# Towards Practical Distributed Video Coding

W.A. Rajitha J. Weerakkody

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

UNIVERSITY OF SURREY

Centre for Communication Systems Research
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, UK

September 2008

Copyright © W.A. Rajitha J. Weerakkody

## SUMMARY

Multimedia is increasingly becoming a utility rather than mere entertainment. The range of video applications has increased, some of which are becoming indispensable in modern lifestyle. Video surveillance is one area that has attracted significant amount of focus and also benefited from considerable research effort for development. However, it is noted that there is still a notable technological gap between an ideal video surveillance platform and the available solutions, mainly in the form of the encoder and decoder complexity balance and the associated design costs. In this thesis, we focus on an emerging technology, Distributed Video Coding (DVC), which is ideally suited for the video surveillance scenario, and fits many other potential applications too. A number of technical advances are proposed to the coding framework, in view of bringing the concept of DVC towards a practical realization targeted at future commercial developments and standardizations.

The research scope includes designing novel coding frameworks, enhancing the existing structures, and designing enhanced algorithms for individual functional elements within the codec framework, improving both encoder and decoder. First, the Slepian-Wolf codec efficiency is improved using a Turbo Trellis Coded Modulation (TTCM) based design. Then a Unidirectional DVC (UDVC) framework is designed involving iterative decoding with *side information* refinement, which enables the suppression of practically hindering reverse feedback channel. Novel approaches for *quantization* and *reconstruction* algorithms are presented to improve the compression efficiency of the codec. A non linear quantisation algorithm is proposed in contrast to the commonly used linear quantiser, whereas the proposed reconstruction algorithm is specifically optimised for the UDVC framework. A refinement algorithm is proposed for improving the quality of side information. Considering the significance of video communications over wireless media in the target applications, a modified decoding algorithm is proposed to improve the DVC codec error resilience so that the channel coding cost is minimised.

The efficiency and effectiveness of all proposed algorithms are verified through simulations and comparing the results with the state of the art. Significant performance gains are recorded in terms of the compression efficiency and the error resilience where applicable.

# ACKNOWLEDGEMENTS

There are many people who have contributed in making this research and the accompanying thesis a reality, to whom I am very grateful.

Special thanks go to Dr Anil Fernando, supervisor of this research, for everything he has done for me to start a PhD research in this very interesting area of Distributed Video Coding and then enabling a smooth flow of progress. It is indeed his guidance, enthusiasm, constructive criticism, and outstanding support that made me finish this project with some remarkable achievements. Secondly I would like to thank Prof. Ahmet Kondoz, my co-supervisor and the director of my research group, for enabling a smooth transfer from Brunel University half way through my research and funding the work through the research centre to keep the financial burden away from me. Also I wish to thank Dr. Tatiana Kalganova for been my co-supervisor at the Brunel University and her support extended to me.

Then I would like to show my appreciation to Dr Hemantha Kodikara Arachchi, Dr. Buddika Adikari, Gishantha Fernando, Murat Badem, Steven Adedoyin and Lasith Yasakethu for their valuable contributions towards this thesis and excellent company.

I would also like to thank the academic and non-academic staff of University of Surrey and Brunel University for everything they have done for me and this project.

I am always mostly grateful as ever to my respectful parents Mr. and Mrs Weerakkody, and family for the undisputed effort taken in making me through the support and guidance extended from the very first day of my life, and a special thank goes to my loving wife Lakshi who accompanied and supported me in everyway throughout the past 3 years.

Finally, thanks to all my friends, for their friendship and support.

# TABLE OF CONTENTS

**SUMMARY** ..... I

**ACKNOWLEDGEMENTS** ..... II

**TABLE OF CONTENTS** ..... III

**LIST OF FIGURES** ..... V

**LIST OF TABLES** ..... VII

**LIST OF ACRONYMS** ..... VIII

**CHAPTER 1** ..... 1

INTRODUCTION ..... 1
1.1 BACKGROUND AND MOTIVATION ..... 2
*1.1.1 DVC application scenarios* ..... 2
*1.1.2 Motivation* ..... 5
1.2 RESEARCH METHOD ..... 5
*1.2.1 Choice of baseline technology* ..... 6
*1.2.2 Simulations and test conditions* ..... 7
*1.2.3 Presentation of results* ..... 7
1.3 OBJECTIVES AND SCOPE OF THE RESEARCH ..... 8
1.4 ACHIEVEMENTS AND CONTRIBUTIONS ..... 8
*1.4.1 Publications* ..... 9
1.5 STRUCTURE OF THE THESIS ..... 11

**CHAPTER 2** ..... 12

LITERATURE REVIEW ..... 12
2.1 VIDEO SIGNALS ..... 12
*2.1.1 Luminance and colour components* ..... 12
*2.1.2 Picture formats* ..... 13
*2.1:3 Picture quality assessment* ..... 15
2.2 VIDEO CODING BASICS ..... 16
*2.2.1 Exploiting spatial correlations* ..... 16
*2.2.2 Exploiting temporal correlations* ..... 17
*2.2.3 Entropy Coding* ..... 18
2.3 CONVENTIONAL VIDEO CODING TECHNOLOGIES ..... 18
*2.3.1 History of video coding* ..... 19
*2.3.2 H.264/AVC Standard* ..... 20
2.4 DISTRIBUTED SOURCE CODING ..... 22
*2.4.1 Slepian-Wolf Theorem* ..... 23
*2.4.2 Wyner-Ziv Theorem* ..... 25
2.5 DISTRIBUTED VIDEO CODING ..... 26
*2.5.1 Complexity balance of the video codecs* ..... 26
*2.5.2 The DVC Codec* ..... 29
*2.5.3 Pixel-Domain DVC Architecture* ..... 30
*2.5.4 Transform-Domain DVC Architecture* ..... 44
*2.5.5 Recent developments in DVC* ..... 46
2.6 CONCLUSION ..... 50

**CHAPTER 3** ..... 51

DISTRIBUTED VIDEO CODEC BASED ON TURBO TRELLIS CODED MODULATION ..... 51
3.1 OVERVIEW OF TCM AND TTCM ..... 51
*3.1.1 Set partitioning* ..... 53
*3.1.2 Turbo TCM (TTCM)* ..... 55
3.2 CHANNEL CODING IN THE SLEPIAN-WOLF CODER ..... 55

3.3 DVC CODEC WITH TTCM ..... 56
*3.3.1 TTCM based DVC encoder* ..... *56*
*3.3.2 TTCM based DVC decoder* ..... *57*
*3.3.3 Codec integration* ..... *60*
*3.3.4 Simulations and Results* ..... *61*
3.4 CONCLUSION ..... 65

**CHAPTER 4** ..... **66**

UNIDIRECTIONAL DVC FRAMEWORK ..... 66
4.1 RATE CONTROL IN DVC ..... 67
4.2 ENCODER RATE CONTROL (ERC) ..... 68
*4.2.1 Pixel Scattering* ..... *69*
*4.2.2 Estimating the correlation noise at the encoder* ..... *74*
4.3 ITERATIVE REFINEMENT OF DECODED FRAMES ..... 75
*4.3.1 Video Distortion in Unidirectional DVC Systems* ..... *76*
*4.3.2 Iterative Refinement* ..... *77*
4.4 CONCLUSION ..... 88

**CHAPTER 5** ..... **90**

DVC PERFORMANCE IMPROVEMENT AT MODULAR LEVEL ..... 90
5.1 NON-LINEAR QUANTISATION ..... 90
*5.1.1 The pdf of Residual frames* ..... *91*
*5.1.2 Non-Linear Transformation* ..... *94*
*5.1.3 Simulation Results* ..... *96*
5.2 ENHANCED RECONSTRUCTION ALGORITHM ..... 99
*5.2.1 Reconstruction model* ..... *100*
*5.2.2 Pixel domain implementation* ..... *101*
*5.2.3 Transform domain implementation* ..... *108*
5.3 SIDE INFORMATION ENHANCEMENT BY ITERATIVE REFINEMENT ..... 112
*5.3.1 The refinement algorithm* ..... *112*
*5.3.2 Simulation Results* ..... *115*
5.4 CONCLUSION ..... 118

**CHAPTER 6** ..... **120**

DVC CODEC OPTIMISATION FOR WIRELESS CHANNELS ..... 120
6.1 DUAL CHANNEL SYSTEM ..... 121
6.2 NOISE MODELLING IN THE DUAL CHANNEL SYSTEM ..... 122
6.3 SIMULATIONS AND RESULTS ..... 127
*6.3.1 Simulation System Configuration* ..... *127*
*6.3.2 Analysis of Results* ..... *132*
6.4 CONCLUSION ..... 138

**CHAPTER 7** ..... **139**

CONCLUSIONS ..... 139
7.1 TECHNICAL CONTRIBUTIONS OF THE THESIS ..... 139
*7.1.1 Improving the Slepian-Wolf coding performance with TTCM* ..... *139*
*7.1.2 Unidirectional coding framework development* ..... *140*
*7.1.3 Enhanced functional modules for the DVC codec* ..... *140*
*7.1.4 Optimising DVC performance over wireless channels* ..... *142*
7.2 GENERAL CONCLUSIONS ..... 143
7.3 SUGGESTED FUTURE WORK ..... 143

**APPENDIX A** ..... **146**

**APPENDIX B** ..... **148**

**APPENDIX C** ..... **151**

**REFERENCES** ..... **156**

## LIST OF FIGURES

*Number* *Page*


Figure 1-1 A Generic Video Surveillance Scenario....3
Figure 1-2 Hybrid video codec scenario for mobile communications....4
Figure 2-1 Sampling pattern for 4:2:0 CIF....14
Figure 2-2 Evolution of video coding standards....19
Figure 2-3: AVC encoder....21
Figure 2-4: AVC Decoder....22
Figure 2-5: Distributed coding of two statistically dependent discrete random sequences....23
Figure 2-6: Achievable rate region following the Slepian-Wolf theorem....24
Figure 2-7: Wyner-Ziv theorem with side information....25
Figure 2-8: Complexity balance of conventional video coding....27
Figure 2-9: Example one-to-many type video communication scenario....27
Figure 2-10: Typical many-to-one application scenario....28
Figure 2-11: Complexity balance of distributed video coding....28
Figure 2-12 DVC codec: high level overview....30
Figure 2-13: Pixel domain DVC codec architecture using turbo coding....31
Figure 2-14: Quantization bin formation for $2^M = 4$....31
Figure 2-15: Bit plane extraction....32
Figure 2-16: Turbo Encoder....33
Figure 2-17: Example of an RSC encoder....33
Figure 2-18 Motion estimation, interpolation and compensation in DVC....36
Figure 2-19: Turbo Decoder....38
Figure 2-20 LLR convergence in turbo decoding....39
Figure 2-21 The Hypothetical channel model for side information....40
Figure 2-22: Reconstruction algorithm (4 quantization levels)....43
Figure 2-23: Transform domain DVC codec architecture using turbo coding....44
Figure 2-24 Transform coefficients....45
Figure 2-25 Quantization matrices....46
Figure 3-1 Typical PSK signal constellations....52
Figure 3-2 Rate 2/3 encoder diagram, and associated trellis for 8-PSK TCM....53
Figure 3-3 Set partitioning for 8-PSK TCM....54
Figure 3-4 Symbol mapping for 8-PSK TCM with a rate 2/3 convolutional encoder....54
Figure 3-5 TTCM Encoder....55
Figure 3-6 DVC encoder block diagram with TTCM....56
Figure 3-7 The RSC encoder in the TTCM encoder....56
Figure 3-8 DVC decoder block diagram with TTCM....57
Figure 3-9 4-PSK set partitioning....58
Figure 3-10 Sourcing the bit streams for symbol mapping....58
Figure 3-11 BPSK signal constellation for symbols when parity bit is punctured....59
Figure 3-12 Schematic diagram of the TTCM decoder....59
Figure 3-13 TTCM based DVC codec....60
Figure 3-14 Performance comparison for *Foreman* sequence....63
Figure 3-15 Performance comparison for *Carphone* sequence....63
Figure 3-16 Performance comparison using Gaussian noise model....64
Figure 3-17 Performance comparison using Laplacian noise model....65

Figure 4-1 Simplified DVC codec ..... 68
Figure 4-2 Randomising bit errors (*Foreman* sequence; first bit plane) ..... 70
Figure 4-3 DVC codec with pixel scattering of the input video frames ..... 71
Figure 4-4 Performance comparison of the modified codec for the *Foreman* sequence ..... 72
Figure 4-5 Performance comparison of the modified codec for the *Claire* sequence ..... 73
Figure 4-6 Comparison of a decoded frame from *Foreman* sequence: ..... 74
Figure 4-7 Correlation Noise Estimation at the Encoder ..... 75
Figure 4-8 Accumulation of bit errors on multiple bit planes (not to scale) ..... 76
Figure 4-9 Frame degradation due to bit errors in decoded bit stream ..... 77
Figure 4-10 Macroblock partitions of the input frames ..... 78
Figure 4-11 Unidirectional DVC codec with the refinement algorithm ..... 79
Figure 4-12 Flag-map for one bit plane ..... 80
Figure 4-13 Flagged pixel map for the frame ..... 81
Figure 4-14 Adjacent pixels available for spatial prediction of flagged pixel "X" ..... 81
Figure 4-15 Temporal search for flag-filling ..... 83
Figure 4-16 Rate distortion analysis for the *Foreman* sequence ..... 85
Figure 4-17 Rate distortion analysis for the *Salesman* sequence ..... 86
Figure 4-18 Effect of iterative refinement (*Foreman* sequence) ..... 86
Figure 4-19 Effect of iterative refinement (*Salesman* sequence) ..... 87
Figure 4-20 Effect of the block length selection at a bit rate of 163 Kbps (*Foreman* sequence). 87
Figure 5-1 The Wyner-Ziv residual frame formation ..... 91
Figure 5-2 Residual frame ..... 91
Figure 5-3 The pdf comparison of original and residual Wyner-Ziv frames ..... 92
Figure 5-4 Linear Quantization bin formation (8 levels) ..... 93
Figure 5-5 Non-Linear Quantization bin formation (8 levels) ..... 93
Figure 5-6 The non-linear transformation function ..... 94
Figure 5-7 DVC codec with Non-linear quantiser ..... 95
Figure 5-8 Performance comparison on non-linear quantiser (*Foreman* sequence) ..... 97
Figure 5-9 Performance comparison on non-linear quantiser (*Coastguard* sequence) ..... 97
Figure 5-10 Performance comparison on non-linear quantiser (*Hall Monitor* sequence) ..... 98
Figure 5-11 Proposed reconstruction model for 4 quantization levels ..... 101
Figure 5-12 Pixel domain unidirectional DVC codec with modified reconstruction block ..... 102
Figure 5-13 Scatter plot for the optimum recon angle for each block and the corresponding BER of the decoded bit stream ..... 104
Figure 5-14 Performance comparison of the proposed algorithm (Pixel domain coding, *Foreman*) ..... 106
Figure 5-15 Performance comparison of the proposed algorithm (Pixel domain coding, *Hall Monitor*) ..... 106
Figure 5-16 Performance comparison of the proposed algorithm (Pixel domain coding, *Foreman* sequence: frame no.3) ..... 107
Figure 5-17 Transform domain unidirectional DVC codec with modified reconstruction block ..... 109
Figure 5-18 Performance comparison of the proposed algorithm (Transform domain coding, *Foreman* sequence) ..... 110
Figure 5-19 Performance comparison of the proposed algorithm (Transform domain coding, *Coastguard* sequence) ..... 111
Figure 5-20 Performance comparison of the proposed algorithm (Transform domain coding) ..... 111
Figure 5-21 DVC codec with proposed side information refinement ..... 113

Figure 5-22 Side information refinement algorithm ..... 113
Figure 5-23 Performance Comparison of refinement algorithm for *Foreman* Sequence ..... 116
Figure 6-1 The dual-channel model ..... 121
Figure 6-2 The system block diagram with the channel transceiver ..... 127
Figure 6-3 Rayleigh fading channel model ..... 128
Figure 6-4 Block diagram of the rake receiver ..... 129
Figure 6-5 Performance comparison for the *Foreman* test sequence for the AWGN channel 133
Figure 6-6 Performance comparison for the *News* test sequence for the AWGN channel ..... 134
Figure 6-7 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, GOP=2 ..... 135
Figure 6-8 Performance comparison for the *News* test sequence for the W-CDMA channel, GOP=2 ..... 135
Figure 6-9 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, GOP=12 ..... 136
Figure 6-10 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, with motion search disabled for H.264/AVC, GOP=12 ..... 136
Figure 6-11 Performance comparison of the proposed technique with the existing decoding algorithm (*Foreman,* W-CDMA channel , GOP=2) ..... 137
Figure B-1 PSNR vs. frame no. plot ..... 148
Figure B-2 PSNR vs. Bit rate plot ..... 149
Figure B-3 Graphics for subjective quality assessment ..... 150

## LIST OF TABLES

*Number* *Page*

Table 2-1 Sampling formats: chrominance resolution as a percentage of luminance resolution13
Table 2-2 Extensions of CIF ..... 14
Table 2-3 Motion estimation in conventional and distributed video coding ..... 35
Table 5-1 Complexity analysis: Average decoding time per frame ..... 117
Table 6-1 Configuration Parameters of the W-CDMA Channel ..... 132

# LIST OF ACRONYMS

| | |
|---|---|
| AVC | Advanced Video Coding |
| AWGN | Additive White Gaussian Noise |
| BER | Bit Error Rate |
| bpp | bits per pixel |
| BPSK | Binary Phase Shift Keying |
| CIF | Common Interchange Format |
| DCT | Discrete Cosine Transform |
| DPCM | Differential Pulse Code Modulation |
| DSC | Distributed Source Coding |
| DVC | Distributed Video Coding |
| ERC | Encoder Rate Control |
| GOP | Group of Pictures |
| IDCT | Inverse DCT |
| iid | Independent and identically distributed |
| ISO | International Standards Organisation |
| ITU-T | International Telecommunications Union - Telecommunications standardisation sector |
| JVT | Joint Video Team |
| LDPC | Low Density Parity Check |
| LLR | Log Likelihood Ratio |
| MAD | Mean Absolute Distance |
| MAP | Maximum A-Posteriori |
| MFSE distance | Minimum Free Squared Euclidean distance |
| MMSE | Minimum Mean Squared Error |
| MPEG | Motion Picture Expert group |
| MSB | Most Significant Bit |
| MSE | Mean Squared Error |
| MSE distance | Minimum Squared Euclidean distance |
| PSK | Phase Shift Keying |
| PSNR | Peak-to-peak Signal to Noise Ratio |
| QCIF | Quarter CIF |
| QPSK | Quadrature Phase Shift Keying |
| RBS | Radio Base Station |
| RCPT | Recursive Convolutional Punctured Turbo |
| RGB | Red, Green, Blue (Colour signals) |
| RSC | Recursive Systematic Convolutional |
| SAD | Sum of Absolute Difference |
| SISO | Soft-In-Soft-Out |
| SNR | Signal to Noise Ratio |

| | |
|---|---|
| SQCIF | Sub QCIF |
| TCM | Trellis Coded Modulation |
| TTCM | Turbo TCM |
| UDVC | Unidirectional DVC |
| VCEG | Video Coding Expert group |
| WCDMA | Wideband Code Division Multiple Access |

*Chapter 1*

# INTRODUCTION

In recent times, video services are increasingly becoming an indispensable utility for the human lifestyle. To enable this tremendous growth in demand, many technological developments have been achieved through research in video coding and compression leading to the standardization efforts of ITU-T VCEG and ISO/IEC MPEG. Almost all of these approaches focused on an encoder intensive coding framework with computational complexity of encoder been 5 to 10 times higher compared to that of the decoder [1]. This complexity is mainly attributed to the source correlation exploitation algorithms, including motion estimation and compensation functions performed in the encoder. This encoder-intensive framework was ideally suitable and motivated by the common one-to-many applications where the encoded video stream from one transmitter is received and viewed by many.

However, this coding paradigm is increasingly challenged by a class of new applications that demand a lower complexity at the encoder possibly at the cost of increased computational burden to the decoder. Some of the application scenarios that would potentially benefit from this modified complexity balance include: wireless low-power surveillance and multimedia sensor networks, wireless PC cameras, disposable video cameras and mobile camera phones. Distributed Video Coding (DVC) has emerged as a new technological breakthrough with computationally very flexible architecture that would ideally fulfil the demands of this increasingly popular class of applications.

DVC is currently identified as a premature technology, yet with immense potential for the future. The research activity pertaining to this thesis intends to push forward the concept of DVC into reality, possibly targeting at future standardization exercises and commercial deployments. This chapter will introduce the background and motivation leading to this research followed by the research environment and the associated objectives and achievements.

## 1.1 Background and motivation

The principles of distributed source coding were originally presented in the 1970s, in the form of Slepian-Wolf theorem [2] published in 1973 and Wyner-Ziv theorem [3] published in 1976. However, the potential applicability of distributed source coding principles for video coding was not realized for three decades. The initiatives in developing a video coding technology based on this concept have been unveiled in 2002. Ironically, it came simultaneously in two independent proposals in [4] and [5]. Since then, DVC research has come a long and productive way gathering momentum on the way through, as evident from the amount of related publications found in literature over the past years. The enthusiasm on this research area could be attributed to a number of attractive and promising features of DVC, which could be identified as:

- The flexibility in manoeuvring the algorithmic and computational complexity of the video encoder, possibly towards ultra low complex encoders that are beyond the scope of conventional video coding technologies

- The notably high error resilience demonstrated by the DVC encoded bit streams in the face of noisy channel conditions [6]

Through the above features, DVC has promised to open-up new horizons for the design of video encoder-sensors that are possibly of very low cost, size and power consumption. This will predictably benefit existing and new application scenarios that demand video sensor deployments in vast numbers, rapid deployment options and usage over less reliable transmission media. A few of the potential applications of DVC that exploit the above features are discussed next.

### *1.1.1 DVC application scenarios*

It is worth identifying some of the potential application scenarios of DVC in order to understand the motivation and the driving force behind this research.

**Video surveillance**

With the interest on public safety at its best; video surveillance has arguably become an irresistible component of modern life over the past years. Ranging from the monitoring of safety critical and high security installations, through urban areas, public transport links, etc., it

has penetrated all the way into domestic monitoring applications. However, it is noted that the cost of video sensors is still a significant problem affecting pervasive deployments in surveillance applications. On this backdrop, it is deemed necessary to look at key characteristics of a video surveillance scenario. These systems are mainly characterised by a large number of video sensors that covers the interested area. These video cameras capture the video data and upload to centralized servers for processing and displaying or storing for the real-time and offline monitoring purposes. A generic set up is illustrated in Figure 1-1. Generally, since the potential of sharing the central processing unit is very high; the cost of encoders is a significant factor affecting the wide utilization of surveillance systems. This situation holds valid for other similar applications such as monitoring of the elders, disabled people, and the children, disaster zone and traffic monitoring as well.

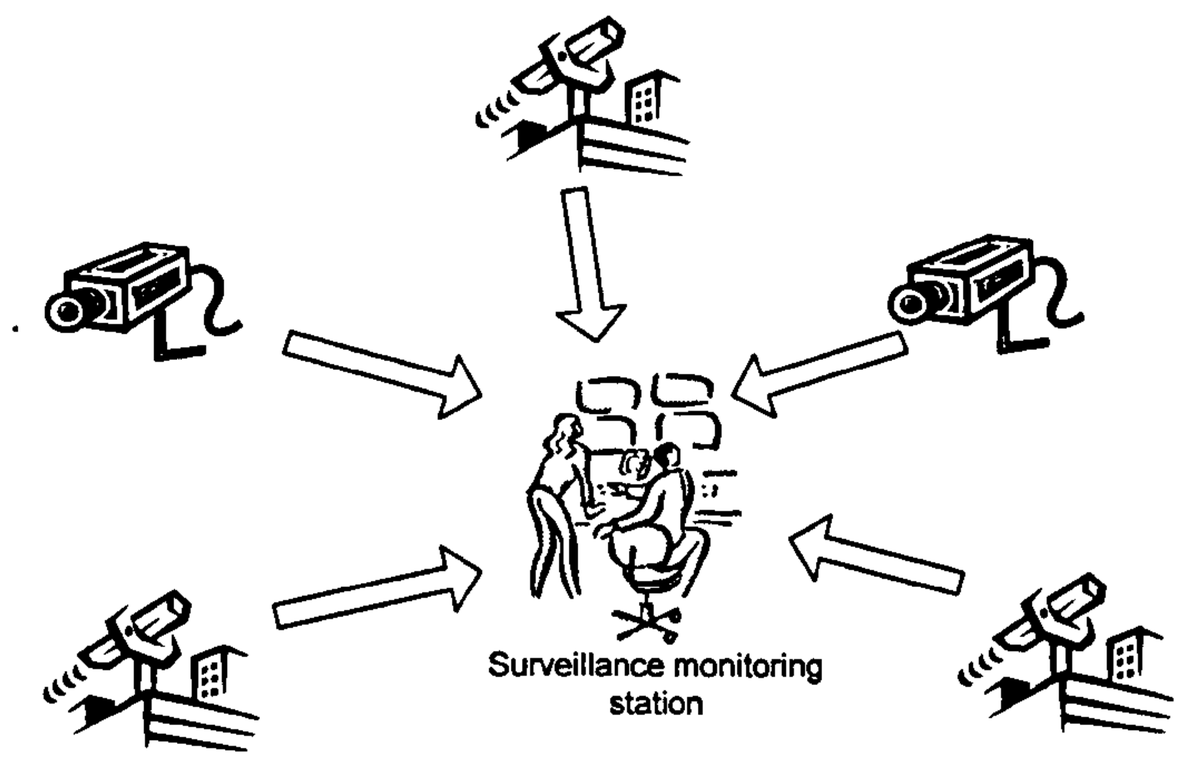



Figure 1-1 A Generic Video Surveillance Scenario

**Wireless Mobile Communications**

Wireless mobile video communication is another application that has a potential demand for a low cost video encoder, since each and every mobile handset has to enclose a video encoder and a decoder. Also it is vital that the codec can be implemented in a relatively small footprint inside a mobile handset due to the fact that the dimensions and the weight of a handset are very important to the consumer, upon whom the marketability of the product depends. The power consumption is yet another key consideration in mobile handsets. Therefore, the optimum video codec for a mobile handset is a one that neither requires a large space nor is

expensive and consumes minimum amount of power.. A possible economical hybrid solution is illustrated in Figure 1-2.

In mobile communications, a very large number of users share a single radio base station (RBS) while only a small percentage are active. The complexity of the video encoder (in video capturing) and decoder (in video display) hardware in the mobile unit is important in deciding the price, size and power consumption of the equipment. A possible solution would be to use a combination of DVC and conventional coding techniques; DVC encoder and a conventional decoder. In this scenario, very simple low cost DVC encoder and the equally low complex conventional decoder are placed in the handset while the complex segments of both approaches are pushed to the RBS side of the link. Transcoding capabilities from DVC to conventional coding formats and vice-versa need to be integrated to the RBS. This would obviously incur an additional cost at the central installation, but it is understood that this investment would be well justified by the fact that the central hardware are heavily shared considering the very low activity ratios apparent in typical mobile usage.

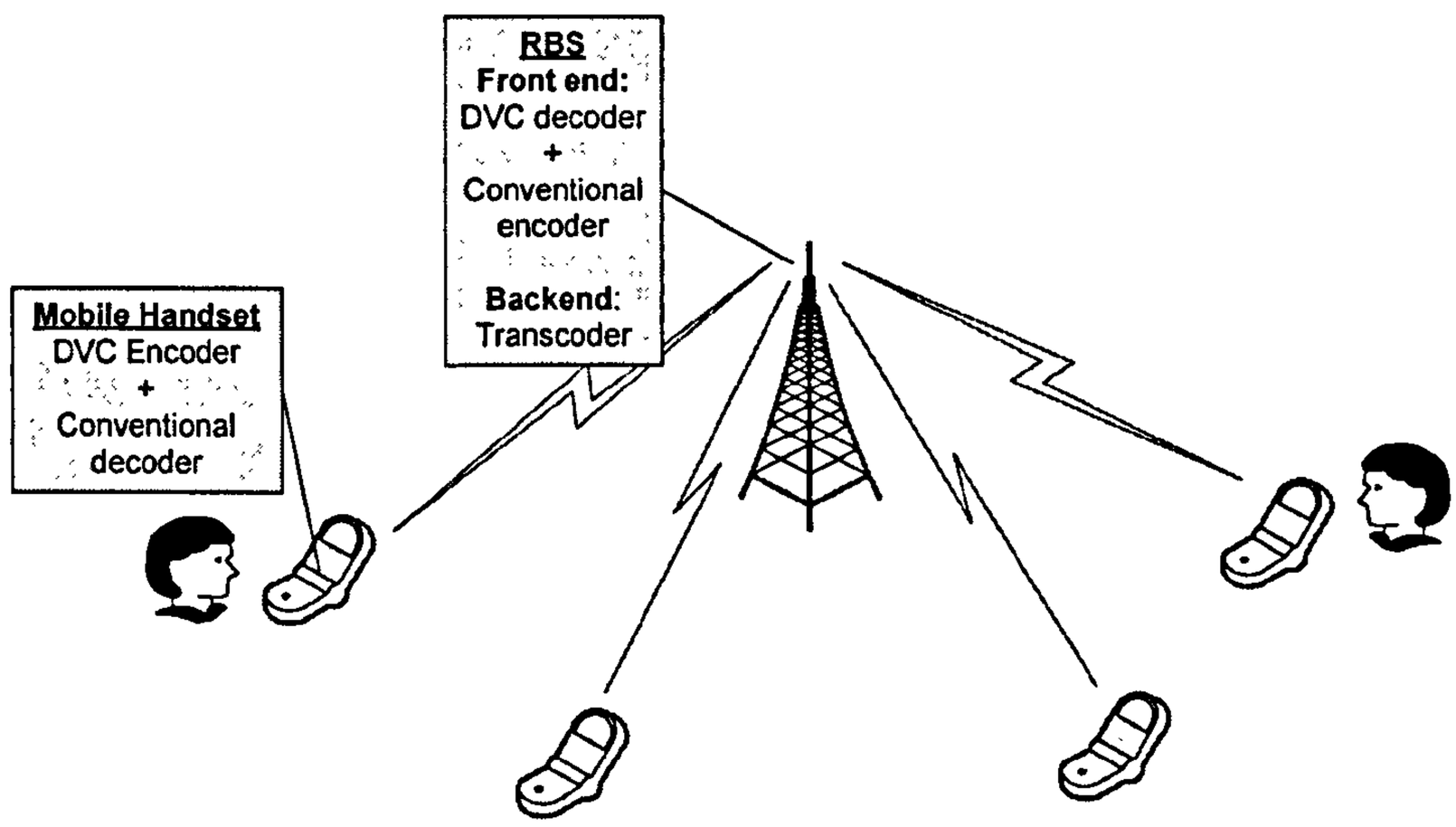



Figure 1-2 Hybrid video codec scenario for mobile communications

**Disposable video camera**

Disposable video camera is another concept that could efficiently exploit the low-encoder-complexity feature of DVC. The underlying motive in the concept of disposable video camera is to produce a video camera at a very affordable cost for one-off use. After using, the camera need to be taken into a designated sales outlet where the data captured into the storage in the unit is processed. Once this data is transferred into a DVD or a similar easily accessible media, the unit needs to be disposed off. Some of the key considerations in this process could be identified as: (i) the centralized equipment used for post-processing the captured data, at possibly the sales outlet, will be re-used with thousands of disposable cameras, (ii) the video encoding algorithm used in the camera need to be efficient enough to minimize the compressed video storage requirements of the device, and most importantly, (iii) the complexity of the video encoder algorithm has to be sufficiently low for the manufacturing cost of the device to be maintained low for it to be affordably priced. Simple encoder structure further helps conservation of battery power.

### *1.1.2 Motivation*

On the background of DVC, this research activity is motivated by a number of factors:

- The attractive features of DVC that are unique in the sense of the capabilities and properties of parallel technologies
- The current and predicted increasing demand for the application domains targeted by DVC
- The challenging nature of the proposed research, particularly in a relatively untouched area as was the case with DVC at the time this research was commenced in 2005.

## 1.2 Research method

This research was commenced at a time that the concept of DVC was only starting to establish; basic codec framework proposed, but no major developments were evident. On this backdrop, some of the design and testing approaches followed throughout this research activity are discussed in this section.

### *1.2.1 Choice of baseline technology*

The work mainly involved computer simulations. The baseline DVC codec was selected as per the latest published literature and the innovative algorithms proposed in this thesis were implemented and tested on top of this baseline codec. As the state of the art evolved through the parallel research activities in the community, the baseline is accordingly updated. As an example of this baseline evolution, consider the *side information*[1] generation algorithms. Pixel level interpolation was an initial and effective proposal for generating side information [4] and thus is followed as the basis for implementing a number of proposed algorithms in this thesis. However when more efficient algorithms (e.g. motion interpolation) were established through peer research, they were adopted for the baseline system. Note here that as long as the proposed innovations are considered to be independent of the evolved factor (the type of side information, in the said example) the original simulations are presented as it is in the thesis. However, if there is a significant dependency is detected, the proposed innovations are re-simulated on the evolved baseline system.

Another key choice of approach in DVC is between the pixel domain and transform domain coding. Both these choices are still active in the research field, even though recent preference in transform domain coding owing to better compression efficiency, albeit the increased computational complexity particularly in the encoder. In this regard, most of the novel concepts and designs discussed in this thesis are implemented on pixel domain DVC architectures, for its drastically low complex, and hence application-favoured feature. However it is understood that these concepts would produce similar performance enhancements in the transform domain frameworks. This fact is verified when some of the selected proposals are tested on both cases as elaborated later in this thesis.

It is also noted that the first half of this research was conducted at Brunel University, where the baseline codec was developed in-house. During the latter half, the research was carried out as a partner of the VISNET II project at the University of Surrey, thus the innovations were implemented over the VISNET II DVC codec. Accordingly the Chapters 3, 4 and 6, and section 5.3 use the in-house codec at Brunel, whereas the sections 5.1 and 5.2 use the VISNET II DVC codec at Surrey.

---

[1] *Side information* is an information stream estimated at the DVC decoder. This process formulates a vital functionality in the DVC codec and a detailed description is given in section 2.5.3.3.

### *1.2.2 Simulations and test conditions*

The innovative algorithms discussed in this thesis are implemented in the computer based baseline codec as discussed in previous section. The choice of test conditions for the simulations is generally based on the reference results in literature to which the proposed results are compared with. Therefore, in keeping the compatibility with the compared results, the adopted test conditions vary significantly throughout this thesis. The appropriate conditions are documented before each simulation result. Some of the key parameters of test conditions are discussed below.

**Video sequence**

There are a number of standard test video sequences that are used in the simulations. All videos are in QCIF format (176×144 pixels; YUV 4:2:0). A full list of sequences and the relevant details are given in the *APPENDIX A*. As explained above, the appropriate sequences are selected for each test case, based on the reference literature considered.

**Averaging properties over sequence**

The PSNR is usually calculated only for the luminance signals in DVC literature and the same practice is followed here.

In some test cases presented in this thesis, the PSNR and bit rate values are averaged only for the Wyner-Ziv frames of the sequence, again following the original practice in DVC literature. However, once H.264/AVC intra frame coding is established as the standard choice of technology for the key frame coding, as evident in the DISCOVER [7] and VISNET [8] European NoE projects, the averages of all frames including both Wyner-Ziv and key frames are considered.

### *1.2.3 Presentation of results*

In this thesis, the test results are usually presented in three formats:

- PSNR vs. Frame number plots showing objective video quality
- PSNR vs. Bit rate plots showing objective video quality
- Snapshots of frames showing subjective video quality

A detailed description of these techniques and corresponding formats are presented in *APPENDIX B*.

