dB)</th>
<th>Theory $\Delta Div$ (dB)</th>
<th>$\epsilon$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philips [109, 110] DF-FSK 140 Mb/s</td>
<td>0.65</td>
<td>-</td>
<td>7.5</td>
<td>$1^a$ $\Delta Div$</td>
<td>0.75</td>
<td>0.70</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>1.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philips [109, 110] DF-FSK 560 Mb/s</td>
<td>idem</td>
<td>-</td>
<td>1.9</td>
<td>1</td>
<td>$\Delta Div$</td>
<td>0.35</td>
<td>0.50</td>
</tr>
<tr>
<td>NEC [76, 99] SF-FSK 400 Mb/s</td>
<td>0.9</td>
<td>-</td>
<td>2?</td>
<td>1</td>
<td>$\Delta Div$</td>
<td>1.0</td>
<td>0.35</td>
</tr>
<tr>
<td>KDD [95, 96] CPFSK 560 Mb/s</td>
<td>0.98</td>
<td>$M=0.5$</td>
<td>1.5?</td>
<td>1</td>
<td>$\Delta Div$</td>
<td>1.5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu [116, 117] DPSK 1.2 Gb/s</td>
<td>2.4</td>
<td>$\tau = T$</td>
<td>2?</td>
<td>2$^b$</td>
<td>total</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rec.noise</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Delta Div$</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Fujitsu [10] CPFSK 2.5 Gb/s</td>
<td>5.0</td>
<td>$M=0.8$</td>
<td>2</td>
<td>$3^c$</td>
<td>total</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau=5T/8$</td>
<td></td>
<td></td>
<td>opt.loss</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta=1.0$</td>
<td></td>
<td></td>
<td>rec.noise</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\Delta Div$</td>
<td>1.2</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Single branch sensitivity measured with optical diversity unit and LO aligned for polarisation diversity reception but second branch disconnected and all signal in one branch. For $\Delta Div$ measurement second branch connected.

$^b$ Single branch measurement with optical diversity unit inserted, but with all light of both signal and LO in the first branch. For polarisation diversity LO realigned for both branches and second branch connected.

$^c$ Single branch measurements are with a 0.3 dB excess loss 50:50 coupler and no second branch. Polarisation diversity measurements with an optical diversity unit inserted and the second branch connected. The increased LO power-splitting is compensated by a 2.5 dB higher LO output power.
Fujitsu has used two alternative methods. For the 1.2 Gbit/s DPSK system [116, 117] they also set-up a complete diversity receiver, including the optical diversity hybrid. However, for the reference measurement they not only direct all signal light, but also all LO-light into one IF-branch. This means that not only $\Delta Div$ is measured, but also the effect of the increased LO-power on the photodiodes in the reference situation, which gives a 2.1 dB additional sensitivity improvement. For the 2.5 Gbit/s CPFSK system\textsuperscript{19} [10] the reference measurement is even more different, since here the optical polarisation diversity hybrid is replaced by a standard 50:50 3 dB coupler. Since the 50:50 coupler has an excess loss of 0.3 dB, and the polarisation diversity an excess loss of 1.0 dB this introduces an additional 0.7 dB excess loss penalty. Furthermore, the LO-power in the case of single-branch reception is 8.5 dBm, yielding a receiver noise sensitivity penalty of 0.5 dB due to the combined effect of the coupler excess loss (0.3 dB), the 0.72% photodiode quantum efficiency and the 14 pA/$\sqrt{\text{Hz}}$ receiver noise. In order to compensate for the increased optical hybrid loss and the power-splitting over four photodiodes, the LO-power for the polarisation diversity reception has been increased to 11 dBm. One can easily calculate the resulting receiver noise penalty to 0.7 dB, 0.2 dB higher than the single-branch situation.

The resulting discrepancy between theory and practice of $\Delta Div$ varies from very low - 0.05-0.15 dB of the Philips and the Fujitsu DPSK experiments - via 0.65-0.7 dB for the NEC and Fujitsu CPFSK measurements to 1.05 dB for the KDD system. The latter value is almost certainly due to pulse distortion and ISI caused by the narrow IF filter, which is in turn a consequence of the low IF. It can be stated in general that well-designed systems have both an average IF and an IF bandwidth of at least twice the bitrate. Lower values introduce spectral distortion, IF-to-baseband crosstalk or a combination of both, all leading to increased sensitivity penalties. This is exactly the case for the KDD system, which should in contrast have had values considerably higher than 2 in view of the MSK modulation.

From these results it can be concluded that it indeed appears possible to confirm the theoretical results of Chapters 4-7 by practical system measurements. However, when accuracies of 0.1 dB are required, both the system design and the measurement method must be optimal. Especially in the case of narrow IF bandwidth (small $m$) and/or high bitrates, the accuracy rapidly decreases due to additional spectral distortion and IF branch imbalance.

\textsuperscript{19}In my opinion this is the best coherent optical system reported to date, combining superior performance with a high level of engineering. The single-branch and polarisation diversity sensitivities are $-46.2$ and $-44.1$ dBm respectively. Also the lasers, long-cavity three-section $\lambda/4$-shifted DFB lasers with 10 mW output power, 2.5 MHz linewidth and 8 GHz $\beta_n$ characteristics, are the best reported so far.
8.6 Concluding remarks

In this chapter two systems of our own group and six systems from other laboratories have been analysed. First, the 140 Mbit/s CMC FSK system has been described in detail, illustrating the design considerations influencing the final receiver performance. Even with the general system choices such as polarisation diversity versus control, the bitrate, the type of modulation and demodulation, etc., there are still many degrees of freedom related to the design of the actual receiver. It has been shown that the design of the IF-stages has a large impact on the overall system performance. Frequency deviation, IF filter bandwidth, central IF, location of Mark and Space frequencies, quadratic or linear IF-detection, and the type of post-detection combining all influence performance figures such as the presence of a linewidth floor, the polarisation dependence, ISI and, of course, the sensitivity penalties $\Delta SNR$ and $\Delta Div$.

A detailed overall receiver sensitivity analysis of the 140 Mbit/s and 1 Gbit/s systems shows the importance of incorporating these IF-penalties. In both systems the IF-penalties make up 20%-30% of the total sensitivity penalty, a figure that may easily exceed 50% when high-power LO-lasers and optical components with lower losses are used. As a result, the 'non-ideal electronics'-penalty has been reduced to a mere 5%, where this so-called 'unexplained penalty' can be attributed to Mark-Space IF-imbalance and/or IF-to-baseband crosstalk. Moreover, it is almost equal to the overall accuracy of the theoretical analysis.

Practical system measurements at 140 Mbit/s confirm the calculated diversity penalty with the extremely low inaccuracy of 0.05 dB, which is a result of the fact that we designed the system for optimum IF - i.e., polarisation-independent - performance. Measurements at the higher bitrate of 560 Mbit/s gave slightly higher inaccuracies, almost certainly due to increased spectral distortion of the IF signal.

Finally six different systems presented in the open literature, all containing measured values of $\Delta SNR$ or $\Delta Div$, have been analysed. The discrepancies vary from very small - 0.1-0.25 dB - to large, 1 dB. Large inaccuracies can, in most cases, be related to increased spectral distortion in narrow IF BPF's, illustrating the importance of optimal IF-design. Accurate analysis was hampered by a lack of exact system design data, especially the IF bandwidth (and thus $m$) is often not reported. Given the limitations of using published data, the analyses of these systems yields results that never disagree with theory, and often validate this by a remarkable agreement.
Chapter 9

Conclusions

9.1 IF-detection theory

In Chapter 3–7 a unified theoretical analysis of IF-detection in coherent optical receivers has been presented, covering all modulation and detection schemes with the exception of synchronous detection. ASK (non-coherent single-filter IF-detection), FSK (non-coherent dual-filter IF-detection) and CPFSK/DPSK (differentially coherent delay-line discriminator IF-detection) cover at least 95% of all systems reported so far. The single exception that could not be treated analytically, apart from a few special cases, was linear IF-detection. Since it has been shown on several occasions that linear detection is not favourable for optimum system performance, especially in relation with polarisation diversity reception, I do not consider this a grave shortcoming.

The results from the analysis that are of direct importance for system engineering are summarised in Table 9-1. They include the shot-noise limited BER expression and the associated sensitivity for determining the ultimate performance of the selected modulation scheme. The next step determines the linewidth-floor in relation to the IF bandwidth expansion factor \( m \). After the pdf’s, the threshold and the BER, the table ends with the compact expressions for the sensitivity penalty \( \Delta SNR \) relative to the ultimate SNR of each modulation scheme, the FSK-ASK sensitivity improvement \( \Delta FSK \) and the polarisation diversity penalty \( \Delta Div \).

Most of the results presented in Table 9-1 are for optimum detection, especially in the case of CPFSK and polarisation diversity. It has been shown in many cases that the model can easily be adapted to cover non-optimum receiver configurations as well. Examples are the IF-branch imbalance in case of polarisation or phase diversity, non-optimum detection efficiency of CPFSK and non-ideal modulation and/or IF-detection of DPSK. Other special cases, not specifically covered so far, can in principle be analysed in the same straightforward way.
Table 9-1: Summary of results obtained with the unified theoretical approach of Chapter 3–7. Results apply in all cases to receivers with heterodyne reception and quadratic IF-detection. For CPFSK the penalties are given for $\delta=1$ only, which yields the same results as for DPSK. $\Delta SNR$ is relative to the ideal $SNR=\rho_{\text{min}}$ at a BER of $10^{-9}$.

<table>
<thead>
<tr>
<th>IF-detection</th>
<th>non-coherent</th>
<th>differentially coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>modulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demodulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ideal $P_e$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SNR=\rho_{\text{min}}$</td>
<td>$\frac{1}{2} \exp(-\rho/4)$</td>
<td>$\frac{1}{2} \exp(-\rho/2)$</td>
</tr>
<tr>
<td>$P_{\text{floor}}$</td>
<td>$\Gamma\left(\frac{m}{2}, \frac{m \pi^2}{4} (1 - b')\right)$</td>
<td>$\gamma = 2\pi \Delta \nu_{IF} T / m$</td>
</tr>
<tr>
<td>$(\Delta\nu_{IF} T)_{\text{max}}$</td>
<td>$0.02 \text{ M}^{1.54}$</td>
<td>$0.04 \text{ M}^{1.54}$</td>
</tr>
<tr>
<td>$p_s(u)$</td>
<td>$p(u</td>
<td>2m)$</td>
</tr>
<tr>
<td>$p_M(u)$</td>
<td>$p(u</td>
<td>2m, 2\rho)$</td>
</tr>
<tr>
<td>decision threshold $\zeta$</td>
<td>$Y = b/2:N = \zeta^2 \rho/4$</td>
<td>$0$</td>
</tr>
<tr>
<td>detection improvement factors</td>
<td>$-\alpha = 1 + \sin \delta \pi / 2$</td>
<td>$\beta = 1 - \sin \delta \pi / 2$</td>
</tr>
<tr>
<td>$P_S$</td>
<td>$\Gamma(m, Y)/T(m)$</td>
<td>$\left(\frac{1}{2}\right)^m \exp(-\rho/2)$</td>
</tr>
<tr>
<td>$P_M$</td>
<td>$1 - Q_m(\sqrt{2\rho}, \sqrt{2Y})$</td>
<td>$\sum_{k=0}^{m-1} \left(\frac{1}{2}\right)^k L_k^{m-1}(-\frac{\rho}{2})$</td>
</tr>
<tr>
<td>$\Delta SNR \text{ [dB]}$</td>
<td>$0.708 \log m + 0.541 \log^2 m$</td>
<td>$1.085 \log m + 0.625 \log^2 m$</td>
</tr>
<tr>
<td>$\Delta FSK \text{ [dB]}$</td>
<td>$2.82 - 0.377 \log m - 0.084 \log^2 m$</td>
<td>$0.883 + 0.376 \log m$</td>
</tr>
<tr>
<td>$\Delta Div \text{ [dB]}$</td>
<td>$0.262 + 0.255 \log m + 0.121 \log^2 m$</td>
<td>$0.383 + 0.376 \log m$</td>
</tr>
<tr>
<td>$\Delta FSK_{Div} \text{ [dB]}$</td>
<td>$2.70 - 0.498 \log m + 0.037 \log^2 m$</td>
<td></td>
</tr>
</tbody>
</table>
Several factors contribute to the success of the model. First of all high accuracy is ensured by using probability density functions throughout the analysis. Accuracies of better than 0.1 dB are a prerequisite for useful penalty analysis of practical systems; this excludes other known methods than the one presented here. At least in the case of quadratic IF-detection no approximations are required, and results are obtained in a fully analytical and highly accurate way.

The second advantage of the unified receiver model developed for the analysis is the absence of linewidth effects. By separating the analysis of the effects of phase noise (the linewidth) and amplitude noise (e.g. the front end equivalent input noise) mathematical complications due to cross-effects can be avoided. It has been made credible that under certain conditions - i.e. large enough IF-bandwidth for all types of IF-detection, and/or short enough delay time in case of differentially coherent IF-detection - linewidth effects may indeed be neglected. However, this does imply that when these conditions are not fulfilled the model and its outcome lose their validity.

The final reason for the high accuracy is that the model is made as realistic as possible. Where simplifications are adopted - such as a rectangular IF bandwidth, white IF noise or integrate-and-dump post-detection filtering - they are often permissible in general and approach the real performance in practical systems. The assumption that the system operates ISI-free (which is required in practice) avoids that exact filter and pulse-form representations complicate the analysis and make it less generally applicable. Inclusion of post-detection filtering is essential for obtaining a realistic system model. Models without this filtering analyse only the - at least for coherent optical communication - hypothetical case of IF matched filtering. Finally, the symbolical notation of the pdf's in the form \( p(u|\theta, \lambda) \) and \( p(u|\vartheta) \) makes the analysis more transparent and compressed by reducing the amount of complicated chi-square-derived formulae. This is also the key to simple analysis of special cases, such as polarisation diversity receivers with imbalance.

For the analysis of single-filter IF-detection one additional item is the decision threshold setting. When accurate real figures for the sensitivity and BER are required, one can not leave this threshold as an independent parameter in the results. This means that the actual threshold setting as a function of \( m \) and the SNR must be derived, which has indeed been achieved.

9.2 Design guidelines for coherent optical systems

The performance of coherent optical systems is strongly related to the characteristics of the lasers that are used in the transmitter and as local oscillator. In practice these characteristics still limit the freedom in designing systems for optimum performance,
through the finite laser linewidth but for example also the modulation characteristics. There are in principle two different approaches to the design of a coherent optical system. The first simply states:

Determine what kind of system has to be constructed, including the modulation format and the type of IF-detection, and find lasers that fit the (linewidth) requirements of such a system.

This means that for decreasing IF bandwidth, often caused by a lack of electrical receiver bandwidth, the linewidth requirements become more stringent. This is valid for all modulation schemes. However, in view of the limited availability of good - i.e narrow linewidth - lasers it appears more realistic to use an alternative approach, stating:

Given the linewidth specifications of the available lasers, determine which combination of modulation format and IF-detection scheme yields optimum performance.

The second approach will now be used as a basis for giving some design guidelines, based on the results of the theoretical analysis as well as the experience gained from designing the systems described in Chapter 8.

Influence of linewidth

There is only one way of avoiding an observable BER-floor in the receiver characteristic, and that is increasing the frequency modulation index.

It is important to realise that all IF-detection schemes analysed, feature in principle the same kind of linewidth dependence. Since an increase in modulation index also requires a widening of the IF bandpass filters, either of the methods suffices for BER-floor reduction. However, the best results are obtained when modulation index and filter configuration are well matched. The first step is then the selection of the frequency modulation index required for a certain level of the BER-floor.

To make the BER-floor unnoticeable at practical BER levels, it should be at $10^{-20}$ or lower.

The second step determines the IF bandwidth (or $m$), which must be large enough to avoid excessive ISI. For high modulation index $M$ an IF bandwidth expansion factor $m$ equal to at least $M+2$, the modified Carson rule, should be used. For small $M$ (below 1) a relatively larger $m$ is required.
9.2. DESIGN GUIDELINES FOR COHERENT OPTICAL SYSTEMS

IF bandpass filtering

The minimum required IF bandwidth in receivers with non-coherent IF-detection follows directly from the maximum tolerable IF linewidth. To be on the safe side it is better to use a wider IF bandwidth. Firstly the effective IF linewidth may fluctuate (e.g. due to optical reflections), giving momentarily higher values, and secondly the shape of the linewidth spectrum does not have to be exactly Lorentzian. Especially deviations in the tails of a Lorentzian spectrum may lead to considerable variations in the BER-floor. A second advantage of wide IF filters is the reduced pulse distortion and ISI. The sensitivity penalty due to the wide IF filtering is not negligible but low enough to be only a minor drawback. However, a practical limitation is given by the required overall receiver bandwidth. Assuming that no IF-baseband spectral overlap is allowed, the upper corner frequency of the optical receiver must be at least \((1 + M + m)/T\) in the case of FSK. A small gain is possible by introducing a limited overlap of the two BPF’s around the central IF, but in general one must accept a bandwidth efficiency in the order of \(1/(M + m)\). Also, increasing the receiver bandwidth invariably leads to higher equivalent input noise values, introducing larger receiver noise penalties and/or requiring larger LO-powers for compensating this.

The minimum IF bandwidth is determined by the maximum IF linewidth.
The maximum IF bandwidth is determined by the tolerable sensitivity penalty due to this excess bandwidth.

For receivers with differentially coherent IF-detection, the IF bandwidth can also be used for linewidth tolerance improvement, as explained above. In particular in high-bitrate (2.5 Gbit/s) heterodyne CPFSK systems it is often not possible to obtain IF bandwidths of more than twice the bitrate; this means that overall Nyquist design of the transmission link will be very critical. Taking the IF bandwidth too narrow would also make a receiver very sensitive to deviations of the average intermediate frequency from its nominal value. CPFSK phase diversity reception (non-synchronous direct conversion reception) is one of the few possibilities for overcoming these bandwidth limitations, yielding a very high bandwidth efficiency of \(2/(2+M)\) (about 0.6-0.7) due to the ‘baseband’ or quasi-homodyne operation.

---

1 The bandwidth efficiency has been defined as the ratio of the bitrate and the receiver bandwidth that is used for the reception. For FSK this is about \(1/(M+m+1)\), for ASK about \(1/(m+1)\) and for high-bitrate heterodyne CPFSK \(1/(2M+1)\). To give an example: The 140 Mbit/s Philips FSK-system described in Chapter 8 has a receiver bandwidth of 2350 MHz, yielding a bandwidth efficiency of 140/2350 or 1/16.8=6%. With a frequency modulation index of 1200/140=8.6 and \(m=1050/140=7.5\), the factor \(1/(M+m+1)\) is equal to 1/17.1. Both methods yield about the same value for the bandwidth efficiency.
IF-branches and polarisation handling

An increasing number of IF-branches directly translates itself into increasing sensitivity penalties due to the additional NxN-noise from every branch. It is clear that this noise-determined penalty is lowest for single branch reception, implying the use of polarisation control. However, polarisation control is anything but a simple solution for practical implementation. When fibre-related operations (squeezing, pulling, rotating, etc.) are used for the control, reliability is usually very low, while the control bandwidth is often a factor 10 to 100 slower than the required 200-500 Hz. Integrated optics solutions, on the other hand, are still useless due to high insertion loss. There are some other ways open\(^2\), which may hopefully lead to practical polarisation controllers.

These drawbacks of polarisation control nearly always lead to polarisation diversity as the preferred receiver scheme. However, it should be remembered that although generally adopted, it is certainly not ideal. First the optical diversity unit introduces excess loss compared to the single 50:50-coupler of the polarisation control receiver. Secondly the LO-power is split over four instead of two photodiodes, giving reduced shot noise and an increased receiver noise penalty. Thirdly, as mentioned above, the second branch introduces a sensitivity penalty due to the additional NxN noise after IF-detection and post-detection filtering. Nevertheless, even with these three penalties, polarisation diversity can certainly be an attractive solution when optimally implemented, with very low polarisation dependence. The main problem, especially at high bitrates, is the required balance between the two IF branches. The overall amplitude-, phase- and noise-characteristics of the two branches must in principal be identical from the optical input of the photodiodes to the post-IF-detection summation point. At high bitrates this is virtually impossible, introducing sensitivity penalties equal to the SNR-imbalance. In particular when coherent optical systems must one day be produced in larger series, these drawbacks of polarisation diversity reception must certainly be borne in mind.