## 1.3 Objectives and scope of the research

The goal of the research scope of this thesis is moving the DVC concept towards a commercially viable technology that is capable of producing competitive solutions for low cost video surveillance and similar applications. It is believed that this could be achieved through the design of an efficient and robust coding framework and enhancing the performance of the constituent functional modules. In view of this objective, the specific objectives of the thesis are:

- Design of robust frameworks for both feedback-based and unidirectional DVC
- Enhancement of the coding efficiency of the Slepian-Wolf codec by considering alternate channel coding techniques
- Establishment of low cost encoder-rate-control algorithms for unidirectional DVC
- Optimisation of the performance of both feedback-based and unidirectional frameworks by enhancing the functional modules
- Enhancement of the error resilience of the DVC codec operating over wireless channels

## 1.4 Achievements and contributions

As a result of the research that is presented in this thesis, number of original achievements and contributions has been made to the field of DVC supported by a number of related publications as listed later in this section. The most significant achievements amongst them can be identified as:

- ✓ Turbo Trellis Coded Modulation (TTCM) based design of the Slepian-Wolf codec with improved coding efficiency

- ✓ Iterative frame refinement algorithm for improving the decoded picture quality in the unidirectional DVC framework, with no computational burden to the encoder, or no additional bit rate consumption

- ✓ Non-Linear quantization algorithm for improving the compression efficiency of the DVC codec

- ✓ Enhanced reconstruction algorithm, optimized for unidirectional coding framework, to enhance the quality of reconstructed video frames. This also is achieved without additional bit rate cost or encoder complexity.

- ✓ Side information optimization by iterative refinement, for improving compression efficiency of the codec

- ✓ Optimisation of the DVC decoding algorithm for improving error resilience of the coding scheme, for robustness in performance when the encoded bit stream is transmitted over error prone wireless channels

### *1.4.1 Publications*

Some selected publications related to the technical proposals discussed in this thesis are listed below. A full list of publications is given in APPENDIX C.

**Journals:**

1. W. A. R. J. Weerakkody, W. A. C. Fernando, A. B. B. Adikari, "Unidirectional Distributed Video Coding for Low Cost Video Encoding," IEEE Trans. Consumer Electronics, vol. 53, issue 2, pp. 788 -795, May 2007.

2. W. A. R. J. Weerakkody, W. A. C. Fernando, A. B. B. Adikari and R. M. A. P. Rajatheva, "4-PSK TTCM for Wyner-Ziv frame coding in DVC," IEICE Electron. Express, Vol. 3, No. 8, pp.178-183, 2006.

**Conference proceedings:**

3. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, " Distributed video coding of Wyner-Ziv frames using Trellis Coded Modulation," Proceedings of IEEE CCECE 2006, pp 314 – 317, Ottawa, Canada, May 2006.

4. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, R.M.A.P. Rajatheva, "Distributed video coding of Wyner-Ziv frames using Turbo Trellis Coded Modulation," Proceedings of IEEE ICIP 2006, Atlanta, USA, Oct 2006.

5. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, "Distributed video coding of pixel interleaved Wyner-Ziv frames," Proceedings of IEEE ICIIS, Sri Lanka, Aug. 2006.

6. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, "Distributed Video Coding of Pixel Interleaved Wyner-Ziv Frames with Iterative Decoding," Proceedings of WPMC 2006, San Diego, USA, September, 2006.

7. W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, "An Iterative Refinement Technique for Side Information Generation in DVC," Proceedings of IEEE ICME 2007, Beijing, China, July 2007.

8. W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, S.Adedoyin, "Map Decoder Design for Distributed Video Coding in Error Prone Wireless Channels," Accepted for the Proceedings of IEEE ICIIS 2007, Sri Lanka, Aug 2007.

9. W.A.R.J. Weerakkody, W.A.C. Fernando, R.M.A.P. Rajatheva, T.Kuganeswaran, K. Muraleetharan, "Distributed Video Coding With MIMO for Video Wireless Communications," Proceedings of IEEE ICSPC 2007, Dubai.

10. W.A.R.J. Weerakkody, W.A.C. Fernando, A.M.Kondoz, "An Enhanced Reconstruction Algorithm for Unidirectional Distributed Video Coding," Accepted for the Proceedings of IEEE ISCE 2007, Algarve, Portugal, April 2008.

**Under Review:**

**Journals:**

11. W.A.R.J.Weerakkody, W.A.C. Fernando, F.Pereira, A. M. Kondoz, "An Enhanced Reconstruction Algorithm for Unidirectional Distributed Video Coding," Submitted to EURASIP journal, submitted May 2008.

12. W.A.R.J. Weerakkody, W.A.C. Fernando, A.M. Kondoz , "Modifying the Turbo Decoder for DVC based Video Codecs," Submitted to IEICE Trans. on Fundamentals, Submitted June 2008.

13. W.A.R.J.Weerakkody, W.A.C. Fernando, M.B.Badem, A.M.Kondoz, "Non-Linear Quantization for Pixel Domain Distributed Video Coding," Submitted to IET Electronics Letters, submitted June 2008.

## 1.5 Structure of the Thesis

This thesis consists of 7 chapters and the rest of the thesis is structured as follows.

The second chapter presents technical review of video coding in general, the underlying principles of DVC and the currently available DVC frameworks, followed by a brief discussion on recent developments. Then the next chapter discusses a novel proposal for the use of Turbo Trellis Coded Modulation (TTCM) for DVC replacing the commonly used turbo coder. The fourth chapter is based on the unidirectional DVC framework that eliminates the reverse feedback channel involved in the basic coding framework. A number of practical unidirectional DVC algorithms are discussed in this chapter. Fifth chapter introduces a number of novel algorithms to the DVC architecture to enhance the coding efficiency. The sixth chapter discusses the implications of the transmission of DVC encoded video streams over wireless channels and introduce an enhanced decoding algorithm to improve the error resilience of the codec. The last chapter summarizes technical contribution of this thesis and gives concluding remarks.

*Chapter 2*

# LITERATURE REVIEW

This chapter presents an overview of the existing technologies that this research is built upon, starting with some fundamental principles of video signals and video coding. Then the discussion is extended to an introduction to the distributed source coding principles, the underlying theoretical base of the DVC technology. An architectural overview of the currently available DVC frameworks is then presented followed by a review of some of the significant and related recent developments in DVC research.

## 2.1 Video Signals

Before studying the video coding and compression methodologies, it is important to understand a few vital terminologies and concepts that are utilised throughout the thesis. Accordingly, this section presents the colour component definitions, picture formats, and picture quality assessment techniques, with particular attention to those adopted in the simulations presented here.

### *2.1.1 Luminance and colour components*

The raw video signals captured by cameras are generated in three colour signals: red, green and blue; thus forming the common name RGB. However, to provide the compatibility with black and white video signal processing and also based on the high correlation of the R, G and B components, the RGB signals are usually transformed into a new colour space, with a luminance component and two chrominance components. As per the popular PAL standard colour system, the $Y$ (luminance), $U$ and $V$ (chrominance) components are computed from the gamma-corrected RGB components ($R'$, $G'$ and $B'$ respectively) using the below formulas [9]:

$$Y = 0.299R' + 0.587G' + 0.114B' \qquad 2\text{-}1$$

$$U = -0.147R' - 0.289G' + 0.436B' = 0.492(B' - Y) \qquad 2\text{-}2$$

$$Y = 0.615R' - 0.515G' - 0.100B' = 0.877(R' - Y) \quad 2\text{-}3$$

It is further noted that for the digital video signals the ITU-R Recommendation BT.601 (formerly CCIR-601) has defined a $YC_bC_r$ colour system, slightly offset from the above formulas [9]. This enforces a range limiting of the $Y$ component to 16-235 quantum levels, and also a chrominance component distribution centred on 128, the grey level, with a range of 16-240.

$$Y = 0.257R' + 0.504G' + 0.098B' + 16 \quad 2\text{-}4$$

$$C_b = -0.148R' - 0.291G' + 0.439B' + 128 \quad 2\text{-}5$$

$$C_r = 0.439R' - 0.368G' - 0.071B' + 128 \quad 2\text{-}6$$

### *2.1.2 Picture formats*

The picture formats utilised for video signals vary significantly based one the requirements of the applications. For example, video conferencing and video telephony usually demand only low degree of resolutions, whereas high definition television inevitably demands larger frame sizes with higher luminance and chrominance resolution. To fulfil these varying requirements, a series of frame formats have been defined. Each frame format generally defines a number of parameters including the spatial resolution for the luminance and chrominance components and the temporal resolution (frame rate). The spatial resolution for the chrominance signals is usually reduced compared to that of the luminance signal following the nature human visual perception. The *sampling format* defines the percentage of chrominance resolution compared to the luminance resolution as listed in Table 2-1.

Table 2-1 Sampling formats: chrominance resolution as a percentage of luminance resolution

| Sampling format | Horizontal % | Vertical % |
|---|---|---|
| 4:4:4 | 100 | 100 |
| 4:2:2 | 50 | 100 |
| 4:2:0 | 50 | 50 |
| 4:1:1 | 25 | 100 |

**CIF**

In worldwide teleconferencing following the ITU-R Recommendation BT.601, it is important to convert between the various frame formats adopted for example between European (625 line, 50Hz) and North American and Far East (525 line 60Hz). The Common Interchange Format (CIF) has been defined to enable this type of conversions between different standards [9]. This defines a sampling format of 4:2:0 as illustrated in Figure 2-1.

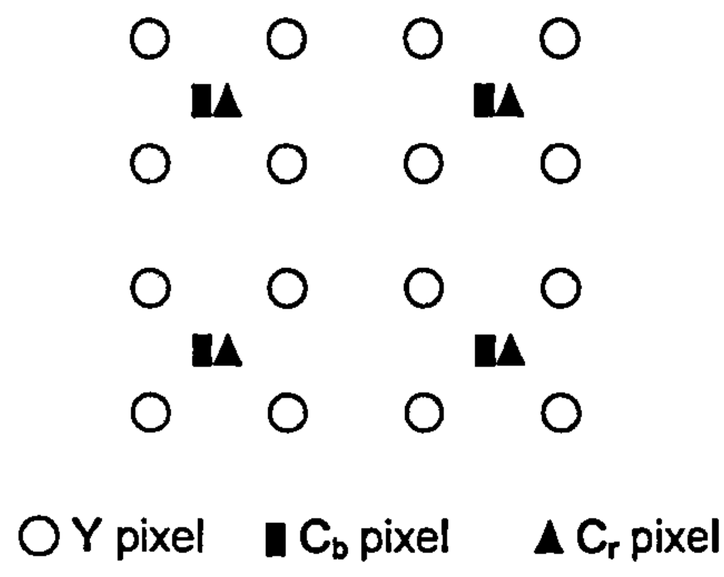



Figure 2-1 Sampling pattern for 4:2:0 CIF

A number of extensions of CIF have been defined for certain applications. These define the spatial resolution (luminance signal) as listed in Table 2-2.

Table 2-2 Extensions of CIF

| Format | Luminance Resolution |
|---|---|
| SQCIF | 128×96 |
| QCIF | 176×144 |
| CIF | 352×288 |
| 4CIF | 704×576 |
| 16CIF | 1408 × 1152 |

The research discussed in this thesis utilises standard test sequences of Quarter CIF (QCIF) frame format for the simulations, as followed commonly in most current DVC literatures.

### *2.1.3 Picture quality assessment*

Since this research focuses on lossy coding approaches, it is useful to get an understanding of how to quantify the resultant distortion in the decoded video streams. There are primarily two approaches to the picture quality assessment:

- Subjective quality assessment: measures the human perception of the video quality. The decoded (degraded, in the case of lossy coding) video stream is shown to a number of subjects and there perception is recorded. Then statistical analysis of the collected results is performed to derive the quality parameters.

- Objective quality assessment: mathematically calculate the quality parameters. There is no variability depending on the perception of the individual users.

The simplest form of objective quality measurement is the ratio of the peak-to-peak of signal to the root mean squared processing noise. Thus this parameter is called the peak-to-peak signal to noise ratio (PSNR) and is computed as:

$$PSNR = 10\log_{10}\left[\frac{255^2}{\frac{1}{N}\sum_{all\ i}\sum_{all\ j}\left(Y_{ref}(i,j)-Y_{prc}(i,j)\right)^2}\right] \qquad 2\text{-}7$$

where $Y_{ref}(i,j)$ and $Y_{prc}(i,j)$ represent the pixel values of the reference and decoded pictures respectively and $N$ is the total number of pixels in the picture. It is noted that there is some criticism about the validity of PSNR as a true representation of the picture quality. This is mainly due to ill-reflection of local quality degradations in small patches in the frame. If the degraded area is small, the PSNR is not significantly penalized but the viewers may be attracted to the degraded patch and perceive it as a significant loss of quality.

Even though PSNR has its own drawbacks, it has been widely utilised in evaluating the coding efficiency mainly due to the simple and straightforward calculation. The same practice has been adopted in this thesis. However, wherever possible, graphical images of the decoded frames have been produced against the reference images to help evaluate the user perception of the processed video signals.

## 2.2 Video Coding Basics

The primary goal in video coding is to reduce the redundant information in order to achieve compression. Three fundamental approaches are usually taken to accomplish this goal:

- Spatial redundancy reduction, exploiting the correlations of similar pixels within a frame
- Temporal redundancy reduction, exploiting the similarities between successive frames
- Entropy coding, for reducing redundancies between compressed data symbols

*Intra frame coding* and *inter frame coding* are two important concepts involved in video coding. Intra frame coding refers to the exploitation of redundancies within each individual frame in a sequence, or in effect coding still images. This effectively means exploiting the spatial correlations. On the other hand, inter frame coding is the process of eliminating the temporal redundancies, i.e. the similarities between the successive frames. A conceptual overview and implementation techniques of each of above is given in the following sections.

### *2.2.1 Exploiting spatial correlations*

Exploiting the spatial correlations typically utilise two techniques: Predictive coding and Transform coding.

**Predictive coding**

This involves the prediction of each pixel or group of pixels using the previously coded information on the same frame (intra frame predictive coding) or adjacent frames (inter frame predictive coding). In general the neighbouring pixels would make the most contribution in this process due to the high spatial correlations. A prediction error is then generated and coded by comparing the predicted picture with the original picture. In concept, this is a lossless coding technique.

Predictive coding is not a common option in DVC, where the source redundancies are exploited at decoder and the original frame is not available in the process, as elaborated later in section 2.5. However, some spatial prediction techniques have been recently proposed to refine the frame estimation at the decoder (e.g. [29][37]) to improve the coding efficiency.

**Transform coding**

In transform coding, the source data in pixels are mapped to a transform domain. In most cases, the image energy in the transform domain is concentrated in the low frequency region, i.e. in only a few transform coefficients. Therefore, transform coding, used in conjunction with *quantisation,* creates an opportunity for data compression. However, since the quantised coefficients (some low significant coefficients are totally discarded in this process) cannot be reconstructed, this is a lossy coding technique.

It is noted that transform coding is a feasible compression method for use in the DVC framework. Even though transform coding adds a computational burden to the DVC encoder, which otherwise has very low complexity, this is treated as a viable cost in many cases.

**Quantisation**

The quantisation of the transform coefficients is a technique used to yield a compression, since the domain transformation alone does not provide any compression. Once the image energy is unevenly distributed in the transform coding, the quantisation levels are defined for each coefficient band to reflect its significance in the reconstructed image. This process introduces the so-called quantisation noise.

In the context of DVC, quantisation plays a significant role, both in the pixel domain and transform domain coding frameworks. The DVC codec incorporates a special *reconstruction* algorithm at the decoder to partly compensate for the quantisation noise. Particular attention to the quantisation process is paid later in this thesis in the design of non-linear quantisation algorithm for DVC, in contrast to the widely used linear scalar quantiser.

### *2.2.2 Exploiting temporal correlations*

Temporal redundancy reduction is an inter frame coding technique that involves estimating the motion in a video sequence.

**Motion estimation**

The motion activity between consecutive frames of a video sequence is typically estimated by block matching. For this purpose the frames are generally divided in to an array of rectangular (can be square) blocks of say $M \times N$ pixels. Then this block on the current frame is compared

with candidate blocks of same dimensions on the reference frame, within a pre-defined window (search window). The search window is larger than the block size and usually centred on the same coordinates with the block on the current frame. Once the *best match* is located, the spatial displacement between the matching blocks on the current and reference frames derives the *motion vector*. There are a number of measures used to derive the *best match*: mean squared error (MSE), sum of absolute difference (SAD) etc.

The block search performed on all blocks on the current frame produces the motion vector map, which is then used to motion compensate with the reference frame to generate the predicted frame.

A motion estimation and compensation process is utilised in DVC for the same purpose; i.e. to exploit temporal correlations, but the design details of the algorithm are notably different mainly due to the unavailability of the *current frame* in the motion estimation process, as discussed in details later in section 2.5.3.3.

### *2.2.3 Entropy Coding*

Entropy coding is a lossless data compression technique that reduces the redundant information in a bit stream, which are usually independent of the spatial or temporal redundancies in a video sequence. The most common entropy coding techniques are Huffman coding and arithmetic coding. These algorithms check for recurring patterns in the bit stream and represent them with code words.

## 2.3 Conventional Video Coding Technologies

The term *conventional* video coding has been utilised across this thesis to identify the video coding approaches that do not follow the distributed source coding principles. Since DVC marked a new dimension to video coding with the independent-encoding-joint-decoding concept that enabled flexible codec architecture, anything that follows the joint-encoding-joint-decoding concept are grouped and labelled as conventional technologies. At this point, an appreciation of the historical perspective of video coding is deemed helpful to set the background of the proposed research. It is noted that the state of the art in conventional video coding technologies is H.264/AVC. Therefore, an overview of the H.264/AVC standard is also added.

### *2.3.1 History of video coding*

Initial attempts of video coding and communications systems date back to the 1960s in the form of an analogue video phone system. But it was not until the mid-late 1980s that a formal and organised video codec standardisation exercise is reported. An illustration of the historical evolution time line of video coding standards is shown in Figure 2-2.

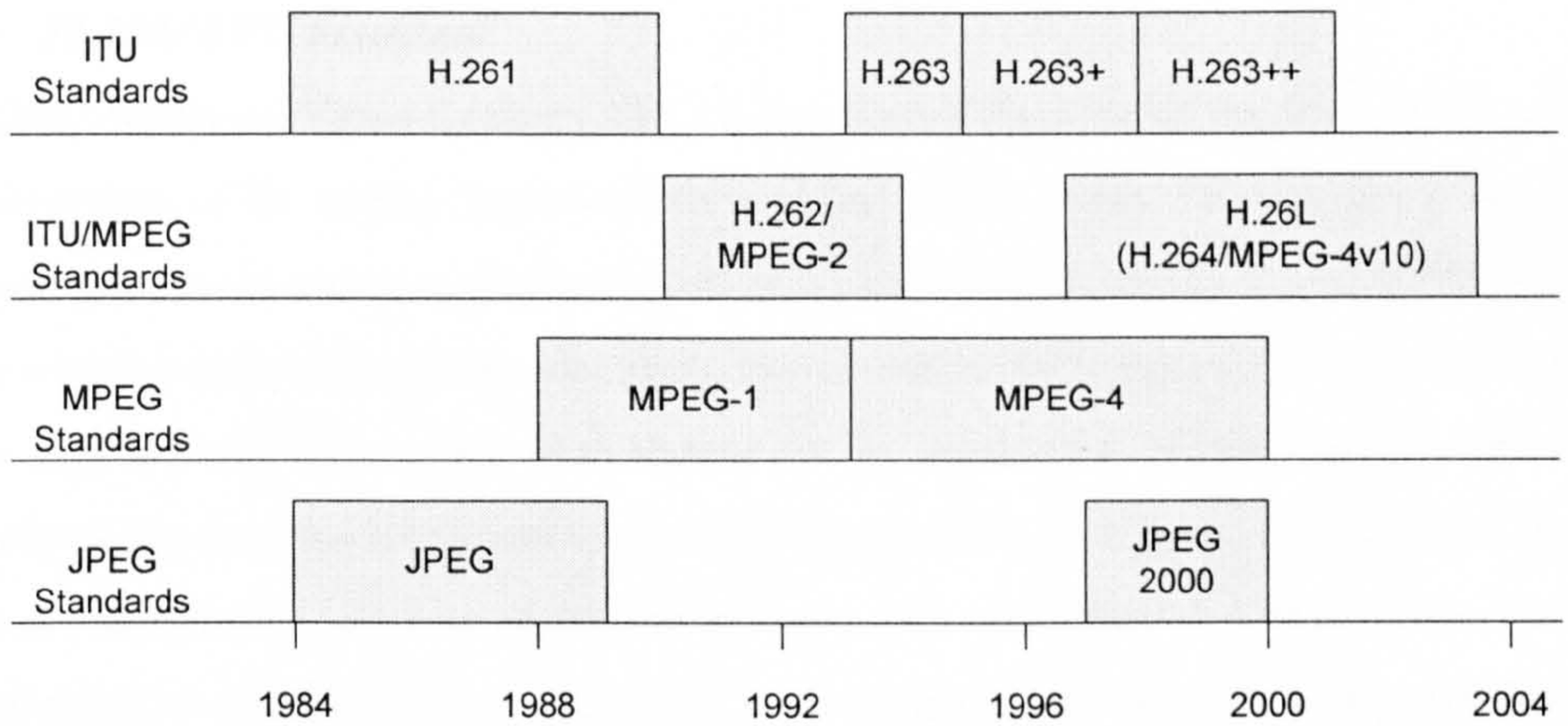



Figure 2-2 Evolution of video coding standards

The H.261 standard was thus referring to a combination of the inter frame differential pulse code modulation (DPCM) and discrete cosine transform (DCT), originally intended at video conferencing applications at 384 kbps; later extended to other bit rates. Motion Picture Expert group (MPEG) started investigations in early 1990s on video coding for storage, such as CD-ROMs. This resulted in the MPEG-1 standard. With MPEG-1 optimised for non-interlaced video at 1.2-1.5 Mbit/s rates, a new generation of standard then emerged for coding interlaced video at higher bit rates (4-9 Mbit/s) called MPEG-2. The latter became a very popular standard utilised for terrestrial and satellite TV broadcasting the DVDs. The same standard was later adopted by the telecommunications standardisation sector of the International Telecommunication Union (ITU-T) under the generic name H.262 due to its potential for scaling down.

The identification of the potential applications in the very low bit rate domain such as 64 Kbit/s or less then paved the way for the MPEG-4 standard. With its object based coding approach, MPEG-4 opened up new horizons for coding scalability and video synthesis. Around the same time ITU-T carried out work on a new standard H.263 for similar target

applications. The evolutions of this standard with improved compression efficiency were named as H.263+ and H.263++. Based on the prevailing achievements, a new collaboration called the Joint Video Team (JVT) was formed in 1997, bringing together video coding expert groups of ITU-T and ISO/IEC MPEG. The joint effort resulted in the H.264/MPEG-4v10 standard, the state of the art in video coding.

### *2.3.2 H.264/AVC Standard*

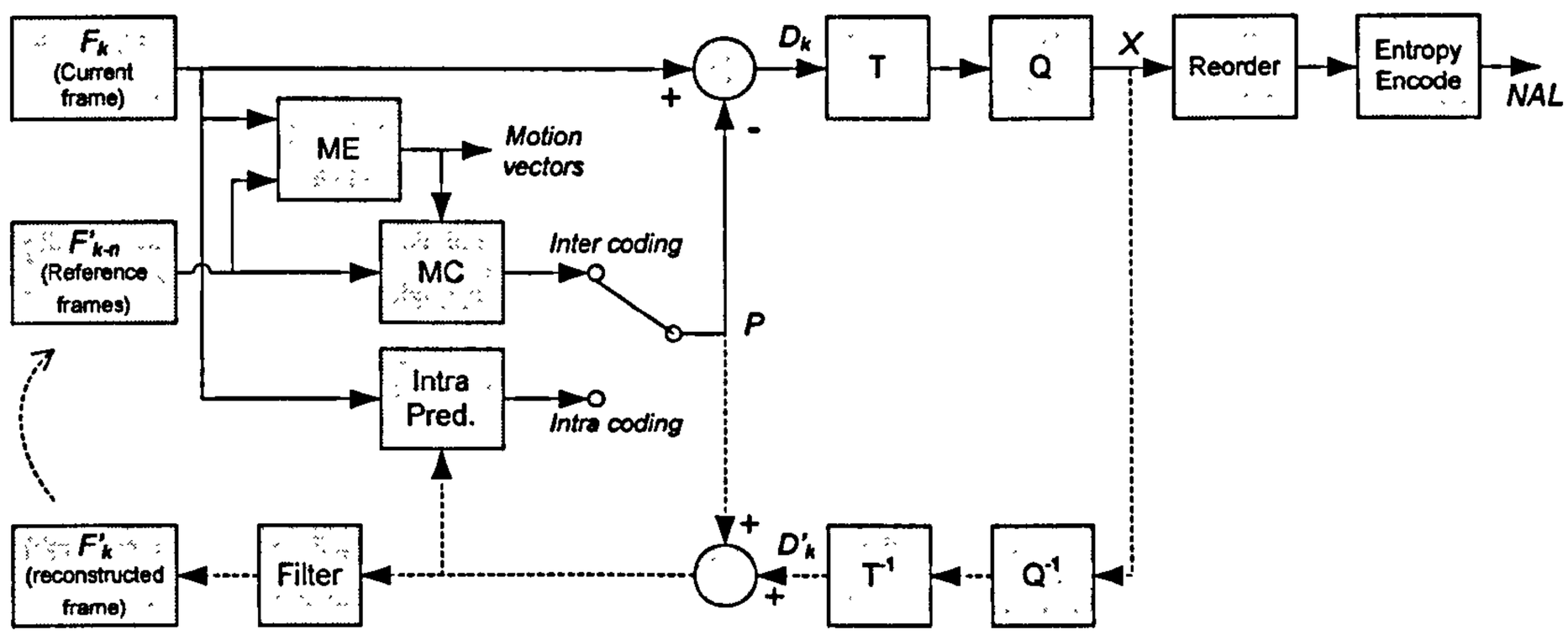
As H.264/Advanced Video Coding (AVC) is considered the state of the art in video coding, a brief overview of its coding framework is justified at this point. Historically, it has been a common practice in video coding standards (such as MPEG1, MPEG2 and MPEG4) not to define a codec (encoder and decoder pair), but to define the syntax of an encoded video bit stream together with the method of decoding this bit stream [10]. H.264 standard follows suite and defines the encoded bit stream syntax and the decoding technique. Typical block diagrams of the encoder and decoder are illustrated in Figure 2-3 and Figure 2-4 respectively. Note that each of these modules have been improved in the internal behaviour when compared with the previous video coding standards. A brief description of the codec architecture is given below.

#### *2.3.2.1 Encoder*

In general, there are two basic types of frames in a group of pictures (GOP): intra coded frames and inter coded frames, based on the adopted source correlation exploitation methods. In either case, the input frame to be encoded ($F_k$) is split into macroblocks and a prediction signal is generated at the encoder. Intra coding exploits the spatial correlations between macroblocks within the frame and inter coding exploits the temporal correlations with the reference frame(s). In order to facilitate this process, the encoder keeps a repository of reconstructed frames generated through a *reconstruction path* (indicated by dashed lines in Figure 2-3) mimicking the decoding algorithm. There are different types of inter coded frames (P, B etc.) that vary in terms of the reference frames and the prediction modes (forward/ backward/ bidirectional) used in the motion estimation and compensation process. Note here that the use of multiple reference frames is a novel feature of the H.264/AVC framework. The original frame to be encoded is then compared with the prediction signal (either intra coded or inter coded) to generate the residual or difference signal ($D_k$) for each macroblock. The residual signal is then transformed (DCT), quantised, reordered and finally entropy coded before passing to the Network Abstraction Layer (NAL), with additional data, as the encoded payload.

The additional data includes: macroblock prediction modes, quantiser step sizes, motion vector information etc.

The reconstruction path, that mimics the decoder flow, involves the inverse quantization and inverse transform of the coefficients and adding to the residual signal to generate a replica of the decoded signal; this is called the reconstructed frame. A filter is used to smooth the blocking artefacts in the reconstructed frame. These frames then go to the frame repository to be used as reference frames in inter frame coding.




| | | | |
|---|---|---|---|
| $F_k$ | - Current frame | $D_k$ | - Difference macroblock |
| $F'_{k-1}$ | - Previous frame | $D'_k$ | - Reconstructed difference macroblock |
| $uF'_{k-1}$ | - Reconstructed macroblock | $T$ | - Block transform |
| $F'_k$ | - Reconstructed filtered macroblock | $T^1$ | - Inverse block transform |
| ME | - Motion estimation | $Q$ | Quantization |
| MC | - Motion Compensation | $Q^1$ | - Inverse Quantization |
| $P$ | - Predicted macroblock | $X$ | - Quantised transform coefficients |

Figure 2-3: AVC encoder

*2.3.2.2 Decoder*

The decoder construction is shown in Figure 2-4 (usually drawn in the reverse direction in line with the reconstruction path of the encoder). The input signal (in the NAL units) are entropy decoded and reordered before the inverse quantization and inverse transform to generate the reconstructed difference signal ($D'_k$). This is then added to the prediction signal, which is

generated by the motion compensation of the reference frame(s) using the motion vectors received from the encoder. It is noted that the prediction signal at the encoder and decoder are necessarily identical. A de-blocking filter is used at the end to suppress the blocking artifacts, to produce the decoded frames ($F_k'$).

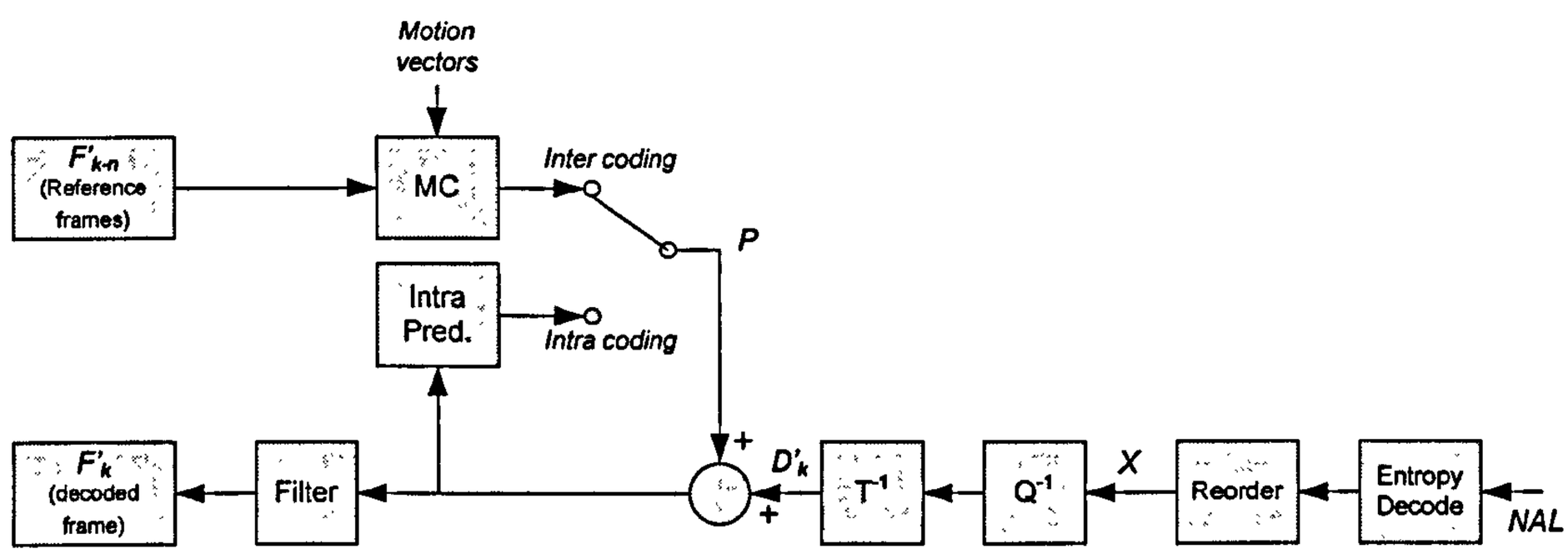



Figure 2-4: AVC Decoder

There are a number of contrasting differences in the DVC framework discussed in this thesis in comparison to the H.264/AVC framework briefly presented above. These changes mainly attribute to the source correlation exploitation mechanism and the resulting complexity balance between the encoder and decoder. In H.264/AVC, the encoder is significantly more complex than the decoder due to: (i) motion estimation and intra prediction search and (ii) decoding path in the encoder. In DVC, the inter frame correlations are exploited at the decoder; thus resulting in a major shift of the complexity balance. This modified video coding framework is discussed in detail in the next sections.

## 2.4 Distributed source coding

DVC is a practical realisation of the distributed source coding (DSC) principles that were originally introduced by Slepian and Wolf in 1973 [2]. The Slepian-Wolf theorem discusses the rate conditions for the independent encoding and lossless joint decoding of statistically dependant discrete random information sequences. Later, this concept was extended for lossy coding with side information at the decoder by Wyner and Ziv [3] and is commonly called the Wyner-Ziv theorem. Wyner-Ziv coding is the practical implementation of this concept as followed in the current DVC frameworks. The DSC concept deviates from conventional source coding, in that the latter involves joint encoding of the statistically dependant information streams accompanied with joint decoding.

The Slepian-Wolf and Wyner-Ziv theorems are presented in detail in the next sub sections.

### *2.4.1 Slepian-Wolf Theorem*

Let's assume that $X$ and $Y$ are two independently and identically distributed (i.i.d.) and statistically dependant discrete random sequences. Also assume that these two sequences are independently encoded with bit rates $R_X$ and $R_Y$, respectively, yet they are necessarily jointly decoded. This jointly decoding process involves the exploitation of the source correlations between them as shown in Figure 2-5.

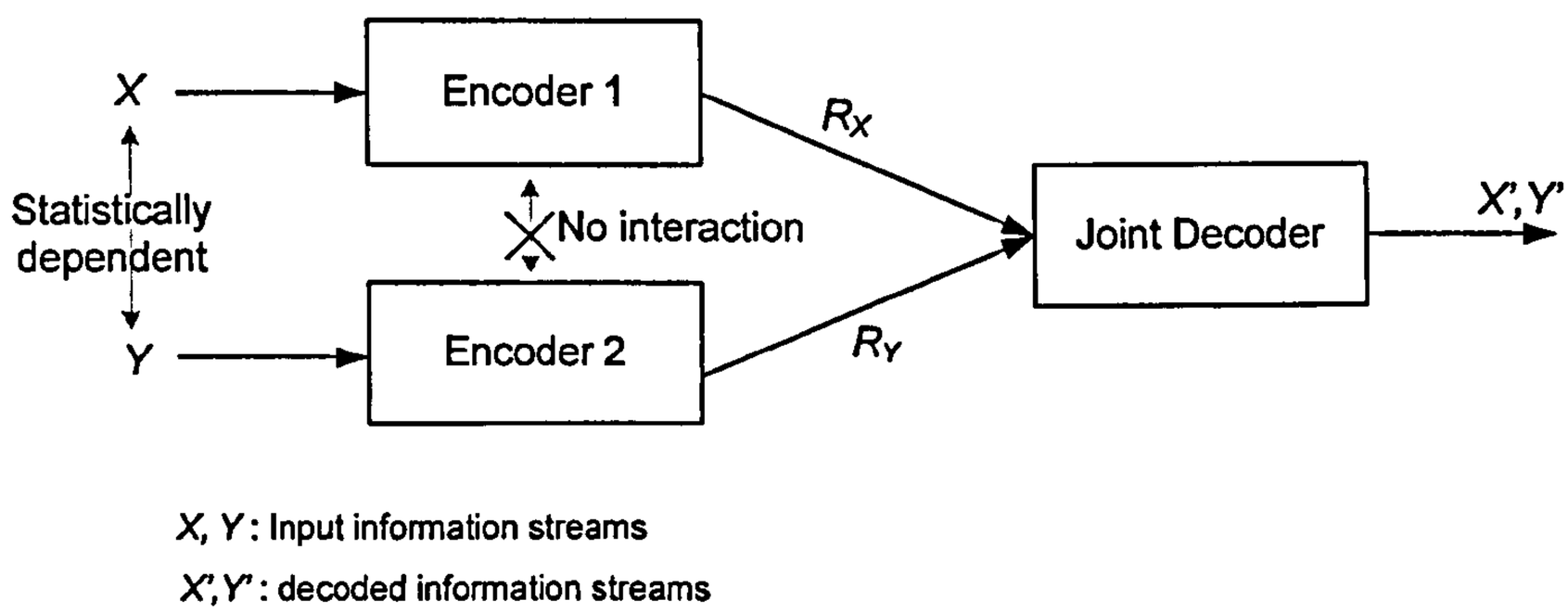



Figure 2-5: Distributed coding of two statistically dependent discrete random sequences

Slepian and Wolf have presented an analysis [2] of the possible rate combinations of $R_X$ and $R_Y$ for the reconstruction of X and Y with an arbitrarily small error probability as shown below:

$$R_X \geq H(X/Y) \qquad 2\text{-}8$$

$$R_Y \geq H(Y/X) \qquad 2\text{-}9$$

$$R_X + R_Y \geq H(X,Y) \qquad 2\text{-}10$$

where $H(X,Y)$ is the joint entropy of $X$ and $Y$, and $H(X|Y)$ and $H(Y|X)$ are their conditional entropies. According to the Slepian-Wolf theorem, if the total overall bit rate

exceeds the summation of individual bit rates $R_X$ and $R_Y$ and if the conditional entropy between $X$ and $Y$ is lower than the summation, then the independently encoded streams could be jointly decoded with arbitrarily small bit error probability. Thus, a lower bound to the bit rate is set to be equal to the summation of individual bit rates.