A similar consideration is also valid in other receiver schemes where multiple IF branches are used, especially phase diversity. The problems that arise from combining phase and polarisation diversity, yielding four IF-branches\(^3\), make this solution impracticable. The conclusion is thus:

*Use preferably only one, otherwise at most two IF-branches.*

\(^2\)A promising alternative for practical polarisation controllers is the use of (e.g., LiNbO\(_3\)) wave plates as proposed by H. Shimizu and K. Kaede, *Endless polarisation controller using electro-optic waveplates*, Elect. Lett., Vol. 24, No. 7, 1988, pp. 412–413.

9.2. DESIGN GUIDELINES FOR COHERENT OPTICAL SYSTEMS

Optical detection

The opto-electronic conversion is the first, and for the overall receiver sensitivity the most critical operation in a coherent optical receiver. Since the receiver sensitivity is defined as the minimum received power that produces a certain BER (usually $10^{-9}$), it implies that at the photodiode the Signal amplitude approaches the receiver noise level as much as possible. The receiver front end must then give maximum conversion of optical power into electrical current, with as low a noise level as possible at the required bandwidth, and with sufficient balancing for RIN-suppression. Concerning the maximum opto-electronic conversion, two simple rules apply. First the excess loss of the optical circuitry should be as low as possible, and secondly the quantum efficiency of the photodiodes should be close to 100%. In contrast to IF-related penalties which increase slowly with e.g. the IF-bandswidth, every dB of optical loss translates directly into a dB sensitivity penalty. And especially in the case of polarisation diversity, when both the Signal and LO light pass through the optical diversity unit, optical loss not only results in a direct sensitivity penalty but also - through the reduced LO-power on the photodiodes - in an increased receiver noise penalty. In many cases a reduction of optical loss (e.g. by selecting more advanced components, optimisation of fibre-to-photodiode coupling or the use of fibre splices instead of connectors in the optical unit) and the selection of high quantum efficiency photodiodes is the fastest way to sensitivity improvement. This consideration is also the major obstacle for the introduction of integrated optical units in practical systems, since these devices - in particular when coupled to single-mode fibres - invariably give high excess losses.

Low-noise optimisation of optical front ends is one of the most difficult electronic tasks in coherent optical system design. It should be remembered that, given a certain technology, the noise-power–bandwidth product is constant, which should be taken into account from the first system specifications. The fastest way to receiver noise reduction is the use of active elements with a higher transconductance (HEMT's for example) or a considerably lower capacitance (integration). When using a certain type of input transistor, the options that remain are parasitics reduction (using chips instead of packaged components) and tuning. However, the latter method can easily give non-uniform amplitude characteristics, which is another type of problem.

If everything else fails, coherent optical receiver designers fortunately always have one last remedy: increase of LO-power. Despite the simplicity of this solution, this view is not commonly endorsed by laser manufacturers.

IF-detection

From this thesis one thing has become manifest:

*Always adopt quadratic IF-detection when the system performance should be predictable.*
Nevertheless, linear IF-detection is widely used, mostly because of its simplicity. Simulations even suggest that linear detection yields a slightly higher receiver sensitivity than quadratic detection, although the differences are usually marginal. The price one has to pay for the ease of receiver design is a total lack of means for accurate theoretical system analysis.

Apart from the convenient fact that quadratic IF-detection facilitates theoretical analysis of the detection process, there are some other very practical reasons. First, quadratic detectors can actually be built, with as high an accuracy as required. This offers the important advantage that the performance of one of the most critical system elements can be accurately predicted. This is not the case when linear IF-detectors are used. In fact, really linear detectors can not be built, leaving the system designer with unknown detection characteristics. A second advantage of quadratic detection is related to polarisation diversity, facilitating direct post-detection summation without the need for ratio combining.

Another important aspect of the IF-detection is the Mark-Space SNR-balance, which is determined by the overall IF-section characteristics. Differences in the noise-level and/or the amplitude characteristic of the spectral regions belonging to the Mark and Space Signals directly yield SNR differences at the (FSK and CPFSK) IF-detector. This means that the detection process is not perfectly symmetrical anymore, or that the frequency discriminator characteristic becomes skewed. The penalties resulting from this skewness are difficult to analyse, and can easily ammount to several dB. It is clear that the problem of Mark-Space SNR-balance becomes more severe for high bitrates and/or large frequency deviations. Noise tuning and amplitude resonances are also not recommendable for obtaining perfect balance. Quasi-homodyne CPFSK phase diversity reception is an alternative method that offers perfect balance at the receiver. At the same time one should not overlook the effects of residual AM when direct frequency modulation of the transmitter laser is employed.

Post-detection filtering

The post-detection filter offers the last possibility before regeneration in the decision circuit for optimising the overall system performance. The three-fold task of the filter consists of suppression of disturbances (e.g. second harmonics of the IF) above the Nyquist frequency, SNR-improvement by filtering the detected baseband spectrum, and finally pulse-form optimisation and ISI minimisation by filtering according to Nyquist’s criterion. The first two tasks can be performed rather easily by simply using a low-pass filter of bandwidth $1/2T$. However, the third function is extremely critical, since both the amplitude-and phase-characteristics of the system transfer function from transmitter input to IF-detector output must be equalised while maintaining the minimum noise bandwidth set by the second requirement. In theory the post-detection filter must also equalise the non-uniform noise characteristics of the IF-detector output, but at high bitrates this function is usually neglected for the sake of simplicity.
9.2. DESIGN GUIDELINES FOR COHERENT OPTICAL SYSTEMS

At high bitrates filters must be very simple and small, making tuneability anything but an easy task. An effective structure is formed by a two-section filter, with a transmissive zero at the clock frequency and a LPF-section for phase- and amplitude-optimisation. It is very hard to predict system and filter performance, so the final adjustments and optimisations must always be performed on a working system.

Data regeneration and timing recovery

Data regeneration is the last analogue step in the receiver and a non-linear operation. It is thus as important as optical detection and IF-detection, which is often overlooked. In principle the two operations of data regeneration (obtaining digital Mark and Space signals from the noisy baseband signal) and timing recovery (removing of jitter, establishing a fixed edge spacing of the digital signal) take place at the same instance in an edge-triggered D-flip-flop. A stable clock signal must be generated using a PLL or an extremely narrow (SAW-)filter, yielding a signal with very low phase noise. The rising edge of the clock signal that triggers the flip-flop must be accurately located in the center of the signal eye-diagram for maximum SNR. The decision circuit itself must generate as little noise as possible, since this directly reduces the SNR. The noise performance of the decision circuit is expressed by the decision ambiguity (in mV).

In particular when the decision circuit has too high an ambiguity, people sometimes separate the regeneration and retiming operations by using a clipping- or slice-amplifier in front of the retiming flip-flop. However, although the actual system performance with a slice-amplifier can be as good as with a standard decision circuit, the same is valid as for the choice of quadratic versus linear IF-detection:

*Do not perform clipping or other limiting operations on the IF and the baseband signals if the receiver performance should be predictable.*

In practice this means that the input signal to the decision circuit must be such that it is at least 10–20 times larger than the decision ambiguity of the circuit, while the amplitude should be low enough to avoid clipping in the stages before the actual flip-flop. This often requires automatic gain control of the decision circuit input voltage.

An important parameter of the decision circuit is the decision threshold setting. In the theoretical analysis the decision threshold has always been treated in relation to IF-detection, but the actual setting takes place in the decision circuit. For FSK, CPFSK and DPSK the threshold is in principle always equal to 0, which is in practice in the center of the eye-pattern. However, due to amplitude imbalance in the IF the IF-detection may become asymmetrical, which means that the threshold setting should be adjusted. The most critical decision threshold setting is for single-filter IF-detection. It has been shown that the dependence of the BER on the threshold is very high, so the setting must be extremely precise. Since the single-filter decision threshold depends on the SNR, in principle some sort of control loop is required. This is however
difficult to realise, since the IF SNR is hard to measure accurately, in particular for low values. The setting is therefore in most cases optimised for large SNR, although this introduces an additional (small) sensitivity penalty in case the SNR is smaller.

Control loops

Two important control loops have to keep the intermediate frequency and the signal levels constant. The Automatic Frequency Control (AFC) that maintains the IF constant, usually has a very small loop bandwidth of a few kHz, thus compensating for 'slow' drift effects due to e.g. the temperature. When using FSK, CPFSK or DPSK demodulation it is very important that the central IF is located in the frequency discriminator zero-crossing (FSK and CPFSK) or at the maximum of the detection curve (DPSK). Deviations introduce asymmetrical IF-detection, resulting in rapidly increasing sensitivity penalties. It is often the best solution to use the same type of frequency discriminator for both IF-detection and AFC. This gives some guarantee that the baseband output is indeed optimised.

The Automatic Gain Control (AGC) must stabilise the amplitude of the signals in the receiver. However, one may ask which signal should be optimised. On the one hand it is very important that the input signal of the IF-detector has a fixed amplitude, in particular for quadratic IF-detection. Large input signals introduce higher harmonics that distort the baseband signal and introduce sensitivity penalties. On the other hand the final receiver sensitivity and BER are determined in the decision circuit, which requires a constant eye-pattern amplitude and threshold setting. By selecting an AGC scheme that stabilises either of the two there is a high probability that problems occur with the other non-linear operation. Therefore:

It is best to use an AGC control scheme that stabilises both the IF and the baseband amplitudes.

General remarks

Coherent optical receivers are not different from most other types of digital communication receivers, but only have their own peculiar boundary conditions. This means that receiver design faces the same problems, in particular front end design, IF-detection, Nyquist filtering and regeneration. However, due to the relatively poor laser characteristics only the simplest schemes can be used, which are in most cases identical to those in use for several decades in the more classical radio receivers. When realising this, and taking into account the special boundary conditions set by lasers and optics, it is now indeed possible to design and construct high-performance coherent optical systems. Systems that can be analysed with a high degree of accuracy.
Appendix A

Review of probability density functions

A.1 Summary of pdf characteristics

Definitions

The pdf of a variable $x(t)$ gives the probability distribution for all possible values of $x$ averaged over a certain time, which implies that the statistics of $x$ are stationary. The important normalisation is

$$\int_{-\infty}^{+\infty} p(x) \, dx = 1 \quad \text{(A.1-1)}$$

The cumulative distribution function $P(x)$ and complementary cumulative distribution function $Q(x)$ are defined as:

$$P(x) = \int_{-\infty}^{x} p(y) \, dy \quad \text{(A.1-2)}$$

$$Q(x) = \int_{x}^{+\infty} p(y) \, dy$$

$$P(x) + Q(x) = 1 \quad \text{(A.1-3)}$$

A joint pdf of multiple variables $x_1, x_2, \ldots, x_n$ should be integrated over all variables to obtain the normalisation to 1. In order to derive the pdf of a single variable from a joint pdf this should be integrated over all unused variables:

$$p(x_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p(x_1, x_2, \ldots, x_n) \, dx_2 \ldots dx_n \quad \text{(A.1-5)}$$
If the variables $x_1, \ldots, x_n$ are independent, the joint pdf can be separated into the product of $n$ independent pdf's:
\[
p(x_1, \ldots, x_n) = p(x_1) \cdots p(x_n)
\]  
(A.1-6)

**Moments**

The $n^{th}$ moment of a distribution is defined as
\[
\mu'_n = E\{x^n\} = \int_{-\infty}^{+\infty} x^n p(x) \, dx
\]  
(A.1-7)

The first moment $\mu = \mu'_1$ is the mean or average value of $x$ while the second moment $\mu'_2$ is a measure for the total power of $x$. The central moments (around $\mu$) are defined by
\[
\mu_n = E\{(x - \mu)^n\} = \int_{-\infty}^{+\infty} (x - \mu)^n p(x) \, dx
\]  
(A.1-8)

The first central moment is always 0, while the second central moment (or variance) is a measure for the AC power of the variable:
\[
\mu_2 = E\{x^2\} - E^2\{x\} = \text{var}(x) = \mu'_2 - \mu^2
\]  
(A.1-9)

**Operations on variables**

**Addition**

If $x_a$ and $x_b$ are two independent variables and $p_a(x)$ and $p_b(x)$ the corresponding probability density functions, the pdf $p_s(x)$ of the sum of $x_a$ and $x_b$ is sought.
\[
x_s = x_a + x_b
\]  
(A.1-10)

The probability that $x_s$ has a certain value $X$ is equal to all probabilities that $x_a$ is equal to $\alpha$ and $x_b$ is equal to $X - \alpha$:
\[
p_s(X) = \sum_{\alpha} p_a(\alpha) p_b(X - \alpha) \Delta \alpha
\]
\[
= \int_{\alpha} p_a(\alpha) p_b(X - \alpha) \, d\alpha
\]  
(A.1-11)

The latter form is the definition of the convolution of $p_a(x)$ and $p_b(x)$ so the shorthand notation of the resulting pdf after addition of the variables will be
\[
p_s(x) = p_a(x) * p_b(x)
\]  
(A.1-12)

A simple graphical illustration of a convolution of two pdf's limited to non-negative
values is given in Figure A.1-1. The pdf \( p_a(x) \) and the folded and shifted pdf \( p_b(X-x) \) will not overlap for \( X \) less than 0, in which case the integration interval in (A.1-11) will be from 0 to \( X \). This result was to be expected since the sum of two positive values will always be positive.

Subtraction

Subtracting two variables can be performed in almost the same way.

\[
x_d = x_a - x_b
\]

\[
p_d(X) = \sum_{\alpha} p_a(X + \alpha)p_b(\alpha) \Delta \alpha
\]

\[
= \int p_a(\alpha)p_b(-X + \alpha) \, d\alpha
\]

The \( p_b \)-component in the integral can be rewritten into \( p_b(-(X-\alpha)) \) which, when compared to the convolution integral (A.1-11), leads to the final expression

\[
p_d(x) = p_a(x) * p_b(-x)
\]

\[
= \int p_a(\alpha)p_b(-(X-\alpha)) \, d\alpha
\]

By definition the above result is identical to the cross-correlation function \( \phi_{ab}(x) \) of the two variables \( x_a \) and \( x_b \).

The cross-correlation of two non-negative variables using the convolution integral (A.1-15) is illustrated in Figure A.1-2, which shows that a double folding of \( p_b(x) \) is
required. The first is due to the fact that \( p_b(-x) \) is used, while the second is due to the convolution. The net effect will be no folding of \( p_b(x) \). However, this means that also for negative values of \( x \) the convolution will be non-zero, even when non-negative pdf's are used. In the latter case the convolution integral (A.1-15) must be split into two separate integrals:

\[
p_d(x) = \int_{0}^{\infty} p_a(\alpha) p_b(-(x - \alpha)) \, d\alpha \quad (x \leq 0)
\]

(A.1-16)

\[
p_d(x) = \int_{x}^{\infty} p_a(\alpha) p_b(-(x - \alpha)) \, d\alpha \quad (x \geq 0)
\]

(A.1-17)

**Multiplication**

When \( x_a \) and \( x_b \) are two independent variables with their corresponding pdf's \( p_a(x) \) and \( p_b(x) \) the resulting variable after multiplication is defined as

\[
x_m = x_a \cdot x_b
\]

(A.1-18)
A.2. NORMAL DISTRIBUTION

When both \( x_a \) and \( x_b \) are positive random variables, the pdf of \( x_m \) is identical to the Mellin convolution of two functions \( p_a(x) \) and \( p_b(x) \):

\[
p_m(x) = \int_0^\infty \frac{1}{\alpha} p_a(\alpha)p_b \left( \frac{x}{\alpha} \right) d\alpha
\]

(Springer and Thompson [105] have proven that in case of more general variables that may assume both positive and negative values, in the above integral \( 1/\alpha \) should be replaced by \( 1/|\alpha| \).

A.2 Normal distribution

The normal or Gaussian distribution is the best documented distribution and needs not be presented here in detail. However, from a point of normalisation a very short summary will be given. The pdf of a normal distribution can be written in its general (non-central) form as

\[
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x-\bar{x})^2}{2\sigma^2} \right)
\]

The first and second moment are then simply given by

\[
\mu = E(x) = \bar{x} \quad (A.2-2)
\]
\[
\mu_2 = \text{var}(x) = \sigma^2 \quad (A.2-3)
\]

When \( \mu = 0 \) and \( \mu_2 = 1 \) the distribution is often referred to as unit normal.

The distribution function itself has no analytical representation since the integral of \( p(x) \) can not be solved in closed form. It is related to the error function or probability integral.

\[
P(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{y^2}{2\sigma^2} \right) dy
\]

\[
= \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{x}{\sqrt{2}\sigma} \right) \quad (x > 0)
\]

with the error function defined by (Gr.8.250-1):

\[
\text{erf}(x) = \Phi(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)dy
\]

---

De complementary distribution function $Q(x)$ is defined by the complementary error function $\text{erfc}(x)$:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) \, dy = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2\sigma^2}}\right)$$  \hspace{1cm} (A.2-6)

$$\text{erfc}(x) \overset{\Delta}{=} 1 - \text{erf}(x)$$  \hspace{1cm} (A.2-7)

$$\simeq \frac{1}{\sqrt{\pi x}} \exp(-x^2)$$  \hspace{1cm} (A.2-8)

### A.3 Chi-square distribution

**Non-central chi-square distribution**

The chi-square distribution in its most general form is called the non-central chi-square distribution\(^2\). It can be considered as a generalised Rice distribution. When $X_1, X_2, \ldots, X_\vartheta$ are identically distributed unit normal variables\(^3\) and $a_1, a_2, \ldots, a_\vartheta$ are constants, then the variable

$$x = \sum_{j=1}^{\vartheta} (X_j + a_j)^2$$  \hspace{1cm} (A.3-1)

is said to be non-central chi-square distributed with $\vartheta$ degrees of freedom and a noncentrality parameter $\lambda$ defined by

$$\lambda \overset{\Delta}{=} \sum_{j=1}^{\vartheta} a_j^2$$  \hspace{1cm} (A.3-2)

The variable $x$ is often represented as $\chi^2_\vartheta(\lambda)$. The pdf of $x$ will be written in the short form $p(x \mid \vartheta, \lambda)$ and the distribution function by $P(x \mid \vartheta, \lambda)$. Although in the above


\(^3\)An alternative way to define the non-central chi-square variable, is as the sum of non-central normal variables:

$$x = \sum_{j=1}^{\vartheta} (X_j - <X>)^2 \quad \text{with}$$

$$<X> = \vartheta^{-1} \sum_{j=1}^{\vartheta} X_j$$

This may sometimes be of use.
definition (A.3-1) \( \vartheta \) must be integer, the pdf is still a non-central chi-square with \( \vartheta \) degrees of freedom if \( \vartheta \) is any positive number. The pdf of \( x \) is identical to

\[
p(x \mid \vartheta, \lambda) = \frac{1}{2} \left( \frac{x}{2} \right)^{\frac{\vartheta-2}{4}} \exp \left( -\frac{x}{2} - \frac{\lambda}{2} \right) I_{\frac{\vartheta-2}{2}} \left( \sqrt{\lambda x} \right) \quad (x \geq 0)
\]

(A.3-3)

where \( I_m(\cdot) \) is the modified Bessel function of the first kind and order \( m \). Since \( x \) is equal to a sum of squares it is limited to positive values. When the variance of each variable \( X_j \) is equal to \( \sigma^2 \) the pdf can be written as [62]

\[
p(x \mid \vartheta, \lambda) = \frac{1}{2\sigma^2} \left( \frac{x}{\lambda \sigma^2} \right)^{\frac{\vartheta-2}{4}} \exp \left( -\frac{x}{2\sigma^2} - \frac{\lambda}{2} \right) I_{\frac{\vartheta-2}{2}} \left( \sqrt{\frac{\lambda x}{\sigma^2}} \right) \quad (x \geq 0)
\]

(A.3-4)

The noncentrality parameter then becomes

\[
\lambda = \sum_{j=1}^{\vartheta} \frac{a_j^2}{\sigma^2}
\]

(A.3-5)

Moments

The \( k^{th} \) moment around zero of the pdf can be calculated using (A.1-7) as

\[
\mu'_k = E\{x^k\} \quad (A.3-6)
\]

\[
= \frac{1}{2} \left( \frac{1}{\lambda} \right)^{\frac{\vartheta-2}{4}} \exp \left( -\frac{\lambda}{2} \right) \int_0^\infty x^{\frac{\vartheta}{2}+k-\frac{1}{2}} \exp \left( -\frac{x}{2} \right) I_{\frac{\vartheta-2}{2}} \left( \sqrt{\lambda x} \right) dx \quad (A.3-7)
\]

Integrals of this form will appear frequently in the analysis. By setting \( I_m(\cdot) = (-j)^m J_m(jx) \) (Gr.8.406-1) and using (Gr.6.643-4) the following standard integral can be derived:

\[
\int_0^\infty y^{m-1+k} \exp \left( -\frac{y}{\alpha} \right) I_{m-1} (\beta \sqrt{y}) = k! \left( \frac{\alpha \beta}{2} \right)^{m-1} \alpha^{k+1} \exp \left( -\frac{\alpha \beta^2}{4} \right) L_k^{m-1} \left( \frac{\alpha \beta^2}{4} \right) \quad (A.3-8)
\]

where \( L_k^{m}(x) \) is the generalised Laguerre polynomial, which is a member of the family of orthogonal polynomials, defined by (Gr.8.970-1)

\[
L_k^{m}(x) = \sum_{i=0}^{m} (-1)^i \left( \begin{array}{c} m + \alpha \cr m - i \end{array} \right) \frac{x^i}{i!}
\]

(A.3-9)
One important remark concerning these polynomials is the following. In the definition of the Laguerre polynomial above, \( x \) should be a positive real variable in order to maintain orthogonality. However, in all results related to chi-square distributions the variable \( x \) will have a negative value. Although this means that the orthogonality disappears, the definition given by (A.3.9) remains valid.

Applying (A.3.8) to the expression for the moments yields

\[
\mu_k' = E\{x^k\} = 2^k k! L^\frac{k}{2} - 1\left(-\frac{\lambda}{2}\right) \quad (A.3.10)
\]

Setting \( k = 1 \) and using (A.3.9) yields the expectation \( E\{x\} = \vartheta + \lambda \), while setting \( k = 2 \) gives the second moment around 0

\[
E\{x^2\} = (\vartheta + 2)(\vartheta + 2\lambda) + \lambda^2 \quad (A.3.11)
\]

The variance of the non-central chi-square pdf can then be calculated as

\[
\text{var}(x) = E\{x^2\} - E^2\{x\} = [(\vartheta + 2)(\vartheta + 2\lambda) + \lambda^2] - [\vartheta(\vartheta + 2\lambda) + \lambda^2] = 2(\vartheta + 2\lambda) \quad (A.3.12)
\]

When the variance of each variable \( X \) is equal to \( \sigma^2 \) the expressions for the mean and variance are given by:

\[
E\{x\} = (\lambda + \vartheta)\sigma^2 \quad (A.3.13)
\]

\[
\text{var}(x) = 2(2\lambda + \vartheta)\sigma^4 \quad (A.3.14)
\]

The pdf for \( \vartheta=1 \) has a singularity for \( x \downarrow 0 \) and is given by:

\[
p(x \mid 1, \lambda) = \frac{1}{\sqrt{2\pi x}} \exp \left(-\frac{x + \lambda}{2}\right) \cosh \left(\sqrt{\lambda x}\right) \quad (A.3.15)
\]

It can be proven that this pdf is valid, since the normalisation integral from 0 to infinity yields exactly 1.

In the limit of either large \( \vartheta \) or large \( \lambda \) the chi-square distribution approaches the normal distribution.

Reproductivity

A very important aspect of the chi-square distribution is its reproducibility. If two variables are independent and chi-square distributed as \( \chi^2_{\vartheta_1}(\lambda_1) \) and \( \chi^2_{\vartheta_2}(\lambda_2) \), then the sum \( \chi^2_{\vartheta_1+\vartheta_2}(\lambda_1+\lambda_2) \) is distributed as \( \chi^2_{\vartheta_1+\vartheta_2}(\lambda_1+\lambda_2) \). In other words: if \( x_1 \) and \( x_2 \) have chi-square pdf's \( p_1(x \mid \vartheta_1, \lambda_1) \) and \( p_2(x \mid \vartheta_2, \lambda_2) \), then the pdf of \( x_s = x_1 + x_2 \) is given by \( p_s(x \mid \vartheta_1+\vartheta_2, \lambda_1+\lambda_2) \). The chi-square distribution is thus reproductive under convolution.

\[
p_1(x \mid \vartheta_1, \lambda_1) * p_2(x \mid \vartheta_2, \lambda_2) = p_s(x \mid \vartheta_1+\vartheta_2, \lambda_1+\lambda_2) \quad (A.3.16)
\]

\[
\text{var}(x) = 2(2\lambda + \vartheta)\sigma^4 \quad (A.3.12)
\]
The characterization of chi-square distributed variables means that if \( x \) is distributed as \( \chi^2_\theta (\lambda) \) and \( x = X_1 + X_2 + \ldots X_\theta \) with all \( X_i \) identically distributed, then each \( X_i \) is distributed as \( \chi^2_1 (\lambda/\theta) \).

**Central chi-square distribution**

The central chi-square distribution\(^4\) can be derived from the non-central one by setting all \( a_j \) (and consequently \( \lambda \)) equal to 0. The variable \( y \) then becomes the sum of central squares

\[
y = \sum_{j=1}^\theta Y_j^2
\]

(A.3-17)

and the pdf can be derived from (A.3-4) as

\[
p(y \mid \theta) = \frac{1}{2\sigma^2 \Gamma(\frac{\theta}{2})} \left( \frac{y}{2\sigma^2} \right)^{\frac{\theta - 2}{2}} \exp \left( -\frac{y}{2\sigma^2} \right) \quad (x \geq 0)
\]

(A.3-18)

It is clear that the reproducibility and characterisation as defined for the non-central chi-square distribution are valid for the central case. The central chi-square distribution is furthermore a special case of the gamma distribution [61]. This means that the distribution functions can be expressed in terms of the (incomplete) gamma functions \( \gamma(\alpha, y) \) and \( \Gamma(\alpha, y) \) through the relations

\[
P(y \mid \theta) = \int_0^y p(z \mid \theta)dz = \frac{1}{\Gamma(\frac{\theta}{2})} \gamma(\frac{\theta}{2}, \frac{y}{2})
\]

(A.3-19)

\[
Q(y \mid \theta) = \int_y^\infty p(z \mid \theta)dz = \frac{1}{\Gamma(\frac{\theta}{2})} \Gamma(\frac{\theta}{2}, \frac{y}{2})
\]

(A.3-20)

The gamma functions are defined by the integral (Gr.8.350)

\[
\gamma(\alpha, y) \triangleq \int_0^y \tau^{\alpha-1} \exp(-\tau)d\tau
\]

(A.3-21)

\[
\Gamma(\alpha, y) = 1 - \gamma(\alpha, y)
\]

(A.3-22)

When \( \alpha \) is integer these functions are written as finite sums and can be regarded as truncated series expansions of the exponential function (Gr.8.352):

---

\(^4\)An extensive overview of most characteristics of the chi-square distribution can be found in: N.L. Johnson, S. Kotz, Continuous univariate distributions - I, Chapter 17, John Wiley & Sons, New York, 1970. [61]
\[
\frac{\Gamma(1 + n, y)}{n!} = \exp(-y) \sum_{k=0}^{n} \frac{x^k}{k!} \\
\frac{\gamma(1 + n, y)}{n!} = 1 - \exp(-y) \sum_{k=0}^{n} \frac{x^k}{k!}
\] (A.3-24)

Moments

The \(k\)th moment around 0 of the central chi-square distribution can be calculated similarly to the non-central case by:

\[
\mu'_k = E\{y^k\} = \frac{1}{2^{n/2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty y^{\frac{n}{2} + k - 1} \exp\left(-\frac{y}{2}\right) dy
\]

Using the standard integral (Gr.3.381-4):

\[
\int_0^\infty x^{\nu-1} \exp(-\mu x) dx = \mu^{-\nu} \Gamma(\nu)
\] (A.3-26)

the moment follows as

\[
\mu'_k = 2^k \frac{\Gamma\left(\frac{\nu}{2} + k\right)}{\Gamma\left(\frac{\nu}{2}\right)}
\] (A.3-27)

This result can also be obtained by setting \(\lambda\) equal to zero in (A.3-10) in which case the Laguerre polynomial reduces to

\[
L_{\frac{k}{2} - 1}(0) = \binom{\frac{\nu}{2} + k - 1}{k} = \frac{\Gamma\left(\frac{\nu}{2} + k\right)}{k! \Gamma\left(\frac{\nu}{2}\right)}
\] (A.3-28)

The mean and variance can easily be derived from the ones given for the non-central distribution or (A.3-27) as:

\[
E\{y\} = \vartheta \sigma^2
\] (A.3-29)

\[
\text{var}(y) = 2 \vartheta \sigma^4
\] (A.3-30)

Again, the pdf for \(\vartheta=1\) has a singularity for \(y \downarrow 0\) and can be derived from (A.3-18) as:

\[
p(y \mid 1) = \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right)
\] (A.3-31)

Using (A.3-26) the normalisation from 0 to infinity indeed gives exactly 1.
A.4. CHI DISTRIBUTION

Relation to non-central chi-square pdf

Apart from the fact that the central chi-square distribution can be obtained from the non-central one by setting $\lambda$ equal to zero, they are also related in a different way. The non-central pdf can namely be regarded as an infinite mixture of central pdf’s weighted by a Poisson distribution\(^5\) with expected value $\lambda/2$:

$$p(x \mid \vartheta, \lambda) = \sum_{j=0}^{\infty} \left[ \frac{(\lambda/2)^j}{j!} \exp \left( -\frac{\lambda}{2} \right) \right] p(x \mid \vartheta + 2j) \quad (A.3-32)$$

In the same way the cumulative distribution functions are related as

$$P(x \mid \vartheta, \lambda) = \sum_{j=0}^{\infty} \left[ \frac{(\lambda/2)^j}{j!} \exp \left( -\frac{\lambda}{2} \right) \right] P(x \mid \vartheta + 2j) \quad (A.3-33)$$

A.4 Chi distribution

Non-central chi distribution

The Chi distribution is much less documented than its ‘parent’ Chi-square distribution. A non-central Chi-distributed variable can be obtained by taking the root of a sum of squares defined by (A.3-1):

$$x = \sqrt{\sum_{j=1}^{\vartheta} (X_j + a_j)^2} \quad (A.4-1)$$

The pdf can be derived from (A.3-4) and (A.3-18) by replacing all $x$ by $\sqrt{x}$ and setting $p_{\chi'}(x) = 2xp_{\chi^2}(x^2)$:

$$p_{\chi'}(x \mid \vartheta, \lambda) = \sqrt{\frac{\lambda}{\vartheta \sigma^2}} \left( \frac{x^2}{\lambda \sigma^2} \right)^{\frac{\vartheta}{2}} \exp \left( -\frac{x^2}{2\sigma^2} - \frac{\lambda}{2} \right) I_{x^2} \left( x \sqrt{\frac{\lambda}{\sigma^2}} \right) \quad (A.4-2)$$

$$x \geq 0$$

The noncentrality parameter $\lambda$ is defined in the same way as (A.3-5).

Moments

Due to the root in the definition of $x$ and $y$, calculating the moments of chi distributions gives more problems than with chi-square distributions. The moments around zero of

\(^5\)Johnson and Kotz [62, chapter 28], see also Abramovitz and Stegun [1, 26.4.25]
Appendix A. Review of Probability Density Functions

the non-central chi distribution are given by the integral

$$
\mu'_k = E\{x^k\} = \lambda^{-\frac{k+2}{2}} \exp\left(-\frac{\lambda}{2}\right) \cdot \int_0^\infty y^{\frac{k}{2}+k} \exp\left(-\frac{y^2}{2}\right) I_{\frac{k+2}{2}}(y\sqrt{\lambda}) \, dy
$$

Using the relation $I_m(x) = (-j)^m J_{-m}(jx)$ and setting $z = y/\sqrt{2}$ the $k^{th}$ moment can be written as

$$
\mu'_k = 2^{\frac{k+2}{4}} \lambda^{\frac{k+2}{2}} (-j)^{\frac{k}{2}} \cdot \int_0^\infty z^{\frac{k+2}{2}} \exp(-z^2) J_{\frac{k-2}{2}}(z\sqrt{2\lambda}) \, dz
$$

With the standard integral (Gr.6.631-1):

$$
\int_0^\infty z^\mu \exp(-\alpha z^2) J_\mu(\beta z) \, dz =
\frac{\Gamma\left(\frac{\nu+\mu}{2}\right)}{\beta \alpha^{\nu/2} \Gamma(\nu+1)} \cdot \exp\left(-\frac{\beta^2}{8\alpha}\right) M_{\frac{k+2}{2}, \frac{k}{2}}\left(\frac{\beta^2}{4\alpha}\right)
$$

(A.4-3)

the integral in the expression for $\mu'_k$ is given by

$$
\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sqrt{-2\lambda}} \exp\left(-\frac{\lambda}{4}\right) M_{\frac{\nu}{2}, \frac{\nu}{2} - \frac{\lambda}{2}}\left(-\frac{\lambda}{2}\right)
$$

(A.4-4)

$M_{a,b}(x)$ is a Whittaker function, which is a particular form of a degenerate hypergeometric function. It is therefore more convenient to eliminate the Whittaker function using the definition (Gr.9.220-2):

$$
M_{a,b}(z) = z^{b+\frac{1}{2}} \exp\left(-\frac{z}{2}\right) \Phi\left(b - a + \frac{1}{2}, 2b + 1; z\right)
$$

(A.4-5)

Taking together all components yields the expression for the $k^{th}$ moment around zero of the non-central chi distribution:

$$
\mu'_k = 2^{\frac{k}{2}} \frac{\Gamma\left(\frac{k+2}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \Phi\left(-\frac{k}{2}, \frac{k}{2} + \frac{\lambda}{2}\right)
$$

(A.4-6)

This result has also been obtained by Park [88]. The degenerate hypergeometric function $\Phi(\alpha; \gamma; z)$ is also known as $_1F_1(\alpha; \gamma; z)$ and is defined as (Gr.9.210-1):

$$
\Phi(\alpha, \gamma; z) = 1 + \frac{\alpha z}{\gamma 1!} + \frac{\alpha(\alpha+1) z^2}{\gamma(\gamma+1) 2!} + \ldots
$$

(A.4-7)

---

6 See Gradsteyn and Ryznik [30], sections 9.21–9.22.
Only when \( k \) is an even integer may the expression (A.4-6) be reduced to a generalised Laguerre polynomial using the relation (Gr.8.972-1):

\[
L_n^\alpha(x) = \binom{n+\alpha}{n} \Phi(-n, \alpha + 1; x) \quad (A.4-8)
\]

The moments around zero are then given by the following expression (which could in this case also be obtained directly by using (Gr.6.631-10) and not (Gr.6.631-1) for solving the integral):

\[
\mu'_k = 2^{k/2}(k/2)! \int_{k/2}^{\infty} \left( -\frac{\lambda}{2} \right)^{k/2} (k \text{ even}) \quad (A.4-9)
\]

Note that this formula is identical to the expression for the moments of the non-central chi-square distribution (A.3-10), but with \( k/2 \) instead of \( k \). This can also be written as

\[
\mu'_k(\chi'_{\theta}) = \mu'_{k/2}(\chi'_{\theta}) \quad (A.4-10)
\]

The second moment around zero is therefore identical to the mean value of the chi-square pdf, being \( \nu + \lambda \). The first moment however, is far more complicated to calculate since the degenerate hypergeometric function in (A.4-6) should be solved. Using the functional relation (Gr.9.212-1)

\[
\Phi(\alpha, \gamma; z) = \exp(z)\Phi(\gamma - \alpha, \gamma; -z) \quad (A.4-11)
\]

and the asymptotic representation for large \( \lambda \) of \( \Phi(\alpha, \gamma; z) \) (Ab.13.1.4), in combination with an approximation by Marcum [75, p.174], we obtain an extended asymptotic expansion:

\[
\Phi(\alpha, b; z) = \frac{\Gamma(b)}{\Gamma(a)} \exp(z)z^{a-b} \left\{ 1 + \frac{(1-a)(b-a)}{z} + O(1, z^{-2}) \right\} \quad (z \gg 1) \quad (A.4-12)
\]

the final approximations for the expected value and the variance (with the variance of each \( X \) equal to \( \sigma^2 \)) are then given by:

\[
E(x) = \left[ \sqrt{\lambda} + \frac{1}{\sqrt{\lambda}} \right] \cdot \sqrt{\sigma^2} \quad (\lambda \gg 1) \quad (A.4-13)
\]

\[
\text{var}(x) = \left[ \frac{2}{\theta} - \frac{1}{\theta^2 \lambda} \right] \cdot \sigma^2 \quad (\lambda \gg 1) \quad (A.4-14)
\]
Central chi distribution

The central chi distribution is defined by

$$ y = \sqrt{\sum_{j=1}^{\vartheta} Y_j^2} \quad (A.4-15) $$

The pdf is then given by

$$ p_{\chi}(y \mid \vartheta) = \frac{y}{\sigma^2 \Gamma\left(\frac{\vartheta}{2}\right)} \left(\frac{y^2}{2\sigma^2}\right)^{\frac{\vartheta-2}{2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad (y \geq 0) \quad (A.4-16) $$

The moments of the central chi distribution can be derived directly from

$$ \mu'_k = \mathbb{E}\{y^k\} = \frac{1}{2^{\frac{\vartheta-k}{2}} \Gamma\left(\frac{\vartheta}{2}\right)} \int_0^\infty y^{\vartheta+k-1} \exp\left(-\frac{y^2}{2}\right) \, dy \quad (A.4-17) $$

With the standard integral \((Gr.3.461-3)\)

$$ \int_0^\infty z^{2n+1} \exp(-p_2 z^2) \, dz = \frac{n!}{2p^{n+1}} \quad (A.4-18) $$

this results in

$$ \mu'_k = 2^{\frac{k}{2}} \frac{\Gamma\left(\frac{\vartheta+k}{2}\right)}{\Gamma\left(\frac{\vartheta}{2}\right)} \quad (A.4-19) $$

This result can also be obtained by setting \(\lambda\) to zero in \((A.4-6)\), knowing that \(\Phi(a, b; 0)\) is equal to 1 (see the definition). Again the relation \((A.4-10)\) is valid.

In general the uneven moments cannot be calculated analytically. However, an approximation by Johnson and Kotz \([61, \text{ch.17, formula (64)}]\) is given as (including the variance \(\sigma^2\) of each \(Y^2\))

$$ \mathbb{E}\{y\} = \sqrt{\vartheta} \left\{1 - \frac{1}{4} \vartheta^{-1} + \frac{1}{32} \vartheta^{-2} + \frac{5}{128} \vartheta^{-3} \ldots\right\} \cdot \sqrt{\sigma^2} \quad (A.4-20) $$

For \(\vartheta=2\) and using the standard integral \((Gr.3.371)\) it can be shown that this is equal to \(\sqrt{\pi/2 \cdot \sigma}\). With the above approximation for \(\mathbb{E}(y)\) the variance can be found to be

$$ \text{var}(y) = \left\{1 - \frac{1}{8} \vartheta^{-1} - \frac{1}{32} \vartheta^{-2} + \frac{217}{8192} \vartheta^{-3} \ldots\right\} \cdot \sigma^2 \quad (A.4-21) $$

For \(\vartheta=2\) this is indeed a good approximation of \((2 - \frac{\pi}{2})\sigma^2\).
Figure A.4-1: Obtaining a chi probability density function by folding a non-central normal probability density function around 0. (Illustrated for $\theta=1$).