In this case, there is a rate region that is achievable with the distributed coding of two statically dependent i.i.d. sources, $X$ and $Y$ for the recovery of the information with an arbitrarily small error probability according to the Slepian-Wolf theorem, which is illustrated in Figure 2-6. Here, the vertical, horizontal and diagonal lines are drawn to represent the Equation 2-8, 2-9 and 2-10 respectively. These lines represent the lower bounds for the achievable rate combinations of $R_X$ and $R_Y$ and thus the shaded area marks the feasible operating region.

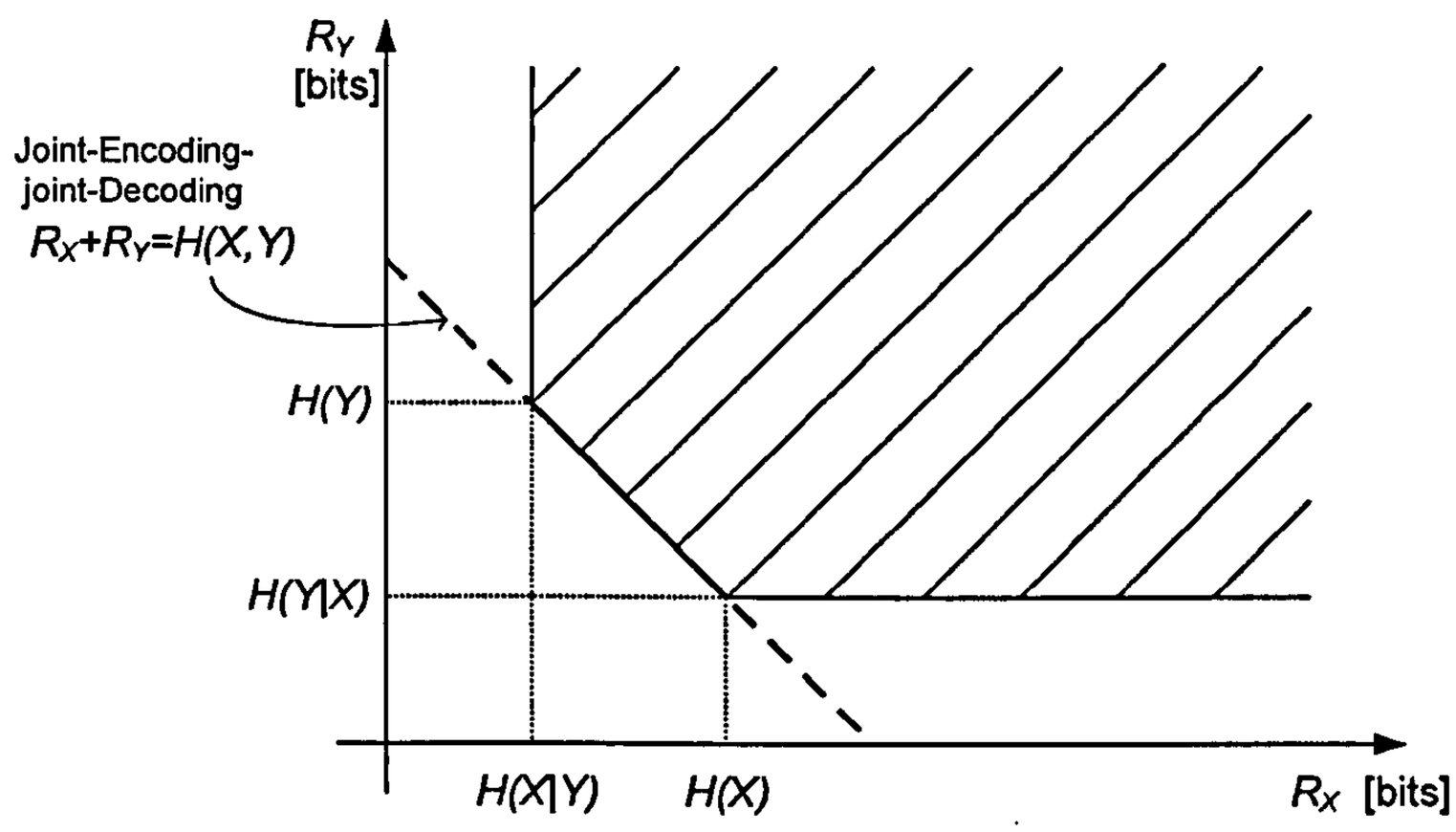



Figure 2-6: Achievable rate region following the Slepian-Wolf theorem

At this point, it is useful to theoretically compare the resulting bit rates and associated distortion of the above independent encoding scenario with the more conventional joint encoding-joint decoding scenario. In the latter, the source redundancies are exploited both at the encoder and decoder, thus the total minimum bit rate is equal to the joint entropy of the two information streams as depicted by the dashed line in Figure 2-6. Therefore, it can be concluded that there should be no theoretical loss of compression efficiency due to the utilization of independent encoding approach for the statistically dependent sequences

compared to joint encoding adopted in the conventional video compression schemes including the state-of-the-art H.264/AVC.

### *2.4.2 Wyner-Ziv Theorem*

Wyner-Ziv theorem [3] is an extension of the Slepian-Wolf theorem for lossy video compression scheme utilizing a statistically dependent secondary information stream at the decoder, which is known as *side information*. This case is illustrated in the Figure 2-7.

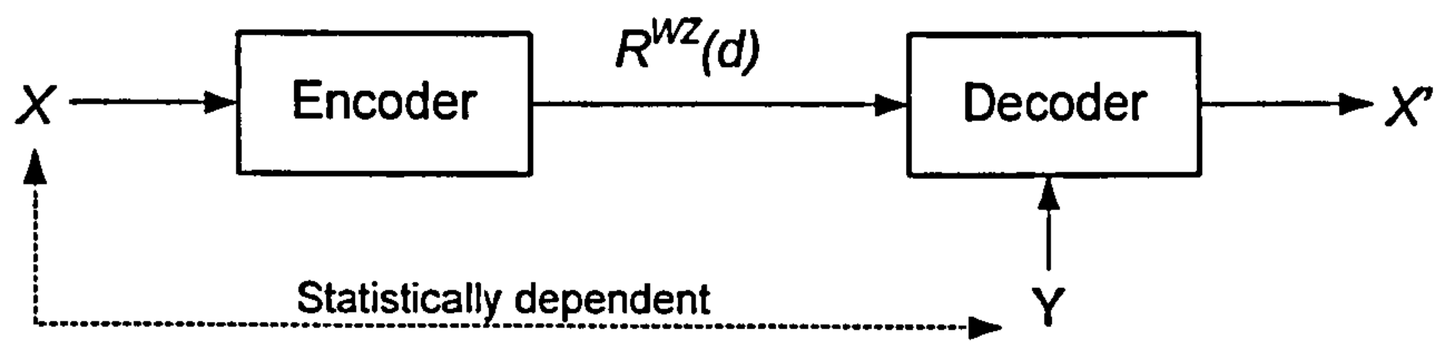



Figure 2-7: Wyner-Ziv theorem with side information

Wyner-Ziv theorem makes two major assumptions, one is that the distortion level, $d$ is finite and acceptable, between the source information $X$ and the corresponding decoded output $X'$. Therefore, it is essentially a lossy compression scheme. The second is that a statistically dependent secondary information stream $Y$, is available at the decoder. Wyner and Ziv attempted to quantify the minimum bit rate to be transmitted from the encoder to the decoder; called the Wyner-Ziv bit rate $R^{WZ}(d)$ for achieving the finite distortion $d$ between the input and output. Accordingly, the Wyner-Ziv theorem states that, if the statistical correlations between $X$ and $Y$ are exploited only at the decoder, the transmission rate increases compared to the case, where the correlations are exploited at both encoder and the decoder for the same average distortion, $d$. Thus, the Wyner-Ziv theorem states:

$$R^{WZ}(d) \geq R_{X|Y}(d), \qquad d \geq 0 \qquad \text{2-11}$$

where $R^{WZ}(d)$ is the Wyner-Ziv minimum encoding rate for $X$ and $R_{X|Y}(d)$ is the minimum rate required to encode $X$, where $Y$ is simultaneously available at the encoder and decoder for the same average distortion, $d$.

When $d$ becomes zero, or in other words when no distortion exists, the Wyner-Ziv theorem is equivalent to its mother theorem's results, i.e., Slepian-Wolf results. Or in simple terms, it emphasise that reconstructing the information sequence $X$ that consists of an arbitrarily small error probability is possible even if the source correlations between $X$ and the side information $Y$ are exploited only at the decoder.

## 2.5 Distributed Video Coding

Initial proposals for DVC were put forward in 2002 as a realization of the distributed source coding principles for the video signals. Since then DVC has evolved significantly over time in terms of the coding framework refinements, improved compression efficiency as well as the enthusiasm in the research community. This section provides an overview of the basic DVC framework of DVC on which this research is built upon. Also, the recent developments in parallel research are reviewed. First, a conceptual analysis of the practical sense of the modified complexity balance scenario as featured in DVC is presented.

### *2.5.1 Complexity balance of the video codecs*

In video compression, removing the spatial and temporal redundancies involves rigorous mathematical computations, necessarily making it a computationally complex entity. In the conventional approach to video coding with joint encoding and joint decoding, as is the case with ISO/IEC MPEG and ITU-T H.26x based solutions, the source coding is performed at the encoder. Thus, the conventional video encoders are usually highly complex. On the other hand, the decoders involve much simpler algorithms allowing it to be computationally inexpensive. A symbolic representation of this complexity balance of the conventional coding structures illustrated in Figure 2-8 and the encoder complexity of such a system is typically expected to be 5~10 times higher than that of the decoder [1]. These types of architectures are mainly motivated by the very common and traditional one-to-many type of video applications including television broadcasting and multicasting. An example video communication scenario with one video encoder and many receivers is shown in Figure 2-9. With multiple decoders, the

minimizing of decoder complexity to facilitate the fabrication of low cost decoders is the dominant consideration and hence, the conventional video coding complexity balance is justified for these scenarios.

However, in contrast to the above scenario, many-to-one type applications such as the video surveillance rely on a large number of video sensors, capturing encoding and uploading to centralised servers as discussed earlier in section 1.1.1, and re-illustrated in Figure 2-10 for clarity. These scenarios necessarily demand low cost video encoder architectures for economically viable deployment.

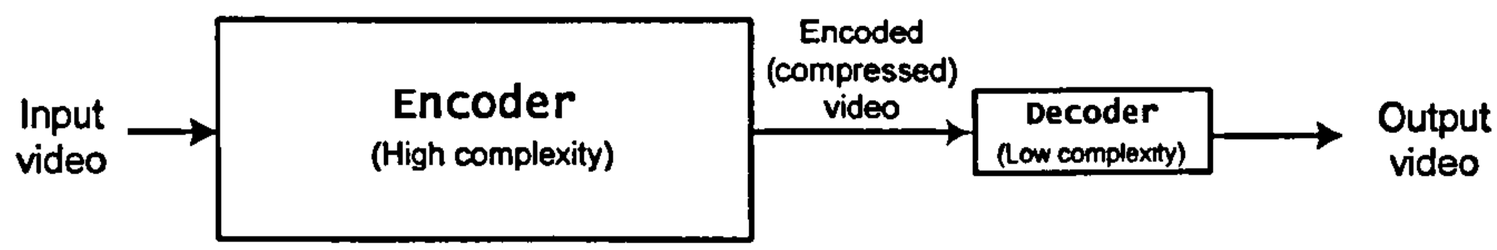



Figure 2-8: Complexity balance of conventional video coding

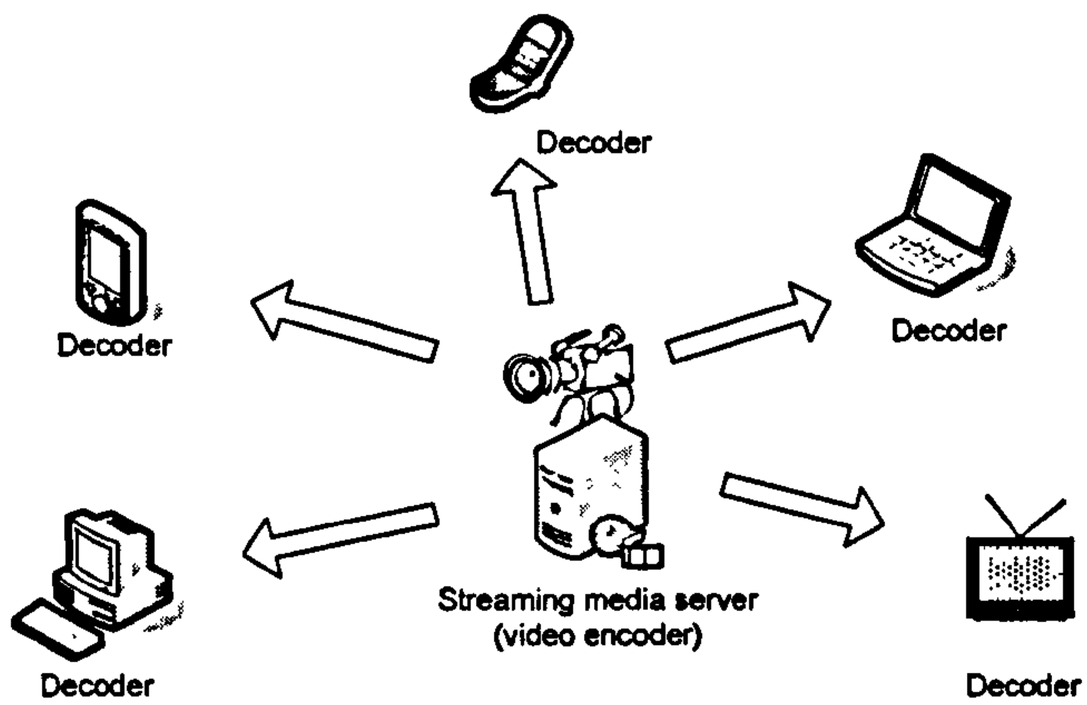



Figure 2-9: Example one-to-many type video communication scenario

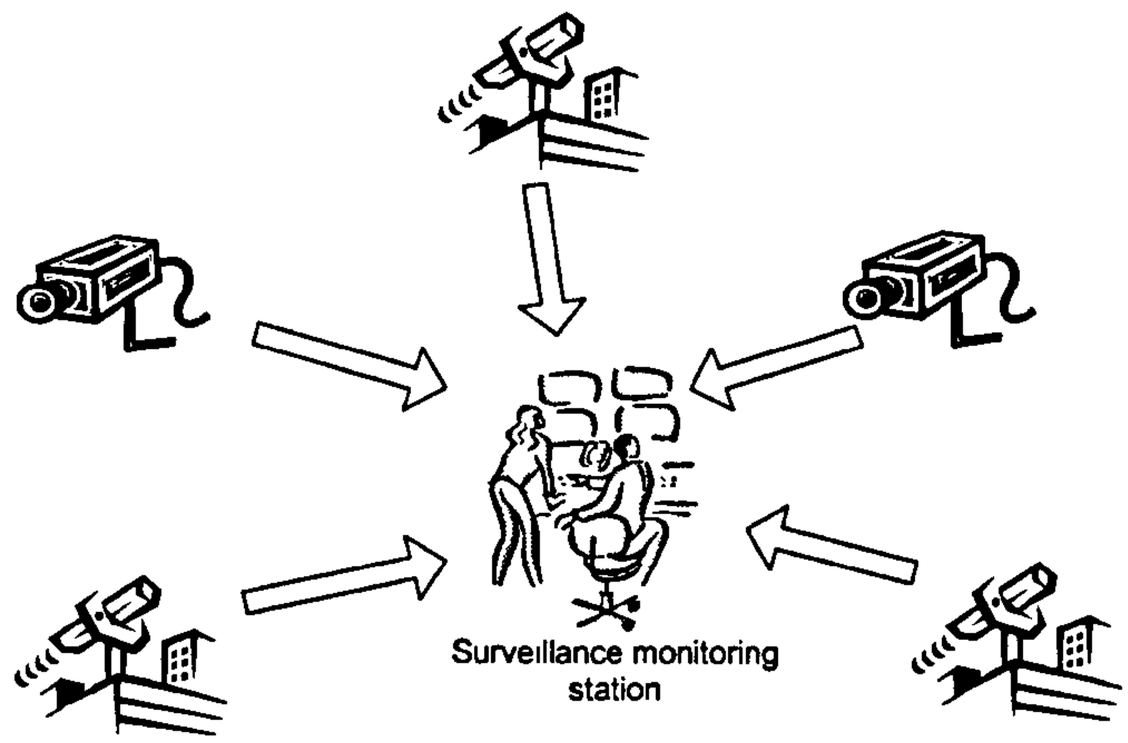



Figure 2-10: Typical many-to-one application scenario

Accordingly, a shifted complexity balance that would ideally benefit the many-to-one scenario is shown in Figure 2-11. The preliminary observations based on DVC algorithms that utilize this modified balance have shown that the encoder complexity could be brought down to negligible levels, i.e., to a simple state machine with two shift registers that are coupled with the bit stream preparation circuitry. However, some additional computations have been added to the DVC encoder to improve performance and robustness; for example *transform coding* and *encoder rate control.* In general, DVC encoders are significantly lower in complexity due to the independent-encoding-joint-decoding property.

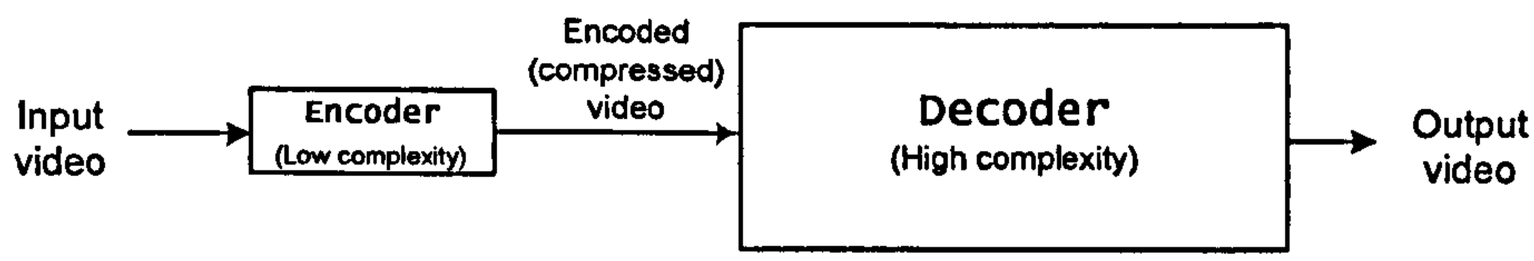



Figure 2-11: Complexity balance of distributed video coding

### *2.5.2 The DVC Codec*

Two independent initiatives can be identified in literature with regard to the development of a DVC framework based on the distributed source coding principals. First, the University of Stanford, USA put forward a proposal in 2002 for pixel domain DVC, introducing two different types of frames called Wyner-Ziv frames and Key frames [4]. This architecture, is commonly identified as the *Stanford DVC framework*. The Stanford research group have further extended the discussion later, also considering transform domain coding [11] and Hash based motion compensation [12]. An architectural description of the DVC codec based on this framework is presented in the next section as relevant to this thesis. The second framework proposal for the DVC codec was presented by the University of Berkeley, CA USA, again in 2002, under the name "Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding" or PRISM, as it is usually referred in the literature [5]. The PRISM encoder uses transform coding followed by the syndrome coding with base and refinement quantisation for the low frequency transform coefficients. High frequency components are more conventionally entropy coded.

The Stanford framework has the advantages of lower encoder complexity and ease of implementation, over the PRISM framework. On the other hand, PRISM is a more robust architecture, and uses no key frames or a reverse channel in the base framework as in the Stanford case. Possibly due to the more flexible architecture, Stanford framework is far more popular than PRISM among DVC researchers and for the same reason, the Stanford framework is adopted as the basis for the research pertaining to this thesis.

A high level overview of the Stanford based DVC codec is illustrated in Figure 2-12. At the DVC encoder, a parity bit stream is generated using a channel encoding tool (typically a turbo encoder) and transmitted to the decoder as the Wyner-Ziv encoded payload. The decoder incorporates the corresponding channel decoding algorithm and a side information stream statistically correlated to the original Wyner-Ziv frame sequence is processed in the decoder with the parity bit stream. The correlation noise in the side information compared to the original Wyner-Ziv sequence is modeled using a Laplacian distribution. The estimation of the side information at the DVC decoder plays a key role in this coding framework.

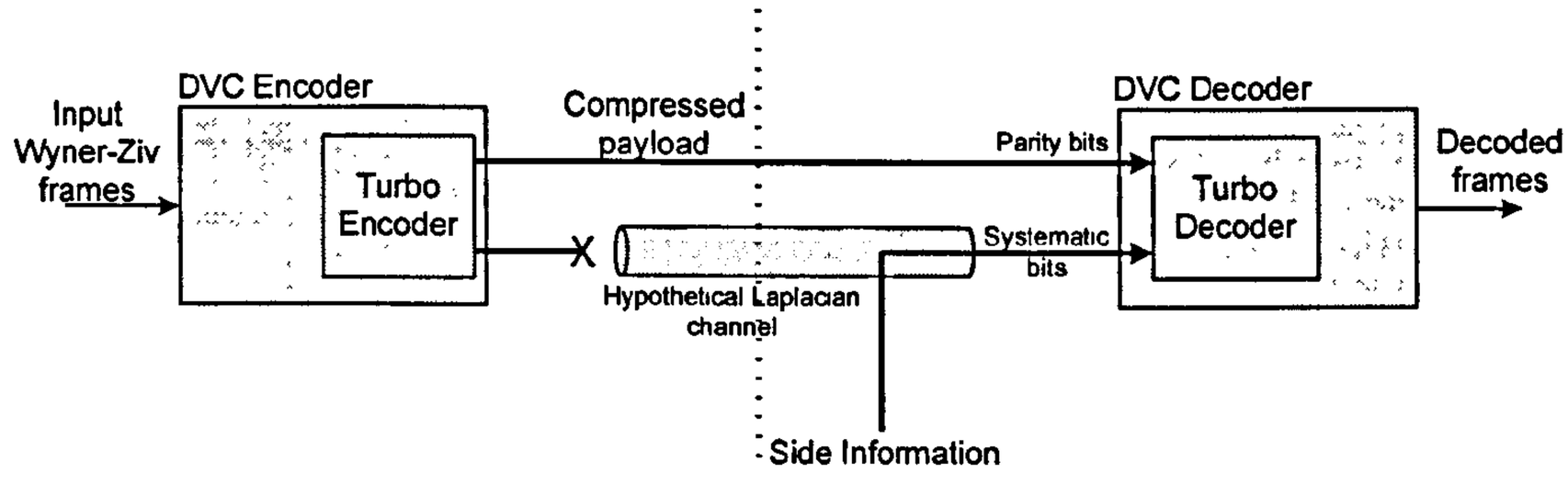



Figure 2-12 DVC codec: high level overview

When the Stanford framework based DVC codec design is concerned, the pixel domain DVC architecture produces the lowest complex encoder, which in complexity sense would best fit the demands of the target applications, while transform domain design allows the encoder to be more complex compared to the previous design with a rewarding gain in the rate-distortion performance. However, it should be noted that the transform domain DVC encoder is still very much less complex compared to conventional video encoders. Discrete cosine transform (DCT) is mostly proposed in the transform domain implementations (e.g.[11][20]). While Wavelet transform has also been utilized in some proposals (e.g.[21]).

The next sections provide an architectural overview of the pixel domain and transform domain DVC codecs.

### *2.5.3 Pixel-Domain DVC Architecture*

A block diagram of the pixel domain DVC architecture, with turbo coding in the Slepian-Wolf codec, is illustrated in Figure 2-13. Figure illustrates the key functional elements of the codec, including quantiser, bit plane extractor, Slepian-Wolf encoder, and key frame encoding mechanism in the encoder side. The decoder structure constitutes the Slepian-Wolf decoder, side information generator, reconstructor, and key frame decoder as dominant functionalities. In the original Stanford DVC framework, there is an essential reverse information path from the decoder to encoder in addition to the forward path. It is used to communicate dynamic parity request messages from the decoder to encoder buffer.

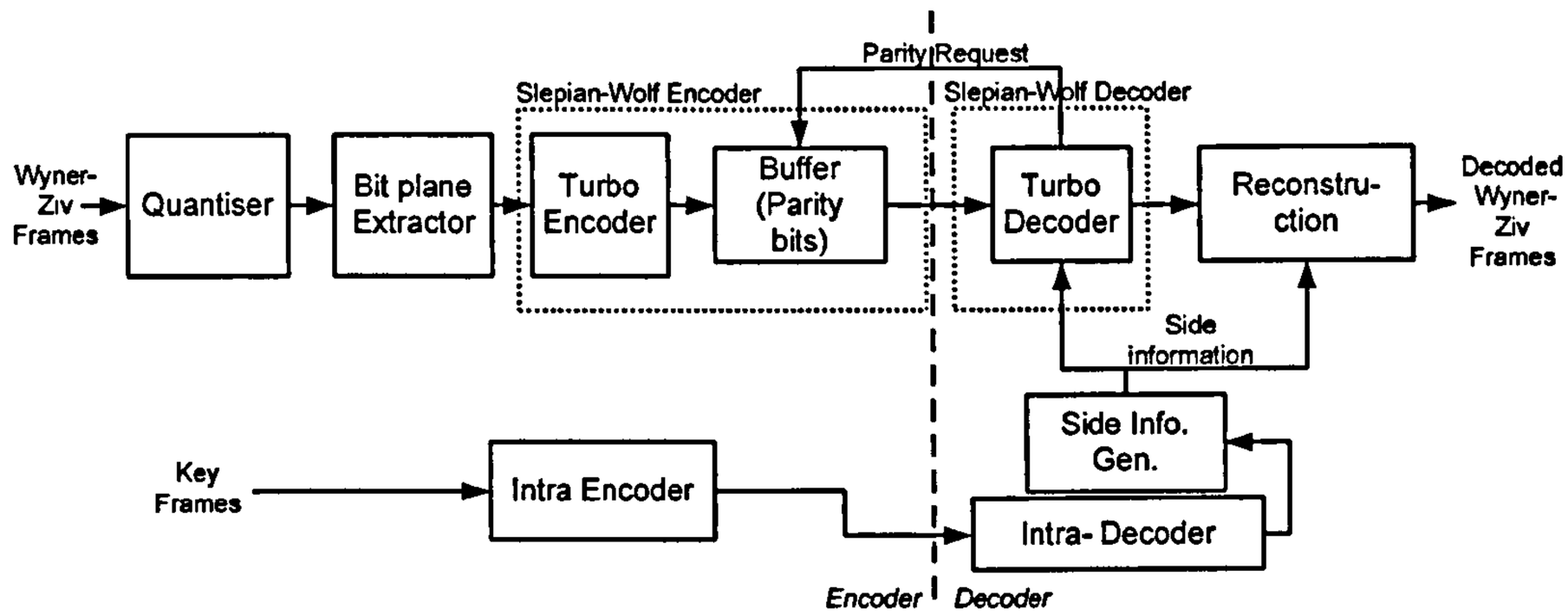



Figure 2-13: Pixel domain DVC codec architecture using turbo coding

The functionalities and implementation of these blocks are discussed in the following sub-sections.

### *2.5.3.1 Quantization and Bit Plane Extraction*

Quantisation is an important function in any signal coding algorithm. In pixel domain DVC, this effectively means limiting the number of discrete levels each pixel takes. In the source unquantised video signal, each pixel is represented by 8 bits, resulting 256 quantization levels. The quantization function is usually described by using a series of quantization bins as illustrated in Figure 2-14. The number of bins is determined by the quantization parameter $M$.

Figure 2-14: Quantization bin formation for $2^M = 4$

Limiting the number of discrete levels adds the *quantisation noise* and thus this is an irreversible lossy coding process. The DVC codec framework incorporates a *reconstruction* function at the decoder to partially compensate for this quantisation noise.

Bit plane extraction is performed in the DVC encoder to prepare the bit stream for Slepian-Wolf coding, by grouping together bits of compatible significance. This process is illustrated in

Figure 2-15 where $N$ and $n$ represent the number of pixels in the frame and the number of bit planes in the quantised signal, respectively.

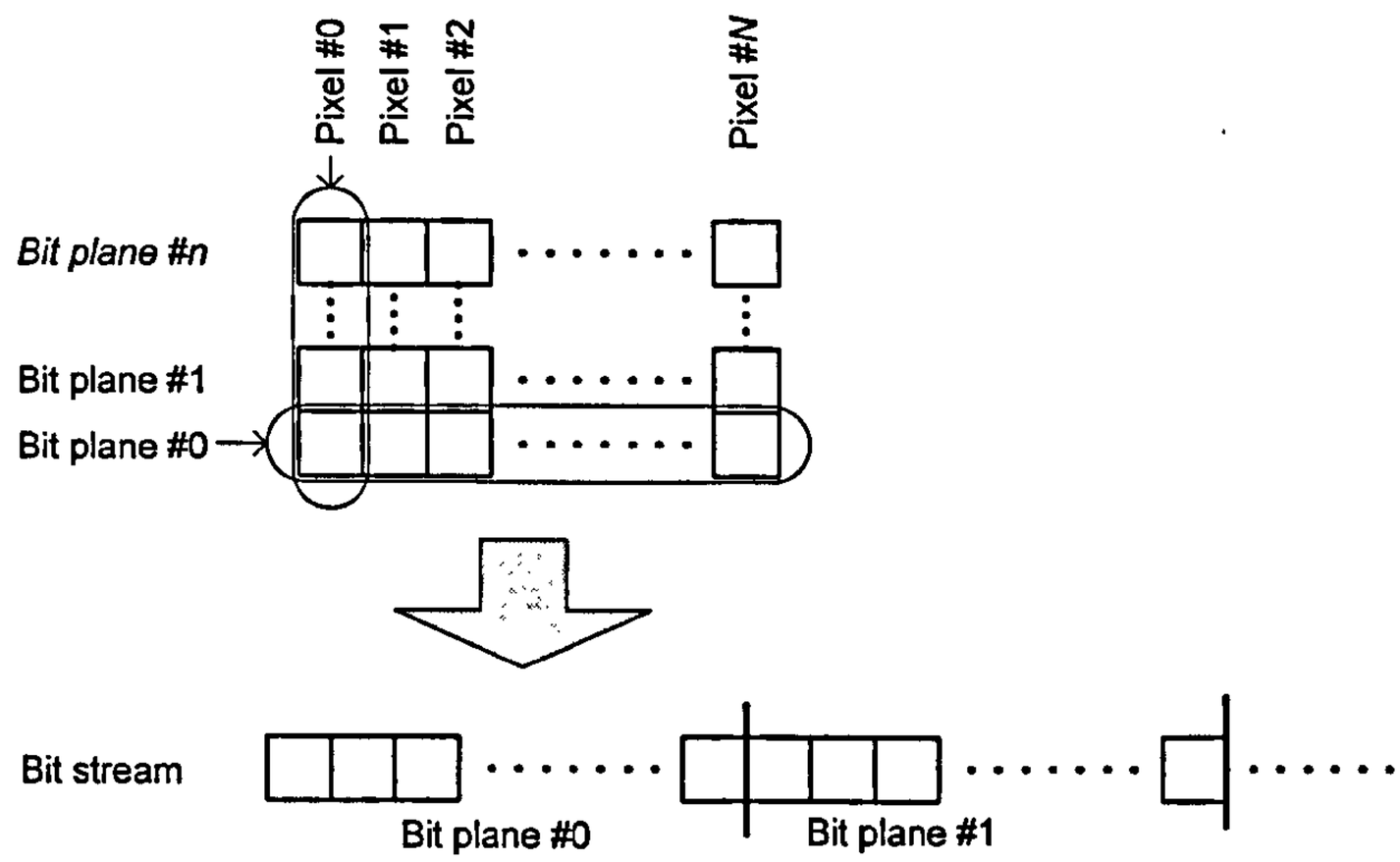



Figure 2-15: Bit plane extraction

In the bit plane extraction process, the most significant bits of all pixels of a frame are extracted and aligned in a row, to form the first bit plane. Then the second level significant bits of all pixels are extracted at the end of first bit plane to form the second bit plane. The same process is continued for all bits of the pixels.

It is noted that the pixel domain quantisation and bit plane extraction processes are closely interlinked in the implementation. In effect, the quantization could be achieved by truncating the final bit stream at a bit plane boundary, enclosing the desired number of bit planes.

#### *2.5.3.2 Slepian-Wolf Encoder*

The Slepian-Wolf coding in the Stanford DVC framework incorporates a channel coding algorithm to correct the noise (with reference to the original Wyner-Ziv bit stream in the encoder) in the side information. Turbo coding was used in its place in the initial DVC proposal by Aaron et al. [4] and the majority of subsequent DVC research followed suite.

**Turbo Encoder**

The concept of turbo coding was presented in 1993 for channel coding in communication systems by Berrou et al [13]. Turbo coding has been the technology choice for the Slepian-Wolf codec in DVC possibly due to the simplicity in implementation and the efficient performance. Turbo encoder consists of two parallel convolutional encoders separated by an interleaver as illustrated in Figure 2-16.

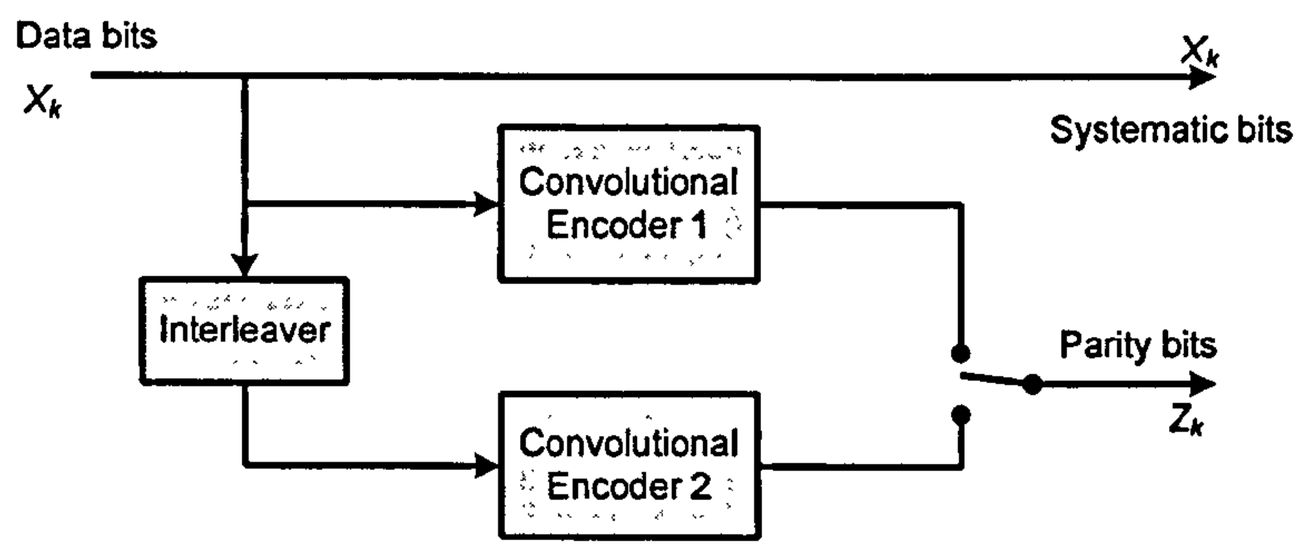



Figure 2-16: Turbo Encoder

The structure of a convolutional encoder is defined by the generator polynomial which takes the form:

$$G(D) = [\,1,\ g_2(D) / g_1(D)\,] \qquad 2\text{-}12$$

where $g_1(D)$ and $g_2(D)$ are the feed-forward and feedback polynomials respectively. The illustration of Figure 2-17 is an example of a rate ½ recursive systematic convolutional (RSC) encoder with a generator polynomial of $(1, 13/15)_{oct}$.