Relation to normal pdf

An interesting link to the normal distribution is the fact that the chi distribution can be seen as a folded normal distribution [61, p.93], which follows from the equality $\sqrt{x^2} = |x|$. This is also illustrated in figure A.4-1. In the non-central case it is now clear why a chi (or chi-square) distribution approaches a normal distribution for large $\lambda$ (and thus mean); the part of the pdf that is folded around the vertical axis becomes negligibly small. In the central case (with folding around the mean of the normal pdf) a central chi or half-normal distribution is obtained. In the one-dimensional case ($\theta=1$) this pdf can be derived from (A.4-16) as

$$p(y) = 2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad (y \geq 0) \quad (A.4-22)$$

which is indeed two times the normal pdf limited to positive values.
Appendix B

Decision threshold calculation

B.1 Exact threshold computation

The threshold equation is given by

$$I_{m-1}(2\sqrt{Y\rho}) = \frac{\exp(\rho)}{\Gamma(m)} (Y\rho)^{\frac{m-1}{2}}$$  \hspace{1cm} (B.1-1)

which can be rewritten using

$$x = 2\sqrt{Y\rho} \triangleq \zeta \rho$$  \hspace{1cm} (B.1-2)

$$Y = \frac{\zeta^2 P}{4}$$  \hspace{1cm} (B.1-3)

$$b = \frac{\zeta^2 A^2}{4}$$  \hspace{1cm} (B.1-4)

into

$$I_{m-1}(x) = \frac{\exp(\rho)}{\Gamma(m)} \left(\frac{x}{2}\right)^{m-1}$$  \hspace{1cm} (B.1-5)

The exact solution can be found numerically using the series expansion of the modified Bessel function (Gr.8.445):

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{1}{k!(\nu + k)!} \left(\frac{z}{2}\right)^{\nu+2k}$$  \hspace{1cm} (B.1-6)

This yields
Figure B.1-1: Illustration of the iterative calculation of the optimum threshold deviation factor $x = \zeta \rho$.

\[
\sum_{k=0}^{\infty} \frac{\left( \frac{x}{2} \right)^{m-1+2k}}{k! \Gamma(m+k)} = \frac{\exp(\rho)}{\Gamma(m)} \left( \frac{x}{2} \right)^{m-1}
\]

\[
\exp\left( -\frac{\rho}{2} \right) \sum_{k=0}^{\infty} \frac{\left( \frac{x}{2} \right)^{2k}}{k! \Gamma(m+k)} = \frac{\exp\left( \frac{\rho}{2} \right)}{\Gamma(m)}
\]

\[
f(x) = z_{\text{ref}}
\]

(The left and right side of the equation have been multiplied by $\exp(-\rho/2)$ in order to avoid overflows during the numerical evaluation). The derivative of $f(x)$ is defined as $g(x)$ and is given by

\[
g(x) \triangleq \frac{df(x)}{dx} = \exp\left( -\frac{\rho}{2} \right) \sum_{k=0}^{\infty} \frac{\left( \frac{x}{2} \right)^{2k+1}}{k! \Gamma(m+k+1)}
\]

Both $f(x)$ and $g(x)$ are of the same magnitude, which means that the function $f(x)$ is extremely steep$^1$. The following iterative operation can now be used to calculate the zero of $f(x) - z_{\text{ref}}$ with a very high accuracy:

\[
x_0 = \left( 1 + \frac{m}{10} \right) \rho
\]

\[
z = f(x_n) - z_{\text{ref}}
\]

\[
x_{n+1} = x_n + g(x_n) \cdot z
\]

---

$^1$To give a numerical indication: for $m=5$ and $\rho=100$, $I_4(x)=1.2 \times 10^{12}$, while both $f(x)$ and $g(x)$ are in the order of $10^{21}$. 
The starting value $x_0$ is chosen to be larger than the exact solution $x_{opt}$. The calculation stops when the accuracy is high enough or when $x$ becomes negative. In all cases the absolute accuracy in $x$ is $10^{-12}$-$10^{-16}$, so the relative accuracy of $\zeta$ - with a value between 1 and 1.5 in all practical cases of interest - will be in the order $10^{-14}$-$10^{-18}$.

### B.2 Iterative threshold approximation

A fast but nevertheless accurate way of obtaining the optimum threshold is by using a simple iteration derived from the threshold equation (B.1-5). It has been shown in the previous section that for normal SNR-values the value of the modified Bessel function is in the order of $10^{50}$. The asymptotic approximation (4-30) may thus be used, although the accuracy decreases with increasing $m$. The threshold equation (B.1-5) can then be rewritten into:

\[
\frac{\exp(x)}{\sqrt{2\pi x}} = \frac{\exp(\rho)}{2^{m-1} \Gamma(m)} x^{m-1} \tag{B.2-1}
\]

\[
\exp(x) = \frac{\sqrt{2\pi} \exp(\rho)}{2^{m-1} \Gamma(m)} x^{m-\frac{1}{2}} = C x^{m-\frac{1}{2}} \tag{B.2-2}
\]

\[
x = (m - \frac{1}{2}) \ln x + \ln C \tag{B.2-3}
\]

where $C$ is determined by $m$ and $\rho$ only, and equal to

\[
C = \frac{\sqrt{2\pi} \exp(\rho)}{2^{m-1} \Gamma(m)} \tag{B.2-4}
\]

---

**Figure B.2.1:** The iteration towards the solution $x_{opt}$ of the threshold equation.
As illustrated in Figure B.2-1 the two functions \( x \) and \((m - \frac{1}{2}) \ln x + \ln C\) converge, so the following iteration can be used:

\[
x_{k+1} = (m - \frac{1}{2}) \ln x_k + \ln C
\]

(B.2-5)

With a starting value \( x_0 = \rho \), four to six iterations are usually enough for a relative accuracy of \(10^{-5}\). Appendix B.4 compares the results of the iteration with the exact results.

### B.3 Analytical threshold approximation

Formula (B.2-1) can also be the basis for an analytical approximation of the threshold. For this purpose \( x \) will be replaced by its equivalent \( \zeta \rho \), which yields:

\[
\frac{\exp(\zeta \rho)}{\sqrt{2\pi}\rho} = \frac{\exp(\rho)}{\Gamma(m)} \left( \frac{\zeta \rho}{2} \right)^{m-1}
\]

(B.3-1)

\[
\exp((\zeta - 1)\rho) = \frac{\sqrt{2\pi}}{2^{m-1}\Gamma(m)}(\zeta \rho)^{m-\frac{1}{2}}
\]

(B.3-2)

Taking the natural logarithm at both sides and grouping some of the terms yields

\[
\zeta - 1 = \frac{\ln \pi + (3 + 2m) \ln 2 + (2m - 1) \ln \zeta \rho - 2 \ln \Gamma(m)}{2\rho}
\]

(B.3-3)

It can be checked that for \( m=1 \) this equation reduces to (4.32). Since \( \zeta \) is always close to 1, \( \ln \zeta \) may be approximated by \( \zeta - 1 \), and the equation can be rewritten into

\[
\zeta - 1 = \frac{\ln \pi + (3 + 2m) \ln 2 + (2m - 1) \ln \rho - 2 \ln \Gamma(m)}{2\rho - 2m + 1}
\]

(B.3-4)

A final simplification can be achieved using an approximation for the natural log of the gamma-function (Gi.8.343):

\[
\ln \Gamma(m) = m \ln m - \frac{1}{2} \ln m - m + \ln \sqrt{2\pi} + O(m)
\]

(B.3-5)

where \( O(m) \) is a remainder that can be neglected. Combining the latter two results gives the final analytical expression for the threshold deviation factor \( \zeta \):

\[
\zeta - 1 = \frac{(2m - 1) \ln \left( \frac{\rho}{2m} \right) + \ln 2 + 2m}{2\rho - 2m + 1}
\]

(B.3-6)

### B.4 Comparison of the results

The results obtained with the different methods are compared in the following table. These methods are: the exact numerical calculation from Appendix B.1, the iteration
B.4. COMPARISON OF THE RESULTS

described in Appendix B.2, the analytical approximation from Appendix B.3, the formula from Schwartz, Bennett and Stein [98, p.291] as detailed in section 4.1.2, and finally the numerical results published for two cases by Foschini, Greenstein and Vannucci [21]. Three receiver settings will be evaluated, all using quadratic IF-detection. The first is \( m=1 \leftrightarrow \rho=80 \), which is a reference case for comparison with older analysis methods without post detection filtering. The other two are \( m=5 \leftrightarrow \rho=100 \) and \( m=8 \leftrightarrow \rho=112.2 \), which are the two cases evaluated extensively by Foschini, Greenstein and Vannucci.

**Table B.4-1:** Comparison of the accuracy of the different methods for calculating the threshold of single-filter receivers with quadratic IF-detection. Errors in \( \zeta \) are relative to the exact solution.

<table>
<thead>
<tr>
<th>method</th>
<th>( \zeta )</th>
<th>( \frac{b}{A^2} = \frac{\zeta^2}{4} )</th>
<th>( Y = \frac{b}{2N} )</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m=1 ), ( \rho=80 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exact</td>
<td>1.039095</td>
<td>0.26993</td>
<td>21.5944</td>
<td></td>
</tr>
<tr>
<td>analytical</td>
<td>1.039118</td>
<td>0.26994</td>
<td>21.5953</td>
<td>2.2 ( 10^{-5} )</td>
</tr>
<tr>
<td>iteration</td>
<td>1.039114</td>
<td>0.26994</td>
<td>21.5952</td>
<td>1.8 ( 10^{-5} )</td>
</tr>
<tr>
<td>Schwartz</td>
<td>1.024695</td>
<td>0.26250</td>
<td>21.0</td>
<td>-1.4 ( 10^{-2} )</td>
</tr>
<tr>
<td>( m=5 ), ( \rho=100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exact</td>
<td>1.164446</td>
<td>0.33898</td>
<td>33.8983</td>
<td></td>
</tr>
<tr>
<td>analytical</td>
<td>1.164484</td>
<td>0.33901</td>
<td>33.9006</td>
<td>3.2 ( 10^{-5} )</td>
</tr>
<tr>
<td>iteration</td>
<td>1.163736</td>
<td>0.33857</td>
<td>33.8571</td>
<td>-6.1 ( 10^{-4} )</td>
</tr>
<tr>
<td>Foschini</td>
<td>1.1313</td>
<td>0.320</td>
<td>32.0</td>
<td>-2.9 ( 10^{-2} )</td>
</tr>
<tr>
<td>( m=8 ), ( \rho=112.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exact</td>
<td>1.219340</td>
<td>0.37170</td>
<td>41.7045</td>
<td></td>
</tr>
<tr>
<td>analytical</td>
<td>1.219239</td>
<td>0.37164</td>
<td>41.6975</td>
<td>-8.3 ( 10^{-5} )</td>
</tr>
<tr>
<td>iteration</td>
<td>1.235997</td>
<td>0.38192</td>
<td>42.8517</td>
<td>1.4 ( 10^{-2} )</td>
</tr>
<tr>
<td>Foschini</td>
<td>1.1781</td>
<td>0.347</td>
<td>38.9</td>
<td>-3.5 ( 10^{-2} )</td>
</tr>
</tbody>
</table>

It is clear that the analytical approximation gives very good results, with an accuracy better than \( 10^{-4} \) up to \( m=8 \). The main sources of inaccuracies are the approximations of the Bessel function and the logarithm of the gamma-function. If required one can use more of the remainders-terms in the series approximations to obtain an even higher accuracy.

The other methods, either the iteration or the results obtained by Schwartz et al. and Foschini et al., are less accurate. The error of the iteration method rapidly increases with increasing \( m \), which is caused by the exponential approximation of the
modified Bessel function. The theoretical formula of Schwartz, Bennett and Stein and the numerical results of Foschini, Greenstein and Vannucci are both inaccurate by 1% or more.
Appendix C

BER of single-filter receiver

C.1 The partial Mark error rate for $m=1$

The generalised expression for $P_M$ with linear and quadratic IF-detection - using the
generalised threshold $Y$ - is given by:

$$P_M = 2 \exp(-\rho) \int_0^{\sqrt{Y}} y \exp(-y^2) I_0(2y\sqrt{\rho}) dy$$

(C.1-1)

Panter [87, p.708] has indicated that this integral can be solved through the method
of successive integration by parts, using the standard integral (Ab.11.3.25):

$$\int_0^\beta z^n I_{n-1}(az) dz = \frac{\beta^n I_n(a\beta)}{a}$$

(C.1-2)

In the definition of integration by parts (Ab.3.3.12)

$$\int f(y)g(y)dy = [f(y)G(y)] - \int f'(y)G(y)dy$$

(C.1-3)

we can now use $f(y) = \exp(-y^2)$ and $g(y) = yI_0(2y\sqrt{\rho})$, yielding:

$$\int_0^{\sqrt{Y}} \exp(-y^2) \cdot yI_0(2y\sqrt{\rho}) dy = \left[\exp(-y^2) \cdot \frac{y}{2\sqrt{\rho}} I_1(2y\sqrt{\rho})\right]_0^{\sqrt{Y}} +$$

$$+ \frac{1}{\sqrt{\rho}} \int_0^{\sqrt{Y}} \exp(-y^2) \cdot y^2 I_1(2y\sqrt{\rho}) dy$$

(C.1-4)

The latter integral is of the same form as the original one, but one order higher.
Successive integration of these integrals results in the following combination of a sum
and a remainder in the form of a rest-integral:

\[
\frac{1}{2} \exp(-Y) \sum_{k=1}^{K} \left( \frac{Y}{\rho} \right)^{\frac{k}{2}} I_k(2\sqrt{\rho Y}) + \\
+ \left( \frac{1}{\rho} \right)^{\frac{K}{2}} \int_{0}^{\sqrt{Y}} \exp(-y^2) \cdot y^{K+1} \cdot I_K(2y\sqrt{\rho})dy
\]  

(C.1-5)

The modified Bessel function with large order \( K \) within the rest-integral can be simplified by taking only the first term of its series expansion, being \((y\sqrt{\rho})^K / k!\) (Ab.9.6.10). The remainder due to the rest-integral can now be written as

\[
O(K) = \frac{1}{K!} \int_{0}^{\sqrt{Y}} \exp(-y^2) y^{2K+1} dy = \frac{1}{2K!} \int_{0}^{Y} \exp(-z) z^K dz = \frac{1}{2K!} \gamma(K+1,Y)
\]  

(C.1-6)

For \( K \) going to infinity \( \gamma(K+1,Y) \) approaches zero, which means that the final expression for \( P_M \) now becomes:

\[
P_M = \exp(-\rho - Y) \sum_{k=1}^{\infty} \left( \frac{Y}{\rho} \right)^{\frac{k}{2}} I_k(2\sqrt{\rho Y})  \\
= \exp(-\rho - Y) \sum_{k=0}^{\infty} \left( \frac{Y}{\rho} \right)^{\frac{k+1}{2}} I_{k+1}(2\sqrt{\rho Y})
\]  

(C.1-7)

This is the same expression as in Panter [87, p.708] except for the typing error in his result. (The term \( A^2 \) within the argument of the modified Bessel function should be \( A \)). The reason for using the second notation will become clear when the general expression for all \( m \) is derived in the next section.

C.2 Partial error rates for quadratic IF-detection

The Space partial error rate

The general expression in \( Y \) of \( P_S \) is

\[
P_S = \int_{Y}^{\infty} \frac{1}{\Gamma(m)} y^{m-1} \exp(-y) dy
\]  

(C.2-1)
With the standard integral (Gr.3.351-2)

$$
\int_a^\infty y^n \exp(-by)dy = \exp(-ab) \sum_{k=0}^{n} \frac{n!}{k!} \frac{a^k}{b^{n-k+1}}
$$

(C.2-2)

this yields

$$
P_S = \exp(-Y) \sum_{k=0}^{m-1} \frac{Y^k}{k!}
$$

(C.2-3)

$$
= \exp(-Y) \left\{ 1 + Y + \frac{Y^2}{2} + \frac{Y^3}{6} + \frac{Y^4}{24} + \frac{Y^5}{120} + \ldots + \frac{Y^{m-1}}{\Gamma(m)} \right\}
$$

(C.2-4)

It would be useful to have an expression without the gamma functions, enabling rapid calculation. This can be obtained by representing the incomplete gamma function as a continued fraction (Gr.8.358):

$$
\Gamma(m, Y) = \exp(-y)Y^m \cdot \frac{1}{y + \frac{1}{1 + \frac{2 - m}{y + \frac{2}{1 + \ldots}}}}
$$

(C.2-5)

When only the first fraction is used we can write $P_S$ as

$$
P_S = \frac{Y^m \exp(-Y)}{\Gamma(m)(Y - m + 1)}
$$

(C.2-6)

The gamma function can finally be eliminated with the help of Stirlings formula (Ab.6.1.37):

$$
\Gamma(z) = \exp(-z)z^{z-\frac{1}{2}}\sqrt{2\pi} \cdot O \left( 1 + \frac{1}{12z} \right)
$$

(C.2-7)

This yields

$$
P_S = \frac{e\sqrt{m/2\pi}}{2\pi} \left( \frac{Y}{m} \right)^m \frac{\exp[-(Y - m + 1)]}{(Y - m + 1)}
$$

(C.2-8)

For $m=1$ this expression reduces to

$$
P_S = \frac{e}{\sqrt{2\pi}} \exp(-Y) = 1.08 \exp(-Y)
$$

(C.2-9)
The inaccuracy in the error-rate (!) is thus 8% for all \( Y \), which means that, due to the steepness of the BER-curve, the inaccuracy in \( Y \) and the SNR will be very low. For increasing \( m \) the inaccuracy reduces rapidly; for example at \( m=6 \) and \( \rho=94 \) the exact and approximate solutions are 8.83 \( 10^{-10} \) and 9.01 \( 10^{-10} \) respectively, yielding a relative error in the error rate of 2.0%.

**The Mark partial error rate**

The generalised expression for \( P_M \) of a single filter receiver with quadratic IF detection and bandwidth expansion \( m \) is given by

\[
P_M = \int_0^Y \left( \frac{y}{\rho} \right)^{m-1} \frac{\text{exp}(-y - \rho)}{\sqrt{2\rho}} I_{m-1}(2\sqrt{\rho}y) \, dy
\]  
\[\text{(C.2-10)}\]

This is the same result as given by Shnidman [102] and Proakis [92]. In fact the expression above can be regarded as the definition of the generalised Marcum's Q-function, so \( P_M \) can be written as

\[
P_M = 1 - Q_m(\sqrt{2\rho}, \sqrt{2Y})
\]  
\[\text{(C.2-11)}\]

\[
Q_m(\alpha, \beta) \overset{\Delta}{=} \int_{\beta}^{\infty} \left( \frac{z}{\alpha} \right)^{m-1} \text{exp} \left( -\frac{z^2 + \alpha^2}{2} \right) I_{m-1}(\alpha z) \, dz
\]  
\[\text{(C.2-12)}\]

\[
Q_m(\sqrt{2\alpha}, \sqrt{2\beta}) \overset{\Delta}{=} \int_{\beta}^{\infty} \left( \frac{z}{\alpha} \right)^{m-1} \text{exp}(-z - \alpha) I_{m-1}(2\sqrt{\alpha z}) \, dz
\]  
\[\text{(C.2-13)}\]

After a change of variable, a form of \( P_M \) more suited for analytical evaluation is obtained:

\[
P_M = 2\rho^\frac{1-m}{2} \text{exp}(-\rho) \cdot \int_0^{\sqrt{Y}} \text{exp}(-y^2) \cdot y^m I_{m-1}(2y\sqrt{\rho}) \, dy
\]  
\[\text{(C.2-14)}\]

The same method as used in the previous section for \( m=1 \) can be applied. With (C.1-2) and (C.1-3) the integral can be written as

\[
\int_0^{\sqrt{Y}} \text{exp}(-y^2) \cdot y^m I_{m-1}(2y\sqrt{\rho}) \, dy = \left[ \text{exp}(-y^2) \frac{y^m}{2\sqrt{\rho}} I_m(2y\sqrt{\rho}) \right]^{\sqrt{Y}}_0 + \\
+ \frac{1}{\sqrt{\rho}} \int_0^{\sqrt{Y}} \text{exp}(-y^2) \cdot y^{m+1} I_m(2y\sqrt{\rho}) \, dy
\]  
\[\text{(C.2-15)}\]

The resulting integral is of the same form, but one order of \( m \) higher. Successive integration leads to a sum plus remainder-integral \( O(K) \):

\[
P_M = \text{exp}(-\rho - Y) \sum_{k=0}^{K} \left( \frac{Y}{\rho} \right)^{\frac{k+m}{2}} I_{k+m}(2\sqrt{\rho Y}) + O(K)
\]  
\[\text{(C.2-16)}\]
C.2. PARTIAL ERROR RATES FOR QUADRATIC IF-DETECTION

\[
O(K) = 2 \exp(-\rho) \left( \frac{1}{\rho} \right)^{K+m} \cdot \int_0^{\sqrt{Y}} \exp(-y^2)y^{K+m+1}I_{K+m}(2y\sqrt{\rho})dy
\]  
(C.2-17)

The remainder can be simplified for large \( K \) by using only the first term of the series expansion of the Bessel function, \((y\sqrt{\rho})^{K+m}/(K+m)!\), yielding

\[
O(K) = 2 \exp(-\rho) \left( \frac{1}{(K+m)!} \right) \int_0^{\sqrt{Y}} \exp(-y^2)y^{2K+2m+1}dy
\]

\[
= \exp(-\rho) \left( \frac{1}{(K+m)!} \right) \int_0^{Y} \exp(-u)u^{K+m}du
\]

\[
= \exp(-\rho) \frac{\gamma(K+m,Y)}{(K+m)!}
\]
(C.2-18)

For large \( K \) the incomplete gamma-function approaches zero, which means that the remainder can be neglected. The final expression for \( P_M \) becomes:

\[
P_M = \exp(-\rho - Y) \sum_{k=0}^{\infty} \left( \frac{Y}{\rho} \right)^{k+m} I_{k+m}(2\sqrt{\rho Y})
\]
(C.2-19)

Shnidman [102, (12)] also gives this formula as one of the elementary forms of the generalised Marcum Q-function.