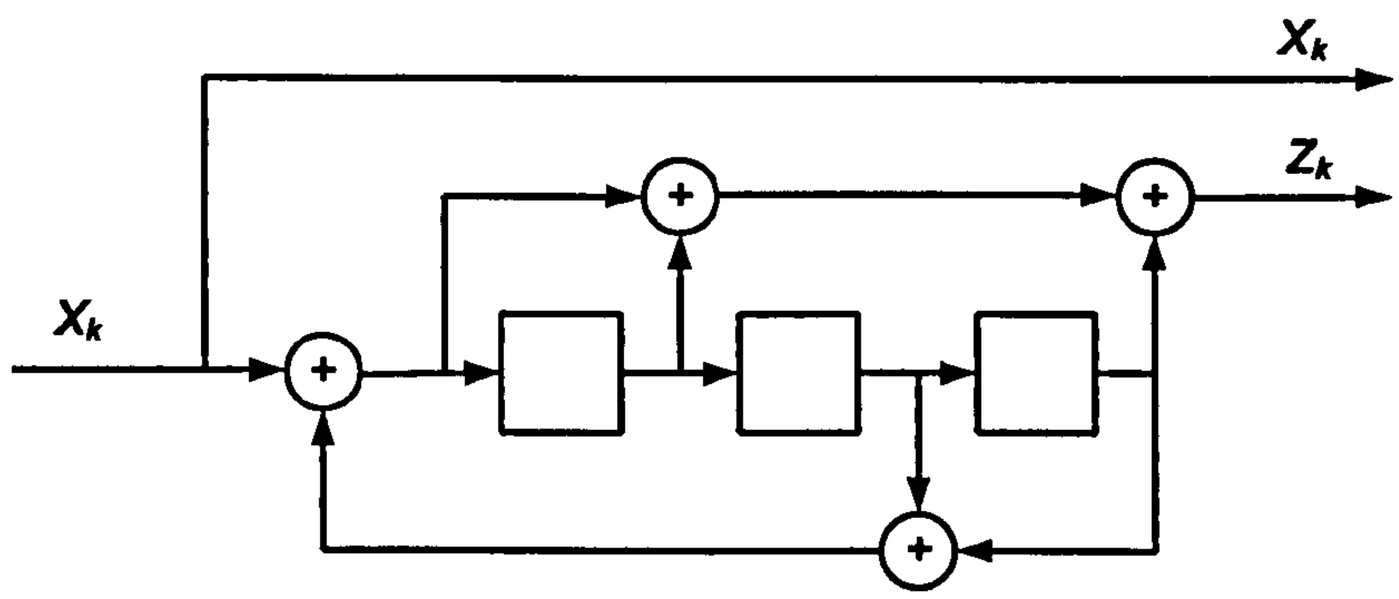



Figure 2-17: Example of an RSC encoder

**Parity Bit Puncturer**

Periodic puncturing of the parity bit stream generated by the turbo encoder essentially plays a key role in determining the compression efficiency of the DVC codec. Through the two convolutional encoders forming the turbo encoder, generate $2n$ number of parity bits for an input block of $n$ bits. Technically, the compressed bit stream or the output of the DVC encoder for the input video sequence is this parity bit stream. Thus, in order to achieve a compression, this parity bit sequence is subject to a periodic puncturing function, where selected parity bits are eliminated from the full stream in a periodic pattern. The parity bit puncturing is usually symmetrical over the two bit streams from the two convolutional encoders.

In essence, the purpose of the parity bits received at the decoder is to clear the correlation noise in the side information with reference to the original Wyner-Ziv frame. Thus effectively the parity bit rate and pattern need to be a reflection of the said correlation noise. In the common DVC framework, this process is realised by involving a closed feedback loop between the encoder and decoder. Accordingly, the parity puncturing for each block starts at an arbitrarily small bit rate and the pattern is dynamically optimized using the reverse channel feedback.

An alternative approach is taken in the unidirectional DVC frameworks to determine the parity puncturing rate by predicting the correlation noise level at the encoder.

#### *2.5.3.3 Side Information*

As referenced in the Wyner-Ziv theorem [3] (see section 2.4.2), side information is a statistically dependant information stream to the original Wyner-Ziv frame sequence to be encoded. Since the side information is necessarily generated at the decoder to comply with the independent-encoding-joint-decoding concept of DVC, a sequence of reference frames, called key frames, are made available at the decoder to facilitate this estimation process. The source correlations in the key frames are exploited in the process. The estimation accuracy, in terms of the high correlation of derived side information frame to the corresponding original Wyner-Ziv frame, is vital in decoding to a sufficiently small error rate at a minimum bit rate, as elaborated in the Wyner-Ziv theorem.

Exploiting the temporal and spatial correlations in a video stream for source coding are a well established functionalities in conventional video coding technologies. Motion estimation and compensation algorithms have been developed to exploit the temporal redundancies. This section presents how exploiting temporal and spatial correlations is adopted in DVC.

**Temporal prediction in DVC**

An initial proposal for generating the side information using the key frames involved a pixel interpolation algorithm [4]. More enhanced algorithms deploy motion estimation techniques. However, the motion estimation and compensation algorithms in DVC contrastingly differ from those used in conventional video coding. In this regard, it is worth at this point to comparatively analyze the distinctive approaches taken in conventional video coding and DVC.

***Motion estimation:***

A one-to-one comparison of the deployment of motion estimation algorithms in conventional and distributed video coding approaches is listed in Table 2-3 below.

Table 2-3 Motion estimation in conventional and distributed video coding

| Conventional video coding | DVC |
|---|---|
| • Performed at encoder<br>• Target frame is available for estimating motion<br>• Actual motion field from reference frame to target frame is estimated. | • Performed at decoder<br>• Target frame (Wyner-Ziv frame) is not available for estimating motion<br>• Motion field is estimated across reference frames and interpolated/extrapolated. |

In DVC, the motion field is estimated between two reference frames (key frames) and then either interpolated or extrapolated based on the temporal distance between frames in the group of pictures (GOP) to derive the vector map for motion compensation. This process is illustrated in Figure 2-18 for a vector interpolation scenario. Note that in some cases,

extrapolation is chosen over interpolation, to avoid future reference frames, analogues to the I-B-I and I-P-I GOP structures in H.264/AVC.

The motion estimation, interpolation and compensation process in Figure 2-18 is summarised below. Note that *k* is the frame sequence number and Wyner-Ziv frame (*k*) is the target frame to be decoded.

1. Motion estimation: A macroblock-wise full motion search is performed between the two adjacent key frames: (*k-1*) and (*k+1*).

2. Motion interpolation: two sets of motion vectors are synthesised by interpolating the motion vector map derived in step 1, to represent the motion field from either of the key frames towards the target Wyner-Ziv frame. Note here that the interpolated vectors are synthesised with reference to the macroblocks on the Wyner-Ziv frame (*k*).

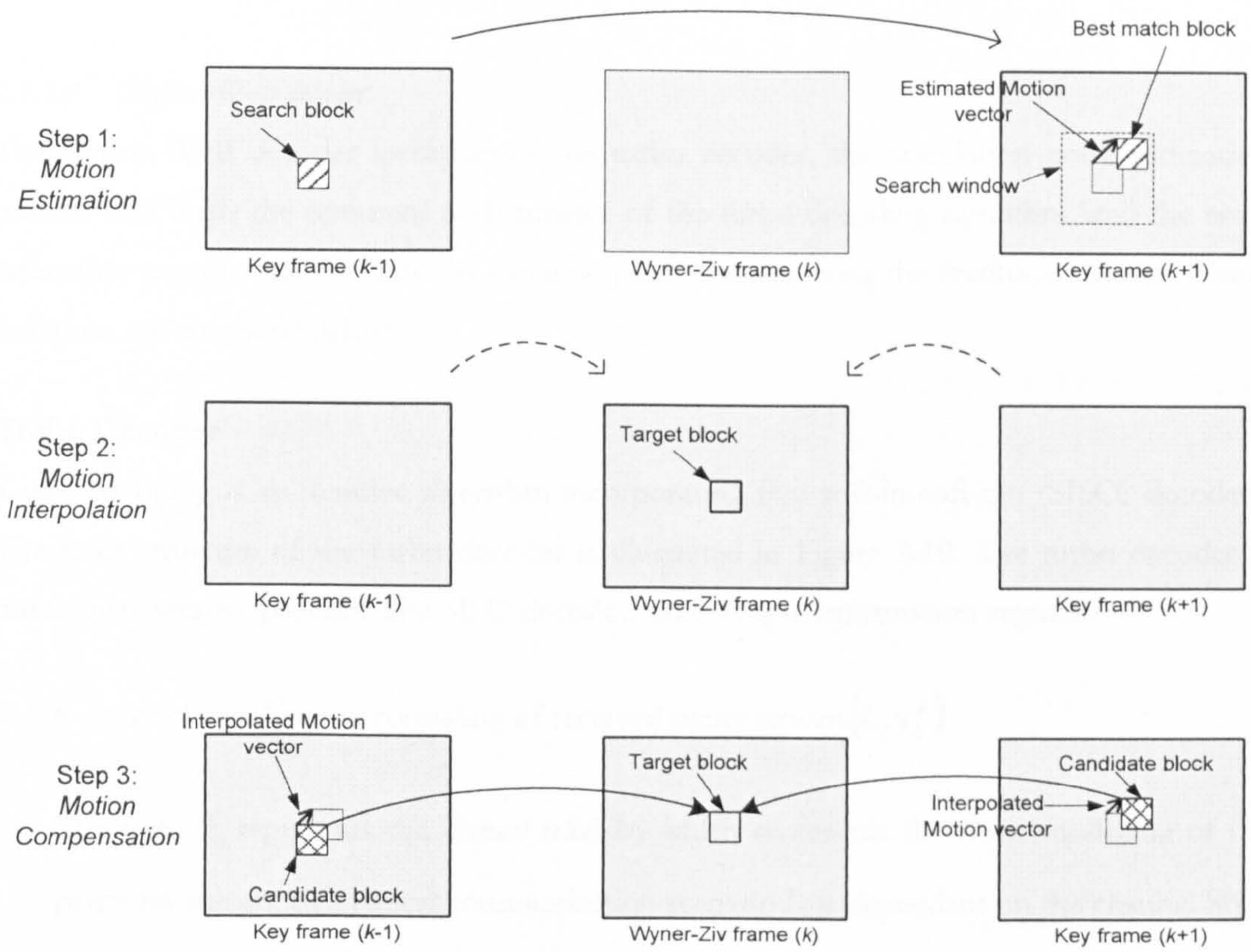



Figure 2-18 Motion estimation, interpolation and compensation in DVC

3. The bi-directional motion compensation is performed using the synthesised motion vector maps generated in step 2.

**Spatial prediction in DVC**

Spatial prediction is not generally a viable option in the side information generation for Wyner-Ziv frames since no spatial information pertaining to the target frame is available at the decoder. However, some degree of spatial correlation exploitation is feasible for the *refinement* of side information, once a version of *initial side information* is generated by temporal means. An appropriate algorithm is proposed in section 5.3 later in this thesis.

It is also noted here that spatial prediction is dominantly an intra frame coding technique in video coding, whereas DVC focuses on inter frame coding while conventional intra frame coders are commonly used for coding the key frames. However, an exception is available in the Wyner-Ziv coding based intra frame coding solution proposed in [14].

#### *2.5.3.4 Slepian-Wolf decoder*

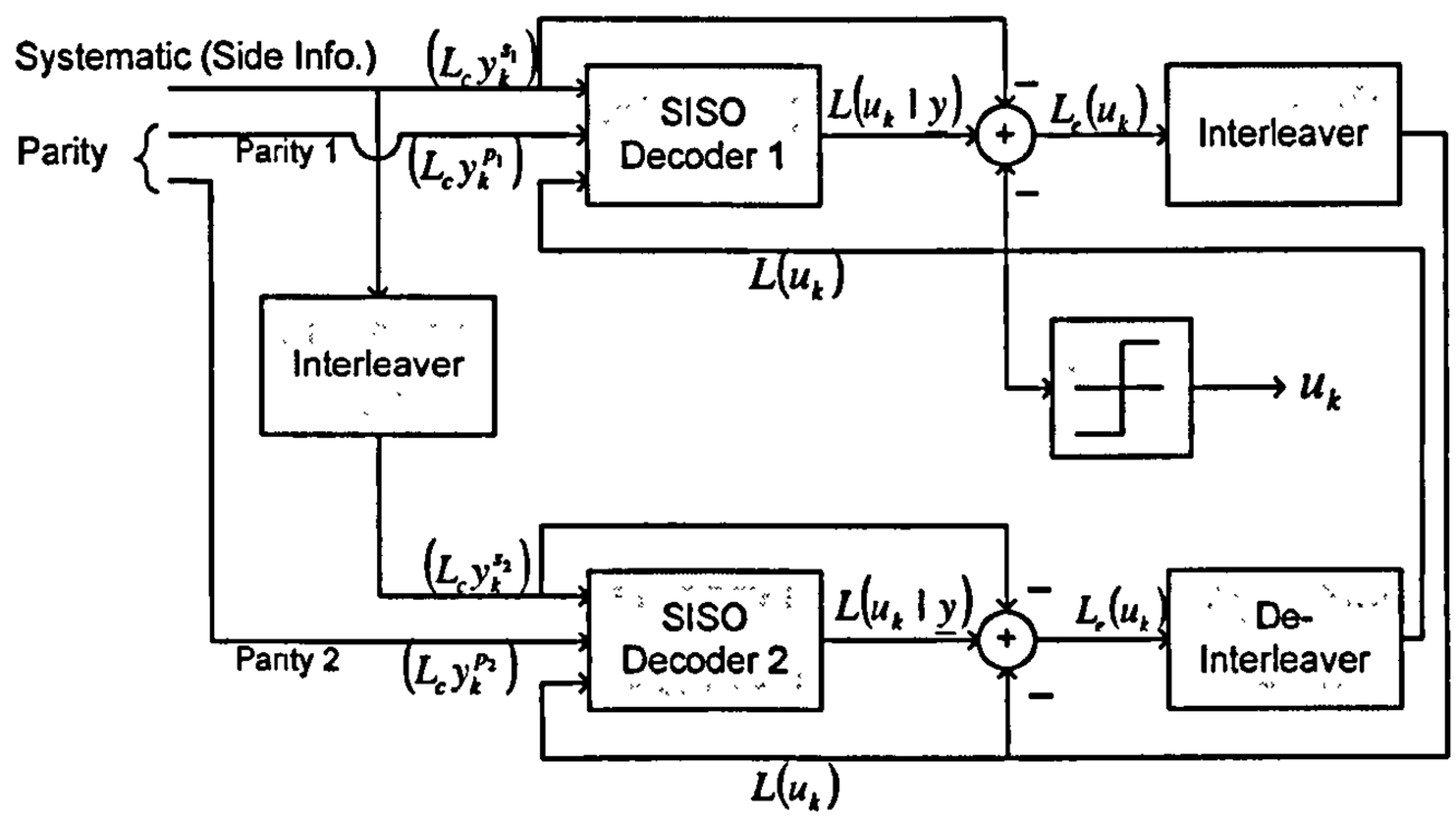
The Slepian-Wolf decoder incorporates the turbo decoder, the correlation noise estimation process that helps the optimum performance of the turbo decoding algorithm, and the error estimation process that facilitate the dynamic parity request using the feedback channel. These functions are discussed below.

**Turbo Decoder**

Turbo decoding is an iterative algorithm incorporating two soft-in-soft-out (SISO) decoders. The basic structure of the turbo decoder is illustrated in Figure 2-19. The turbo decoder is running an iterative process. The SISO decoders take 3 input information streams:

- soft channel inputs consisting of received parity stream $\left(L_c y_k^p\right)$

  The term $L_c$ represents the *channel reliability* which represents the noise modelling of the parity bit stream. In a typical communication scenario $L_c$ is dependant on the channel SNR and the fading amplitude; thus derived through *channel estimation.* In DVC research, however, usually an error free channel is assumed. A detailed analysis of the parity




$(L_c y_k^p)$ - Parity bits

$(L_c y_k^s)$ - Systematic bits

$L(u_k \mid \underline{y})$ - *A-posteriori* information

$L(u_k)$ - Log likelihood ratio (LLR) of data bit $u_k$

$L_e(u_k)$ - Extrinsic LLR

Figure 2-19: Turbo Decoder

transmission over wireless channels, and related noise modelling is presented in Chapter 6 of this thesis. $y_k^p$ is the received punctured parity bit stream.

- systematic bits stream $(L_c y_k^s)$

In the turbo decoding in Slepian-Wolf codec, the systematic bit stream is formulated from the side information. A discussion of the side information estimation process is presented later in this section. Again, $L_c$ represents the channel reliability. However, since the *channel* is hypothetical with regard to the side information, correlation noise modeling is performed to estimate the $L_c$ parameter for the side information.

- and the *a-priori* information $L(u_k)$ from the other constituent SISO decoder in the turbo decoder

In case of rate compatible punctured turbo (RCPT) codes that are used in DVC, parity bits are punctured; therefore, at the decoder; zeros are mapped in to the punctured positions (Note that, 0s and 1s in the bit stream are mapped to *-1* and *+1*). In the first iteration of the decoder,

there is no *a-priori* information about the received information; therefore, the log likelihood ratio (LLR) $L(u_k)$ for each bit is initialized to a common value.

In the algorithm, once the first iteration of the SISO decoder is completed, the A-posteriori LLR $L(u_k \mid \underline{y})$ is subtracted by the corresponding a-priori information $L(u_k)$ and input systematic information $(L_c y_k^s)$ to derive the extrinsic LLR for the bit $L_e(u_k)$.Then the extrinsic LLR is interleaved and used in the next iteration of other constituent SISO decoder as a-priori information $L(u_k)$. Once the desired number of iterations is gone, the decoded soft information is sent through a hard decision filter to obtain the bit stream.

The LLR of the sequence usually converges after a number of iterations of the turbo decoder; effectively, the LLR distribution gradually moves away from the hard decision boundary, as illustrated in Figure 2-20. After a fixed number of iterations (it is generally set to be between 4 and 8 for most cases), the iterative turbo decoder completes its decoding of one data block. And a hard decision decoding is performed to extract the decoded output bit stream.

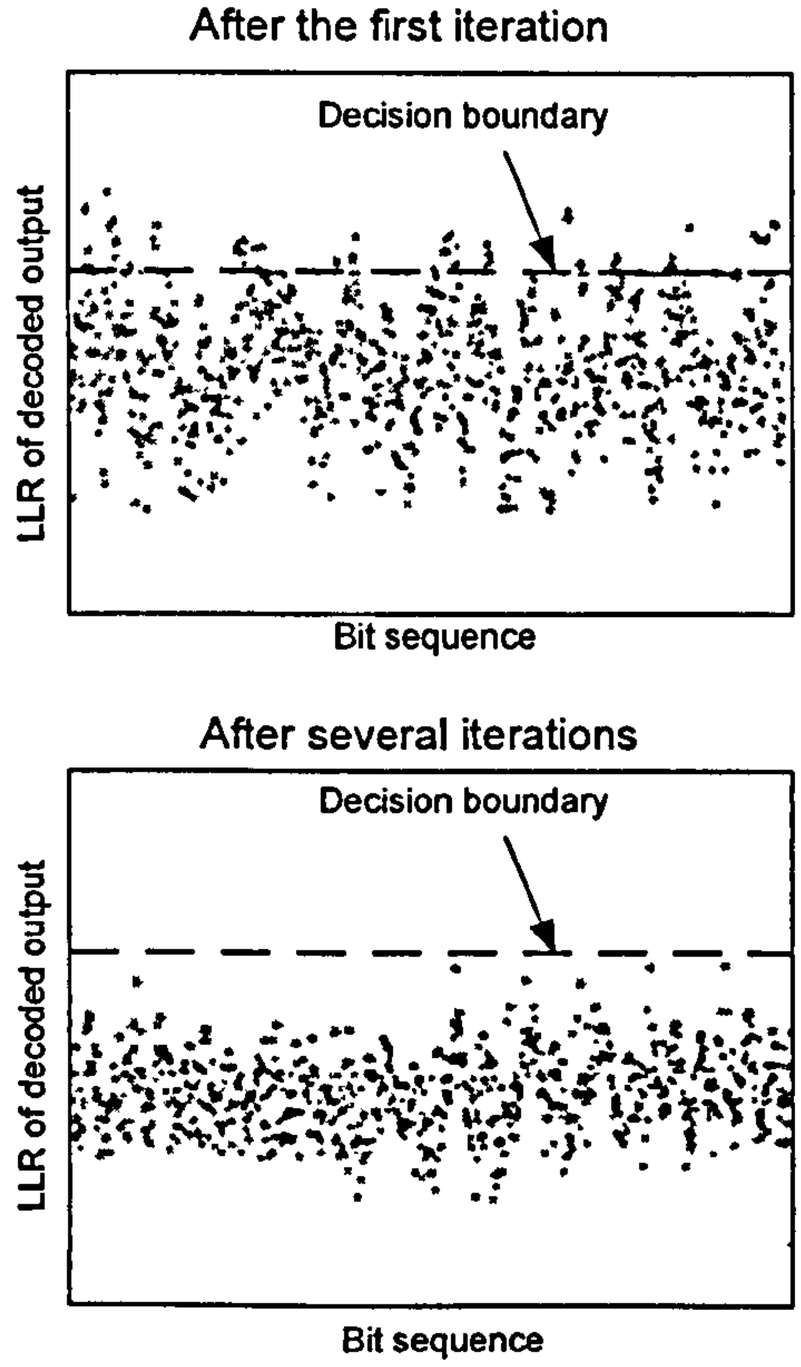



Figure 2-20 LLR convergence in turbo decoding

**Correlation Noise Modelling**

The side information estimated at the decoder as discussed in section 2.5.3.3 is used in the Slepian-Wolf decoder. This side information is usually modelled as a noisy version of the original Wyner-Ziv frame, or in effect the systematic information stream of the turbo encoder utilised at the Slepian-Wolf encoder. The hypothetical channel between the original Wyner-Ziv frame and the side information is illustrated in Figure 2-21.

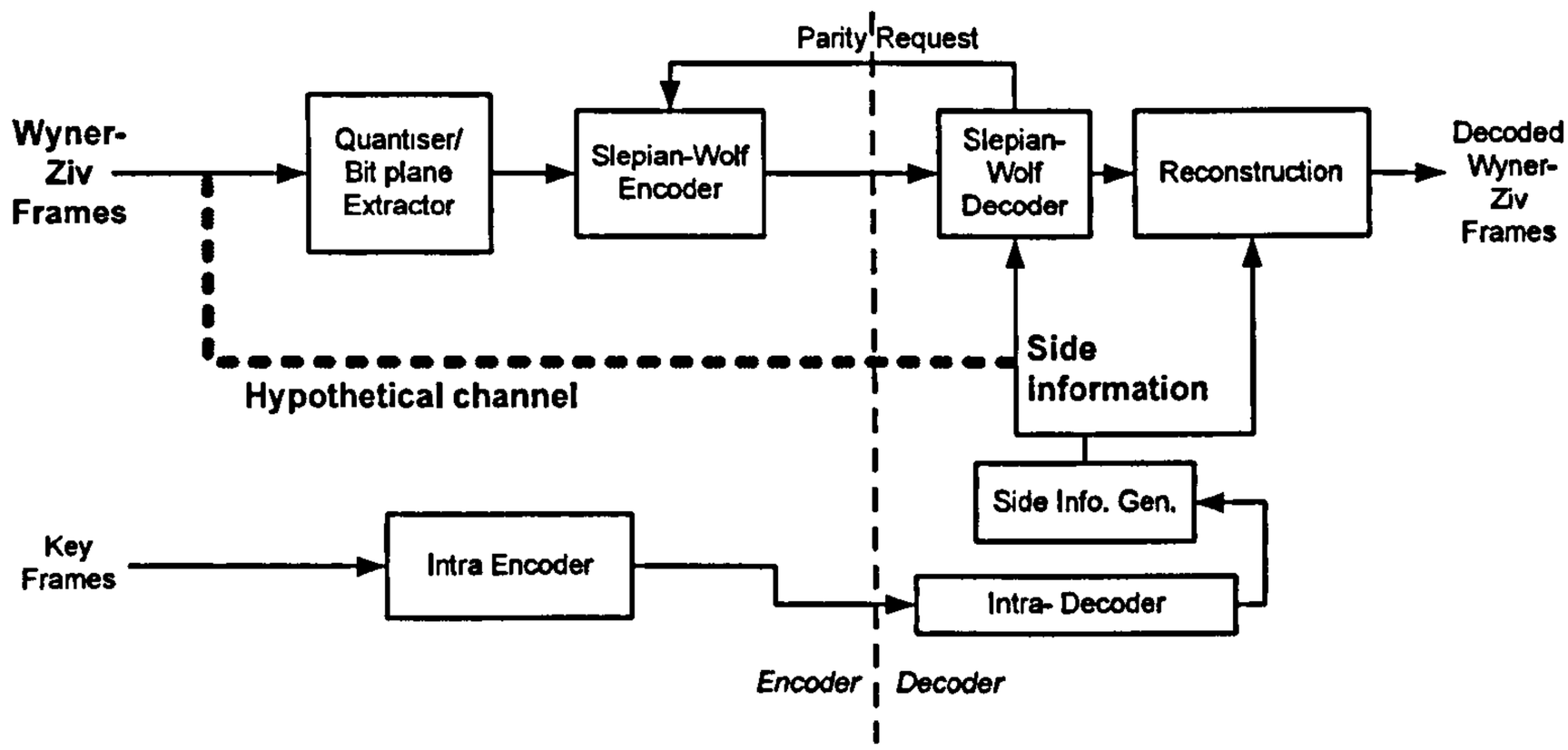



Figure 2-21 The Hypothetical channel model for side information

Thus, the estimation of the statistical properties of this hypothetical channel applied to the systematic bit stream, or in effect modelling the correlation noise between the side information and the original Wyner-Ziv frame, is vital for the efficient performance of the turbo decoder. The estimation accuracy of this statistical noise model affects the parity bit rate the decoder takes to clear the noise in side information to generate the decoded bit stream.

The correlation noise ($N$) in side information can be expressed as:

$$N = X - Y \qquad 2\text{-}13$$

where $X$ and $Y$ represent the original Wyner-Ziv information and side information respectively. It is common to assume that this correlation noise is Laplacian distributed as:

$$f(N) = \frac{\alpha}{2} e^{\alpha|N|} \qquad 2\text{-}14$$

where $\alpha$ is the Laplacian distribution parameter [15]. However, other researchers have reported that either a Gaussian or Laplacian noise model could be utilised here [16]. The dynamic estimation of this noise is a major research problem. In many DVC proposals, offline and perfect noise estimations that compare the original Wyner-Ziv frame and side information have been utilised. Commonly the Laplacian parameter is estimated at sequence level, but due to the local variations of the distribution, a frame level or macroblock level estimation produce better results. An analysis of the corresponding performance gains is reported in [15]. The same paper proposes an algorithm for the dynamic estimation of the noise parameters at the decoder based on a pixel domain DVC framework.

**Error Estimation**

In the Stanford DVC framework, with the feedback channel, the estimation of the *quality* of the output of turbo decoder plays a critical role. In the Slepian-Wolf coder, for each block of data, parity bit request messages are dynamically sent to the encoder buffer. The decision criterion for sending a parity request message is the *quality* of the decoded output being above a preset threshold. One possible approach to define the *quality* of the decoded output is in terms of its BER with reference to the original Wyner-Ziv bit stream. Accordingly, a perfect error estimation technique (using the original Wyner-Ziv bit stream it self for the comparison) has been utilised in initial DVC proposals [4]. Even though this is an impractical approach, it has been followed in many subsequent works, as a workaround solution.

In addition to the original purpose of dynamically terminating the parity request loop between the decoder and the encoder buffer in the Slepian-Wolf codec, the error estimation of the decoded block fulfills an important functionality in the refinement algorithms [17][18] discussed later in this thesis. In these scenarios, the decoded blocks with a noise level above threshold are forwarded to refinement.

A practical solution to the problem of terminating parity request loop is proposed in [19]. This involves the monitoring of Logarithm of the A Posteriori Probability (LAPP) ratio in the SISO decoders. It is known that, higher the absolute value of LAPP ratio for a bit, the confidence in making the hard decision for the bit is higher. Thus the average LAPP ratio is calculated for

each block of bits after each iteration and is compared against a preset threshold. Usually the absolute LAPP ratio increases after each iteration, and once it goes above the threshold, the parity request loop is terminated. Note here that a similar approach has been conventionally tested in turbo coding albeit in a different context; i.e. to terminate the internal iterations loop between the SISO decoders.

#### *2.5.3.5 Reconstruction: Inverse Quantization*

The function of the Reconstruction module incorporated in the DVC decoder is to compensate for the quantization noise introduced at the encoder. Effectively, when the pixel domain DVC coding framework is concerned, reconstruction increases the pixel precision by increasing the number of bits per pixel, which was reduced in the quantization. The decoded bit stream is *inverse bit plane extracted* (the inverse process of bit plane extraction process discussed earlier in this chapter) to generate the decoded frame of quantised pixels. Then the side information frame for the corresponding Wyner-Ziv frame is utilized in the reconstruction module.

**Reconstruction algorithm**

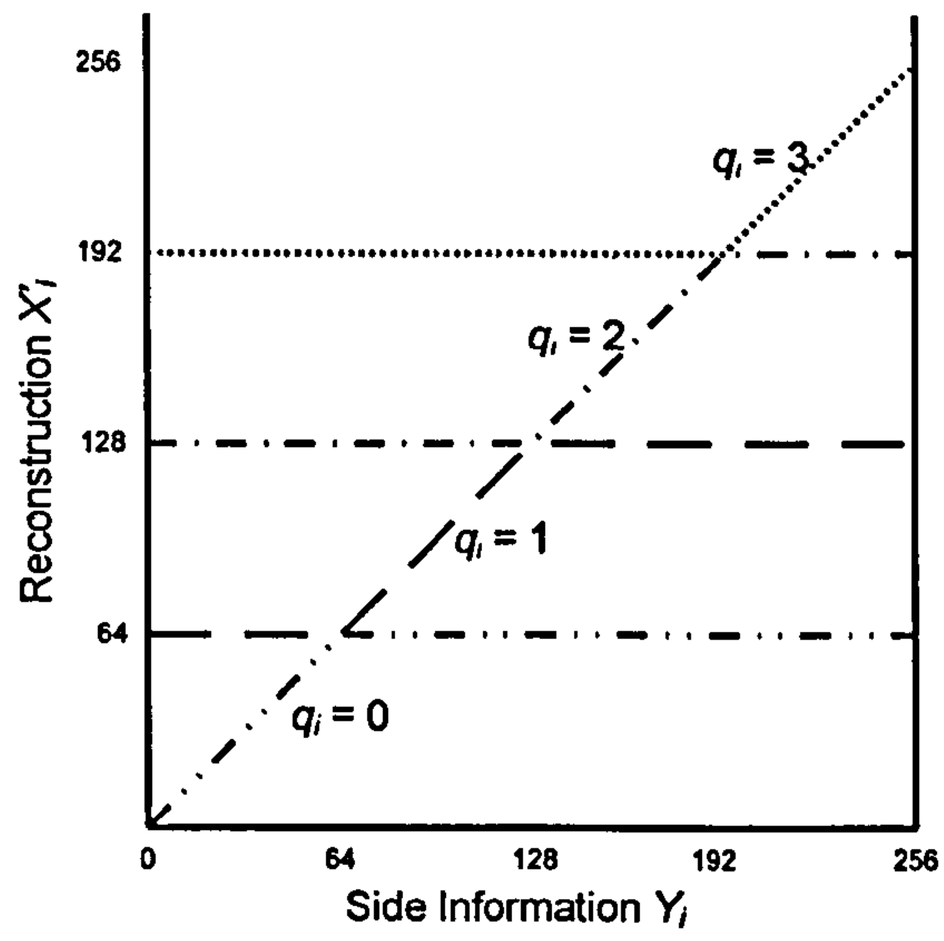
An initial solution for pixel-by-pixel reconstruction is incorporated to the DVC codec by Aaron et al in the original Stanford proposal [4]. This has been followed in majority of subsequent DVC proposals, possibly due to its simplicity in implementation, and very effective performance. A technical description of the Stanford reconstruction algorithm follows.

Let's assume $x$ as the original pixel to be encoded, $y$ as the corresponding side information pixel value and $\hat{x}$ to be the reconstructed pixel after decoding. In this algorithm, the reconstructed pixel $\hat{x}$ is calculated considering the corresponding side information value $y$ and the corresponding quantization bin boundaries as:

$$\hat{x} = \begin{cases} z^{-}, & y < z^{-} \\ y, & z^{-} < y < z^{+} \\ z^{+}, & y > z^{+} \end{cases} \qquad 2\text{-}15$$

where $z^{-}$ and $z^{+}$ represent the lower and upper edges of the decoded quantization bin.

A graphical presentation of this algorithm for four quantization bins is illustrated in Figure 2-22.




Where

$X'_i$ - Reconstructed pixel value (output of the reconstruction function)

$Y_i$ - Corresponding pixel value in side information

$q_i$ - Decoded quantised pixel value

Figure 2-22: Reconstruction algorithm (4 quantization levels)

The reconstructed pixel value, $X'_i$ is derived as $X'_i = E(X_i \mid q_i, Y_i)$, where $X_i$ is the original unquantized pixel value, $q_i$ is the decoded symbol and $Y_i$ is the corresponding unquantized pixel value that is taken from the side information. If a given pixel of the side information, $Y_i$ is within the reconstruction bin of the corresponding quantised pixel $q_i$, then $X'_i$ takes the value of $Y_i$ . If $Y_i$ resides outside the bin, the function clips the reconstruction towards the boundary of the bin closest to $Y_i$.

If the side information is highly correlated with the original Wyner-Ziv frame (that is the property of *good* side information), there will be a higher hit rate into the appropriate reconstruction bin, so that this reconstruction algorithm produces very good results and the number of pixels getting clipped off will be lower. However, where the motion content of a

particular video sequence is high, side information is generally expected to be less close to the original frame due to the estimation errors. This will result a higher frequent occurrences of falling out of the bin. This would create irritating artefacts particularly when the quantization is coarse. This unfolds a drawback of the Stanford reconstruction algorithm: the non-reflection of the statistical properties of the information, particularly the spatial correlations within the frame.

### *2.5.4 Transform-Domain DVC Architecture*

Use of transform coding is an established concept in video coding. It helps the compression efficiency by reducing the spatial redundancies within the frames. The same effects have motivated the incorporation of transform coding in the DVC codec. Discrete cosine transform (DCT) has been most commonly used [11][20], however wavelet transform has also been proposed in literature [21]. An illustration of the transform domain DVC codec is presented in Figure 2-23.

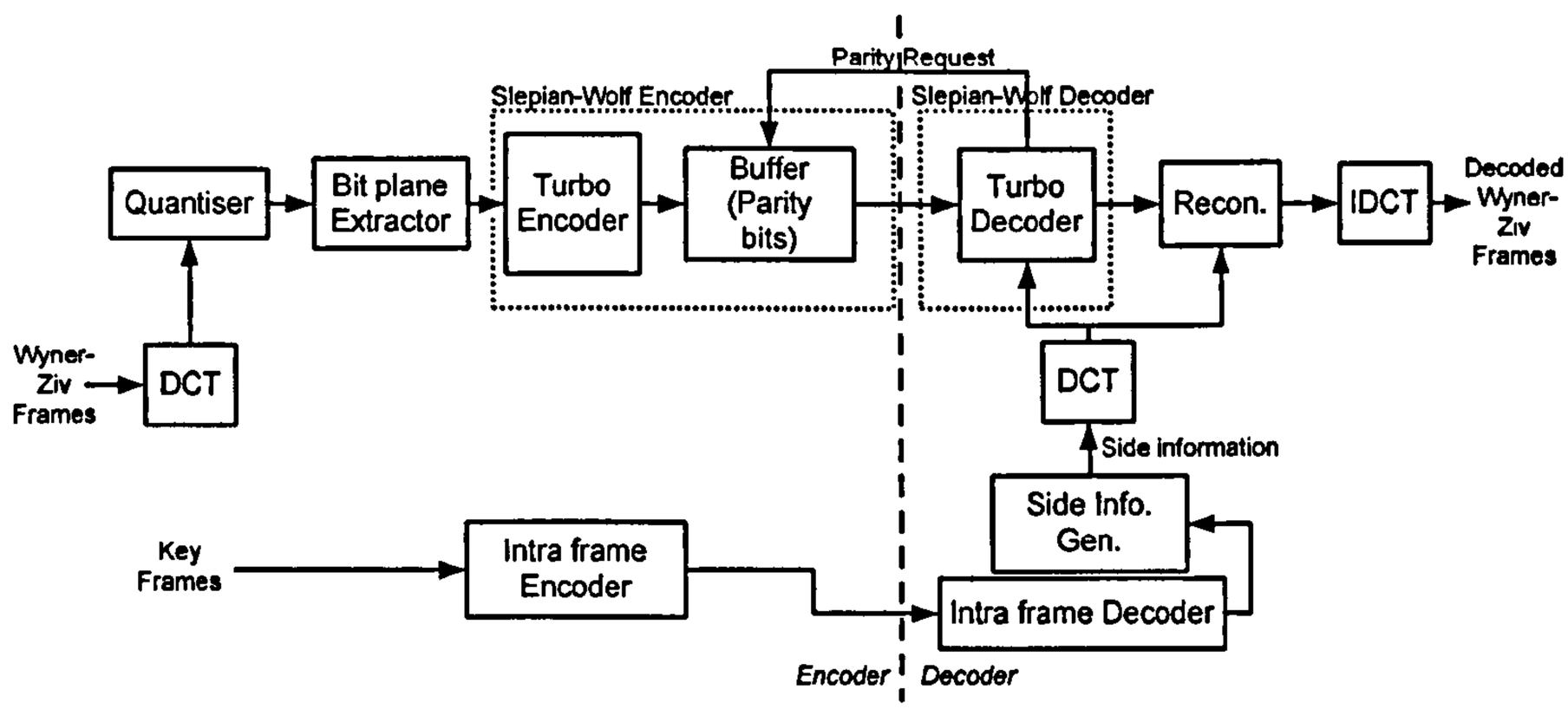



Figure 2-23: Transform domain DVC codec architecture using turbo coding

Except for the additional DCT and IDCT modules and some functional difference in the quantisation and reconstruction algorithms, the codec architecture is very similar to that of the pixel domain codec. The functionalities of the contrasting blocks are discussed below.

**Discrete cosine transform (DCT) and IDCT**

The DCT function in DVC utilizes the integer 4×4 block based transform as defined in H.264/AVC [22]. This involved a two dimensional matrix transform achieved through two one dimensional transforms in the horizontal and vertical directions for each row and column of the 4×4 sample matrix from the source picture. This utilizes the 4×4 transform matrix [22]:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \qquad 2\text{-}16$$

Applying the matrix multiplication to all non-overlapping 4×4 blocks of the picture involve only 16 bit arithmetic that can be implemented using only additions subtractions and shifts, which is not too computer intensive for the low complex DVC encoder to bear. This will transform each 4×4 block of pixels into a 4×4 block of DCT coefficients, each coefficient representing some degree of spatial correlation within the pixel block. The significance of information carried in each coefficient varies in terms of the image energy. The top-left coefficients, which correspond to the low frequency components of the carried information (see Figure 2-24), are more significant considering the human visual perception; human eye is less sensitive to the high frequency components of spatial information. This feature is exploited in the transform domain quantization described in the next section.

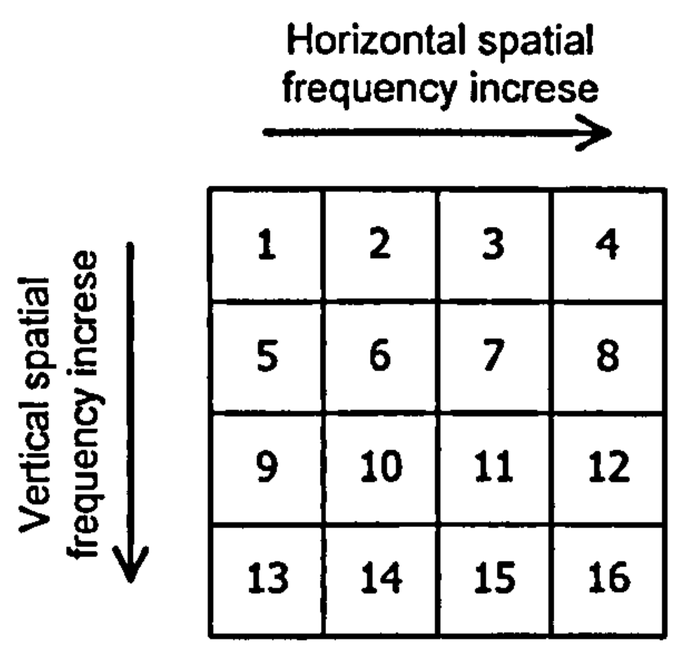



Figure 2-24 Transform coefficients

Then the inverse DCT (IDCT) performed at the end of decoder, utilizes the below inverse transform matrix as defined in the H.264/AVC standard [22]:

$$\tilde{H}_{inv} = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix} \qquad 2\text{-}17$$

**Transform domain quantisation**

Each DCT coefficient is quantised for compression. The number of quantization levels for each coefficient band is determined by its significance in terms of the effect to the human visual perception as discussed above. Accordingly, a set of quantization matrices are defined as shown in Figure 2-25, which reads the number of quantization levels allocated for each band [20]. Each matrix in Figure 2-25 yields one rate distortion point in the codec performance evaluation.

| | | | |
|---|---|---|---|
| 16 | 8 | 0 | 0 |
| 8 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |

Q1

| | | | |
|---|---|---|---|
| 32 | 8 | 0 | 0 |
| 8 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |

Q2

| | | | |
|---|---|---|---|
| 32 | 8 | 4 | 0 |
| 8 | 4 | 0 | 0 |
| 4 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 |

Q3

| | | | |
|---|---|---|---|
| 32 | 16 | 8 | 16 |
| 16 | 8 | 16 | 0 |
| 8 | 16 | 0 | 0 |
| 16 | 0 | 0 | 0 |

Q4

| | | | |
|---|---|---|---|
| 32 | 16 | 8 | 4 |
| 16 | 8 | 4 | 4 |
| 8 | 4 | 4 | 0 |
| 4 | 4 | 0 | 0 |

Q5

| | | | |
|---|---|---|---|
| 64 | 16 | 8 | 8 |
| 16 | 8 | 8 | 4 |
| 8 | 8 | 4 | 4 |
| 8 | 4 | 4 | 0 |

Q6

| | | | |
|---|---|---|---|
| 64 | 32 | 16 | 8 |
| 32 | 16 | 8 | 4 |
| 16 | 8 | 4 | 4 |
| 8 | 4 | 4 | 0 |

Q7

| | | | |
|---|---|---|---|
| 128 | 64 | 32 | 16 |
| 64 | 32 | 16 | 8 |
| 32 | 16 | 8 | 4 |
| 16 | 8 | 4 | 0 |

Q8

Figure 2-25 Quantization matrices

### *2.5.5 Recent developments in DVC*

Two pioneering DVC proposals bearing distinct frameworks have been proposed by the Universities of Stanford [4] and Berkeley [5]. The Stanford framework has been followed in majority of the subsequent research works on DVC and a number of structural enhancements are presented in literature, many of them focusing on the side information estimation process. A few selected notable contributions are listed below as relevant to the research activities discussed in this thesis.

**Side information improvement**

Side information generation has been the most widely focused aspect in DVC research, understandably due to its large impact to the rate distortion performance of the codec. Between 2002 and 2004, very limited literature has been published in this area, in fact DVC as a whole. Among them Aaron et al. proposals are dominant which considers simple pixel interpolation [4] and motion interpolation /extrapolation techniques [23] to predict the side information. Hash based motion compensation is proposed in [24]. Since 2005, a marked increase of research activity in DVC is noted. In 2005, Tagliasaccchi et al. proposed a motion compensated temporal filtering technique [25]. Natário et al. proposed an algorithm to generate side information based on motion field smoothening to provide improved performance [26]. Ascenso et al. presented another scheme using motion interpolation to derive the side information [27]. Artigas et al. proposed an iterative generation process for motion compensated side information [28].

In 2006, from our research group, Adikari et al. presented a sequential motion estimation technique using luminance and chrominance components of the frame [29]. Kubasov et al. presented a Mesh-based motion compensated interpolation approach [30]. A technique to improve performance by sending additional hash information to help side information was proposed in [31]. A modified recursive search block matching approach was discussed by Chien et al. in [32].Anew interpolation strategy with a, multi-estimation mode framework was published in [33]. Tagliasaccchi et al. further devised a Kalman filtering problem for comparison of three different scenarios: motion estimation at the encoder; motion interpolation at the decoder; and motion extrapolation at the decoder [34]. Dalai et al proposed a non-stationary noise model for the side information in [35].

In 2007, A novel motion compensated frame interpolation method for improving side information was proposed in [36] using edge information of decoded frames to generate more accurate interpolated frame. In [37], the authors propose a spatial splitting of the frames to enable exploiting the spatial correlations at the decoder to enhance the side information generation process. Yixuan et al. present a full search for improving side information [38]. Then a rate-distortion analysis of motion-compensated interpolation at the decoder was presented by Tagliasacchi in [39].

In some more recent work in 2008, a side information refinement by exploiting spatio-temporal correlation was presented by Ko et al. [40]. Another refined algorithm was proposed by Feng et al. in [41]. Badem et al. proposed a side information refinement using motion estimation in DC domain for transform-based DVC in [42].

**Improving the coding framework**

A number of proposals have been reported attempting to enhance the codec performance by modifying the coding framework. Accordingly, in 2006, Ascenso et al. suggested an adaptive GOP length scheme based on the motion activity of the frames [43]. Also, Tonomura et al. proposed a JPEG 2000 based DVC scheme in [44]. In 2007, an inverse bit plane decoding order was presented by Vatis et al. in [45]. In some other work, a scalable DVC solution was proposed in [46]. A block-adaptive coding scheme for transform domain DVC was proposed by Chien et al. in [47]. More recently in 2008, Martinez et al. presented a machine learning based approach to address the problem of reverse channel in [48].

**DVC based Video Communications**

Due to the premature nature of DVC research, the transmission channel noise is vastly ignored. A study of the error resilience of the DVC framework was presented by J.Pedro et al in [80] considering a packet based network. Several aspects of video compression and transmission in relation to wireless broadband networks were discussed using the PRISM architecture in [49].

**Reconstruction in DVC**

A basic reconstruction model was incorporated in the original Stanford proposal [4]. More enhanced approaches for this purpose were presented in [50] and [51] known as the minimum mean square error (MMSE) techniques. In [50], Vatis et al. made use of the average statistical distribution of the transform coefficients. In other notable reconstruction proposal [51], Kubasov et al. developed an algorithm to compute the reconstructed value $\hat{x}$ as the expectation $E[x \mid x \in [z^-, z^+), y]$ of the random variable $x$, where $z^-, z^+$ are the reconstruction bin boundaries.

**Other notable contributions**

There are numerous other literature proposing very effective modifications to the DVC codec for improving and analysing the performance. Ishwar et al. [52] presented an information

theoretic study of video codecs based on source coding with side information. Wu et al. proposed a non-uniform scalar quantiser based DVC implementation in [53]. A low complex DVC decoder using Wyner-Ziv coding with universal prediction was introduced in [54]. A critical rate distortion analysis of DVC and conventional coding schemes in the light of motion compensation was presented in [55].

Whereas turbo coding is the most common coding tool used in the Slepian-Wolf coder in the Stanford framework, other approaches have also been presented. As an example, Schonberg et al. discussed an the capabilities of achieving the Slepian-Wolf boundaries using Low Density Parity Check (LDPC) back in 2002 [56]. Some increasing attention on LDPC coding for this purpose has been recently evident. As an example, Westerlaken et al. discussed an LDPC based DVC design in [57]. Wang et al. presented a lattice vector quantisation (LVQ) based DVC scheme with LDPC in [58]. Hao Wang et al. prepared performance comparisons of different channel codes for DVC considering turbo codes and LDPC codes [59].

An analysis of the decoder complexity in DVC was presented by Fang et al. in [60].

While it is common to use DCT in the transform domain coding framework, some DVC proposals using wavelet coding are presented in [61], [62] and [63].

Some discussions are found in literature on the topic of correlation noise modelling in relation to the side information. In 2006, Dalai et al. presented the paper [64] on improving the codec integration through better correlation noise modelling. In [65] Lin-Bo et al. proposed a novel algorithm to model the non-stationary correlation noise statistics to improve the coding performance. Also, they have presented a related discussion on dynamic estimation of the virtual channel model in [66].

Two different aspects of the quantization in DVC are presented in [67] and [68], both of which are mutually independent of the quantization-related discussion in this thesis. [67] presented a nested scalar quantization and [68] presented a Lloyd-Max-based Non-Uniform Quantization Scheme.

In other different aspects of DVC, Maitre et al. presented a 3D scene modeling scheme for DVC [69]. Considering a multiview DVC scenario, a side information generation by fusion approach was proposed in [70]. In more recent work in 2008, a hybrid spatial and temporal

error concealment technique was proposed by Ye et al. in [71]. In [72], a zero motion skip and efficient DCT coefficient encoding algorithm was proposed based on an analysis of the zig-zag scan, run length code, and skipped macroblock techniques used in conventional video coding to improve the compression efficiency. In [73], Bernardini presented a performance evaluation of the DVC schemes considering a wavelet code and a non-feedback structure.

## 2.6 Conclusion

This chapter presented a review of the existing video coding technologies as related to the proposed scope of this thesis. A brief overview of the basics of video coding and the conventional video coding technologies was followed by an introduction to the distributed source coding principles governing the DVC paradigm. A detailed description of the existing DVC framework was then presented to help in understanding how and where the proposed technical proposals fit into the codec framework. This was followed by a review of the recent developments in DVC research.

The next four chapters present the technical proposals pertaining to this thesis starting with a TTCM based solution for the Slepian-Wolf codec in the DVC framework.

*Chapter 3*

# DISTRIBUTED VIDEO CODEC BASED ON TURBO TRELLIS CODED MODULATION

In the DVC codec architectures based on the Stanford framework, the Slepian-Wolf coder incorporates a channel coding technique, to generate a parity bit stream for the Wyner-Ziv frames. In this regard, turbo coding has been the most commonly utilized technique [4][20]. Other options considered in related research include Low-Density Parity Check (LDPC) codes [57]. In this chapter, an alternative novel approach for Slepian-Wolf coding is presented utilizing Turbo Trellis Coded Modulation (TTCM). Trellis coded modulation (TCM) and TTCM have been long-known in relation to communications applications for its high coding gain. Here, its efficient features are exploited to improve the compression efficiency of DVC.

The motivation of the proposed research, the implementation aspects in using TTCM in the Slepian-Wolf codec, and simulation results are discussed in the following sections. First, an overview of TCM and TTCM is presented.

## 3.1 Overview of TCM and TTCM

Trellis coded modulation [74] is a combined coding and modulation scheme, which is identified in communications as a bandwidth efficient signalling scheme. Even though the proposed DVC implementation does not use it as a modulation scheme in the same context, an understanding of the basic concepts and terminology is useful in realizing the proposed technique.

TCM combines ordinary rate $R = k/(k+1)$ binary convolutional codes with an $M$-ary signal constellation (where $M = 2^{k+1} > 2$). In doing this, a coding gain can be achieved without increasing the symbol transmission rate. The key feature here is that the redundant bits introduced by coding do not constitute extra symbols; but simply expand the size of the signal constellation. This inevitably yields a bandwidth saving; vital in a communications system. Some typical phase shift keying (PSK) two dimensional signal constellations are illustrated in

Figure 3-1. Amplitude modulation (AM) can also be used in generating TCM signal constellations.

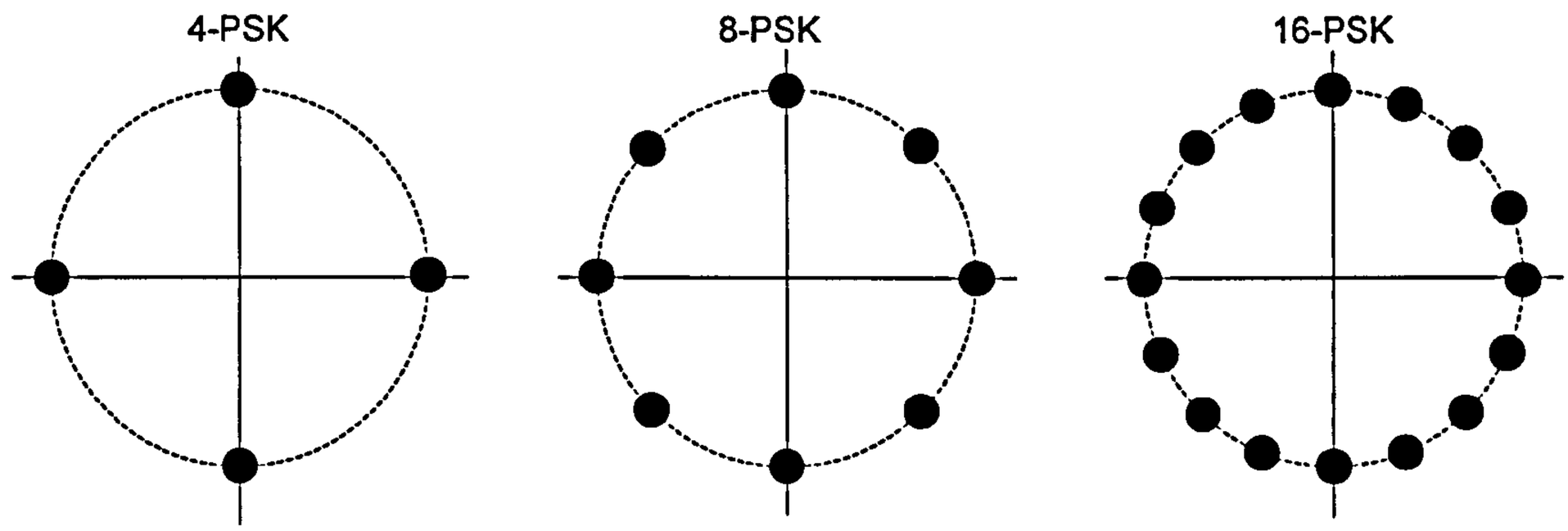



Figure 3-1 Typical PSK signal constellations

**Convolutional Encoder for TCM**

A rate 2/3 convolutional encoder structure that could be used to generate the redundant (parity) bits to expand the signal constellation for 8-PSK TCM is illustrated in Figure 3-2, together with the associated trellis diagram. In the encoder diagram in Figure 3-2, $U_i(D)$ and $V_i(D)$ correspond to the input and output bits of the encoder respectively.

Note here that high rate encoders, typically with only one redundant bit per symbol, are commonly utilized in TCM. This helps in meeting the energy constraints per signal point, in expanding the signal constellation in comparison to uncoded symbols. Here, *minimum squared Euclidean* (MSE) distance ($d^2_{\min}$) and *minimum free squared Euclidean* (MFSE) distance ($d^2_{free}$) are important parameters in determining the type of signal constellation.

In decoding, the Viterbi algorithm metric for two dimensional signal constellations is the distance in the Euclidean space. Therefore, the mapping of symbols on to the Euclidean space is very important in TCM. This problem is addressed by a concept called *mapping by set partitioning* [74].

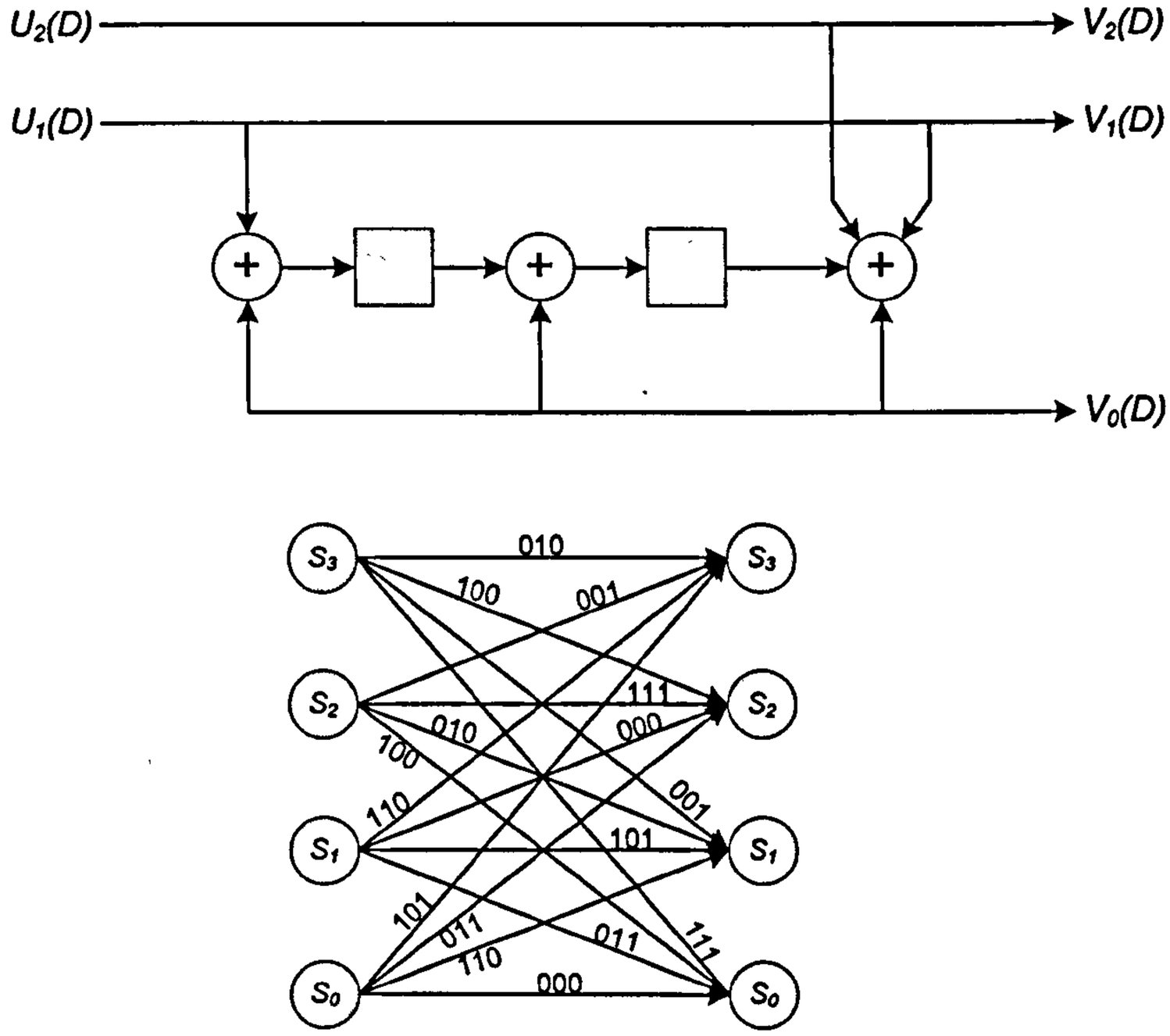



Figure 3-2 Rate 2/3 encoder diagram, and associated trellis for 8-PSK TCM

### *3.1.1 Set partitioning*

Set partitioning is the process of assigning binary labels to the signal points in the chosen constellation (for e.g. 8-PSK) representing the encoder output blocks. The design criterion in set partitioning is the maximization of the MFSE distance $d^2_{free}$ of the overall TCM system. This process involves the successive partitioning of the signal set into smaller subsets of equal size, and thereby generating a tree structure where each signal point corresponds to a unique path through the tree. An example set partitioning tree for 8-PSK TCM is shown in Figure 3-3.

It can be observed from Figure 3-3 that the minimum Euclidean distance, $d_i$ (and thus MSE distance $d^2_{\min}$) between the corresponding signal points at each level increases as we move down the tree. This yields an important property of resilience to noise of the signal bit that represents the set partitioning level in the tree. Thus the signal bits corresponding to lower levels in the tree are in effect better protected. Conversely, the signal bits that need better protection could be placed at the bottom of the set partitioning tree.

Symbol mapping is the process of generating the modulated symbols (e.g. 8-PSK) for the encoded bit stream, following the set partitioning algorithm discussed above. Figure 3-4 illustrates this process.

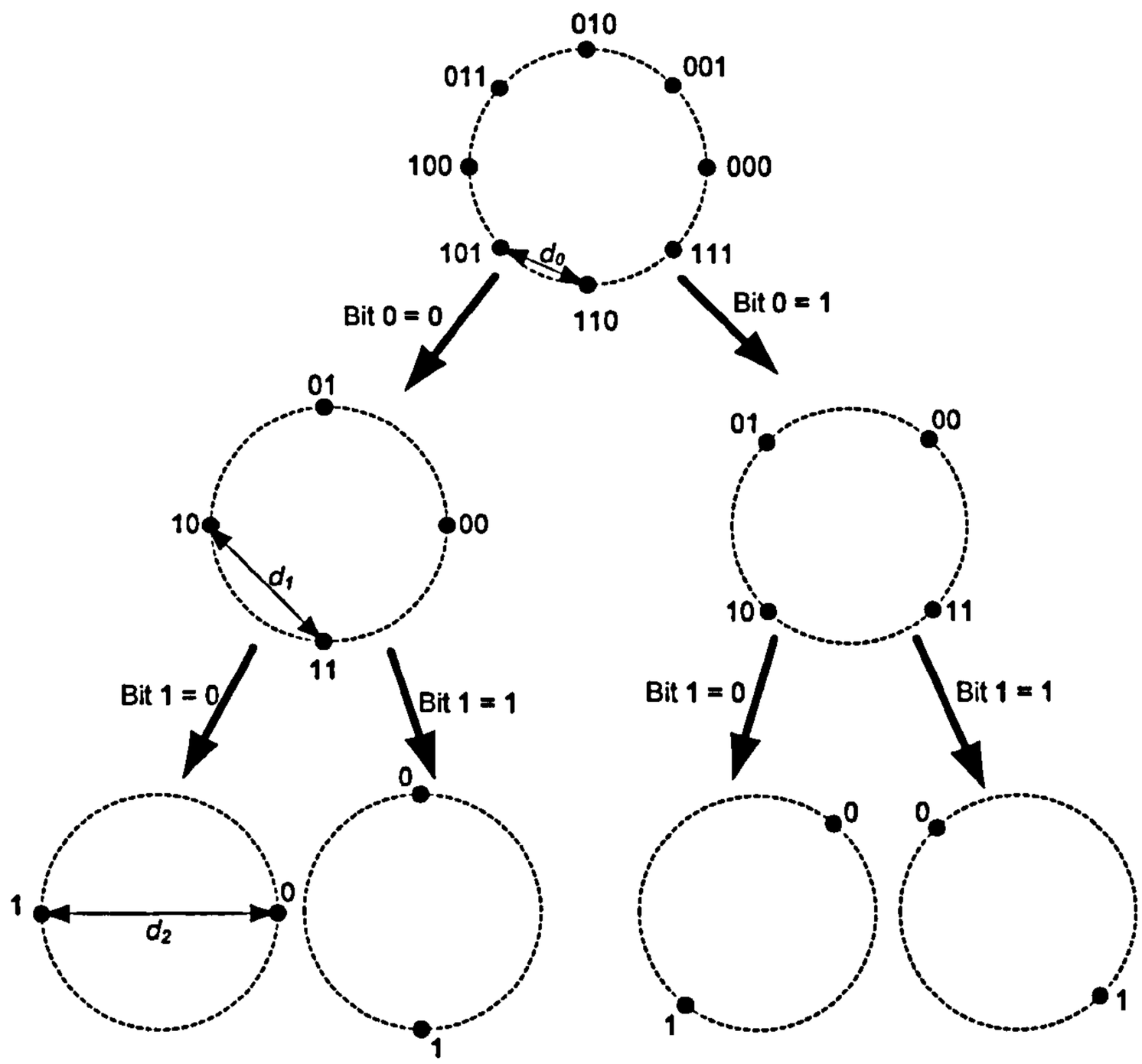



Figure 3-3 Set partitioning for 8-PSK TCM

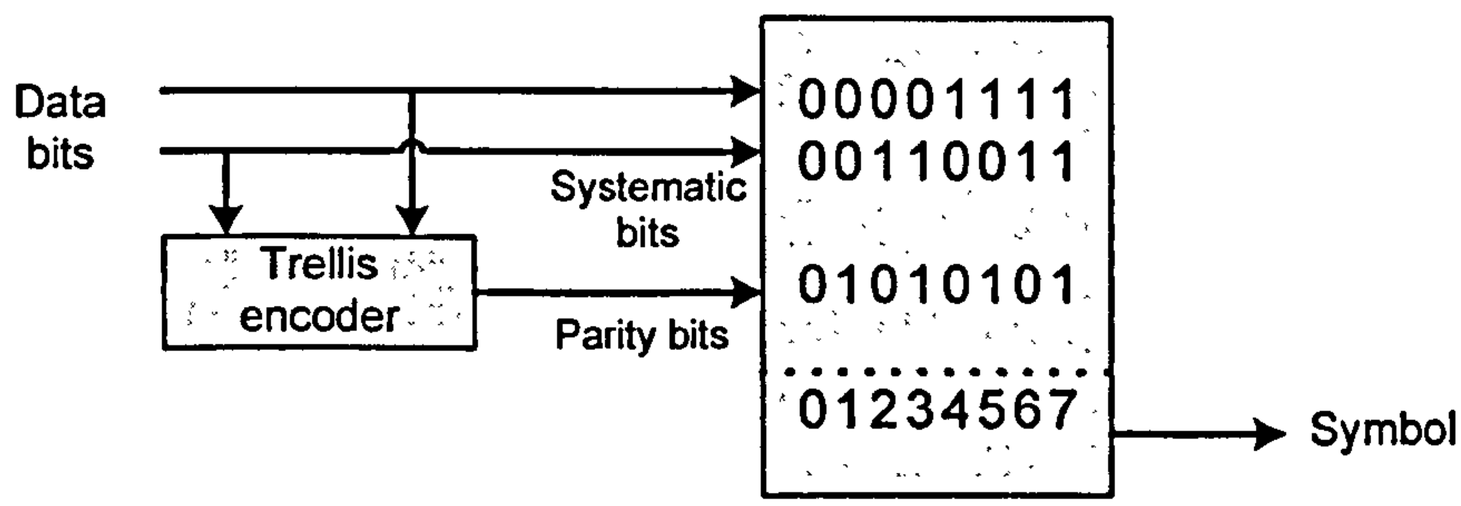



Figure 3-4 Symbol mapping for 8-PSK TCM with a rate 2/3 convolutional encoder

### *3.1.2 Turbo TCM (TTCM)*

TTCM encoder incorporates two parallel TCM encoders separated by an interleaver as illustrated in Figure 3-5. Both TCM encoders generate parity bits at identical rate and usually a symmetric puncturing pattern is used to select parity bits for transmission to the decoder.

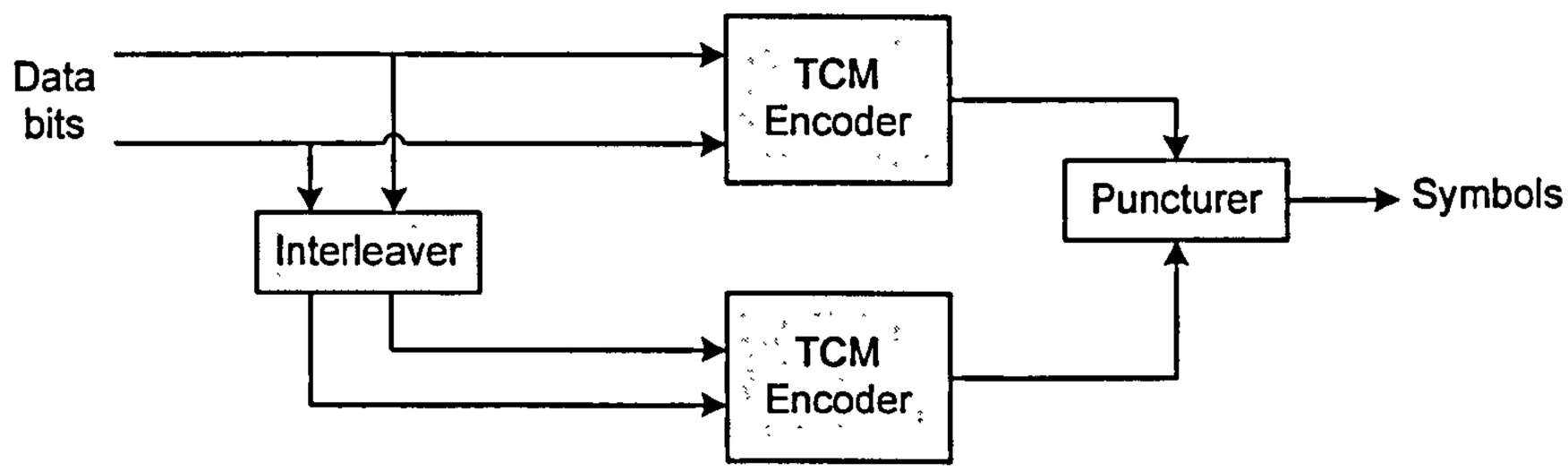



Figure 3-5 TTCM Encoder

## 3.2 Channel coding in the Slepian-Wolf coder

Based on the understanding of the TTCM technology as discussed above, let's focus on the design and implementation aspects of introducing it into the DVC framework. It is noted here that incorporating a channel coding technique in the Slepian-Wolf coder is somewhat different in context to the conventional usage of channel codes, even though the same encoding and decoding principles apply.

When turbo coding is concerned, a typical communication channel scenario will involve the transmission of both systematic and parity bit streams produced in the encoder, to the decoder over the channel. However DVC with turbo coding only transmit the parity stream over the channel; systematic bit stream is sourced locally through side information. Yet, the same decoding principles as in conventional turbo coding holds since the side information is modelled as the original systematic bit stream passed over a hypothetical channel. On this backdrop, the use of TTCM in the Slepian-Wolf is in itself a unique situation owing to the joint coding and modulation aspect of TTCM.

In the proposed solution presented in this thesis, the above problem of joint encoding and modulation scenario is managed by distributing the TTCM encoding process over the DVC encoder and decoder; the parity bit stream is generated at the DVC encoder, whereas mapping

to the signal constellation is performed at the DVC decoder. The proposed implementation is discussed in the next section.

## 3.3 DVC codec with TTCM

The design of the Slepian-Wolf encoder and decoder incorporating TTCM principles is presented in this section.

### *3.3.1 TTCM based DVC encoder*

A pixel domain DVC framework with the feedback channel as discussed earlier in chapter 2 is utilised as the basis for the implementation. As illustrated in Figure 3-6, the bit stream for the Slepian-Wolf encoding is prepared as normal by quantisation and bit plane extraction. Then a turbo trellis encoder, with two recursive systematic convolutional (RSC) encoders, is used to generate a parity bit stream. A rate 1/2 RSC encoder with a constraint length ($K$) of 4 and a generator polynomial of $[11\ 02]_{oct}$ as shown in Figure 3-7 is used in the implementation.