Alternative notations

Using the series expansion of the modified Bessel functions(B.1-6) the expression for \( P_M \) can be rewritten into

\[
P_M = \exp(-\rho - Y) \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{Y^{m+k+j} \rho^j}{(m+k+j)! j!}
\]
(C.2-20)

\[
= \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{\infty} \frac{Y^{m+k+j}}{(m+k+j)!}
\]

The latter sum can be transformed in the following way:

\[
\sum_{k=0}^{\infty} \frac{Y^{m+k+j}}{(m+k+j)!} = \sum_{k=m+j}^{\infty} \frac{Y^k}{k!} = \sum_{k=0}^{\infty} \frac{Y^k}{k!} - \sum_{k=0}^{m+j-1} \frac{Y^k}{k!}
\]

\[
= \exp(Y) - \sum_{k=0}^{m+j-1} \frac{Y^k}{k!}
\]
(C.2-21)
This yields

\[ P_M = 1 - \exp(-\rho - Y) \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{m+j-1} \frac{Y^k}{k!} \quad (C.2-22) \]

This is also one of the elementary forms of \( P_M \) given by Shnidman [102, (8)], like (C.2-19). The last sum in (C.2-19) is related to the definition of the incomplete gamma-function (A.3-25), leading to the more compact notation

\[ P_M = 1 - \exp(-\rho) \sum_{j=0}^{\infty} \frac{\rho^j \Gamma(m+j,Y)}{j! \Gamma(m+j)} \quad (C.2-23) \]

Formula (C.2-22) can now be rewritten using the relation

\[
\sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{m+j-1} \frac{Y^k}{k!} = \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{l=0}^{m-1} \frac{Y^l}{l!} + \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=m}^{\infty} \frac{Y^k}{k!} \\
= \exp(\rho) \sum_{l=0}^{m-1} \frac{Y^l}{l!} + \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k + m - 1)!} \quad (C.2-24)
\]

This leads to the expression

\[ P_M = 1 - \exp(-Y) \sum_{l=0}^{m-1} \frac{Y^l}{l!} \\
- \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k + m - 1)!} \quad (C.2-25) \]

Since the first term is identical to \( P_S \) the following expression is derived:

\[ P_M = 1 - P_S - \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k + m - 1)!} \quad (C.2-26) \]

It can easily be checked that for \( \rho = 0 \) \( P_M \) reduces to \( 1 - P_S \) as expected. For \( m \) going to infinity \( P_S \) approaches 1, while the last sum in (C.2-26) goes to zero, leading to a \( P_M \) approaching 0. Finally, for \( Y = 0 \), the last term goes to zero, while \( P_S \) becomes again 1. So, with the threshold at 0 \( P_M \) becomes indeed 0.
C.3 The BER for quadratic IF-detection

With (C.2-26), and in the case of equal a priori probabilities for Mark and Space, the overall BER is given by

\[ P_e = \frac{1}{2} \left\{ 1 - \exp(-\rho - Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k+m-1)!} \right\} \tag{C.3-1} \]

The validity of this result can be checked by differentiation to \( Y \):

\[ \frac{dP_e}{dY} = \exp(-Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=1}^{j} \frac{Y^{k+m-1}}{(k+m-1)!} \]

\[ - \exp(-Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \sum_{k=0}^{j-1} \frac{Y^{k+m-1}}{(k+m-1)!} \]

\[ = \exp(-Y) \sum_{j=1}^{\infty} \frac{\rho^j}{j!} \left\{ \frac{Y^{j+m-1}}{(j+m-1)!} - \frac{Y^{m-1}}{(m-1)!} \right\} = 0 \tag{C.3-2} \]

Since the term between brackets is equal to 0 for \( j=0 \), the sum can be rewritten into

\[ \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \left\{ \frac{Y^{j+m-1}}{(j+m-1)!} - \frac{Y^{m-1}}{(m-1)!} \right\} = 0 \]

\[ \sum_{j=0}^{\infty} \frac{(\rho Y)^j}{j! \Gamma(m+j)} = \frac{\exp(\rho)}{\Gamma(m)} \]

\[ \sum_{j=0}^{\infty} \frac{(\sqrt{\rho Y})^{m-1+2j}}{j! \Gamma(m+j)} = \frac{\exp(\rho)}{\Gamma(m)} (\rho Y)^{\frac{m-1}{2}} \]

\[ I_{m-1}(2 \sqrt{\rho Y}) = \frac{\exp(\rho)}{\Gamma(m)} (\rho Y)^{\frac{m-1}{2}} \tag{C.3-3} \]

This is the threshold equation (4-79) and (B.1-1), which proves that for the optimum threshold the overall BER is indeed minimal.
C.4 Linear IF-detection

Pdf for a transmitted Space and $m=2$

With $p_x(u|2)$ given by (4-9) the convolution for $m=2$ is given by

$$p(u) = \frac{1}{N^2} \int_0^u y(u-y) \exp \left( -\frac{y^2 - (u-y)^2}{2N} \right) dy \quad (C.4-1)$$

or

$$p(u) = \frac{1}{N^2} \int_0^u y(u-y) \exp \left( -\frac{(y-u/2)^2}{N} \right) dy \quad (C.4-2)$$

After the transformation $y := y - u/2$ we obtain

$$\frac{1}{N^2} \int_{-u/2}^{u/2} \left( \frac{u^2}{4} - y^2 \right) \exp \left( -\frac{y^2}{N} \right) dy \quad (C.4-3)$$

This can be split into two separate integrals, the first of which can be solved directly with (Gr.3.321-2):

$$\int_0^u \exp(-q^2 x^2) dx = \frac{\sqrt{\pi}}{2q} \Phi(q) \quad (C.4-4)$$

$\Phi(q)$ is the error function defined in Appendix A.2. This yields for the first integral

$$\frac{1}{N^2} \exp \left( -\frac{u^2}{4} \right) \cdot \frac{u^2 \sqrt{\pi N}}{4} \Phi \left( \frac{u}{2\sqrt{N}} \right) \quad (C.4-5)$$

The second integral,

$$\frac{2}{N^2} \exp \left( -\frac{u^2}{4} \right) \int_0^{u/2} y^2 \exp \left( -\frac{y^2}{N} \right) dy$$

must be transformed with $z := y^2$, yielding the integral

$$\int_0^{u^2/4} \sqrt{z} \exp \left( -\frac{z}{N} \right) dz \quad (C.4-6)$$

This can be solved with (Gr.3.381-1):

$$\int_0^u x^{\nu-1} \exp(-\mu x) dx = \mu^{-\nu} \gamma(\nu, \mu u) \quad (\text{Re } \nu > 0) \quad (C.4-7)$$

which yields

$$\frac{2}{N^2} \exp \left( -\frac{U^2}{4} \right) \cdot \frac{N^{3/2}}{\gamma \left( \frac{3}{2}, \frac{u^2}{4N} \right)} \quad (C.4-8)$$
Finally, the incomplete gamma function can be reduced using (Gr.8.356-1) into two components:

\[
\gamma \left( \frac{3}{2}, \frac{u^2}{4N} \right) = \frac{1}{2} \gamma \left( \frac{1}{2}, \frac{u^2}{4N} \right) - \sqrt{\frac{u^2}{4N}} \exp \left( - \frac{u^2}{4N} \right)
\]  

(C.4-9)

Knowing that \( \gamma \left( \frac{1}{2}, x^2 \right) = \sqrt{\pi} \Phi(x) \) (Gr.8.359-4) the pdf of the convolution of two central chi pdf's with two degrees of freedom is given by

\[
p(u) = \frac{u}{2N} \exp \left( - \frac{u^2}{2N} \right) + \\
\frac{1}{2} \sqrt{\frac{\pi}{2N}} \exp \left( - \frac{u^2}{2N} \right) \left\{ \frac{u^2}{N} - 1 \right\} \Phi \left( \frac{u}{\sqrt{2N}} \right)
\]  

(C.4-10)

By integrating this pdf from 0 to infinity it can be proven that the normalisation of \( p(u) \) gives exactly 1. However, as presented here, \( p(u) \) is not yet the pdf of the I&D post-detection filter output. To obtain this pdf the weighting \( 1/m \) of the filter, see the definition (3-71), must be included in the pdf. The weighting function reduces the total post-detection output signal, and thus influences both signal and noise. Due to the correlated signals and uncorrelated noise the SNR \( \rho \) should be reduced by \( m \), and \( u \) then becomes \( u/\sqrt{2} \). This yields the final pdf of the post-detection filter output of a single-filter receiver with linear IF-detection and a transmitted Space:

\[
p_{11}(u|4) = \frac{u}{N} \exp \left( - \frac{u^2}{N} \right) + \\
\sqrt{\frac{\pi}{2N}} \exp \left( - \frac{u^2}{2N} \right) \left\{ \frac{u^2}{N} - 1 \right\} \Phi \left( \frac{u}{\sqrt{2N}} \right)
\]  

(C.4-11)

The partial error rate for a Space and \( m=2 \)

The partial error rate \( P_S \) for \( m=2 \), with linear IF-detection and a transmitted Space consists of three separate integrals, obtained from (C.4-11). The first one is

\[
\int_{b}^{\infty} \frac{u}{N} \exp \left( - \frac{u^2}{N} \right) du = \frac{1}{2N} \int_{b}^{\infty} \exp \left( - \frac{z}{N} \right) dz
\]  

(C.4-12)

which yields straightforwardly \( \frac{1}{2} \exp(-2Y) \), with \( Y \) again equal to the normalised threshold \( b_{in}^2/2N \). The second integral is

\[
\frac{1}{N} \sqrt{\frac{\pi}{2N}} \int_{b}^{\infty} \frac{u^2}{2N} \exp \left( - \frac{u^2}{2N} \right) \Phi \left( \frac{u}{\sqrt{2N}} \right) du
\]  

\[
= \frac{1}{2N} \sqrt{\frac{\pi}{2N}} \int_{b}^{\infty} \sqrt{z} \exp \left( - \frac{z}{2N} \right) \Phi \left( \sqrt{\frac{z}{2N}} \right) dz
\]  

(C.4-13)
This integral can not be solved directly. However, since we are only integrating over the tail of \( p(u) \) it is valid to assume that the error function \( \Phi(\cdot) \) can be approximated by 1. The integral can then be solved with (Gr.3.381-3)

\[
\int_0^\infty x^{\nu-1} \exp(-\mu x) \, dx = \mu^{-\nu} \Gamma(\nu, \mu u) \quad (\Re \nu > 0)
\] (C.4-14)

This yields for the integral \((2N)^{3/2} \Gamma(\frac{3}{2}, Y)\), which can be reduced with (Gr.8.356-2) and (Gr.8.359-3) in the same way as done for the pdf into \((2N)^{3/2} [\sqrt{\pi} - \sqrt{\pi} \Phi(\sqrt{Y}) + \sqrt{Y} \exp(-Y)]\). This results in the complete solution of the second integral (C.4-13) as

\[
\frac{\pi}{4} - \frac{\pi}{2} \Phi(\sqrt{Y}) + \sqrt{2\pi Y} \exp(-Y)
\] (C.4-15)

The third integral, which also contains an error function that must be approximated to 1, needs to be rewritten as

\[
\sqrt{\frac{\pi}{2N}} \int_0^b \exp \left( -\frac{u^2}{2N} \right) \, du - \sqrt{\frac{\pi}{2N}} \int_0^\infty \exp \left( -\frac{u^2}{2N} \right) \, du
\] (C.4-16)

The first integral can be solved with (Gr.8.250-1) - the definition of the error function - or (Gr.3.321-2), and yields \( \sqrt{\pi N/2} \Phi(\sqrt{Y}) \). The second integral simply gives \( \sqrt{\pi N/2} \). The complete third integral thus yields

\[
\frac{\pi}{2} \Phi(\sqrt{Y}) - \frac{\pi}{2}
\] (C.4-17)

Combining all three partial integrals yields the expression for \( P_S \):

\[
P_S = \frac{1}{2} \exp(-2Y) + \sqrt{2\pi Y} \exp(-Y)
\] (C.4-18)
Appendix D

BER of dual-filter receiver

D.1 Probability density functions

Quadratic IF-detection

The negative tail of $p(u)$ can be found with

$$p(u) = \int_0^\infty p_1(y|2m, \lambda) p_0(y - u|2m) \, dy \quad (u \leq 0) \quad (D.1-1)$$

The pdf's $p(u|2m, \lambda)$ and $p(u|2m)$ are given by (4-70) and (4-74), so the pdf of the IF-detector output can be written as:

$$p(u) = \left(\frac{1}{2N}\right)^2 \left(\frac{1}{2AN}\right)^{m-1} \frac{1}{\Gamma(m)} \exp \left( -\frac{A^2}{2N} + \frac{u}{2N} \right) \cdot \int_0^\infty y^{\frac{m-1}{2}} (y - u)^{m-1} \exp \left( -\frac{y}{N} \right) I_{m-1} \left( \frac{A}{N\sqrt{y}} \right) \, dy \quad (D.1-2)$$

This integral cannot be found in either Gradsteyn and Ryznik [30] or Abramowitz and Stegun [1], so it cannot be solved directly. The term that poses most problems for directly solving the integral is $(y-u)^{m-1}$. We can use e.g. the binomial expansion (Gr.1.111) of the $(y-u)$-term

$$(y-u)^{m-1} = \sum_{k=0}^{m-1} \binom{m-1}{k} y^k (-u)^{m-k-1} \quad (D.1-3)$$

This yields a sum of integrals:
\[ p(u) = \left( \frac{1}{2N} \right)^2 \left( \frac{1}{2AN} \right)^{m-1} \frac{1}{\Gamma(m)} \exp \left( -\frac{A^2}{2N} + \frac{u}{2N} \right) \]
\[ \cdot \sum_{k=0}^{m-1} \binom{m-1}{k} (-u)^{m-k-1} \]
\[ \cdot \int_0^\infty y^{m/2+k-1/2} \exp \left( -\frac{y}{N} \right) I_{m-1} \left( \frac{A}{N \sqrt{y}} \right) dy \]  

(D.1-4)

The integrals in (D.1-4) can be solved, although it takes several steps to arrive at the proper result. First, using (Gr.8.406) the modified Bessel function is converted in a normal Bessel function through \( I_{m-1}(x) = (-j)^{m-1} J_{m-1}(jx) \). Using (Gr.6.634-4) the standard integral (A.3-8) is obtained, yielding for the integrals

\[ \int_0^\infty y^{m/2+k-1/2} \exp \left( -\frac{y}{N} \right) I_{m-1} \left( \frac{A}{N \sqrt{y}} \right) dy = \]
\[ = k! \left( \frac{A}{2} \right)^m N^{k+1} \exp \left( -\frac{A^2}{4N} \right) L_k^m \left( -\frac{A^2}{4N} \right) \]  

(D.1-5)

(D.1-6)

\( L_k^{m-1}(x) \) is a generalised Laguerre polynomial defined by (A.3-9). As in Appendix A.3 the Laguerre polynomial is not orthogonal due to its negative argument, although the definition remains valid. Note by the way the relation with the moments of a non-central chi-square distribution; \( p(u) \) is an exponential pdf multiplied with a series of weighted moments around zero. Combining (D.1-4) with (D.1-6), and writing out the combination yields the expression for \( p(u) \):

\[ p(u) = \frac{1}{4N} \left( -\frac{u}{4N} \right)^{m-1} \exp \left( -\frac{A^2}{4N} + \frac{u}{2N} \right) \cdot \]
\[ \cdot \sum_{k=0}^{m-1} \frac{1}{\Gamma(m-k)} \left( -\frac{N}{u} \right)^k L_k^{m-1} \left( -\frac{A^2}{4N} \right) \]  

(u \leq 0, \ Mark)  

(D.1-7)

For a transmitted Space \( u \) should be replaced by \(-u\). For \( m=1 \) and \( u = z \) (no post-detection filtering) (D.1-7) reduces to

\[ p(z) = \frac{1}{4N} \exp \left( -\frac{A^2}{2N} + \frac{z}{2N} \right) \quad (z \leq 0) \]  

(D.1-8)

This pdf is illustrated in Figure D.1-1. It shows that for \( z \leq 0 \) the approximated pdf (D.1-8) is indeed identical to the exact solution, calculated with Mathematica\(^1\)

---

\(^1\)Stephen Wolfram, Mathematica, A system for doing mathematics by computer, Addison-Wesley, Redwood City (CA), USA, 1988.
by performing the integration (D.1-2) numerically. For \( z > 0 \) the pdf (D.1-8) may certainly not be used, since the exact solution features a discontinuity in the first derivative.

The expression (D.1-7) is exact, but difficult to analyse in this form. An approximation leading to a more compact expression can be made using (A.4-8) and (A.4-12) for large SNR:

\[
p(u) = \left( \frac{1}{4N} \right)^{m-1} \exp \left( -\frac{A^2}{4N} + \frac{u}{2N} \right) \cdot \\
\sum_{k=0}^{m-1} \binom{m-1}{k} (-u)^{m-k-1} \left( \frac{A}{2} \right)^{2k}
\]

D.1-9

The sum is exactly the binomial expansion of \((A^2/4 - u)^{m-1}\), so the approximated pdf becomes

\[
p(u) = \left( \frac{1}{4N} \right)^{m-1} \exp \left( -\frac{A^2}{4N} + \frac{u}{2N} \right) \left( \frac{A^2}{4} - u \right)^{m-1}
\]

D.1-10

\((u \leq 0 \ , \ SNR \gg 1 \ , \ Mark)\)

Although this approximation is exact for \( m = 1 \), it gives deviations for larger values of \( m \) and should therefore not be used for evaluation of the BER.
The pdf of the NxN noise

Setting $A$ to 0 in $p(u)$ yields the pdf of the dual-filter FSK receiver with only noise at the input. Using (A.3-28) gives

$$p(u) = \left( \frac{1}{4N} \right)^m (-u)^{m-1} \frac{1}{\Gamma(m)} \exp \left( \frac{u}{2N} \right) \cdot \sum_{k=0}^{m-1} \frac{\Gamma(m+k)}{k! \Gamma(m-k)} \left( -\frac{N}{u} \right)^k$$

(D.1-11)

The sum is related to the definition of the spherical Bessel functions of the third kind and order integer plus one half (Gr.8.468):

$$K_{n+1/2}(z) = \sqrt{\frac{\pi}{2z}} \exp(-z) \sum_{k=0}^{n} \frac{(n+k)!}{k!(n-k)!} (2z)^{-k}$$

(D.1-12)

This yields

$$p(u) = \frac{1}{\sqrt{\pi N}} \frac{1}{\Gamma(m)} \left( \frac{1}{4N} \right)^m (-u)^{m-1/2} K_{m-1/2} \left( -\frac{u}{2N} \right)$$

(u ≤ 0)

(D.1-13)

Since the pdf must be symmetrical around $u=0$, the total pdf becomes

$$p(u) = \frac{1}{\sqrt{\pi N}} \frac{1}{\Gamma(m)} \left( \frac{1}{4N} \right)^m |u|^{m-1/2} K_{m-1/2} \left( \frac{|u|}{2N} \right)$$

(D.1-15)

For $m=1$ this reduces to

$$p(u) = \frac{1}{4N} \exp \left( -\frac{|u|}{2N} \right)$$

(D.1-16)

The variance of the NxN noise is calculated with

$$\text{var}(u) = 2 \cdot \frac{1}{\sqrt{\pi N}} \frac{1}{\Gamma(m)} \left( \frac{1}{4N} \right)^m \int_0^\infty y^{2m-1/2} K_{m-1/2} \left( \frac{y}{2N} \right) dy$$

(D.1-17)

With (Gr.6.561-16)

$$\int_0^\infty y^\mu K_\nu(ay)dy = \left( \frac{a}{\alpha} \right)^{\mu+1} \Gamma \left( \frac{1+\mu+\nu}{2} \right) \Gamma \left( \frac{1+\mu-\nu}{2} \right)$$

(D.1-18)

the resulting expression for the variance becomes

$$\text{var}(u) = \frac{8}{\Gamma(m+1)} N^2 = 8mN^2$$

(D.1-19)

This is exactly two times the NxN variance of a single filter output (4m^2N^2) divided by the post-detection filter improvement factor $m$. 

D.2 Bit-error rate

Quadratic IF-detection

Since the contributions of $P_S$ and $P_M$ are identical, calculating one of the two suffices. By replacing $-u$ by $u$ in (D.1-7) the positive tail of the Space pdf is used.

\[
P_S = \left(\frac{1}{4N}\right)^m \exp\left(-\frac{A^2}{4N}\right) \sum_{k=0}^{m-1} \frac{1}{\Gamma(m+k)} N^{k} L_k^{m-1} \left( -\frac{A^2}{4N} \right) \cdot \int_0^\infty u^{m-k-1} \exp\left(-\frac{u}{2N}\right) \, du
\]

(D.2-1)

These integrals can be solved with (Gr.3.351-3)

\[
\int_0^\infty u^n \exp(-\alpha u) \, du = n! \alpha^{-n-1}
\]

(D.2-2)

and, since $P_e = 0.5P_S + 0.5P_M \equiv P_S$, yields

\[
P_e = \left(\frac{1}{2}\right)^m \exp\left(-\frac{A^2}{4N}\right) \sum_{k=0}^{m-1} \left(\frac{1}{2}\right)^k L_k^{m-1} \left( -\frac{A^2}{4N} \right)
\]

(D.2-3)

\[
= \left(\frac{1}{2}\right)^m \exp\left(-\frac{A^2}{4N}\right) \sum_{k=0}^{m-1} \left(\frac{1}{2}\right)^k \sum_{j=0}^{k} \left( \begin{array}{c} m+k-1 \\ k-j \end{array} \right) \frac{1}{j!} \left(\frac{A^2}{4N}\right)^j
\]

(D.2-4)
## Appendix E

### List of symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>unit (dimension)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a(t)$</td>
<td>amplitude of information bearing IF Signal&lt;sup&gt;1&lt;/sup&gt;</td>
<td>V</td>
</tr>
<tr>
<td>$A$</td>
<td>maximum amplitude of information bearing IF Signal</td>
<td>V</td>
</tr>
<tr>
<td>$b$</td>
<td>decision threshold</td>
<td>V</td>
</tr>
<tr>
<td>$B$</td>
<td>one-sided bandwidth of IF filter</td>
<td>Hz</td>
</tr>
<tr>
<td>$C(m)$</td>
<td>BER deterioration factor</td>
<td></td>
</tr>
<tr>
<td>$C(m</td>
<td>\delta)$</td>
<td>BER deterioration factor</td>
</tr>
<tr>
<td>$d(t)$</td>
<td>delayed IF signal</td>
<td>V</td>
</tr>
<tr>
<td>$D$</td>
<td>electrical field responsivity</td>
<td>$\text{m}^2/\text{V}\Omega$</td>
</tr>
<tr>
<td>$\Delta D_{\text{iv}}$</td>
<td>diversity penalty</td>
<td>dB</td>
</tr>
<tr>
<td>$E$</td>
<td>electrical field amplitude</td>
<td>V/m</td>
</tr>
<tr>
<td>$\mathbf{E}$</td>
<td>electrical field vector</td>
<td>V/m (2)</td>
</tr>
<tr>
<td>$\mathbf{E}$</td>
<td>electrical field amplitude vector</td>
<td>V/m (2)</td>
</tr>
<tr>
<td>$E{\cdot}$, $&lt;\cdot&gt;$</td>
<td>mean value, expectation</td>
<td></td>
</tr>
<tr>
<td>$\text{erf}(\cdot)$</td>
<td>error function or probability integral</td>
<td></td>
</tr>
<tr>
<td>$\text{erfc}(\cdot)$</td>
<td>complementary error function</td>
<td></td>
</tr>
<tr>
<td>$f$</td>
<td>frequency</td>
<td>Hz</td>
</tr>
<tr>
<td>$f_d$</td>
<td>carrier-to-peak FSK frequency deviation</td>
<td>Hz</td>
</tr>
<tr>
<td>$f_m$</td>
<td>half period of frequency detection curve</td>
<td>Hz</td>
</tr>
<tr>
<td>$\mathcal{F}{\cdot}$</td>
<td>Fourier transform</td>
<td></td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>peak-to-peak FSK frequency deviation</td>
<td>Hz</td>
</tr>
<tr>
<td>$\Delta F_{\text{FSK}}$</td>
<td>FSK over ASK sensitivity improvement</td>
<td>dB</td>
</tr>
<tr>
<td>$G(f)$</td>
<td>two-sided power density spectrum</td>
<td>W/Hz</td>
</tr>
</tbody>
</table>

<sup>1</sup>Throughout the thesis 'Signal', i.e. capitalized, denotes only the information bearing signal component. When lower-case 'signal' is used this refers to Signal plus noise or noise only.

295
APPENDIX E. LIST OF SYMBOLS

\( h \)  
Planck's constant \( 6.63 \cdot 10^{-34} \) Js

\( h(t) \)  
impulse response

\( H \)  
magnetic field amplitude A/m

\( H(f) \)  
transfer function

\( H_{em}(\cdot) \)  
Hermite polynomial

\( I_n(\cdot) \)  
modified Bessel function of the first kind and order \( n \)

\( J_n(\cdot) \)  
Bessel function of the first kind and order \( n \)

\( k \)  
number of phase diversity branches

\( K_{kkk}(\theta) \)  
phase diversity amplitude factor

\( K_{kkk}(\theta, \kappa, \Delta\theta) \)  
phase diversity amplitude factor with imbalance

\( K_{n}(\cdot) \)  
spherical Bessel function of the third kind and order \( n \)

\( l(\cdot) \)  
likelihood ratio

\( l_0 \)  
threshold value of likelihood ratio dB

\( L_o(\cdot) \)  
optical coupler excess loss

\( L_n(z) \)  
generalized Laguerre polynomial of order \( n \)

\( L_n(z) \)  
Laguerre polynomial of order \( n \)

\( m \)  
IF bandwidth expansion factor

\( M \)  
FM modulation index

\( M_n(s) \)  
Laplace transform of \( p(u) \)

\( M_{n,b}(\cdot) \)  
Whittaker's function

\( M \)  
 optical carrier matrix \((2 \times 2)\)

\( n \)  
fibre refractive index

\( n_1(\cdot) \)  
one-sided noise spectral density in IF W/Hz

\( n(\cdot) \)  
noise component of IF voltage V

\( n(t) \)  
noise component of complex IF voltage \( V \) (2)

\( N \)  
IF noise variance in bandwidth \( 1/T \) W

\( N_m \)  
IF noise variance in bandwidth \( m/T \) W

\( O(\cdot) \)  
remainder in an approximation

\( p(\cdot) \)  
probability density function (pdf)

\( p(x \mid \vartheta) \)  
central \( \chi^2 \) pdf of variable \( x \)

\( p_x(z \mid \vartheta) \)  
central \( \chi^2 \) pdf of variable \( x \)

\( p_n(u \mid 2m) \)  
resulting pdf after \( m \)-fold convolution of \( p_x(z \mid 2) \)

\( p(x \mid \vartheta, \lambda) \)  
non-central \( \chi^2 \) pdf of variable \( x \)

\( p_{\chi^2}(x \mid \vartheta, \lambda) \)  
non-central \( \chi^2 \) pdf of variable \( x \)

\( p_n(u \mid 2m, \lambda) \)  
resulting pdf after \( m \)-fold convolution of \( p_{\chi^2}(z \mid 2, \lambda) \)

\( P \)  
optical power W

\( P(x) \)  
cumulative normal distribution function

\( P(x \mid \vartheta) \)  
cumulative central \( \chi^2 \) distribution function

\( P(x \mid \vartheta, \lambda) \)  
cumulative non-central \( \chi^2 \) distribution function

\( P_e \)  
total symbol error probability

\( Pr(X) \)  
probability of \( X \)

\( P(0) \)  
a priori SPACE symbol probability
\[
\begin{array}{ll}
P(1) & \text{a priori MARK symbol probability} \\
P_S & \text{a posteriori SPACE symbol error probability} \\
P_M & \text{a posteriori MARK symbol error probability} \\
q & \text{electron charge } 1.6 \times 10^{-19} \\
Q(x) & \text{complementary cumulative normal distribution function} \\
Q(x | \theta) & \text{complementary cumulative central } \chi^2 \text{ distribution function} \\
Q(x | \theta, \lambda) & \text{complementary cumulative non-central } \chi^2 \text{ distribution function} \\
Q(\alpha, \beta) & \text{Marcum's Q-function} \\
Q_m(\alpha, \beta) & \text{generalised Marcum's Q-function} \\
r(t), r & \text{amplitude, envelope of IF voltage} \\
\text{rect}(x) & \text{rectangular unit function non-zero from } -x/2 \text{ to } x/2 \\
R & \text{photodiode responsivity} \\
R(\tau) & \text{autocorrelation function} \\
\text{RIN} & \text{Relative Intensity Noise} \\
s & \text{complex frequency in Laplace transforms} \\
s(t) & \text{IF voltage} \\
s(x) & \text{complex IF voltage} \\
S(t) & \text{Signal component of IF voltage} \\
S_x(t) & \text{Signal component of complex IF voltage} \\
\text{sinc}(x) & \sin(x)/x \text{ function} \\
\text{SNR}_o & \text{Signal-to-Noise Ratio of detector output} \\
\Delta\text{SNR} & \text{sensitivity penalty} \\
t & \text{time} \\
\Delta t & \text{sample time interval} \\
T & \text{bittime of data } (1/T \text{ is the bitrate in bit/s}) \\
\text{trap}(x, y) & \text{trapezoidal function with topwidth } x \text{ and bottomwidth } y \\
u(t), u & \text{output voltage of post detection filter} \\
U_M & \text{decision region for Mark} \\
U_S & \text{decision region for Space} \\
\text{var}(\cdot) & \text{variance} \\
\nu(t), \nu & \text{in-phase component of delayed IF noise voltage} \\
\omega(t), \omega & \text{quadrature component of delayed IF noise voltage} \\
W_E(\omega) & \text{spectral density of electrical field} \\
W_P(\omega) & \text{spectral density of optical power} \\
W_\phi(\omega) & \text{spectral density of instantaneous frequency} \\
x(t), x & \text{in-phase component of IF noise voltage} \\
y(t), y & \text{quadrature component of IF noise voltage} \\
Y & \text{normalised general threshold} \\
z(t), z & \text{output voltage of IF detectors} \\
Z_0 & \text{free space impedance } (120\pi \approx 377) \\
\end{array}
\]
APPENDIX E. LIST OF SYMBOLS

\(\alpha\) detection improvement factor
\(\beta\) detection improvement factor
\(\gamma\) IF linewidth-to-bandwidth ratio \(\text{rad}\)
\(\gamma(n, x)\) complementary incomplete Gamma function
\(\Gamma(x)\) Gamma function
\(\Gamma(n, x)\) incomplete Gamma function
\(\delta\) detection efficiency
\(\delta(x)\) Dirac delta- (impulse) function
\(\Delta(x)\) two-sided triangle function between \(-x\) and \(x\)
\(\Delta'(x)\) one-sided triangle function between 0 and \(x\)
\(\zeta\) threshold deviation factor
\(\eta\) quantum efficiency of photodiode \(\%\)
\(\theta\) phase of a voltage or field \(\text{rad}\)
\(\vartheta\) degrees of freedom of \(\chi\) or \(\chi^2\) pdf \(\text{rad}\)
\(\kappa\) diversity IF SNR imbalance (voltage-based)
\(\lambda\) noncentrality parameter of \(\chi\) or \(\chi^2\) pdf
\(\mu\) first moment of a distribution
\(\mu'_n\) \(n^{th}\) moment of a distribution
\(\mu_n\) \(n^{th}\) central moment of a distribution
\(\Delta\nu\) laser linewidth \(\text{Hz}\)
\(\xi\) IF Signal-to-Noise Ratio
\(\Xi(t)\) data signal, \(\pm 1\)
\(\rho\) IF Signal-to-noise Ratio in bandwidth \(1/T\)
\(\sigma^2\) variance of a variable
\(\tau\) time delay \(\text{s}\)
\(\phi\) equivalent polarisation split angle \(\text{rad}\)
\(\phi_{ab}(\cdot)\) cross-correlation function \(\text{rad}\)
\(\varphi\) phase of an optical signal \(\text{rad}\)
\(\dot{\varphi}\) frequency of an optical signal \(\text{rad/s}\)
\(\Delta\varphi\) accumulated phase change in time interval \(\tau\) \(\text{rad}\)
\(\Phi\) polarisation misalignment angle \(\text{rad}\)
\(\Phi(x)\) alternative notation of the error-function \(\text{erf}(x)\)
\(\Phi(a, b, z)\) confluent hypergeometric function
\(\chi_\theta\) central chi variable
\(\chi'_\theta(\lambda)\) non-central chi variable
\(\chi^2_\theta\) central chi-square variable
\(\chi'^2_\theta(\lambda)\) non-central chi-square variable
\(\psi\) DPSK modulation depth phase error \(\text{rad}\)
\(\Psi\) CPFSK phase mismatch \(\text{rad}\)
\(\omega\) angular frequency \(\text{rad/s}\)
\(\omega_d\) CPFSK angular frequency deviation \(\text{rad/s}\)
**Frequently used subscripts:**

<table>
<thead>
<tr>
<th>Subscript</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td>central intermediate frequency</td>
</tr>
<tr>
<td>(div)</td>
<td>(polarisation) diversity</td>
</tr>
<tr>
<td>(eff)</td>
<td>effective</td>
</tr>
<tr>
<td>(eq)</td>
<td>equivalent</td>
</tr>
<tr>
<td>(i)</td>
<td>in-phase component of a complex phasor</td>
</tr>
<tr>
<td>(I)</td>
<td>in-phase IF-branch</td>
</tr>
<tr>
<td>(IF)</td>
<td>intermediate frequency signal or circuits</td>
</tr>
<tr>
<td>(j)</td>
<td>(j^{th}) sample by post-detection filter</td>
</tr>
<tr>
<td>(lo)</td>
<td>local oscillator</td>
</tr>
<tr>
<td>(lin)</td>
<td>linear IF-detection</td>
</tr>
<tr>
<td>(m)</td>
<td>function of the relative IF bandwidth</td>
</tr>
<tr>
<td>(M)</td>
<td>transmitted MARK signal</td>
</tr>
<tr>
<td>(n)</td>
<td>belonging to the total complex noise</td>
</tr>
<tr>
<td>(N x N)</td>
<td>Noise-times-Noise</td>
</tr>
<tr>
<td>(q)</td>
<td>quadrature component of a complex phasor</td>
</tr>
<tr>
<td>(qdr)</td>
<td>quadratic IF-detection</td>
</tr>
<tr>
<td>(Q)</td>
<td>quadrature IF-branch</td>
</tr>
<tr>
<td>(rec)</td>
<td>receiver</td>
</tr>
<tr>
<td>(s)</td>
<td>signal</td>
</tr>
<tr>
<td>(s.b.)</td>
<td>single IF-branch</td>
</tr>
<tr>
<td>(sum)</td>
<td>for summed output of diversity branches</td>
</tr>
<tr>
<td>(S)</td>
<td>transmitted SPACE signal</td>
</tr>
<tr>
<td>(S x N)</td>
<td>Signal-times-Noise</td>
</tr>
<tr>
<td>(S x S)</td>
<td>Signal-times-Signal</td>
</tr>
<tr>
<td>(0)</td>
<td>lower FSK intermediate frequency</td>
</tr>
<tr>
<td>(1)</td>
<td>upper FSK intermediate frequency</td>
</tr>
<tr>
<td>(\parallel)</td>
<td>for parallel input polarisation</td>
</tr>
<tr>
<td>(\perp)</td>
<td>for perpendicular input polarisation</td>
</tr>
</tbody>
</table>

**Two special types of references:**

(Gr.0.643-4) Gradsteyn and Ryznik [30] formula 4, section 6.643
APPENDIX E. LIST OF SYMBOLS
Appendix F

List of abbreviations

AGC  Automatic Gain Control
AFC  Automatic Frequency Control
AM   Amplitude Modulation
AMI  Alternate Mark Inversion
APC  Automatic Polarisation Control
APD  Avalanche Photo-Diode
ASK  Amplitude Shift Keying
AWGN Average White Gaussian Noise
BB   Broad-Band
BER  Bit Error Rate
BPF  Band Pass Filter
CID  Channel IDentification
CMC  Coherent Multi-Channel
CPFSK Continuous-Phase Frequency Shift Keying
DBR  Distributed Bragg Reflector (laser)
DFB  Distributed Feed-Back (laser)
DPSK Differential Phase Shift Keying
EDFA Erbium-Doped Fibre Amplifier
EXOR Exclusive OR
FDM  Frequency Division Multiplexing
FM   Frequency Modulation
FSK  Frequency Shift Keying
HEMT High Electron Mobility Transistor
HPF  High Pass Filter
IC   Integrated Circuit
I&D  Integrate and Dump
IF   Intermediate Frequency
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISI</td>
<td>Inter-Symbol Interference</td>
</tr>
<tr>
<td>LED</td>
<td>Light-Emitting Diode</td>
</tr>
<tr>
<td>LNA</td>
<td>Low-Noise Amplifier</td>
</tr>
<tr>
<td>LO</td>
<td>Local Oscillator (laser)</td>
</tr>
<tr>
<td>LPF</td>
<td>Low Pass Filter</td>
</tr>
<tr>
<td>MESFET</td>
<td>MEtal Semiconductor Field Effect Transistor</td>
</tr>
<tr>
<td>MGF</td>
<td>Moment Generating Function</td>
</tr>
<tr>
<td>OFDM</td>
<td>Optical Frequency Division Multiplexing</td>
</tr>
<tr>
<td>OPLL</td>
<td>Optical Phase Lock Loop</td>
</tr>
<tr>
<td>PBS</td>
<td>Polarisation Beam Splitter</td>
</tr>
<tr>
<td>PCM</td>
<td>Pulse-Code Modulation</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PIN</td>
<td>Positive Intrinsic Negative (photodiode)</td>
</tr>
<tr>
<td>PLL</td>
<td>Phase Lock Loop</td>
</tr>
<tr>
<td>PMF</td>
<td>Polarisation Maintaining Fibre</td>
</tr>
<tr>
<td>PRBS</td>
<td>Pseudo Random Binary Sequence</td>
</tr>
<tr>
<td>PSK</td>
<td>Phase Shift Keying</td>
</tr>
<tr>
<td>RACE</td>
<td>Research and development in Advanced Communications technologies in Europe</td>
</tr>
<tr>
<td>RACOS</td>
<td>RAised COSine</td>
</tr>
<tr>
<td>SAW</td>
<td>Surface Acoustic Waves</td>
</tr>
<tr>
<td>SMF</td>
<td>Single Mode Fibre</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SOP</td>
<td>State-of-Polarisation</td>
</tr>
</tbody>
</table>
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Summary

Fibre-optic transmission of signals with coherent optical detection is the optical equivalent of the traditional heterodyne reception in radio, TV and most other systems that use the 'ether' for transmission. Apart from the carrier frequencies, heterodyning in the radio and optical domain are to a large extent identical. The classical digital modulation schemes with amplitude-, frequency- and phase-shift keying (ASK, FSK and PSK) can thus be employed. Heterodyne and homodyne detection, diversity reception, non-coherent and synchronous IF-detection; almost all techniques known for many years from radio transmission can also be applied in coherent optical systems. However, there is one important limitation, imposed by the stability of the optical oscillators. Where the linewidth (related to the phase stability) of electrical oscillators is in the order of hertz, the value for an optical laser oscillator is usually in the order of megahertz. For proper reception this requires receiver configurations that are optimised in a way completely different to radio receivers.