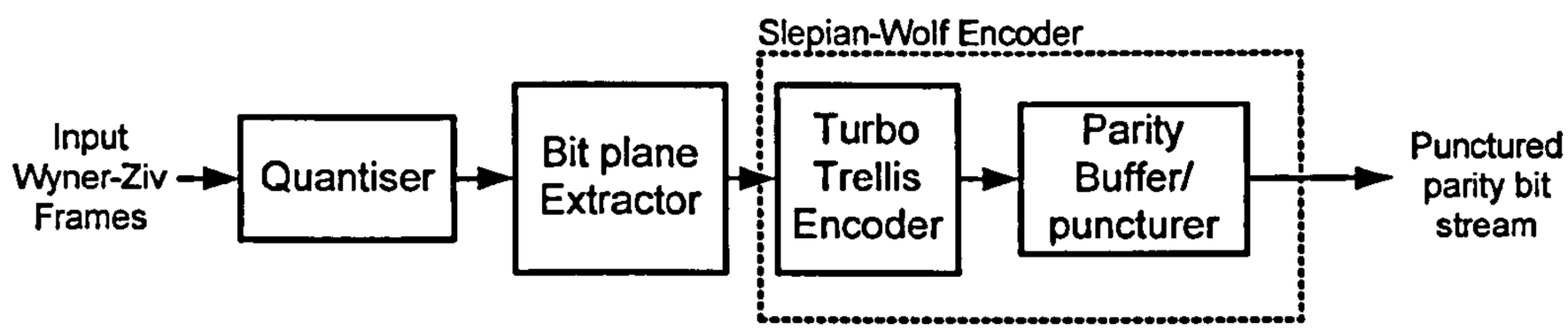



Figure 3-6 DVC encoder block diagram with TTCM

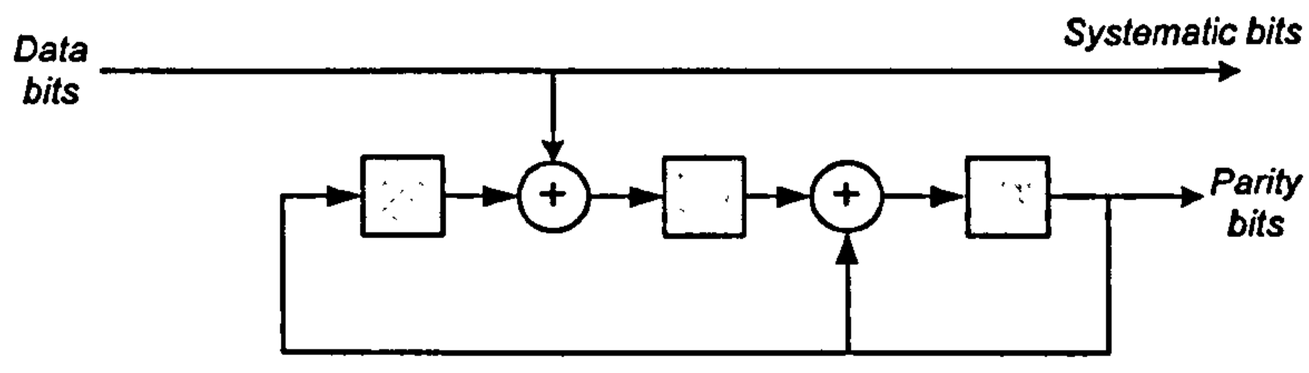



Figure 3-7 The RSC encoder in the TTCM encoder

Unlike in a conventional TTCM scenario in communications, the parity bit stream generated in the turbo trellis encoder is stored in a buffer and the systematic bit stream is discarded. The parity bit stream is then punctured at regular intervals and transmitted to the decoder in chunks

depending on the parity request messages received dynamically over the reverse feedback channel.

### *3.3.2 TTCM based DVC decoder*

The DVC decoder architecture is modified from the existing pixel domain DVC decoder framework to incorporate the signal mapping and TTCM decoding functionalities. The modified DVC decoder block diagram is shown in Figure 3-8. Each individual block is described below.

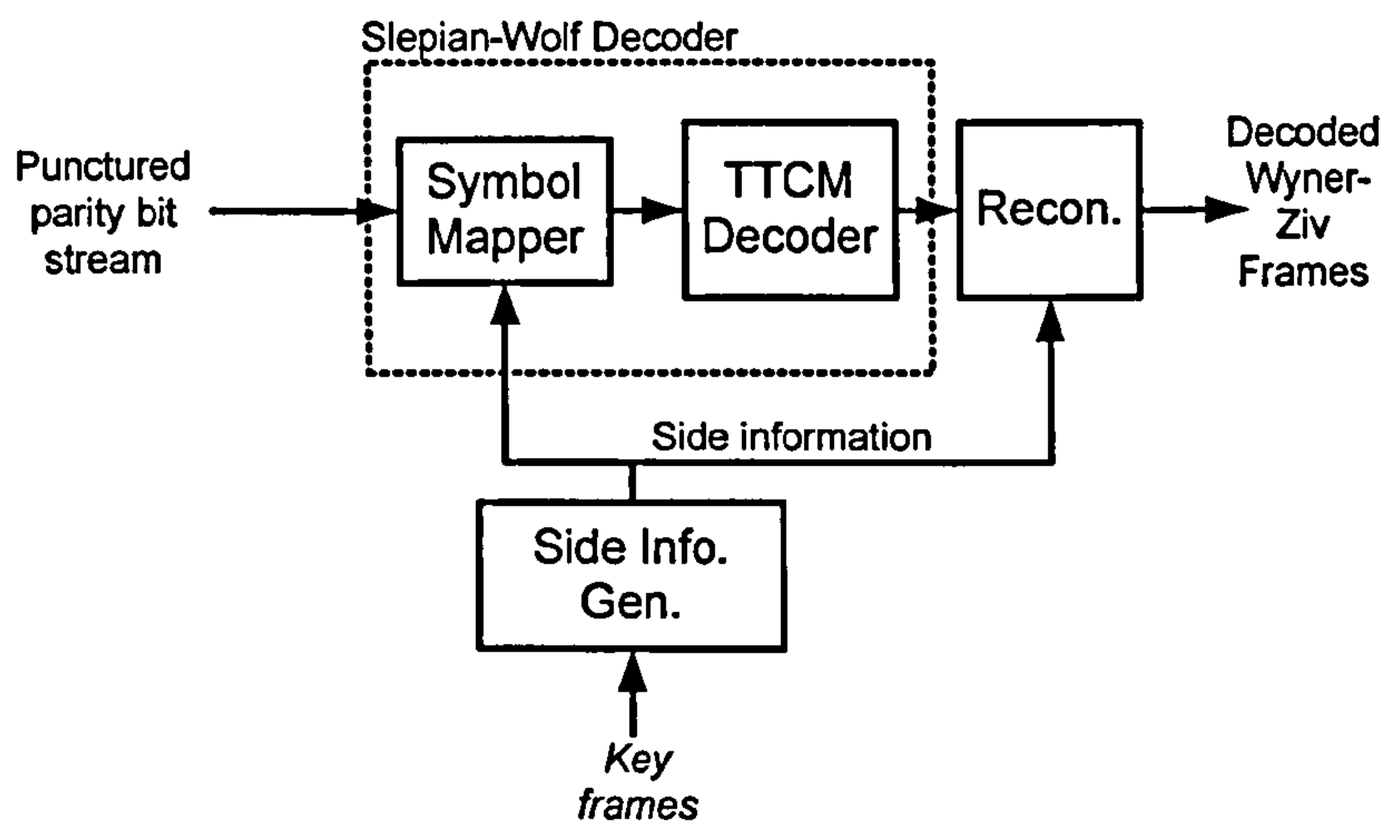



Figure 3-8 DVC decoder block diagram with TTCM

#### *3.3.2.1 Symbol mapping*

The considered implementation incorporates a 4-PSK signal constellation. In this case, 1 data bits and 1 parity bit would be enclosed into one PSK symbol. The associated set partition is shown in Figure 3-9. The systematic data bit to be protected would be placed at the second level with a higher Euclidean distance ($d_1$) between two constellation points compared to first level ($d_0$).

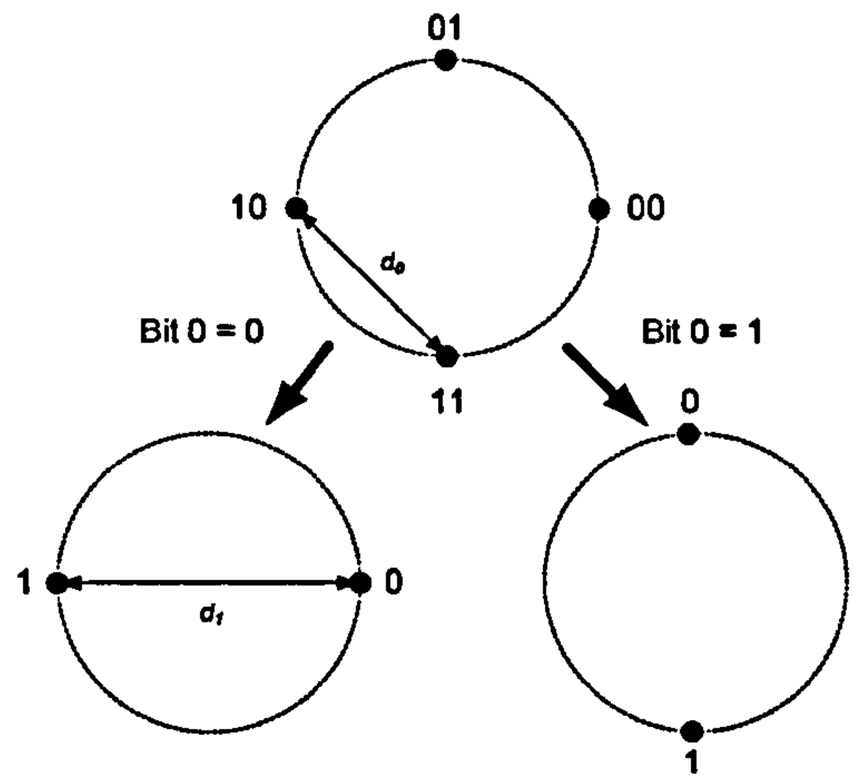
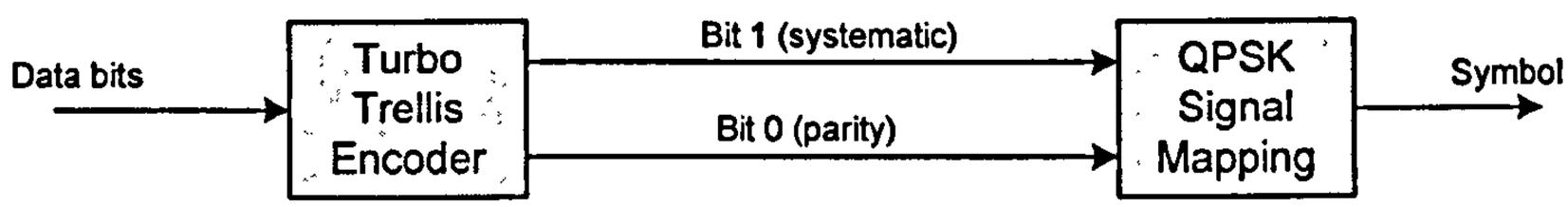
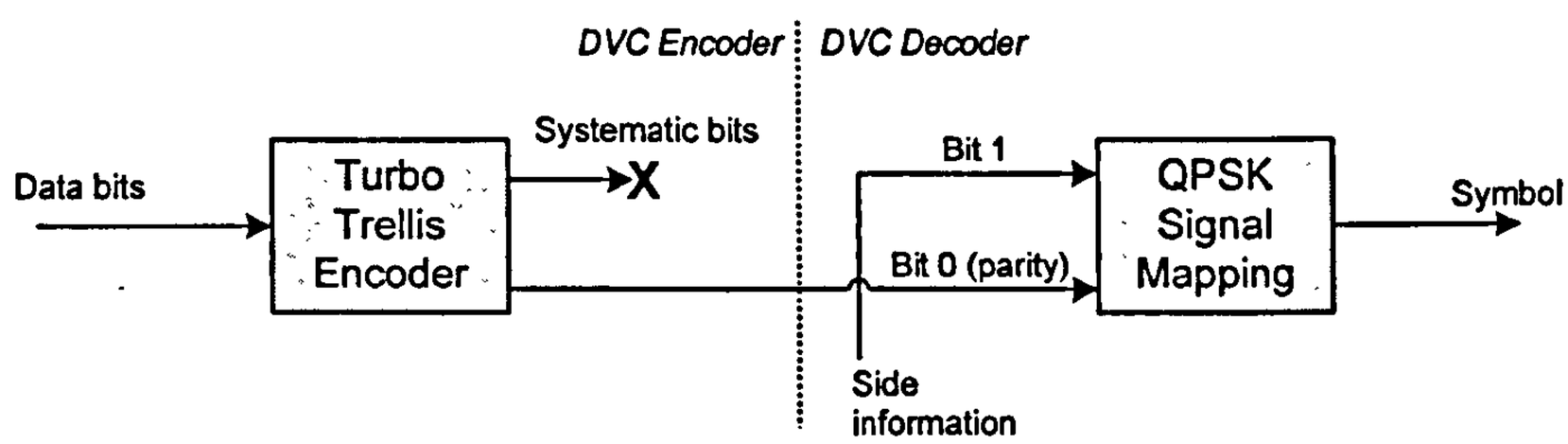



Figure 3-9 4-PSK set partitioning

In mapping to the 4-PSK symbols, the systematic bit stream is sourced from the side information, locally generated at the decoder. A comparison of the bit stream sourcing in the conventional TTCM case and in the DVC scenario is illustrated in Figure 3-10. First, Figure 3-10 (a) sows both systematic and parity bits from the trellis encoder. Then Figure 3-10 (b) shows the modification proposed for the DVC codec.




(a) Conventional TTCM scenario




(b) DVC scenario

Figure 3-10 Sourcing the bit streams for symbol mapping

It is noted that in DVC, the parity bit stream is subjected to periodic puncturing at the encoder before transmission to the decoder to achieve bit rate compression. This means that the systematic (side information) and parity bit streams are not balanced at the symbol mapping module, and thus the input cannot be mapped to the 4-PSK signal constellation. This problem is addressed by introducing a dual-mapper function, which falls back to a binary PSK (BPSK) modulation scheme when the parity bit is not available due to puncturing as shown in Figure 3-11.

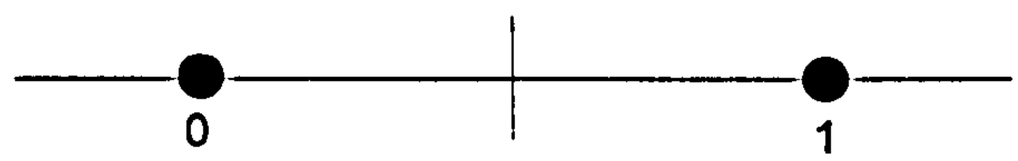



Figure 3-11 BPSK signal constellation for symbols when parity bit is punctured

### *3.3.2.2 TTCM Decoding*

The TTCM decoder incorporates a non-binary symbol based MAP algorithm [74]. The schematic of the TTCM decoder is presented in Figure 3-12. Since the parity bit stream is punctured at the encoder for shrinking the bit rate, the symbol mapping needs to be adapted as discussed in the previous section, by identifying the punctured bit positions. The hard decision decoding is performed after a number of soft iterations of the TTCM decoder as shown in Figure 3-12.

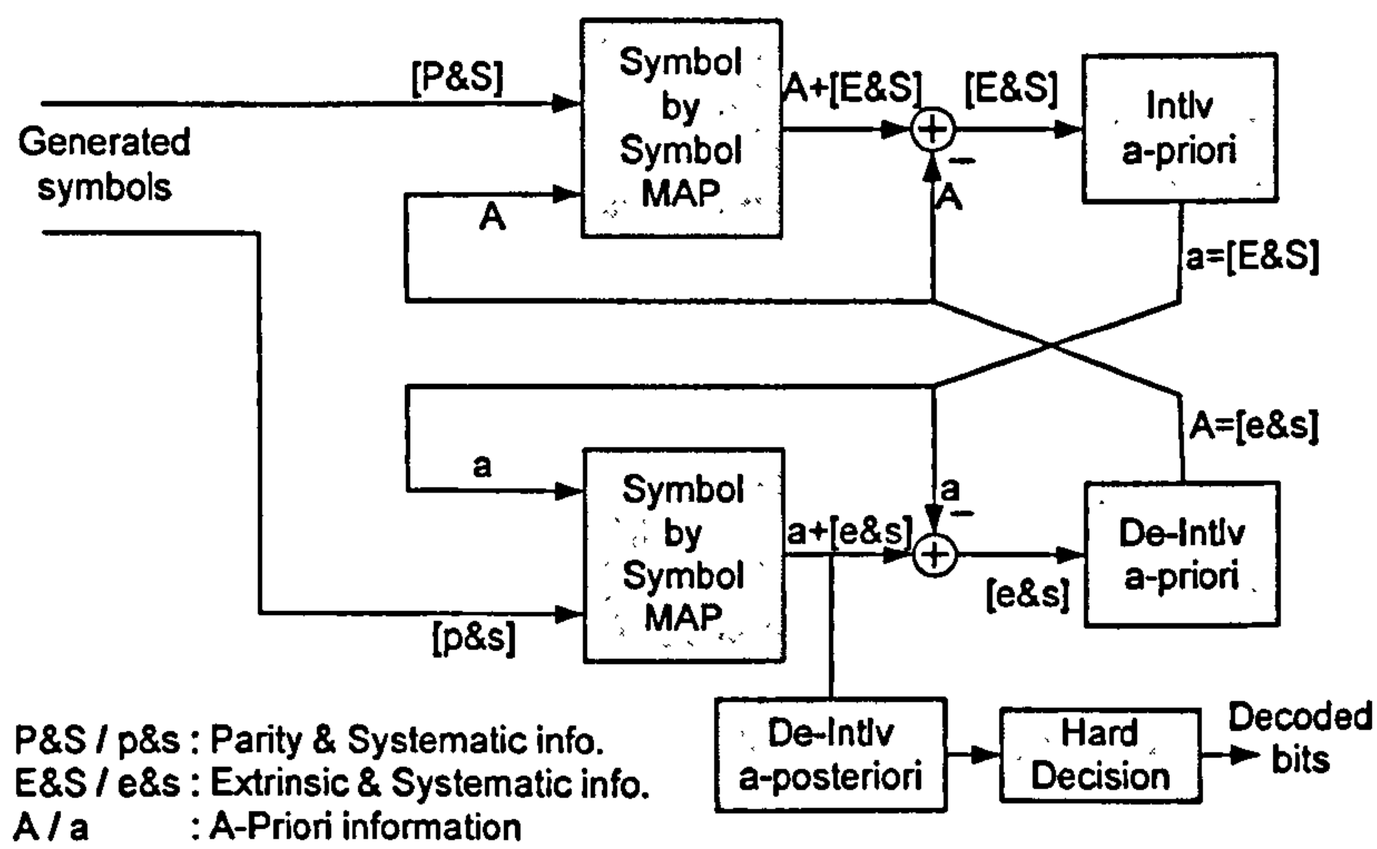



Figure 3-12 Schematic diagram of the TTCM decoder

### *3.3.3 Codec integration*

The integration of the full DVC codec incorporating the TTCM technology is illustrated in Figure 3-13.

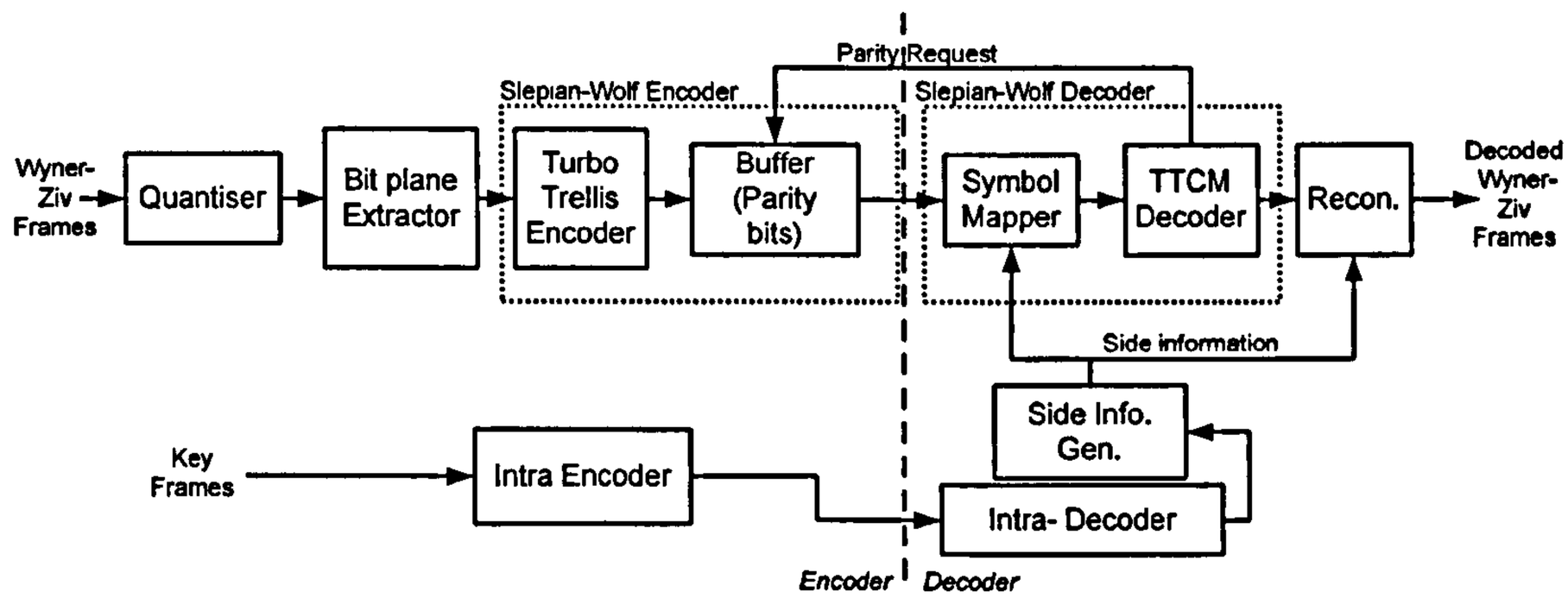



Figure 3-13 TTCM based DVC codec

As in the common Stanford DVC framework, the pixel domain quantiser is used to achieve the different rate-distortion points for the codec. The bit plane extraction is performed to generate the bit stream for the Slepian-Wolf encoder. The turbo trellis encoder generates the parity bit stream, which is then punctured and transmitted to the decoder. In this framework, the key frames shall be transmitted to the decoder using existing intra frame coding techniques (e.g. H264/AVC intra coding). During the simulations presented in the next section, the key frames are assumed to be available lossless for the generation of side information.

At the decoder, the side information is generated by exploiting the temporal correlations in the adjacent intra-decoded key frames. A pixel interpolation algorithm is utilized in this implementation used for the simulations presented here. Accordingly, the pixel position $(i, j)$ of the side information $Y_m$ is derived using the intra decoded key frames $X'_m$ as:

$$Y_m(i,j) = \tfrac{1}{2}\left(X'_{m-1}(i,j) + X'_{m+1}(i,j)\right) \qquad 3\text{-}1$$

where $m$ is the frame number. The signals are mapped to a combination of 4-PSK and BPSK symbols using the side information and parity bit streams, and the TTCM decoding is performed as discussed earlier in this chapter. The TTCM decoder runs for 6 iterations. Error

estimation is performed on the decoded bit stream similar to that utilized in the common feedback based DVC framework, as discussed earlier in chapter 2. The parity request messages are sent to the encoder based on the BER derived in the error estimation being above a pre-set threshold. A frame reconstruction function is finally used to smooth reduce the quantization noise in the decoded output.

As discussed earlier in chapter 2, the side information generated by the temporal interpolation of key frames is modelled using the original Wyner-Ziv frame information and a hypothetical channel noise. Most commonly, this noise is said to follow a Laplacian model, yet a resemblance with either Laplacian or Gaussian distributions is reported in [16]. However, estimation of the statistical properties of this correlation noise is an open research problem. In this regard, the sensitivity of the TTCM decoder in the DVC codec to the estimation accuracy is also tested in the simulations assuming both Laplacian and Gaussian noise models.

### *3.3.4 Simulations and Results*

The DVC codec with the TTCM in Slepian-Wolf coding is simulated for test video sequences using the test conditions listed below.

**Test conditions**

- Video sequences: *Foreman*, *Carphone*; QCIF (176×144); YUV 4:2:0; 100 frames; Wyner-Ziv frame rate: 15fps; GOP size:2

- Bit rate and PSNR calculated and displayed only for Wyner-Ziv frames, and averaged over the sequence.

- PSNR is calculated for the luminance signals only.

- Each rate-distortion point is achieved for pixel domain quantiser parameter: $M$, where $2^M = \{2,4,8,16\}$ defines the number of quantization levels.

The rate-distortion performance results for the *Foreman* and *Carphone* sequences are shown in Figure 3-14 and Figure 3-15 respectively.

The simulation results of the proposed DVC codec are compared with other video coding techniques currently available including:

- DVC codec proposed by Aaron et al [4] using a turbo decoder with side information by average interpolation.

- H.263+ and H.264/AVC codecs

The results show up to 6dB improvement in terms of the PSNR compared to that of a turbo codec with equivalent side information generation [4] as used in our implementation. The lower relative improvement for the *Carphone* sequence may be a result of the lower motion activity level of the sequence, thus the correlation noise in the side information is less and accordingly the room for improvement in the Slepian-Wolf coding performance is less. Other results using the conventional coding approaches, including state-of-the-art H.264/AVC codec are presented for the completeness in comparison. These are based on the redundancy exploitation at the encoder and it is understood that DVC rate-distortion performance still lags behind them by a notable margin.

Note here that enhanced side information estimation algorithms will significantly improve the overall performance of the codec. Further, it is noted that the simulations presented here involve RSC encoders of constraint length of 4 whereas the codec proposed in [4] uses length of 5. Shorter constraint length means reduced computational complexity.

**Sensitivity to correlation noise estimation**

The sensitivity of the TTCM based coding algorithm to the accuracy of the correlation noise estimation is tested. This analysis is considered as significant since a precise estimation of the correlation noise in the side information is still an open research problem.

As per the experimental observations, this noise in the considered side information stream proved to sufficiently closely resemble both Gaussian and Laplacian distributions. The statistical distribution parameters depend on a number of parameters; mainly the motion activity level and locality in each frame. As a result, when using the video codecs with side information, some sensitivity of the performance to the estimation accuracy of the noise distribution related parameters utilized in the maximum a posteriori (MAP) decoding algorithm

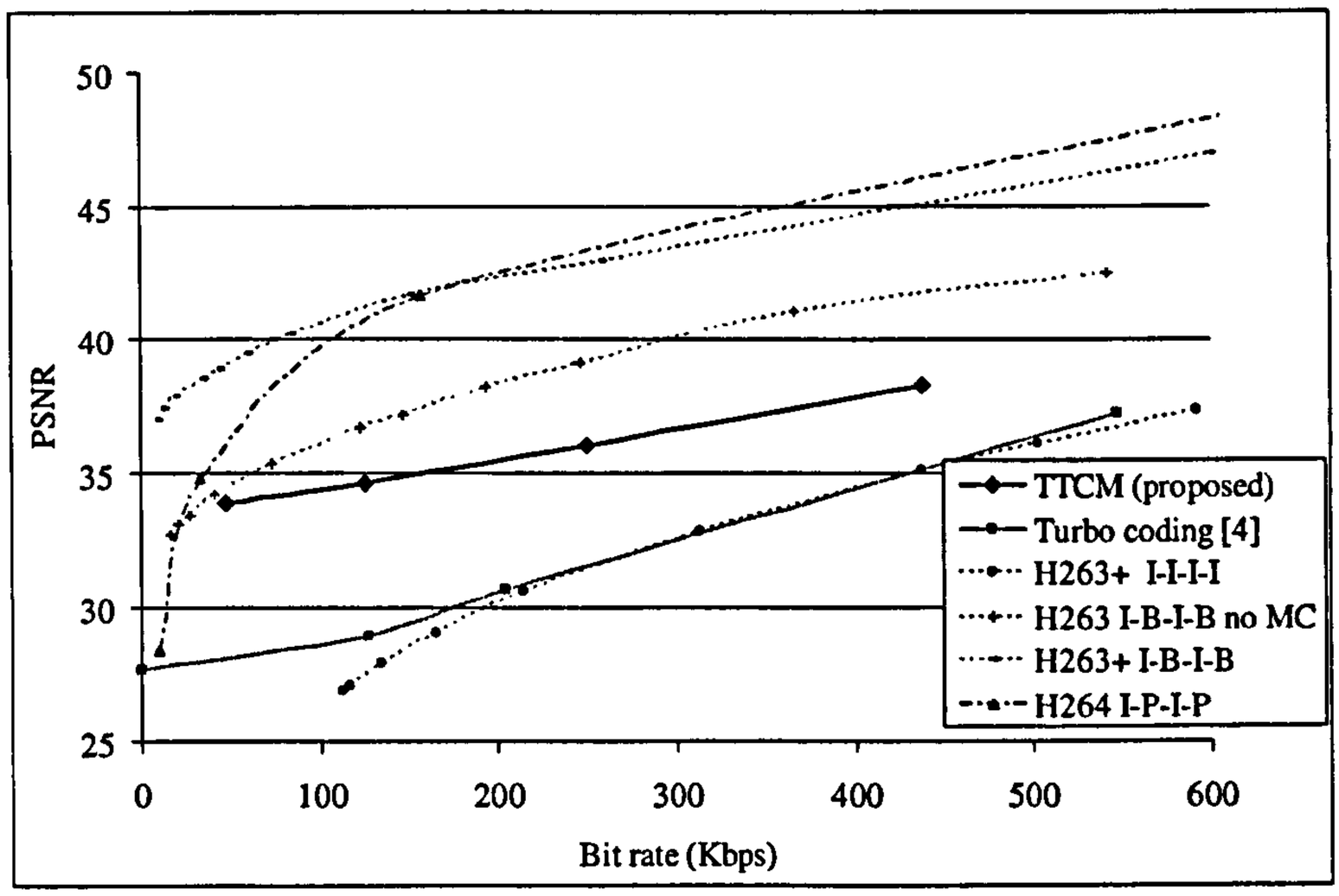



Figure 3-14 Performance comparison for *Foreman* sequence

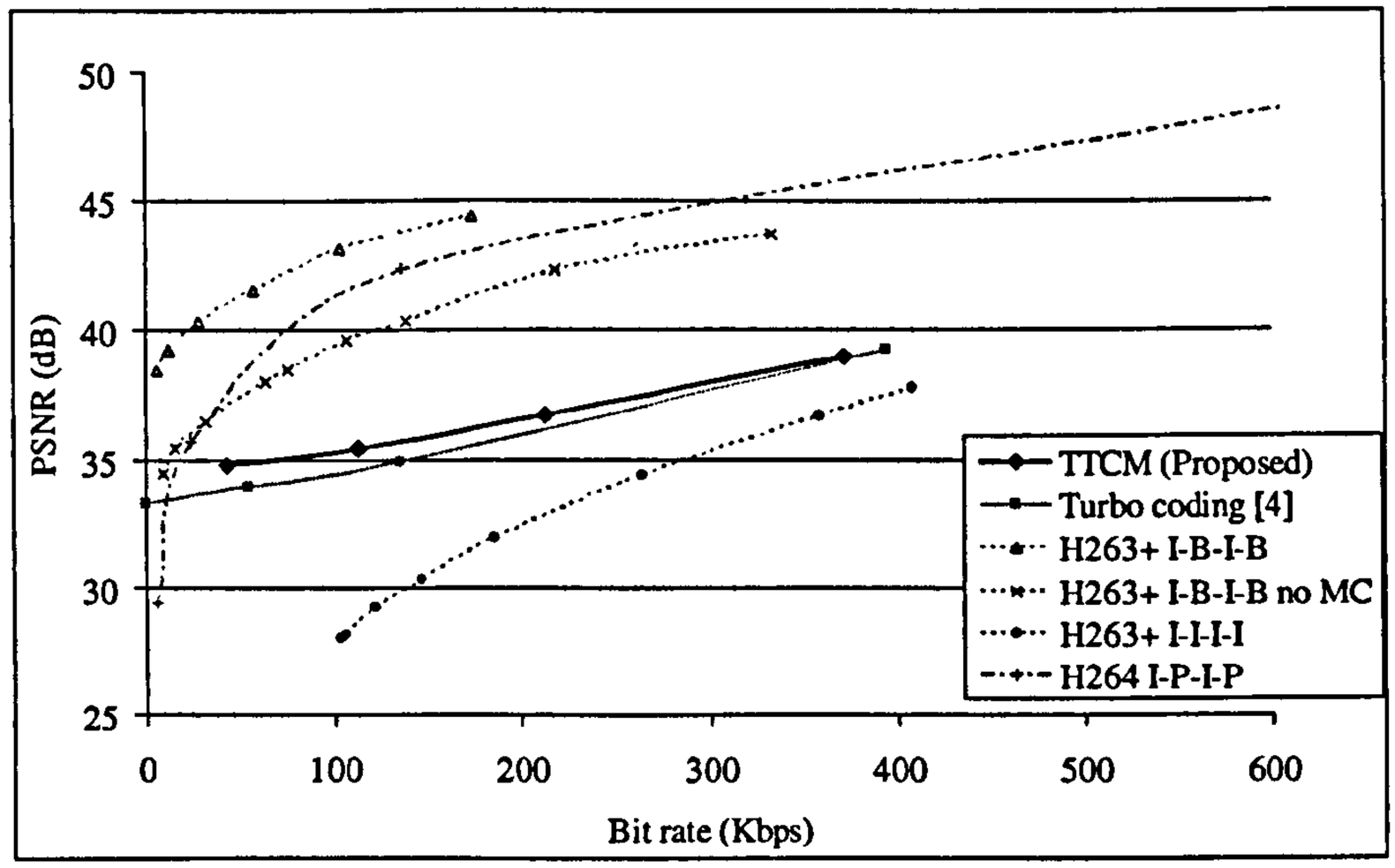



Figure 3-15 Performance comparison for *Carphone* sequence

is naturally expected. However, the TTCM based decoder used in this implementation shows a significant immunity to the sub-optimal estimations.

The simulation setup is described below:

- The decoded picture quality is maintained at a constant level by fixing the quantisation level (4 quantisation levels), side information quality and the decoding BER threshold.
- The correlation noise parameters used in the MAP decoder are manually varied. Effectively this becomes the only independent variable.

Thus, in the closed loop feedback based Slepian-Wolf codec, the bit rate becomes a function of the statistical noise parameters used in the MAP decoder. The bit rate is measured in terms of average bits per pixel (bpp) and the results for the *Foreman* sequence are illustrated in Figure 3-16 and Figure 3-17 for the Gaussian and Laplacian noise models respectively.

These results indicate that in addition to the bit rate for TTCM being lower than the corresponding value for turbo coding, the earlier is very stable over the range of noise variance parameter tested.

It is believed that this feature of the TTCM based DVC codec presented in this chapter would be very beneficial in realisation of a practical DVC codec, when used over dynamically varying motion levels since complex algorithms for estimating the correlation noise may be avoided.