From a detection point of view, the principal function of a (coherent optical) digital receiver is to obtain a maximum receiver sensitivity at a pre-defined bit-error rate. The main problem associated with the large phase noise of the IF signal is the appearance of a floor in the bit-error rate characteristic, introducing rapidly increasing sensitivity penalties. The first task of the system designer, when designing a coherent optical receiver for a certain modulation scheme, is therefore the selection of a configuration that suppresses the floor to an acceptably low level. Then, given the selected type of receiver, it is essential to identify all factors that degrade system performance or, in other words, introduce sensitivity penalties.

The aim of the thesis is to provide tools for the design of coherent optical receivers. Since the evaluation of sensitivity penalties due to the optics in a receiver is now well understood, all effort is concentrated on the penalties related to the electronics. This requires a detailed analysis of the IF-detection process in particular, which should include IF-filtering, the actual IF-detection, post-detection filtering and signal regeneration. A unified theory of IF-detection is therefore proposed. It benefits from a
very compact notation which hides the underlying mathematics and makes that many
different receiver configurations can easily be analysed. For each of the modulation
schemes ASK, FSK, Continuous-Phase FSK (CPFSK) and Differential PSK (DPSK),
the performance of single-branch heterodyne, polarisation diversity heterodyne and
phase diversity homodyne receivers can be evaluated. However, for an accurate and
fully analytical evaluation of the receiver performance it is essential to separate two
effects, the first of which is the influence of the phase noise.

Noticeable bit-error floors are not allowed in operational systems, so the conditions
that lead to unobservable floors are derived first. To this end, existing theories are ex-
tended in order to yield practically applicable results. It is found that a larger linewidth
requires an increasing frequency deviation, and consequently wider IF bandpass filters.
At the same time the effect of the linewidth can be reduced by proper post-detection
filtering. Using the wider IF filters, linewidth effects will not be observed and the
subsequent analysis can be made as if the IF carrier were without phase noise.

The second effect is then the influence of the additive Gaussian noise in the IF
section of the receiver. Due to the large IF filter bandwidth the signal-to-noise ratio
before IF-detection can be much lower than in traditional radio receivers. The analysis
without phase noise is completely analytical and exploits probability density functions
throughout. It yields very compact expressions for the receiver sensitivity penalty as a
function of the IF bandwidth and the efficiency of the detection process. In single-filter
ASK receivers one additional parameter is of major importance: the decision threshold
setting. For this, a very accurate approximation has been derived, allowing a realistic
and complete analysis of ASK receivers.

Finally, the validity of these theoretical results has been verified through the eval-
uation of measurements on two systems. The first is a 140 Mbit/s and 560 Mbit/s
FSK system with polarisation diversity reception and accurate dual-filter quadratic IF-
detection. The second is a CPFSK phase diversity system with differentially coherent
IF-detection at 1 Gbit/s. A detailed comparison of the measured sensitivity values
and the predicted receiver performance yields differences that are well within or only
slightly above the measurement uncertainties. The thesis ends with a set of design
guidelines for coherent optical systems, resulting from both the theoretical analysis
and practical experience in system design.
Samenvatting

Signaaltransmissie via glasvezel in combinatie met coherente optische detectie is het optische equivalent van de traditionele heterodyne ontvangst zoals die wordt toegepast in radio, TV en de meeste andere transmissiesystemen die gebruik maken van de ‘ether’. Afgezien van de draaggolffrequenties zijn heterodyne detectie in het radiodomein en het optische domein voor een zeer groot deel identiek. De klassieke digitale modulatiemethoden met behulp van amplitude-, frequentie- en fase-versleuteling kunnen dus worden toegepast. Heterodyne en homodyne detectie, diversiteitontvangst, niet-coherente en asynchrone middenfrequentdetectie; vrijwel alle bekende technieken uit de radiotransmissie kunnen weer worden gebruikt in coherente optische systemen. Er is echter één belangrijke beperking, veroorzaakt door de stabiliteit van de optische oscil- latoren. Terwijl de fasestabiliteit (of lijnbreedte) van elektrische oscillatoren in de orde is van enkele hertz, is de overeenkomstige waarde voor optische laser-oscillatoren gewoonlijk ter grootte van megahertz. Voor goede ontvangst betekent dit dat de ontvangerconfiguraties geheel anders worden geoptimaliseerd dan het geval is bij radioontvangers.

De belangrijkste kwaliteitseis van een (coherente optische) digitale ontvanger is de ontvangstgevoeligheid bij een voorgeschreven bitfoutenkans. Een groot probleem daarbij zit in de faseruis van het middenfrequente signaal, die een ‘vloer’ in de bitfoutenkarakteristiek en een snel toenemende reductie van de ontvangstgevoeligheid kan veroorzaken. Bij het ontwerp van een ontvanger voor een bepaald modulatieschema is de voornaamste taak dan ook de selectie van een configuratie die de vloer onderdrukt tot een voldoende laag niveau. Gegeven de zo geselecteerde ontvanger moet vervolgens worden bepaald welke factoren het uiteindelijke systeemgedrag beperken en de ontvangstgevoeligheid reduceren.

Dit proefschrift biedt een aantal gereedschappen die kunnen dienen bij het ontwerp van coherente optische systemen. Aangezien gevoeligheidsreducties ten gevolge van de optiek in de ontvanger vandaag de dag goed worden begrepen, is alle aandacht geconcentreerd op de prestaties van de elektronica. Dit vereist een gedetailleerde analyse van met name het middenfrequente detectieproces. Middenfrequentfiltering,
de eigenlijke detectie, post-detectie filtering en signaalregeneratie in het besliscircuit moeten daarbij allemaal worden meegenomen. Hiertoe is een algemene theorie van het middenfrequente detectieproces opgezet. Deze theorie heeft het voordeel van een zeer compacte notatie, die de gebruiker van de onderliggende wiskunde afschermt. Dit maakt het mogelijk op eenvoudige wijze zeer uiteenlopende ontvangerconfiguraties te analyseren. Voor elk van de modulatieschemas ASK, FSK, continuafase-FSK (CPFSK) en differentiële PSK (DPSK) is daarom de werking geanalyseerd van heterodyne ontvangers met een enkele middenfrequent tak, polarisatiediversiteit-ontvangers en fasediversiteits homodyne ontvangers. Voor een nauwkeurige en volledig analytische evaluatie van de ontvangerwerking is het daarbij echter van belang dat twee effecten worden gescheiden. Het eerste is de invloed van de faseruis.

Zichtbare vloeren in de bitfoutenkans karakteristiek zijn niet toegestaan in operationele systemen, zodat allereerst de condities waaronder deze vloeren onzichtbaar zijn dienen te worden gevonden. Hiertoe zijn bestaande theorieën uitgebreid teneinde nuttige resultaten te krijgen. Het blijkt dat een grotere lijnbreedte een toenemende frequentiewaaivariatie vertoont, wat kan worden vertaald in bredere middenfrequentfilters. Tegelijkertijd kan de invloed van de lijnbreedte worden gereduceerd door de juiste post-detectie filtering. Door brede middenfrequentfilters te gebruiken kunnen zichtbare lijnbreedteffecten worden onderdrukt en kan de analyse worden voortgezet alsof de middenfrequent draaggolf zonder faseruis is.

Het tweede effect is vervolgens de invloed van de additieve Gausschische ruis in de middenfrequentiesectie van de ontvanger. Ten gevolge van de grote filterbandbreedte zal de signaal-ruisverhouding voor middenfrequentedetectie veel lager zijn dan in traditionele radioontvangers. De analyse zonder meename van de faseruis is volledig analytisch en maakt enkel gebruik van kansdichthedensfuncties. Dit levert zeer compacte uitdrukkingen voor de ontvangergevoeligheidsreductie als functie van de middenfrequentbandbreedte en de efficiëntie van het detectieproces. In één-filter ASK-ontvangers komt daar nog een belangrijke parameter bij: de beslissdrempel. Hiervoor is een zeer nauwkeurige benadering afgeleid die ook van ASK-ontvangers een realistische en complete analyse mogelijk maakt.

Sommaire

La transmission par fibre optique avec détection optique cohérente est l'équivalent optique de la réception traditionnelle hétérodyne en radio, TV et la plupart des systèmes qui utilisent 'l'ether' pour la transmission. A part les fréquences porteuses, les principes d'hétérodyne dans le domaine radio et le domaine optique sont en grande partie identiques. Les méthodes de modulations digitales classiques d'amplitude, ou de fréquence, ou de phase (ASK, FSK et PSK) peuvent donc être employées. La détection hétérodyne et homodyne, la réception diversity, la détection à fréquence intermédiaire non-cohérente et synchronisée, presque toutes les techniques connues depuis plusieurs années dans la transmission radio, peuvent être également appliquées dans les systèmes optiques cohérents. Pourtant, il y a une limitation importante, imposée par la stabilité des oscillateurs optiques. Si la stabilité de phase (ou la largeur de ligne) des oscillateurs électriques est dans l'ordre de quelques hertz, alors la valeur pour l'oscillator laser optique est habituellement dans l'ordre de megahertz. Pour une bonne réception, cela demande des configurations de récepteur optimalisées d'une manière totalement différente comparée aux récepteurs radios.

La plus importante exigence de qualité d'un récepteur digital (optique cohérent) est d'obtenir une sensibilité de réception maximale à un taux d'erreurs bien défini. Le problème majeur est relié avec le bruit de phase du signal à fréquence intermédiaire, qui fait apparaître un plateau dans la caractéristique du taux d'erreurs. Cela introduit des peines de sensibilités qui augmentent rapidement. La tâche principale en dessinant un récepteur pour une certaine méthode de modulation, est la sélection d'une configuration qui supprime ce plateau jusqu'à un niveau assez bas. Étant donné la configuration du récepteur, il faut déterminer quels sont les éléments qui limitent la performance du système et réduisent la sensibilité de réception.

Cette thèse offre un nombre d'outils qui peuvent servir à dessiner des systèmes optiques cohérents. Puisque les réductions de sensibilité causées par l'optique dans le récepteur sont aujourd'hui bien comprises, toute l'attention est dirigée vers l'électronique. Cela demande particulièrement une analyse détaillée du process de détection à fréquence intermédiaire. Le filtre à fréquence intermédiaire, la détection propre, le
filtrage post-détection et la régénération du signal doivent tous être compris dans cette analyse. A cet effet une théorie unifiée de la détection à fréquence intermédiaire a été élaborée. Cette théorie a l’avantage d’une écriture compacte, qui masque la complexité mathématique pour l’utilisateur. Cela facilite énormément l’analyse des configurations de réception diverses. Pour chacun des systèmes de modulation ASK, FSK, FSK à phase continue (CPFSK) et PSK différentiel (DPSK), la fonction est analysée pour des récepteurs hétérodynes à une seule branche de fréquence intermédiaire, des récepteurs diversity de polarisation et des récepteurs homodynes diversity de phase. Pour une évaluation analytique précise et complète du récepteur il est essentiel de dissocier deux aspects. Le premier est l’influence du bruit de phase.

Des plateaux visibles dans la caractéristique du taux d’erreurs ne sont pas admisibles dans des systèmes opérationnels; de sorte qu’il se doit de trouver les facteurs qui rendent ces plateaux invisibles. A cet effet des théories existantes sont répandues afin d’atteindre des résultats utiles. Il semble qu’une plus grande largeur de ligne exige une déviation de fréquence croissante, ce qui peut être traduit en filtres de bandes à fréquence intermédiaire plus larges. En même temps l’influence de la largeur de ligne peut être réduite par un filtrage post-détection adéquat. Par l’utilisation de larges filtres de bandes à fréquence intermédiaire, les effets causés par la largeur de ligne peuvent être rendus invisibles, et l’analyse peut être continuée comme si l’onde porteuse à fréquence intermédiaire était sans bruit de phase.

Le deuxième effet est l’influence du bruit additive à spectre continu et uniforme dans la section de fréquence intermédiaire du récepteur. A cause du large filtre de bande, le rapport signal-bruit avant la détection de fréquence intermédiaire sera plus bas comparé aux récepteurs radio traditionnels. L’analyse sans la considération du bruit de phase est complètement analytique, et n’utilise que des densités de probabilités. Cela donne des expressions très compactes pour la réduction de la sensibilité de réception comme fonction de la largeur de bande à fréquence intermédiaire et l’efficacité du proces de détection. Dans les récepteurs ASK à un seul filtre un important paramètre s’ajoute: le seuil de décision. Pour cela une approximation très précise est dérivée, qui rend possible l’analyse complète et réaliste des récepteurs ASK.

Pour finir les résultats de l’analyse théorique sont vérifiés par l’évaluation des mesures de système. Un premier système contient un récepteur diversity de polarisation FSK 140-560 Mbit/s avec détection précise à fréquence intermédiaire parabolique à deux filtres. Le deuxième système utilise une réception diversity de phase CPFSK avec détection à fréquence intermédiaire différentielle de 1 Gbit/s. Une comparaison détaillée des mesures et de la sensibilité de réception calculée donne des différences qui tombent à l’intérieur ou très peu à l’extérieur des précisions de mesure. La thèse se termine sur un nombre de directives pour la conception d’un système optique cohérent, résultat de l’analyse théorique et de l’expérience pratique dans la création des systèmes.
Account of previous work related to this thesis

List of publications by the author

Papers


Letters


321
[L4] P.W. Hooijmans and M.T. Tomesen, "Analytical and practical evaluation of


Conference papers

[C1] P.W. Hooijmans, M.T. Tomesen and A. van de Grijp, "A 2.3 GHz low noise
balanced receiver for FSK heterodyne reception, using commercially available
components", Proc. 15th Eur. Conf. on Opt. Comm., ECOC 89, Gothenburg,

[C2] P.W. Hooijmans, M.T. Tomesen and A. van de Grijp, "A linewidth and bitrate
flexible FSK heterodyne system, using a frequency discriminator and biphase
coding", Proc. 15th Eur. Conf. on Opt. Comm., ECOC 89, Gothenburg,

[C3] P.W. Hooijmans, M.T. Tomesen, P.P.G. Mols, A. van de Grijp and C.K. Wong,
"Coherent multichannel system for TV distribution with fast and direct channel
THG5, p. 170.

laser in a coherent multichannel receiver", Proc. 7th Eur. Fiber Optic Conf.,

multi-channel system, for simultaneous transmission of 140 Mbit/s TV and 560
Sept. 1990, paper WeG2.2, pp. 457–460.

demodulation in a FSK heterodyne polarisation diversity transmission exper-
iment", Proc. 16th Eur. Conf. on Opt. Comm., ECOC 90, Amsterdam,

L.L. Kanters, G.D. Khoe, K.G. Wright and T.J.B. Swanenburg, "A 1 Gbit/s
FSK homodyne phase diversity system, employing a high-speed silicon bipolar
LIST OF PUBLICATIONS


Reports


Contributions by others

Both the theoretical analysis and the practical design and evaluation of coherent optical systems have been part of the author's tasks within the group 'Wideband Communication Systems' of the Philips Research Laboratories in Eindhoven, The Netherlands. In particular the construction of the different coherent receivers has been a major joint effort by myself and 5-6 colleagues. The different contributions of colleagues and students are listed below.
CONTRIBUTIONS BY OTHERS

- Chapter 2: Theorem 2.1 about the importance of non-linear operations was developed together with ing. M.T. Tomesen.

- Chapter 3: Many of the tools, but the especially the 'intuitive' checks for the analysis in section 3.1 to 3.4, have been developed and refined in many discussions with ing. A. van de Grijp.

- Chapter 4: The explanation of the threshold behaviour in section 4.2.2 is by ing. A. van de Grijp.

- Chapter 6: The effects of IF imbalance in diversity receivers were originally measured and investigated by ing. M.T. Tomesen, whereafter the theory was incorporated in the analysis.

- Chapter 7: CPFSK phase diversity receivers have been investigated theoretically by R.A.J.C.M. van Gils and F.A.J. Dumont, students of prof.ir. G.D. Khoe of Eindhoven University of Technology, in their respective master's theses. They were supervised for this part by the author. Results are used in section 7.4.

- Chapter 8: The construction of the 140 Mbit/s CMC systems has been the major activity within the group from 1988 to 1992. The original system concept, including the biphase coding, was for a large part proposed by ing. A. van de Grijp. Persons involved in construction of the transmitters were ir. C.K. Wong and ing. L.L. Kan ters. The receivers were produced jointly by K.G. Wright, B.Sc. (polarisation diversity unit), ir. P.P.G. Mols (DBR characterisation, tuning software, receiver hardware, system integration), ing. M.T. Tomesen (IF electronics, system integration and measurements) and the author. Lasers for the systems were obtained from Philips Optoelectronics Center (POC) through ir. W. Nijman and dr. P.I. Kuindersma. All sensitivity and BER-measurements used in section 8.2 were performed by ing. M.T. Tomesen.

The 1 Gbit/s phase diversity system has always been an activity of graduation students of prof.ir. G.D. Khoe of Eindhoven University of Technology: R.A.J.C.M. van Gils, F.A.J. Dumont and P.H.G.M. Thijsen in chronological order, under the supervision of ing. M.T. Tomesen and the author. They all contributed to the construction and improvement of the system, while the latest sensitivity data and penalty analysis are from the work of Dumont. Measurements on the high-speed IC's used in this system were performed by J.J.E.M. Hagaraats from Delft University of Technology.
Connectivity matrix

The following table provides a good insight in the major and minor relations between the papers, letters, conference papers and reports published by the author and his colleagues (listed on pages 321-324), and the contents of the different chapters of this thesis.

<table>
<thead>
<tr>
<th>CHAPTER 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 *Biphasic ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2 *Analytical ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L3 *140 Mbit/s ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4 *Multibitrate ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5 *Fast channel ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1 *2.3 GHz ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2 *Linewidth ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3 *Coherent ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4 *Tunable DBR ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5 *Multi-bitrate ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6 *Perfect ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7 *1 Gbit/s ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8 *Local loop ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9 *Reliability ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10 *Dynamic ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11 *CPFSK ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12 *Novel ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1 *Part I ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2 *Part II ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R3 *Part III ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4 *Part IV ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5 *Nyquist ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6 *1 Gbit/s ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7 *Front-ends ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R8 *Biphasic ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R9 *QbIC ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R10 *Frequency ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R11 *Phase ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R12 *AFC ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R13 *Coherent ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R14 *Coherent ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

● = Direct relation to contents of chapter
○ = Minor relation to contents of chapter
Acknowledgements

I want to thank people at the Philips Research Laboratory, in particular my director ir. B.L.A. (Ben) Waumans and group leader Dr. T.J.B. (Teun) Swanenburg, who gave their permission for the preparation of this Ph.D. thesis. Through them I want to thank Philips Research for all the support received.

My gratitude goes to prof.dr. J.C. Arnbak for his willingness to act as my promotor. I admire him for the preciseness with which he read the large piles of manuscript, despite his many other occupations. His comments always led to more fundamental solutions and approaches.

Both the theoretical and practical analyses presented in this thesis have been part of the daily activities in the 'Wideband Communication Systems'-group, and were not started with the intention of resulting in a thesis. It was only after many reports, papers, patents and, above all, practical results had accumulated that I had the idea of collecting them into a thesis.

Concerning the said daily work, special credit goes to two colleagues, both high-frequency electronics specialists, exceptionally skilled in the art. Bram van de Grijp has been the spiritual father of all coherent optical transmission activities in the group until his retirement in 1990. It was his idea to start the theoretical analysis, and his unique experience in systems design has ensured that my theoretical exercises never lost touch with reality. His careful reading of the manuscript and the following discussions, lavishly laced with trappist beers, have solved the last remaining problems. Mark Tomesen has been a friend and my closest colleague for the past five years, not only in the lab but also on the football field. Together we built all the electronics for the coherent systems, where his insight in analogue electronics almost invariably led to optimum performance. Most system measurements presented in this thesis are his work, giving results of uncomparable accuracy.

Within the CMC RACE-program we have been building the coherent optical systems as a team. Next to Mark Tomesen the colleagues are Peter Mols (local oscillator
tuning, software, receiver units), Kieran Wright (polarisation diversity optics), Ching Kwok Wong (transmitter laser) and Lars Kan ters (temperature control and transmitter units). Special credit goes to them, since without their contributions I could not have presented system results.

Measurements on the 1 Gbit/s systems and components have always been performed by graduation students from Delft and Eindhoven Universities of Technology; René van Gils, Rick Dumont, Hans Hageraats and Pieter Thijssen in order of chronology. I want to thank them for their contributions and the positive effect they had on the working environment.

Many of the colleagues mentioned above, i.e. Peter Mols, Mark Tomesen and René van Gils, as well as Geert Depovere, have carefully read through the manuscript, for which I admire and thank them. John Reid has scanned the complete manuscript for English monstrosities, a gargantuan task by itself. René van Gils was so kind to prepare all the Mathematica pictures in this book, while Alphons van der Heijden has given me considerable Apollo and B_{1OTeX} support.

The whole 'atmosphere' in the group, inspiringly led by Teun Swanenburg in his own special way, has been essential for me to make this thesis. Working in such a team has been a fantastic experience, making my period at Philips Research an unforgettable one.

Finally I want to thank my family, in-laws and friends (especially Peer Wolkers) for their interest and moral support during these years. I thank my parents for my education and for paying my studies, and especially my father - an analogue electronic engineer himself - for his stimulating interest in my results.

The greatest burden of supporting me has rested with my wife Christine and our boys, Pierre and René. I sincerely thank them for their patience when they once again saw me vanishing upstairs. I adore Christine for the enormous support she gave me. To them I dedicate this book.
Curriculum Vitae

Pieter Hooijmans was born in Voorburg (ZH), December 14, 1958. After finishing Atheneum at the Veurs College, Leidschendam, he entered the study of Electrical Engineering at the Delft University of Technology, Delft, The Netherlands in 1977. He belonged to the first group of students that followed the new avionics graduation program, jointly offered by the departments of Electrical Engineering and Aerospace Engineering. He graduated in February 1984 at the Microwave Laboratory, under the supervision of prof. ir. L. Krul on a master's thesis on X-band SLAR and C-band scatterometer calibration for remote-sensing applications.

In March 1984 he joined the Royal Netherlands Air Force for his military service, during which he taught radar and microwave engineering and set up training courses at the Luchtmacht Elektronische en Technische School (LETS) in Schaarsbergen. He left the Air Force in August 1985 as a second lieutenant.

From September 1985 he has been employed at the Philips Research Laboratory in Eindhoven, in the group 'Wideband Communication Systems' headed by Dr. T.J.B. Swanenburg. After a short time working on a 1.2 Gbit/s direct detection system, he became involved in the design and construction of the first coherent optical transmission system of the group. Since then, most of his activities have been part of the European RACE-projects 1010 CMC 'Coherent Multi-Channel system' and 2065 COBRA "Coherent Optical systems implemented for Business Traffic Routing and Access", resulting in the systems presented in this thesis. Over the past years his main activities have been the design and construction of high-frequency electronics for the coherent optical systems, system set-up and evaluation, and theoretical system analysis. Work on this thesis was started in January 1991.
Index
Index

2x2 optical coupler 38
3x3 optical coupler
  as 90°-hybrid
  see I&Q IF signals
  in phase diversity 39, 150, 155–158, 187, 234–239
4x4 optical coupler 150, 187, 234
90° electrical hybrid 165, 187
90° optical hybrid 39–40, 187, 233–236

Addition, effect on pdf 70, 258
AFC
  see Automatic Frequency Control
AGC
  see Automatic Gain Control
Amplitude modulation, residual 117, 254
Amplitude Shift Keying (ASK) 30, 83
  bit-error rate 91–94, 105–107
  detection penalty 109, 140, 155
  linewidth effects 73–82
  linewidth requirement 171
  modulation 30
  phase diversity 38, 150–158
  polarisation diversity 138–144
  ultimate performance 30, 94
  with imbalance 142
Analogue biphase demodulation
  see Biphase
ASK
  see Amplitude Shift Keying
Autocorrelation of IF-detector output 62
Automatic Frequency Control (AFC) 187, 256
Automatic Gain Control (AGC) 97, 209, 255, 256
Avalanche Photodiode (APD) 2
Average white Gaussian noise
  see White noise, channel

Balanced receiver 29, 253
  realisation 208, 224, 232
Bandpass filter (BPF), IF
  definition 48
  minimum width 74, 82, 251
  multiple 118
  noise bandwidth 48
  realisation 209
Bandwidth efficiency 36, 132, 187, 251
Bandwidth expansion factor, IF
  definition 49, 183
  determining 79–82, 123, 214
  in practical system 207, 214, 228
Bayes decision rule 72
BER-floor
  definition 74
  from laser fm 228
  from linewidth 74–82, 168–173
  unnoticeable 80, 250
BER deterioration factor 128, 180
Bessel functions, modified
  asymptotic expansion 87
  in pdf’s 263
  in threshold eqn. 86, 100, 273
  series expansion 89, 273
  spherical 126, 292

332
INDEX

Biphase, linecoding 214–216
  sensitivity penalty 216, 225
Birefringence
  see Glass fibre
Bit-error rate (BER)
  definition 21, 72
  floor
  see BER-floor
Bitrate-distance product 1–4

Carson rule 33, 164
Channel spacing 36, 206
  see Coherent Multi-Channel
Characteristic function 112
Chi-square distribution 60, 85, 175, 262–267
  central 265–267
  moments 100, 266
  pdf 100, 265
non-central 262–265
  moments 100, 263, 290
  pdf 100, 262
reproductivity 70, 264
Chi distribution 59, 84, 111, 267–271
  central 270–271
  moments 56, 270
  pdf 55, 270
non-central 267–269
  moments 56, 267
  pdf 55, 267
Clipping of signals
  see Limiting
CMC
  see Coherent Multi-Channel
Coherence time, of light 25
Coherent IF-detection
  see Synchronous detection
Coherent Multi-Channel (CMC) 4, 36
  channel selection 14
  system 206
Coherent optical detection 3–6, 12–14, 35–45, 256
  first experiments 3
  historical comparison 16
  principle 12
  proposed 3
  ultimate performance 37
Combining 135–138
  equal-gain 135–139, 144, 158, 225, 226
  maximal ratio 43, 135
  realisation 214
  selection 135
Common Mode Rejection ratio (CMRR) 29, 208, 225
Continuous Phase FSK (CPFSK) 32, 173–194
  BER 177
  comparison with DPSK 202
  comparison with FSK 171, 177, 179, 183, 202
detection penalty 181
  linewidth requirements 166–173, 194
  phase diversity 186–194, 251
  polarisation diversity 184–186
  system experiments 232–239
  ultimate performance 33, 166, 181
Conversion loss, of detector 209
Convolution of pdf’s 70, 77, 111, 122, 124, 258
CPFSK
  see Continuous Phase FSK
Cross-connecting 5
Cross-correlation 120, 124, 176, 177, 259
  relation to convolution 122, 259
Crosstalk, IF-to-baseband 40, 48, 165, 209, 210, 239
Cumulative distribution function 257
DBR
  see Laser
Decision circuit 255
  ambiguity 255
non-linear operation 22, 50
optimisation 255
realisation 217
sensitivity 23
threshold 71, 217
Decision theory 71–72, 119–121
and likelihood ratio 71, 121
approximation 88, 101, 275, 276
accuracy of 103
deviation factor 87, 101
influence of IF imbalance 142
influence on linewidth floor 77
normalised 86
Decorrelation, of delayed signal 174
Deep fade 133
Degenerate hypergeometric function 268
asymptotic expansion 269
Degree of freedom, of pdf 55, 262
Delay-and-cross multiply discriminator
40, 186–194
integrated circuit 232
Delay-line discriminator 152–177, 183–184, 194
non-optimum 182, 200
Detection efficiency of CPFSK 162, 165
non-optimum 181–133, 189
optimum 178, 232
Detection improvement factor 176
and SNR imbalance 191
Deterioration factor, BER
see BER deterioration factor
DFB
see Laser
Dialogue transceiver 221
Differentially coherent IF-detection 33–35, 44, 161, 202
Differential partial error rates 89
Differential PSK (DPSK) 34, 194–203
BER 197

comparison with CPFSK 197, 199, 202
comparison with FSK 120, 171, 202
direct modulation 203
linewidth requirements 166–173
modulation depth 198
phase diversity 38, 197
polarisation diversity 197
ultimate performance 34, 166
with non-optimum (de)modulation
197, 202
Direct detection, receivers 1–3
historical comparison 17
Direct laser modulation
DPSK 203
FSK 162, 206, 214
Distribution function 261
complementary 168, 262
Diversity reception 133
phase 38–40, 134
ASK 38, 150–158
CPFSK 186–194, 232–239, 251
DPSK 38, 197
FSK 39
measurements 232–239
penalty 152, 155
polarisation 41, 134–138, 252
ASK 138–144
CPFSK 184–186
DPSK 197
FSK 144–150
measurements 226–230
penalty 141, 147, 226, 242–244
sensitivity penalty 184, 225
DPSK
see Differential PSK
Dual-filter detection 32, 117, 183–184, 289–293
decision theory 119
practical system 206–214, 221
Dynamic range 211
INDEX

Equalisation, of laser fm characteristic 214, 233
Equivalent input noise 23
  measuring of the 24, 209, 223
  tuning of the 24, 208
Erbium-Doped Fibre Amplifier (EDFA)
  see Optical amplifiers
Error function 114, 261, 286
  complementary 168, 262
Error margin, of measurements 239

Fabry-Perot laser
  see Laser, Fabry-Perot
Fading 133
False alarm
  see Probability of false alarm
Fibre, glass
  see Glass fibre
Finite-time integrator
  see Integrate and Dump filter
Flip-flop
  as decision circuit
  see Decision circuit
Flip-flop, as decision circuit 71, 217,
  233
Floor
  see BER-floor
Fourier transform 26, 61–67, 112
Fourier transform 65
Frequency deviation
  see Frequency modulation index
Frequency discriminator
  for phase diversity
    see Delay-and-cross multiply
    realisation 212, 232
Frequency modulation characteristic 214,
  228
Frequency modulation index 31, 118,
  162, 233
  influence on BER-floor 173
  influence on linewidth floor 250
Frequency Shift Keying (FSK) 30, 117
  BER 127–129
    comparison with ASK 123, 129,
    148
    comparison with CPFSK 171, 183
    detection penalty 129, 147
    linewidth requirements 123, 171
    phase diversity 39
    polarisation diversity 144–150
    system 206–221
    system experiments 221–232
    ultimate performance 32, 129
    with imbalance 148
Front end
  see Balanced receiver
FSK
  see Frequency Shift Keying
Gamma distribution 265
Gamma function 105
  approximation 276, 281
  incomplete 78, 93, 105, 107, 177,
  265
    as continued fraction 281
    reduction of 287
Gaussian noise
  see White noise, channel
Generalised Laguerre polynomial
  see Laguerre polynomial
Glass fibre 1
  bandwidth 3
  birefringence 9, 41, 134
  dispersion 2
  introduction 1
  loss 1
  properties 134
Harmonics of IF signal 40, 211, 226
HEMT, GaAs 208, 224, 253
Hermite polynomial 113
  approximation 114
Heterodyne detection 35
  for bandwidth measurements 210
Homodyne
quasi-homodyne
see Diversity reception, phase
Homodyne detection 36, 150
IF-detection 22, 58–64, 84, 253
diode for 58, 210
quadratic 59, 209, 210–214, 253
IF bandwidth
see Bandwidth expansion factor, IF
IF branches, number of 39, 118, 131, 141, 158, 191, 252
IF characteristics, flatness of 226, 254
see Imbalance, Mark-Space
Image-rejection receivers 36
Imbalance, IF 142–144, 153, 155, 158, 184–186, 191–193, 234, 238, 252
measured 227, 238
Imbalance, Mark-Space 225, 231, 254
Impulse response 65–69, 70
In-phase IF signal
see I&Q IF signals
Incomplete gamma function
see Gamma function
Integrate and Dump (I&D) filter 65–70
not matched filter 65, 109
theory 65
Integrated circuits 2, 159, 253
gallium-arsenide 2
silicon 232
Integrated optics 253
Integration by parts 93, 106, 279, 282
Inter-channel interference 206
Inter-Symbol Interference (ISI) 50, 163
elimination of 51, 66–69, 217, 249
Intermediate frequency (IF)
generation 12
signal, definition of 51
I&Q IF signals 39, 40, 150
90° hybrids for 233
Isolator
see Optical isolator
Laguerre polynomial 125, 178, 263, 269, 290
orthogonality for negative argument 125, 264
Laplace transform 76
Laser 1–6
DBR 24
in system experiments 206, 233
DFB 2, 24
in system experiments 206, 214
Fabry-Perot 2, 17
intensity noise
see RIN
linewidth
see Linewidth, laser
phase noise
see Linewidth, laser
Light
as EM-field 9–14
as particles 11, 16, 25
frequency of the 12
Likelihood ratio 72
dual-filter detection 119–121
single-filter detection 86
Limiting of signals 166, 255
Linear IF-detection 59, 84, 111–115, 286–288
Linecoding 214–216
biphase
see Biphase
Linewidth-to-bandwidth ratio, IF 76, 168, 237
Linewidth-to-bitrade ratio 79, 123
in practical system 207, 225, 237
Linewidth, laser 24–27
definition 26
effects 26, 73–82, 122, 166–173
Lorentz spectrum 26, 73
measured in systems 207
INDEX

Local oscillator (LO) 12, 206
output power 222, 233, 253
tuning of 206
Lorentz spectrum
see Linewidth, laser

Marcum Q-function 92
approximation 97
definition 92
generalised 106, 282
series expansion 283
Matched filtering 65
at IF 82, 84, 120
Mellin convolution 261
MESFET, GaAs 208
Minimum Shift Keying (MSK) 33, 163, 170
Modified Bessel function
see Bessel function
Modulation depth
see DPSK, modulation depth
Modulation index
see Frequency modulation index
Moment Generating Function (MGF) 77
Moments, of distribution 258
Multi-bitrate experiments 227–230, 240
Multi-level signals 215
Multichannel reception
see Coherent Multi-Channel
Multiplication, effect on pdf’s 260

Noise 22–30
baseband 59, 211, 230
channel
see White noise, channel
receiver 15, 223
shot
see Shot noise
tuning 208

Noise floor 22, 211, 230
Non-central chi-square pdf
see Chi-square distribution
Non-central chi pdf
see Chi distribution
Non-central normal pdf
see Normal distribution
Non-coherent IF-detection 6, 45
Non-linear operation 22, 58
Non-uniform IF characteristics
see SNR Mark-Space imbalance
Non-uniform laser characteristics
see Equalisation
Noncentrality parameter 55, 176, 190
definition 262, 267
relation to SNR 183
Normal distribution 53, 119, 261
folded 271
Normalised threshold
see Decision threshold
Nyquist criterion
on vestigial symmetry 64
sampling 50, 63

Optical amplifiers 2
Optical Frequency Division Multiplexing (OFDM) 3–4, 206
and CMC
see CMC
Optical isolator 135
Optical Phase Lock Loop (OPLL) 4, 17, 36
Orthogonal signalling 120, 150, 202
Outage 133

Penalty, sensitivity 5, 16
from biphase coding 216, 225
from decision circuit 225
from diversity 140, 147, 155, 184,
225, 242–244
from excess loss 43, 222, 253
from IF bandwidth 109, 129, 181, 225
    measured 230–232, 240–242
from I&Q generation 235
from linewidth 225, 237
from LO power splitting 43
from Nyquist filter 225
from receiver noise 16, 224, 237
from responsivity 16, 222, 253
from RIN 29, 224, 238
from SNR imbalance 235, 238
    system measurements 221
Phase detector for DPSK 194
Phase diversity
    see Diversity reception
Phase mismatch, IF 153, 155–158, 193
Phase noise
    from laser
        see Linewidth, laser
Phase noise, in CPFSK 44, 165
Phase rotation, IF, DPSK 198–202
Phase Shift Keying (PSK) 33, 37
    linewidth 161
Photo detection 10–14, 23–24, 208, 253
    linear process 11
PIN diode 10, 208
Poisson distribution 107, 267
Polarisation beam splitter 41, 135
Polarisation control 41, 159, 252
    for CPFSK phase diversity 232
Polarisation diversity
    see Diversity reception
Polarisation diversity unit 41, 222
    modified for phase diversity 187, 233
Polarisation Maintaining Fibre (PMF) 134–135
Polarisation split angle, effective
    see Polarisation, State of - (SOP)
Polarisation, State of - (SOP) 9–14, 41
    alignment 12–15
beam splitter 41
dependence of sensitivity 226–230
fluctuations 134
    and linear detection 111
    effect on linewidth floor 74–82, 168–173
    function 254
    improvement factor 100, 127, 292
    integrate and dump 65–70, 249
    pdf of output 176
    raised cosine (RACOS) 64, 216
    realisation 216
Post-detection summation
    see Combining
Power density spectrum
    of IF-detector output 61–64
    of IF noise 52
    of IF signal 52
    of laser phase noise 26, 73
Probability density function (pdf) 249, 257–271
    after addition 70, 77, 258
    after subtraction 122, 259
    normalisation of 257, 264, 266, 287
    of a product 260
    of detector output 59
    of I&D filter output 70, 287
    of IF signal 53–58
    with linewidth 76, 168
Probability integral 261
    see Error integral
Probability of detection 73
Probability of false alarms 98
Q-function
    see Marcum Q-function
Quadratic IF-detection 59, 85, 253
    as power detector 86
    dynamic range of 59, 138, 211
    for diversity 42, 137, 226
realisation 209, 210–214
Quadrature IF signal
see I&Q IF signals
Quantum efficiency
see Responsivity
Quasi-homodyne
see Diversity reception, phase, CPFSK

RACE 207
Raised cosine (RACOS) filter
see Post-detection filter
Ratio combiner
see Combining, maximal ratio
Rayleigh distribution 55
Receiver sensitivity
see Sensitivity, receiver
Regeneration, signal 50, 216–217, 255
Relative Intensity noise (RIN) 27–30
cancellation 28, 208, 225, 235
definition 27
penalty 29, 224
Reproductivity, of pdf 70, 264
Responsivity 11, 208, 222, 253
Rice distribution 55, 262
RIN
see Relative Intensity Noise

Schottky diode 210
Selectivity
see Coherent Multi-Channel
Sensitivity, receiver 5, 15
definition 15
measurements 220–221
penalty
see Penalty
polarisation dependence 221, 226–230
Shot noise 15
definition 23
limit 5, 16, 222
measuring the 24, 222
of IF-detector diodes 211

Signal-to-Noise Ratio (SNR) 14–16
definition 54
imbalance
see Imbalance, IF
Mark-Space imbalance 225, 231, 254
Single-filter detection 30, 83, 279–288
line width effects 73–82
of FSK 32, 83
Single-mode fibre (SMF)
see Glass fibre
Slice amplifier 255
Squared IF noise
influence of 62, 100, 109, 126, 131, 141, 226
variance of 60
variance of 126
Squared IF noise, variance of 292
State of polarisation (SOP)
see Polarisation
Stirling formula 281
Stubs, IF filters with 211
Subtraction, effect on pdf’s 122, 124, 259
Superposition principle 50
Suppression characteristic 62
Synchronous detection 17, 30, 32, 33, 43
biphase demodulation as
see Biphase

Threshold
see Decision threshold
Threshold equation 86, 100, 273, 285
approximation 88, 101, 275
accuracy of 103
Transconductance 208
Transimpedance amplifier 208, 211
Tuned front ends
see Equivalent input noise
White noise 52, 249
  channel 15, 24, 48
Whittaker function
  see Degenerate hypergeometric function
Wiener-Kinchine theorem 62