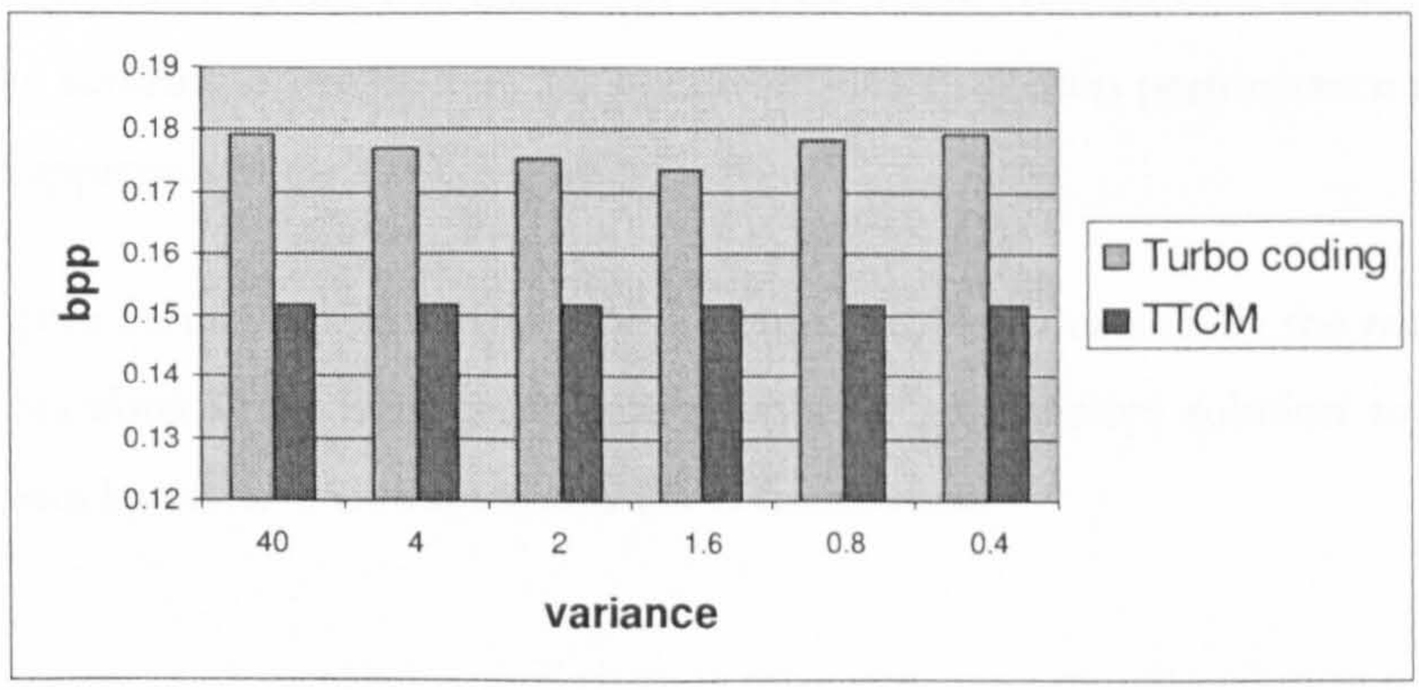



Figure 3-16 Performance comparison using Gaussian noise model

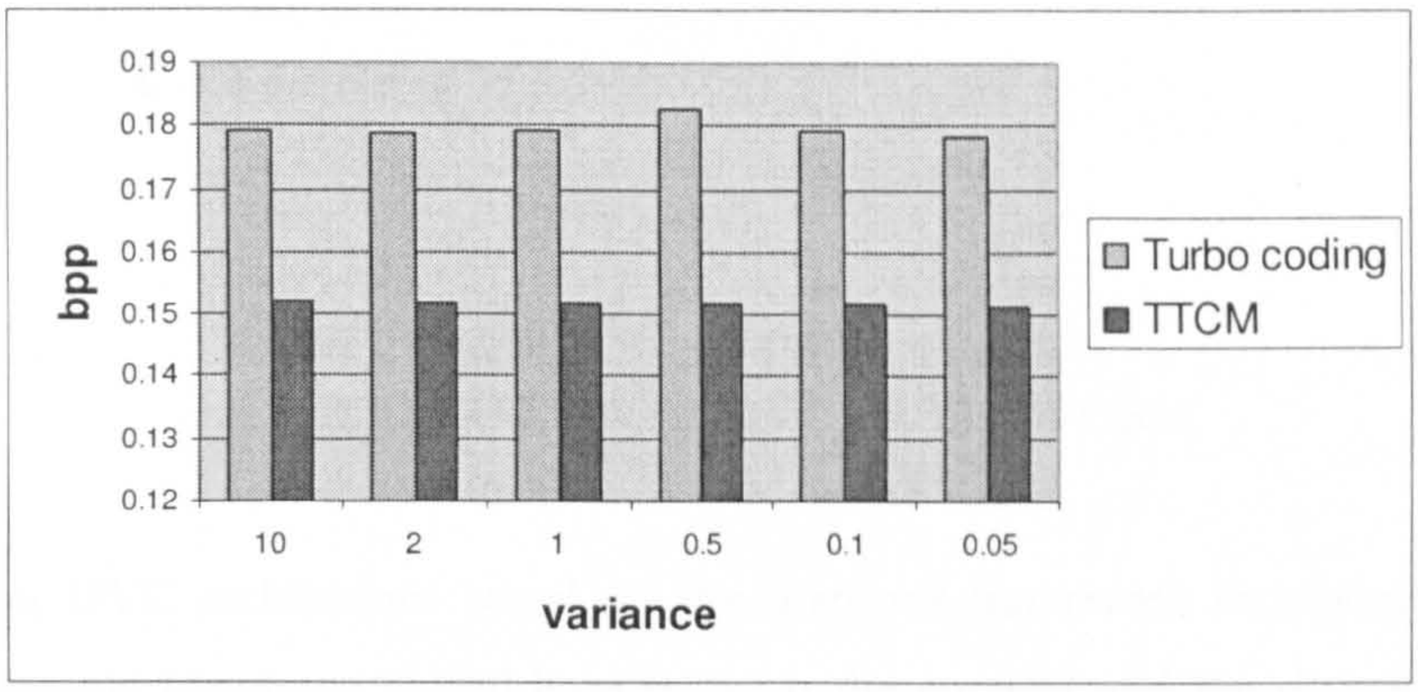



Figure 3-17 Performance comparison using Laplacian noise model

### 3.4 Conclusion

A channel coding algorithm is incorporated within the Slepian-Wolf codec in the Stanford DVC framework., Turbo coding has been the dominant choice of channel coding technology while LDPC also has been tried. In this regard, a novel approach was proposed in this chapter by introducing the TTCM concept into DVC. This proposal was motivated by the high coding gain of TTCM. In designing the TTCM based Slepian-Wolf codec, notably, the TTCM encoder was arranged so that the parity bit stream generated at the encoder is punctured and transmitted as a bit stream to the Slepian-Wolf decoder for mapping into the complex symbols in conjunction with the side information. A set-partitioning technique was utilised in mapping to the signal constellation for optimising the decoding performance. The performance of the proposed TTCM based codec was compared with the more conventional turbo coding based approach. The simulation results depict a significant rate distortion performance gain in using the proposed approach in the DVC codec.

The next chapter focuses on identifying the potential problems caused by the *reverse channel* in the existing Stanford DVC framework, and proposing an effective solution to this reverse channel problem based on a Unidirectional DVC framework.

*Chapter 4*

# UNIDIRECTIONAL DVC FRAMEWORK

The common DVC architecture based on the Stanford framework incorporates a reverse (feedback) channel forming a closed loop between the encoder and the decoder [4]. In this feedback based implementation, as discussed in detail in chapter 2, the decoding process starts with arbitrarily small amount of parity bits per block (the block size in the Slepian-Wolf coder is an implementation-dependant parameter) and error estimation is performed on the output of the decoder. If the derived bit error rate (BER) is above a pre-set threshold, a parity bit request message is sent to the parity buffer held at the encoder. The reverse feedback channel is utilised to communicate these dynamic parity request messages from the decoder to the encoder ensuing in an optimum dynamic rate-control implementation. However it is observed that this dynamic feedback mechanism is a practical hindrance in a number of practical scenarios. Some of the drawbacks of this closed loop coding framework can be identified as:

- There is a variable delay in decoding each block, based on the number of iterations the closed loop takes to achieve the desired BER. If the transmission delay between the transmitter and receiver is large, the above delay will prohibit any real time coding operation.

- The closed loop is feasible only in real-time coding scenarios. This does not permit continuous on-site data storage for offline processing, which is a very common necessity of practical video coding applications.

- Not suitable when a reverse communication channel is not available or having one is highly expensive.

Unidirectional DVC is a concept that proposes a practical solution to this problem. This chapter discusses practical considerations in designing unidirectional DVC algorithms and present appropriate solutions. Here, two distinct approaches are taken in designing a unidirectional DVC codec as described below.

1. Encoder rate control (ERC): Conceptually, ERC is the process of taking the rate decisions at the encoder, without receiving supporting information from the decoder. A few ERC algorithms are presented in this chapter.

2. Decoded frame refinement: This is a novel concept that proposes dynamic iterative refinement of decoded frame sequence at no additional bit rate burden and no interaction with the parity pattern selection or any ERC processes involved. An implementation algorithm is presented later in this chapter.

It is understood that any ERC algorithm will add some computational complexity to the otherwise extremely low complex DVC encoder. Decoded frame refinement, on the other hand does not affect the low encoder complexity feature. The additional processing power demanded in the decoder, in the latter option, could be considered a practically less of a burden in the many-to-one type video communication systems targeted by DVC.

Now let's start with a deeper discussion of the concept of *rate control* as applicable to DVC, for the ease of understanding the challenges faced in ERC. A few implementation algorithms for ERC are presented in section 4.2. Then the section 4.3 proposes an advanced algorithm for decoded frame refinement, by iteratively exploiting the temporal and spatial correlations in the source video sequence.

## 4.1 Rate Control in DVC

In the existing Stanford DVC framework, the rate control process is based on the closed loop feedback mechanism as discussed earlier in chapter 2. Eliminating the feedback path to design a more practically realistic coding framework is a challenging task, mainly due to the constraints imposed in the independent encoding scenario as specified in the distributed source coding principles [2][3]. Reaching the rate conditions and boundaries depicted in the Slepian-Wolf theorem [2] in a practical sense with independent encoding of statistically dependant information streams is yet a far away goal for the researchers.

Figure 4-1 shows a simplified block diagram of the DVC codec to illustrate the rate control mechanism as discussed below.

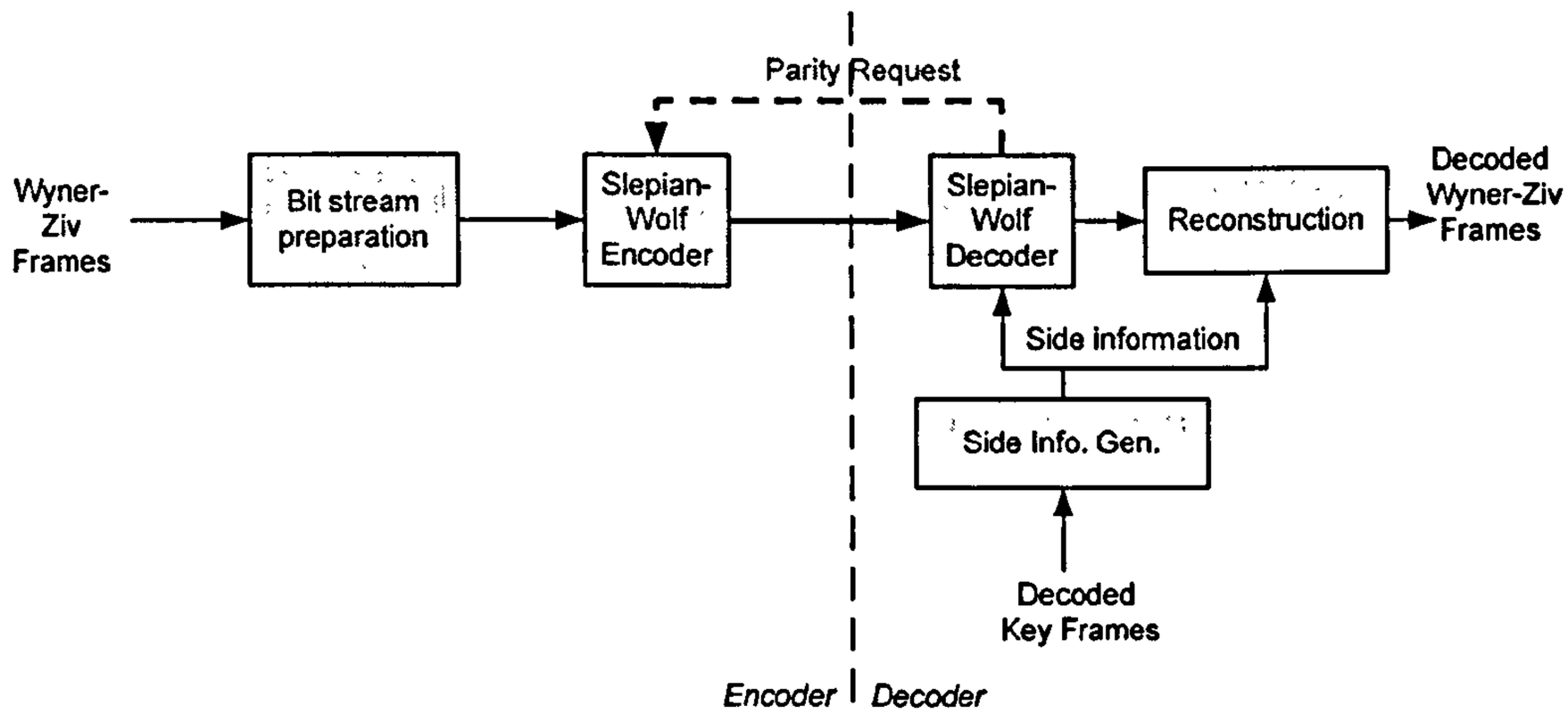



Figure 4-1 Simplified DVC codec

In the DVC codec based on turbo coding, as considered in this discussion, puncturing of the parity bit stream generated by the turbo encoder plays a significant role in determining the rate of video compression. In this scenario, rate control effectively means the determination of the optimum parity puncturing pattern for each block of bits. The optimum parity puncturing pattern is noted to be a function of the correlation noise between the side information and the original Wyner-Ziv bit stream for the block. The higher the correlation noise: higher will be the parity bit rate necessary to decode the block for the desired accuracy. However, there are no means of directly evaluating this correlation noise as the original Wyner-Ziv bit stream and the side information bit stream are only available at the encoder and decoder respectively; they are not co-located. The solution to this problem incorporated in the existing DVC codecs is *decoder rate control* or what we more commonly identify as the *reverse channel feedback*.

ERC is proposed as a more practically viable solution to this rate control problem as discussed earlier in this chapter. However, there are a number of key challenges to be realized in designing an efficient ERC algorithm. The next section elaborates on these challenges and some innovative solutions adopted in addressing them.

## 4.2 Encoder Rate Control (ERC)

Even though ERC is conceptually very appealing, it has its own obstacles barring straight forward implementation:

- Computing an estimate of the correlation noise at the encoder, in order to determine the optimum parity puncturing pattern, necessitates the regeneration of a replica of side information (usually available at decoder) at the encoder.

- Side information estimation is usually a highly computationally expensive process involving motion estimation, interpolation and compensation. This contradicts with the low complexity feature demanded by the targeted application scenarios.

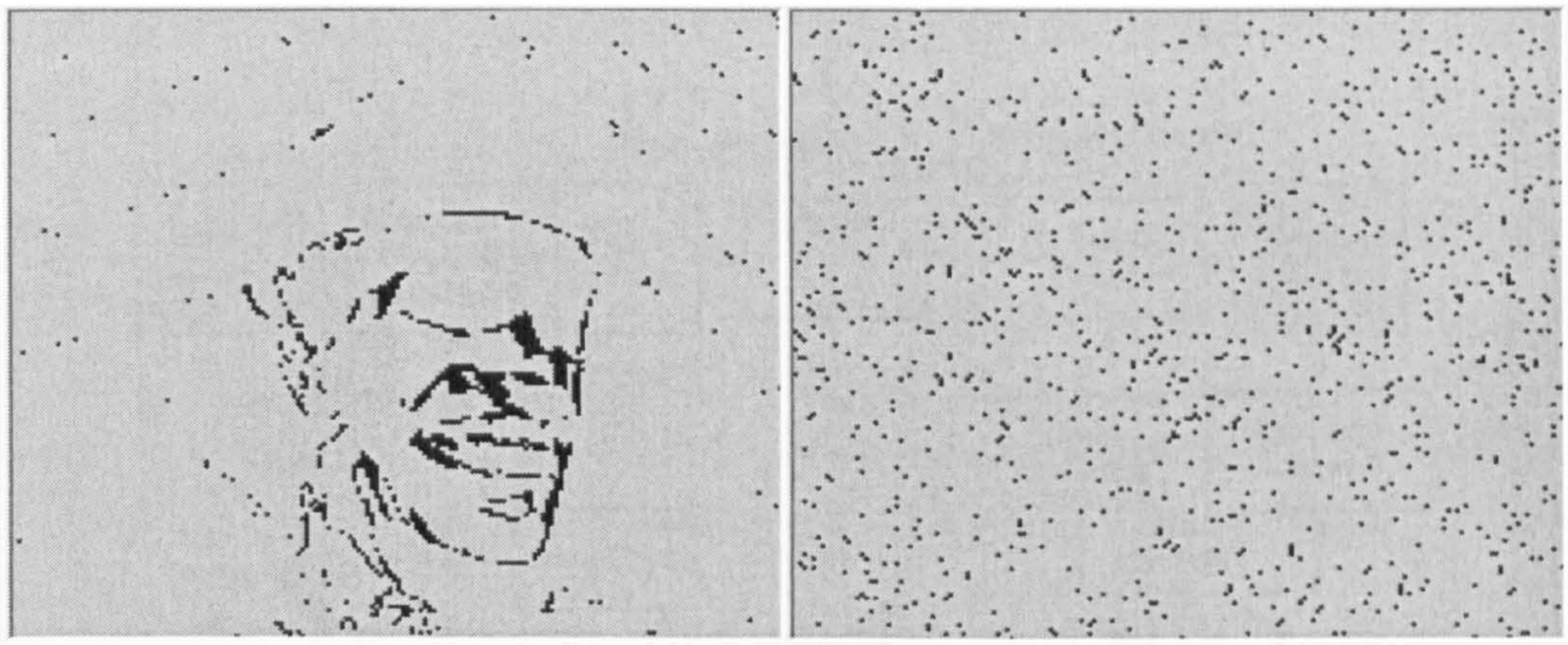
In the light of above developments, the objective of this section is to design an extremely low complex encoder for unidirectional DVC. In view of achieving this a few low complexity algorithms are presented herein. First, a pixel scattering algorithm is introduced to randomise the errors within each block in a fixed rate parity transmission setup. Then a low complexity technique is proposed to estimate the correlation noise at the encoder utilizing the adjacent key frames and thereby determine the parity bit pattern.

### *4.2.1 Pixel Scattering*

Let's consider a pixel domain DVC codec with no reverse channel. The parity bit puncturing pattern is fixed at a pre-determined rate for each block of bits, in its simplest form. This structure results in low rate distortion performance since the parity bits are not allocated efficiently, not reflecting the usually localized error distribution in each frame. Pixel scattering is a low cost approach taken to address the problem of localised errors in the bit stream to be encoded that leads to sub optimal performance in the Slepian-Wolf coding process. Use of interleavers is an established concept in communications, and here this is adapted for randomising the bit errors caused by localised motion in the video sequence as shown in Figure 4-2. The figure illustrates the residual error distribution of the side information with reference to the original Wyner-Ziv frame, for a single bit plane.

In Figure 4-2 (a), the error distribution usually follows the motion activity in the frame. Localized motion activity on a still background is a very common observation in a surveillance video footage. Splitting a single bit plane in to multiple blocks for Slepian-Wolf coding helps to localize the bit errors, yet the fixed parity allocation bars the efficient exploitation of this property as:

- Blocks with high BER get insufficient parity bits.



(a) (b)

Figure 4-2 Randomising bit errors (*Foreman* sequence; first bit plane)

- Blocks with low (or zero) BER wastes parity.
- Also, non-random noise distribution within the blocks is not an ideal scenario for turbo coding.

The introduction of pixel interleaver results in the randomization of bit errors as shown in Figure 4-2 (b); thus improving the turbo decoding performance by improving the utilization of parity bits.

The block diagram of the unidirectional DVC codec incorporating the proposed functionality is shown in Figure 4-3.

In the proposed system, the Wyner-Ziv frames are interleaved in pixel domain before the bit stream is prepared to be sent through the encoder. The interleaver module uses a pseudo-random interleaver index of length ($w*h$) for the luminance pixels where $w$ and $h$ represent the width and the height of the frame. The quantization and bit plane extraction is performed as usual and turbo encoded. Then a predetermined fixed-interval puncturing algorithm is enacted

over the parity bit stream generated in the turbo encoder. The punctured parity bit stream is transmitted to the decoder.

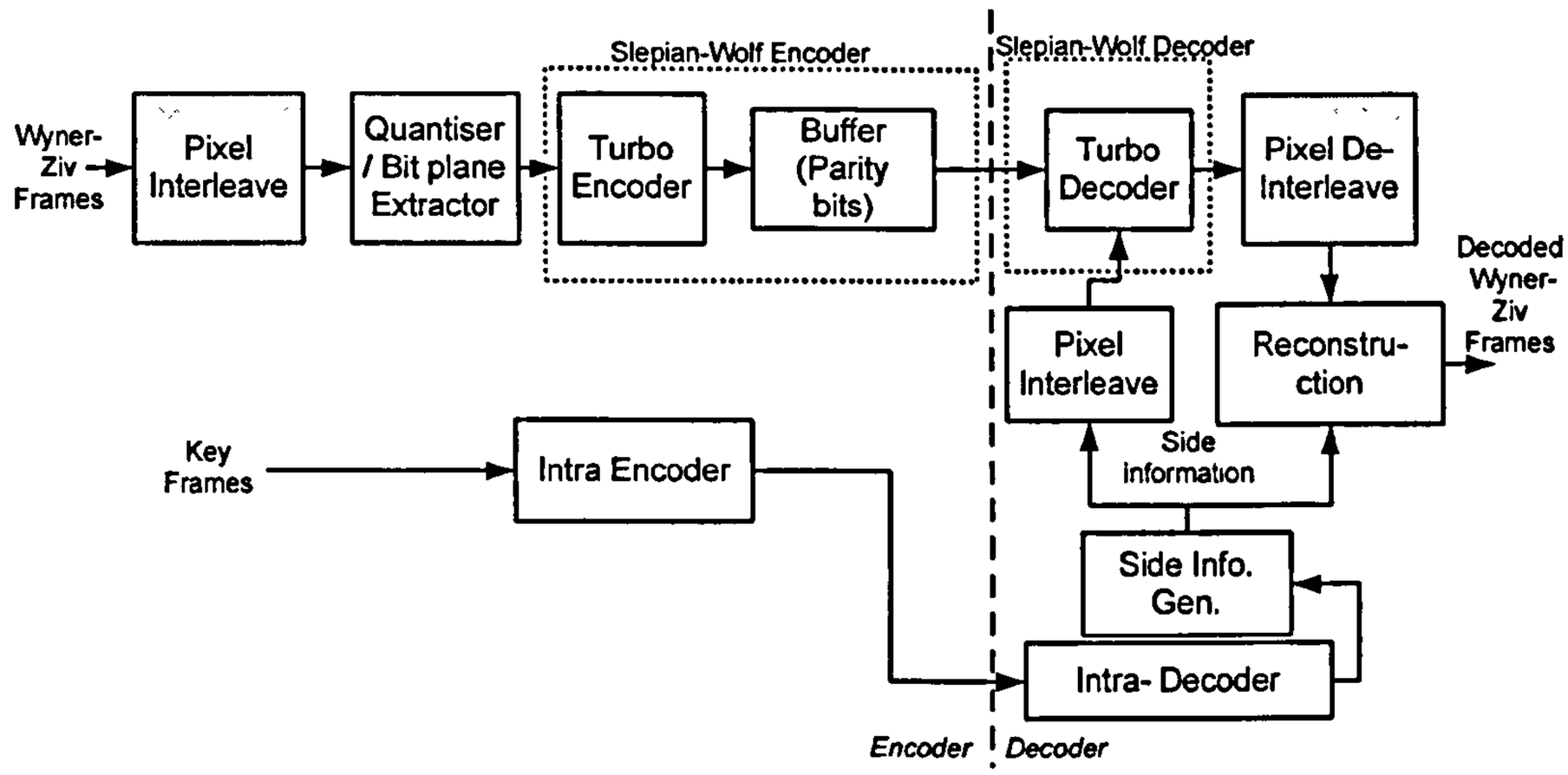



Figure 4-3 DVC codec with pixel scattering of the input video frames

The decoder incorporates an identical pixel interleaving process over the side information stream and a de-interleaver is performed over the decoded bit stream before the video frame is reconstructed.

**Simulation Test Conditions**

- Video sequences: *Foreman* (medium/high motion), *Claire* (low motion); QCIF (176×144); YUV 4:2:0; 100 frames; Wyner-Ziv frame rate: 15fps

- PSNR calculated and displayed only for Wyner-Ziv frames

- PSNR is calculated for the luminance signals only.

- Fixed puncturing interval is set to achieve a constant average bit rate of 1 bit per pixel.

- Reference codec for comparison: A unidirectional codec with fixed parity puncturing as in the proposed codec, but no pixel scattering, with motion extrapolated side

information as in [26]. This enables a fair comparison of the effects of the proposed contribution of this section.

**Simulation Results**

The modified DVC codec is tested for a number of test video sequences and the results for the *Foreman* and *Claire* sequences and are shown in Figure 4-4 and Figure 4-5 respectively.

It is observed that the proposed codec shows consistently improved performance over each sequence. An improvement average PSNR of 1.9dB for *Foreman* and 5dB for *Claire* is observed when compared with the codec without the pixel scattering process and the actual gain on each frame is found to be dependent on the statistics of the frame. In the case of *Claire*, the bit errors in side information is highly localized and hence the effect of randomization is larger compared to *Foreman*. Note here that the rate-distortion results are not presented for varying bit rates due to inconsistency of the reference result at higher bit rates. When higher order bit planes are coded with a fixed parity rate, the residual errors in the decoded bit stream tends to become very significant and a catastrophic effect creeps in to the video quality and PSNR.

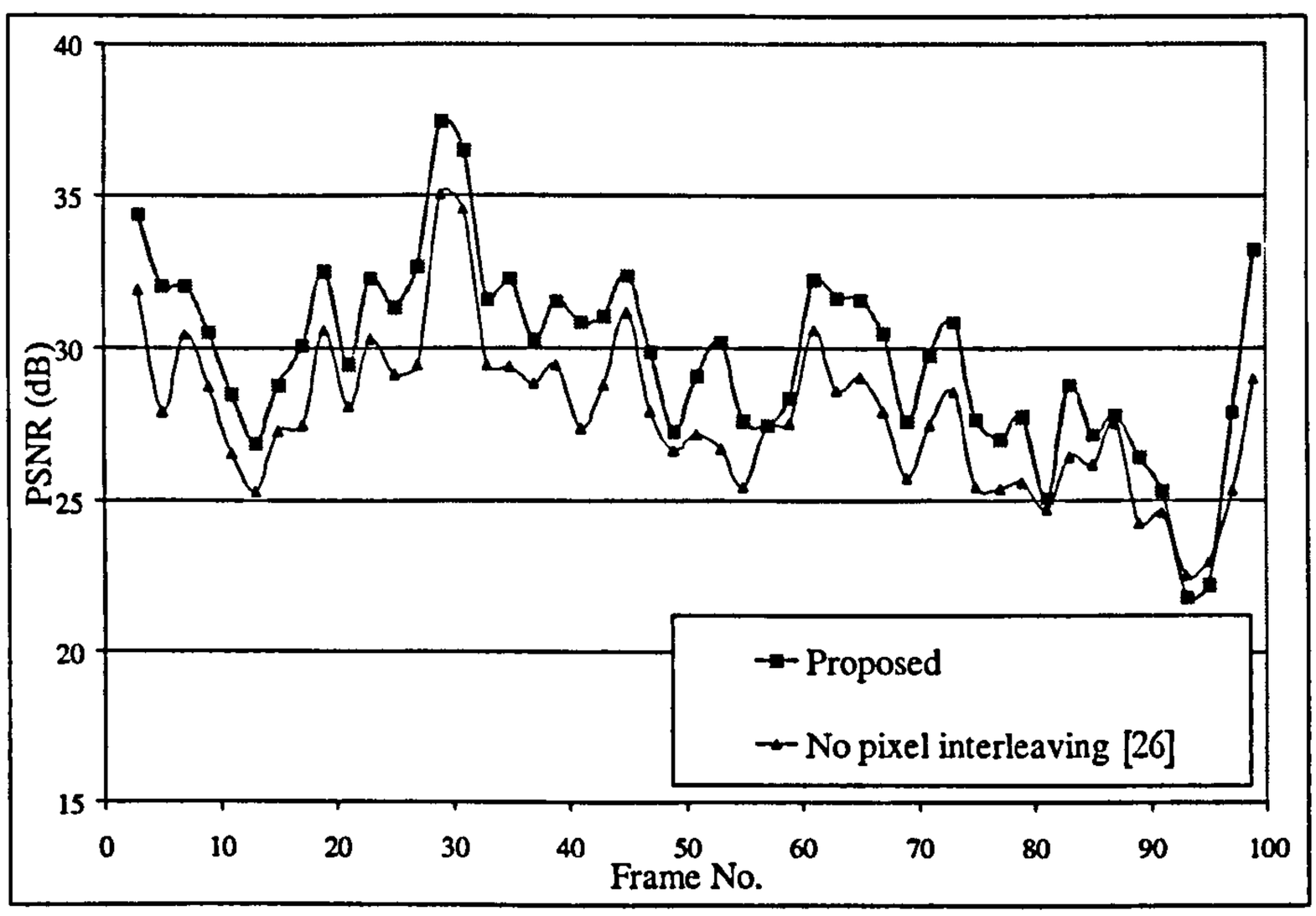



Figure 4-4 Performance comparison of the modified codec for the *Foreman* sequence

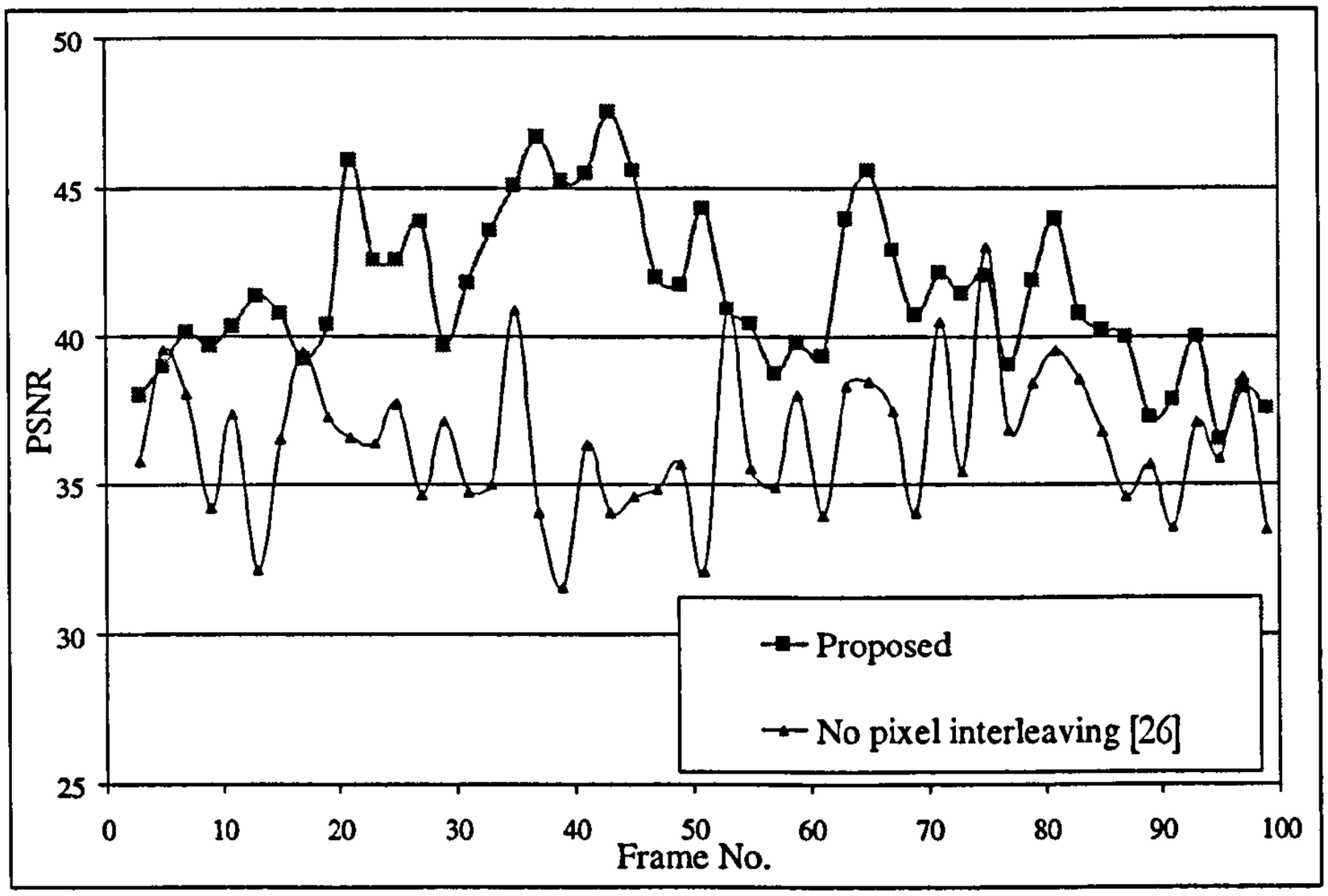



Figure 4-5 Performance comparison of the modified codec for the *Claire* sequence

Figure 4-6 illustrates the improvement achieved by the proposed modification to the DVC codec architecture for a frame selected from the *Foreman* sequence. The apparent disturbance observed in the frame in Figure 4-6 (a) is due to the high motion level concentrated in that locality, where the corresponding coding blocks suffer from insufficient parity bit rate. Figure 4-6 (b) shows significantly fewer disturbances since the effects of motion are now spread over the frame and hence the available parity bits are efficiently utilized in the decoding operation.

Thus, it is evident that the improved performance of the proposed codec is primarily due to the exploitation of local motion effects on the image very commonly evident in video surveillance applications. The areas of zero or low motion levels on the frame can generally be predicted to a very high accuracy in the side information generation process at the decoder. When the side information has less correlation noise compared to the original Wyner-Ziv frame, the decoder needs fewer parity bits for accurate decoding. Any excess amount of parity bits pertaining to these regions is wasted. On the other hand, the regions of high motions, as in the face of the speaker in the *Claire* sequence, are less accurately predicted in side information and hence the decoder inevitable needs large amounts of parity information. Thus the proposed pixel scattering process improves the efficiency of parity utilization.

(a) (b)

Figure 4-6 Comparison of a decoded frame from *Foreman* sequence:

(a) Not pixel interleaved

(b) Pixel interleaved (proposed)

### *4.2.2 Estimating the correlation noise at the encoder*

Even though the pixel scattering algorithm produces a significant performance gain when implemented over a fixed parity coding framework, an intelligent solution need to generate an estimate of the correlation noise of side information at the encoder. This will inevitably enhance the parity pattern allocation so that the overall codec performance can be improved. Note that the pixel scattering could still function as a useful tool in this scenario.

In ERC, however, generating an estimation of the correlation noise of the side information at the encoder itself formulates an important problem. Any algorithm must not be processor and memory intensive, in order to harness the benefits of DVC in targeted practical applications. A possible option is to generate a gross representation of side information using computationally inexpensive means and then use this to approximate the correlation noise. Block diagram of a possible algorithm for the encoder is shown in Figure 4-7.

In this algorithm, a pixel by pixel interpolation of the two adjacent key frames is performed to generate a predicted frame at the encoder. The correlation noise is estimated by considering this predicted frame and the original Wyner-Ziv frame information. The parity puncturing pattern is then computed based on this estimation.

The simulations carried out on this algorithm, for standard video sequences proved very stable behaviour. This ERC algorithm is utilised as the base setup for the enhanced *reconstruction* proposal presented in section 5.2 later in this thesis. Accordingly, the rate-distortion results for this algorithm are illustrated in section 5.2 (the reference result) for the given test conditions.

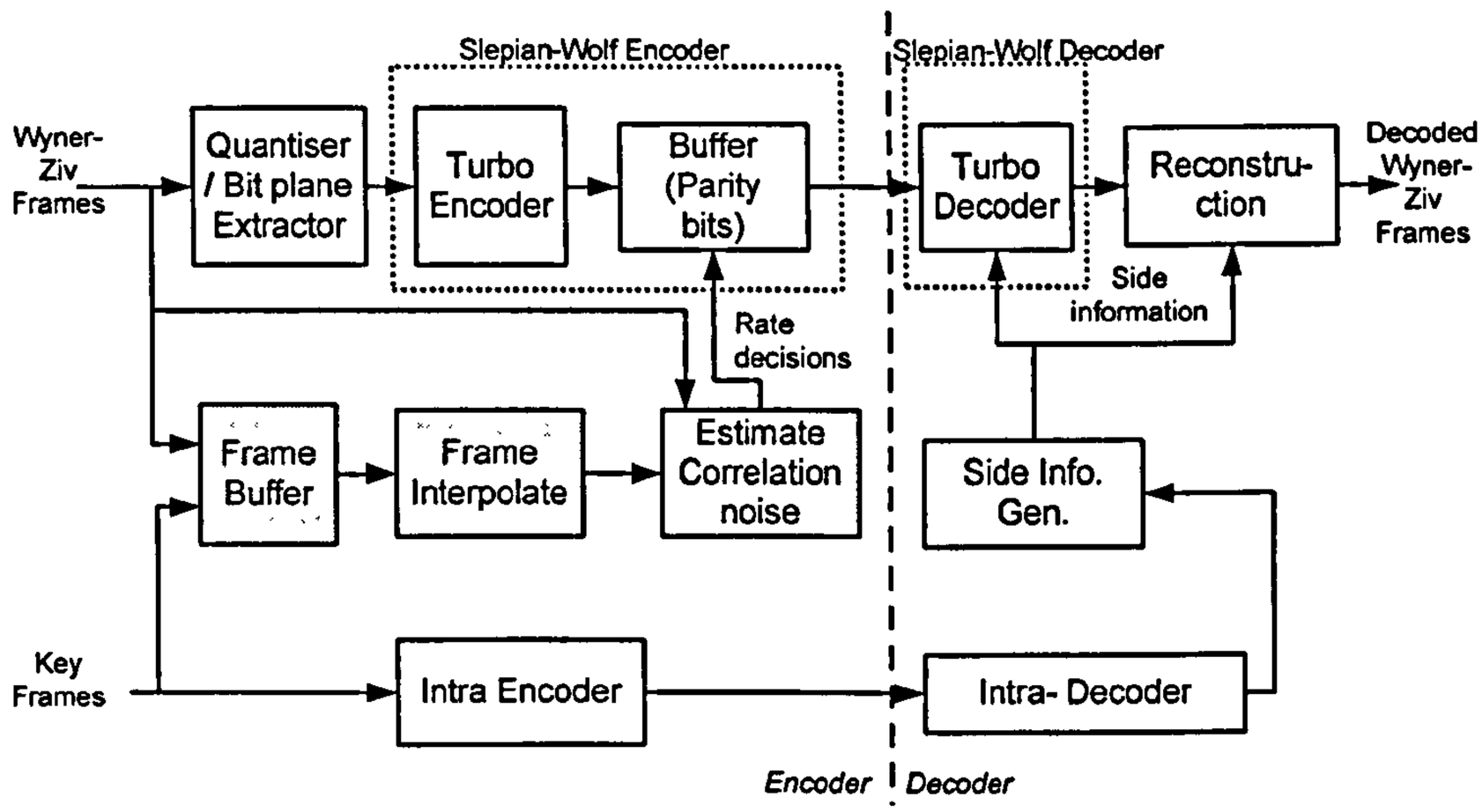



Figure 4-7 Correlation Noise Estimation at the Encoder

## 4.3 Iterative Refinement of Decoded Frames

It is noted that any form of ERC algorithm is effectively a compromise of the computational complexity of the encoder, which otherwise is extremely low in the considered DVC framework. Decoded Frame Refinement is an innovative approach proposed in this thesis to solve this problem, with no additional bit rate or complexity burden to the encoder. Accordingly, in this section, an implementation algorithm is presented that improves the decoded picture quality by iterative refinement in a unidirectional DVC system.

To further understand the motivation for this approach and establish the background that opened up the opportunity to realize the concept of decoded frame refinement, it is worth at this point to further analyze the type of video distortion occurring in unidirectional DVC schemes and the possible causes.

### *4.3.1 Video Distortion in Unidirectional DVC Systems*

The common problem in unidirectional DVC systems is the sub-optimality in estimating the correlation noise at the encoder due to the low complexity constraint imposed. Further, since this estimation is only a statistical process that does not precisely identify the positions of bit errors in the side information block, this sub-optimality easily creeps into the determination of the parity puncturing pattern. There are two possible results of this scenario:

1. Overestimate the parity rate, affecting the compression rate of the codec

2. Underestimate the parity rate, affecting the decoded picture quality due to the uncorrected errors

Let's consider the second scenario in details as this formulates the background for the refinement algorithm presented in this section.

When the parity bit rate is underestimated, the turbo decoding process is affected and the output tends to contain high BER. In the bit plane wise coding framework, the possibly uncorrelated residual bit error positions in each bit plane creates a catastrophic effect by corrupting a significant number of pixels in the reconstructed frame as illustrated in Figure 4-8 for a pixel domain coding set-up.

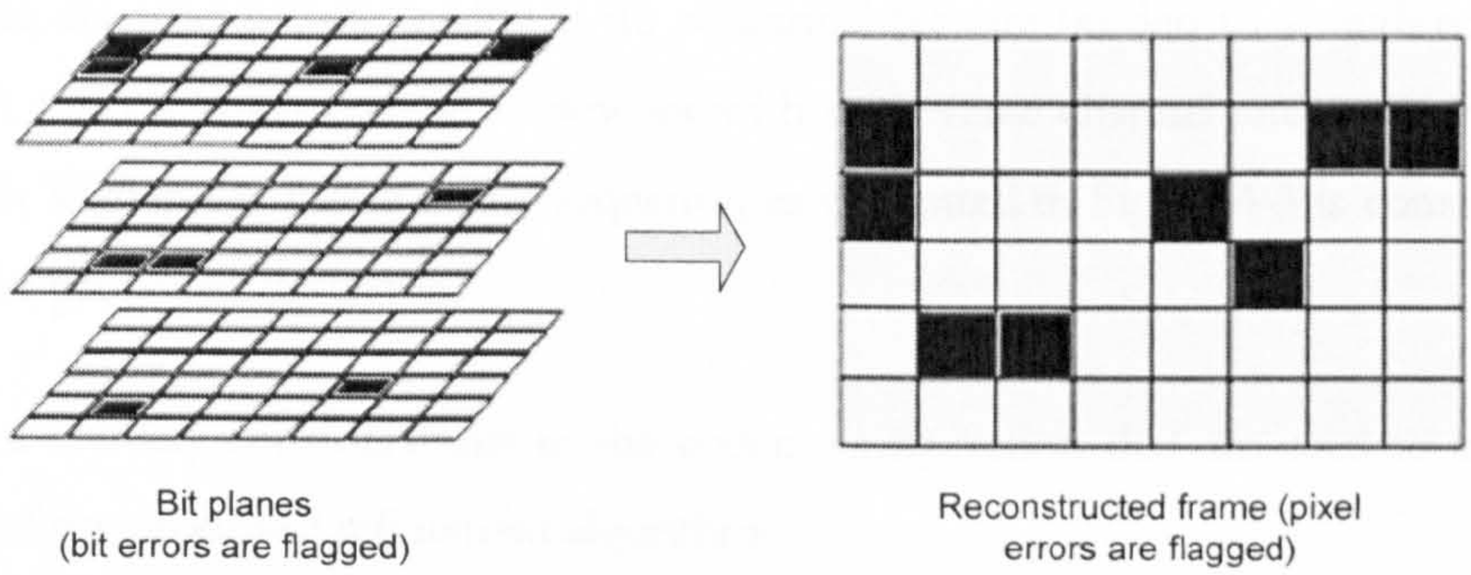



Figure 4-8 Accumulation of bit errors on multiple bit planes (not to scale)

A few frames corrupted by this scenario are shown in Figure 4-9.

Considering the characteristics of the above decoded noisy frames, it is understood that if the noisy bit positions could be located to a reasonable probability, spatial correlations in the frame

Figure 4-9 Frame degradation due to bit errors in decoded bit stream

could be very effectively utilized to improve the picture quality with no additional bit rate burden. A practical solution designed to address this problem is discussed in the next sub-section.

### *4.3.2 Iterative Refinement*

This section presents an iterative refinement algorithm incorporated to the DVC decoder to enhance the decoded picture quality at no additional bit rate burden to a unidirectional DVC solution. A pixel domain coding framework with no reverse channel and a pre-interleaver to scatter each Wyner-Ziv frame in the sequence, as presented in Figure 4-3 is considered as the basis for the proposed algorithm.

There are a few key considerations in the codec configuration that are used to influence the efficiency of the proposed refinement algorithm:

- Before the Slepian-Wolf encoding, the input bit stream is split into blocks of bits that correspond to rectangular macroblocks as illustrated in Figure 4-10. Thus, the block length in the Slepian-Wolf encoder is equal to the number of pixels falling into a macroblock. This length is selected to be significantly smaller than the number of pixels in the frame. This blocking-out helps in localization of errors and enables the flagging process elaborated later in this section. In this case, the block size is noted to

be a compromise of the turbo coding strength and the error localization efficiency. The rectangular macroblock arrangement gives flexibility in optimizing the block size.

- The pre-interleaver used to scatter the images before Slepian-Wolf encoding and then the de-interleaving after Slepian-Wolf decoding, in conjunction with the macroblock partitioning, significantly improves efficiency of the spatial and prediction process by scattering the corrupted pixels. This phenomenon is further elaborated later in this section.

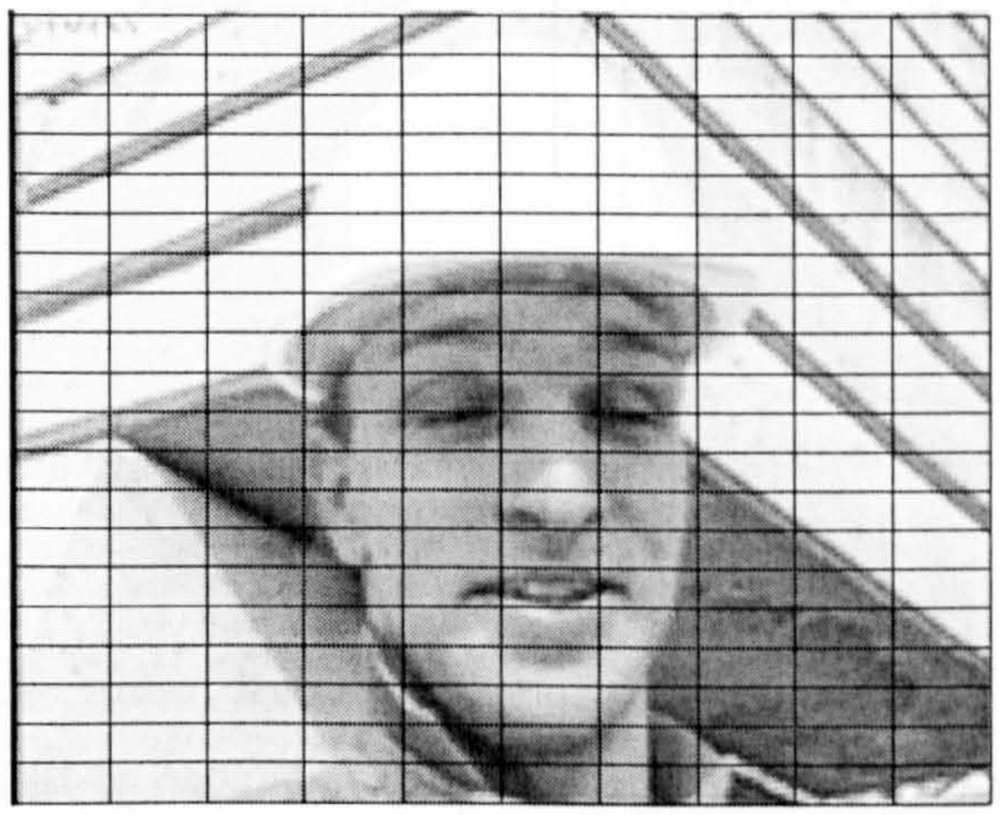


Figure 4-10 Macroblock partitions of the input frames

The codec involves an *initial* side information generation process which is used for the first iteration of the decoder, used with the punctured parity bit stream. The decoder generates the initial side information using the previously decoded Key-frames using a simple frame interpolation algorithm [4]. It is noted that the selection of the initial side information generation algorithm is independent of the proposed iterative decoding architecture and will have its own merits.

The full block diagram of the codec incorporating the refinement algorithm is illustrated in Figure 4-11.

The main structural enhancement proposed in this paper is that the Slepian-Wolf decoder is proposed to run a number of iterations for each bit plane of each frame in addition to the

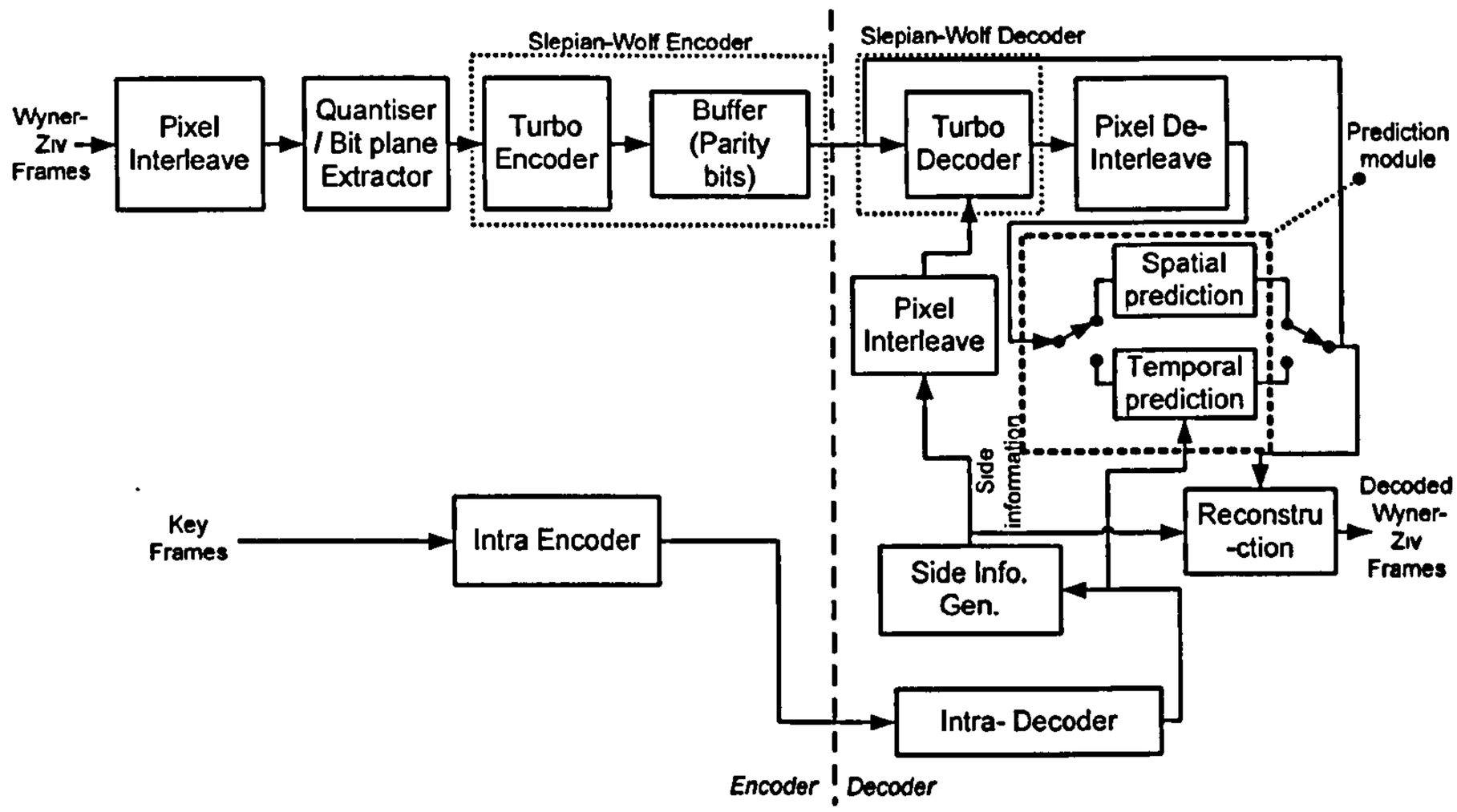



Figure 4-11 Unidirectional DVC codec with the refinement algorithm

internal iterations in the turbo decoder. During each Slepian-Wolf decoder iteration, the decoded output is refined by exploiting either spatial or temporal correlations in the current frame and the adjacent frames of the sequence and iteratively used as side information. The initial side information is used in the first iteration of each bit plane. The functional description of the proposed codec is given below.

#### *4.3.2.1 Encoder*

The Wyner-Ziv frame to be encoded is initially pixel scattered as discussed in section 4.2.1 earlier in this chapter. Then the pixel domain quantization and bit plane extraction is performed at macroblock level. The most significant bit plane is coded first. A Slepian-Wolf coding block is formed by each bit plane of a macroblock. The parity bit stream generated in the turbo encoder is punctured at predetermined fixed intervals for each block and transmitted to the decoder.

#### 4.3.2.2 *Decoding and Flagging*

The initial side information is used in the first iteration of the decoder loop for each block. Then error estimation is performed on the decoded bit stream block as explained earlier in section 2.5.3.4 of this thesis. If the estimate yields a BER exceeding a predetermined error threshold, the bits in the block are *flagged.* For the simulations discussed in this section, the error threshold is set to zero. When the first decoder iteration for all blocks in the first bit plane is completed the flag-map for the frame (current bit plane only) appears as shown in Figure 4-12. The black patches correspond to the macroblocks with BER higher than the threshold.

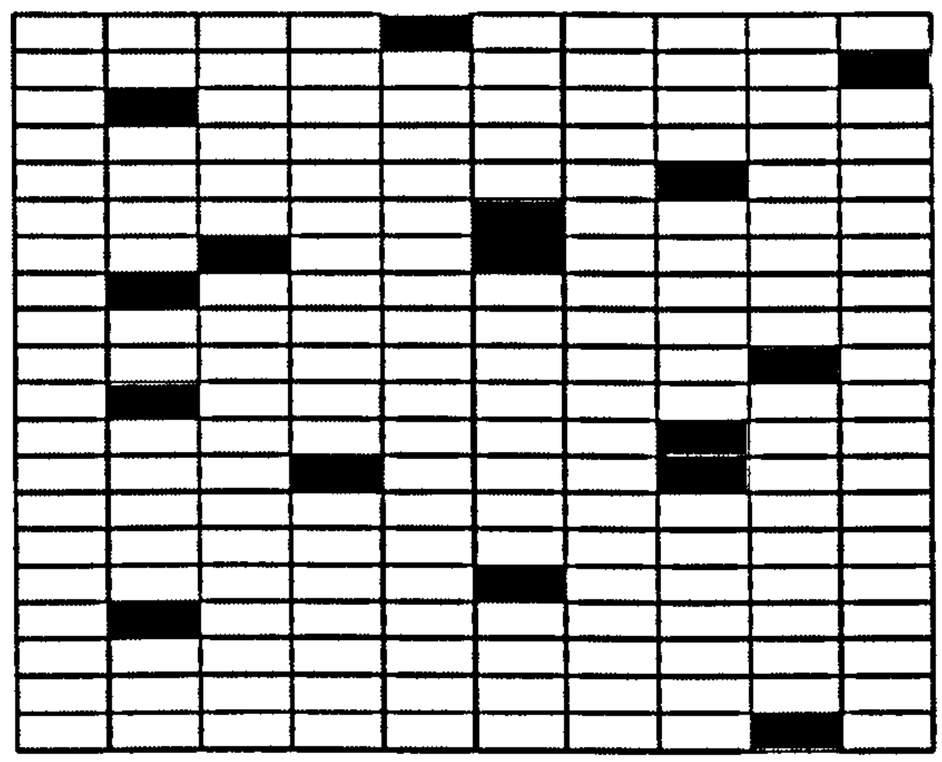


Figure 4-12 Flag-map for one bit plane

Then the frame is re-created by concatenating all bit planes decoded so far. In this process, any residual flags that were carried forward from the previously decoded bit planes are also marked. The re-created frame is then de-interleaved and the flagged-pixel-map for the reconstructed frame appears as shown in Figure 4-13. Note that the flagged are now scattered due to the de-interleaver.

This frame is then forwarded into the prediction module to obtain predictions for the pixel positions flagged on the current bit plane. The prediction module incorporates spatial and temporal prediction mechanisms as described below.

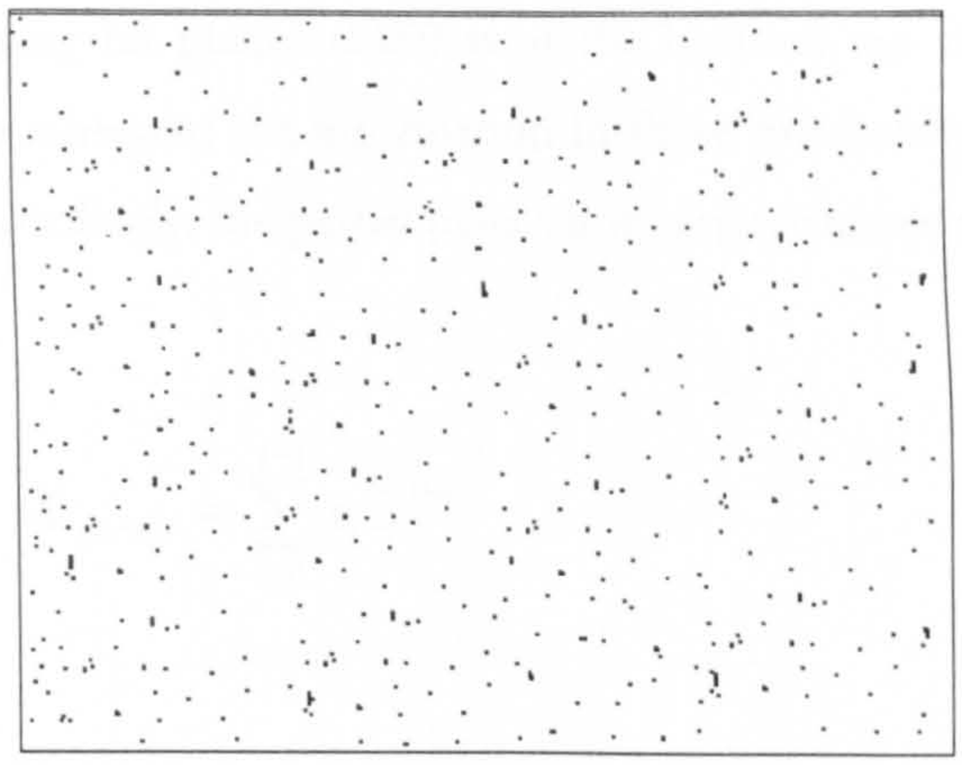


Figure 4-13 Flagged pixel map for the frame

#### *4.3.2.3 Spatial Prediction*

The spatial prediction is carried out for the flagged pixel positions using the spatial correlations of these with the adjacent pixels. Due to the pixel scattering operation performed within the frame, the flagged bits are now scattered and hence up to 8 adjacent pixels are available for spatially predicting the current bit. Figure 4-14 illustrates the corresponding pixel locations.

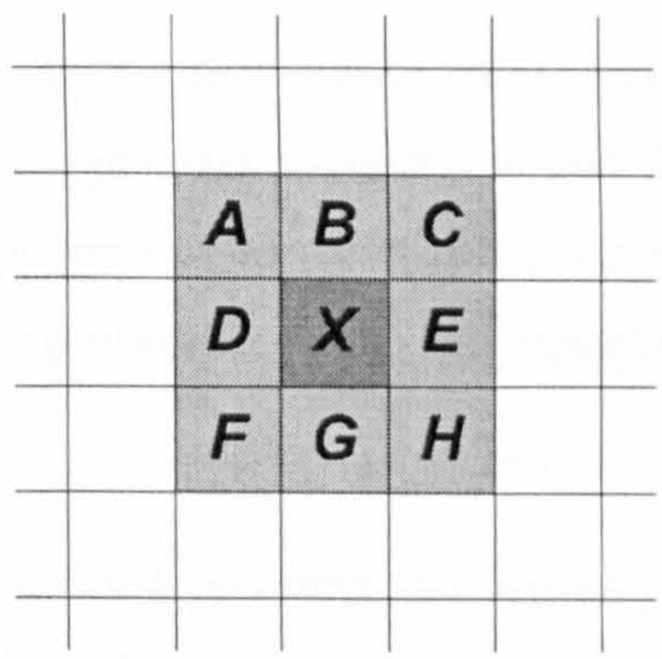



Figure 4-14 Adjacent pixels available for spatial prediction of flagged pixel "*X*"

Assume that $x_j$ is the current bit plane value of the pixel *X*, and it is flagged. The adjacent pixels to the pixel under consideration (*X*) are labeled (*A*..*H*) and the corresponding bits of the current processing bit plane of these pixels are labeled (*a*..*h*). Since the bit planes are decoded

sequentially, the underlying bit planes (MSB is at the bottom) are already decoded. Now a *partial pixel* is defined to represent the information in these previously decoded underlying bit planes of each pixel. Accordingly, the partial pixel values are calculated for the current pixel ($X$) as below:

$$X_{partial} = \sum_{i=0}^{j-1} x_i 2^{7-i} \qquad 4\text{-}1$$

where $j$ is the current bit plane number and $x_i$ denotes the $i^{th}$ bit plane value of pixel $X$. Similarly, the partial pixel values are calculated for the adjacent pixels ($A..H$). Now, the spatial prediction for the current bit $x_j$ is computed by averaging the current bit plane values of the adjacent pixels, of which the partial pixel value ($A_{partial}..H_{partial}$) is the same as that of the current pixel ($X_{partial}$). Pixel level averaging is observed to be a simple and very effective technique to fill a *hole* or a flagged bit position in an array of bits. Any flagged bits in the adjacent pixels are not considered in the prediction.

#### *4.3.2.4 Temporal Prediction*

A motion search algorithm utilizing the previous key frame is used for exploiting the temporal correlations in predicting the flagged bits described above. In this process luminance and chrominance (Y, U & V) components of the video frame are used by the search algorithm to determine the motion vectors as proposed in [75]. The temporal search process is illustrated in Figure 4-15.

A search block of size 8×8 is defined enclosing the flagged pixel to be filled. Then the best match for the block is searched on the reference key frame. The flagged pixels are excluded in calculating the mean absolute difference (MAD) in the block matching. Once the best match block is located on the reference frame, the flagged bit positions in the 8×8 block on the current Wyner-Ziv frame are filled with the corresponding information from the reference key frame.

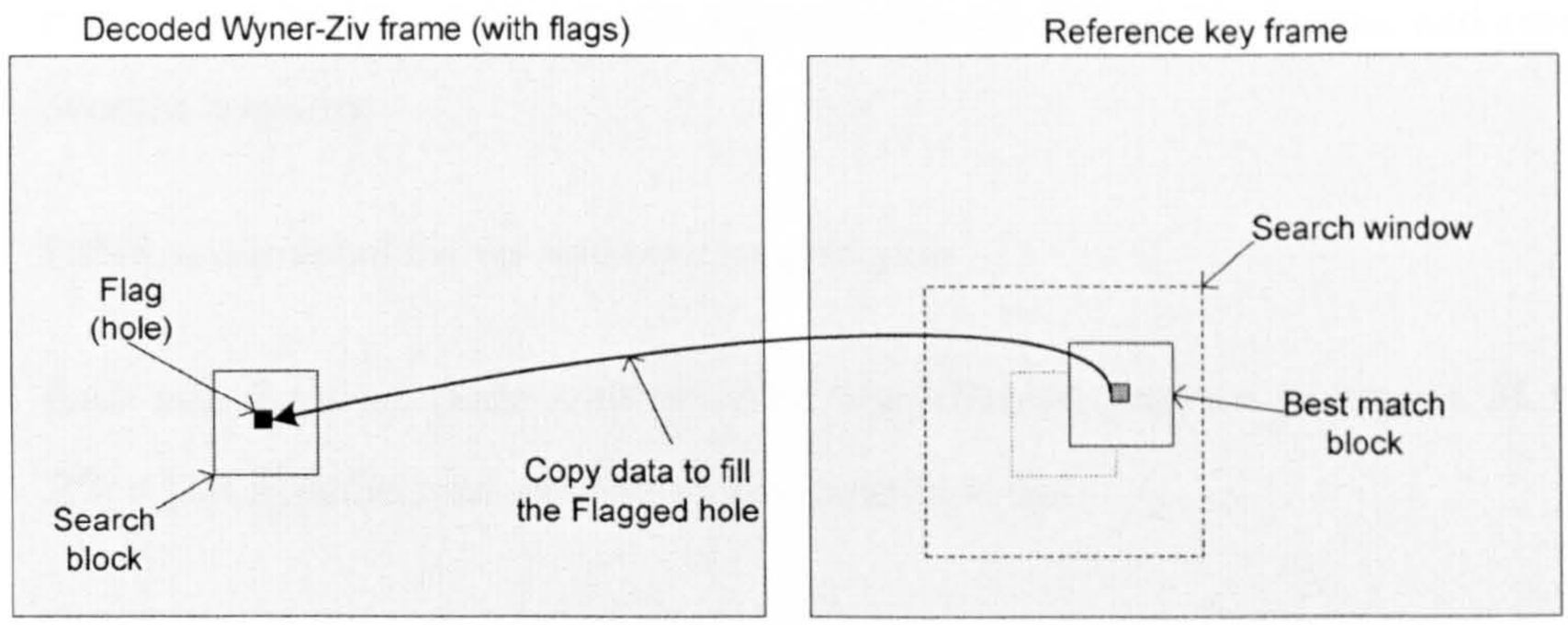



Figure 4-15 Temporal search for flag-filling

#### *4.3.2.5 Updating Flagged Pixels and Iterative Decoding*

During each iteration loop of the proposed decoder, either spatial or temporal correlations are exploited to fill the flagged bit positions. It was established experimentally that the alternate use of the spatial and temporal prediction processes during each iteration of the decoding process yields optimum results. This observation could be explained by the fact that the two techniques exploit a distinct set of redundancies in the bit stream and consecutively exploiting the same type of redundancies is less effective. Once the predictions for the flagged bits in the current bit plane are completed, the next decoder iteration runs, using the same parity bits and the flag-filled bit stream as the new updated side information. After the predetermined number of iterations for one biplane, the decoding process moves to the next bit plane. This process continues until all bit planes are iteratively decoded. Finally, the reconstruction function allows the suppression of some of the quantization noise and produces the final decoded video stream.

#### *4.3.2.6 Simulation Results*

The R-D performance of the unidirectional DVC codec with the refinement algorithm discussed above is tested for standard video sequences using the test conditions listed below.

**Test Conditions**

- Video sequences: *Foreman* (medium/high motion), *Salesman* (low motion); QCIF (176×144); YUV 4:2:0; 100 frames; Wyner-Ziv frame rate: 15fps; GOP size:2

- Bit rate and PSNR calculated and displayed only for Wyner-Ziv frames, and averaged over the sequence.

- PSNR is calculated for the luminance signals only.

- Each rate-distortion point is achieved for pixel domain quantiser parameter: $M$, where $2^M = \{2,4,8\}$ defines the number of quantization levels.

The performance of the proposed DVC codec is compared with H.264-Intra frame coding (profile: main, v10.1) and MPEG-2 Intra-frame coding scheme (TM5 bit rate control algorithm) [76]. The consideration of intra-frame coding schemes for comparison is primarily based on complexity aspects of the encoder. It is understood that the encoders of inter-frame coding schemes of conventional video coding approaches are computationally very expensive and hence could be an over-burden for a low cost video camera, which is the target application domain in this discussion. The considered intra-frame encoders have moderate complexities, which fact justifies their use as references. It is noted that the computational complexity of the DVC encoder is still significantly low compared to these intra-encoders since the earlier does not involve any iterative tasks. The turbo encoder contained within the DVC encoder is not more than a few shift registers making the simple state machine, and an interleaver. The remainder is simply the bit stream preparation from the input frame with the added pixel interleaver. Therefore a significant gain in terms of cost reduction is promised in building the DVC encoder into low end video sensors in addition to the attractive R-D performance.

**Analysis of Results**

Figure 4-16 and Figure 4-17 illustrate the performance comparison of the proposed codec with MPEG-2-Intra coding and H.264-Intra coding for the *Foreman* and *Salesman* sequences respectively.

The results are shown for the bit rates up to 600 Kbps since any higher would be practically not sensible for the low resolution QCIF frame format considered. A significant gain of up to 14dB is shown by the proposed codec. The difference in the relative gain evident for the two sequences could be explained considering the properties of the scene. For the *Foreman* sequence, the performance is superior to the reference schemes by a significant margin at low bit rates, but becomes in par with the competitors, with the increase of bit rate. The main

reason for this observation is that the performance of the proposed DVC codec is partially upper bound by the initial side information stream during the reconstruction process. Further enhancement to the initial side information generation algorithm will contribute to further enhancement of the performance margin. The frames of the *Salesman* sequence, in contrast, have a still, yet cluttered background which yields a very good side information stream for DVC, while not being a favourable condition for any intra-frame coding scheme. Since the side information is of very high objective quality, the output the reconstruction function shows a very high PSNR (approx 43dB in this case) even at very low bit rates, and consequently the incremental gain from additional bit rate is small. However, still background is a very common practical scenario in video recording, particularly in video surveillance applications.

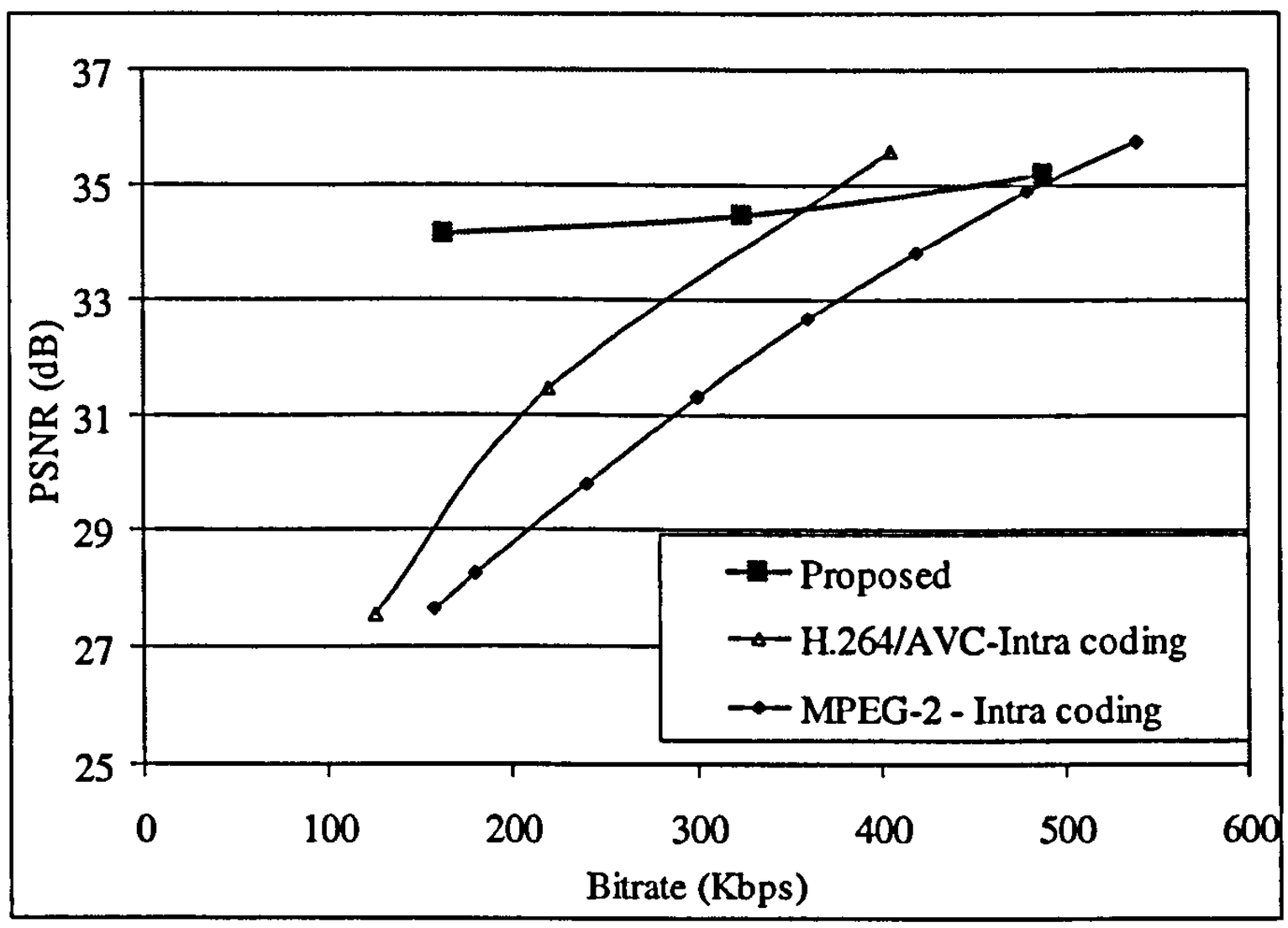



Figure 4-16 Rate distortion analysis for the *Foreman* sequence

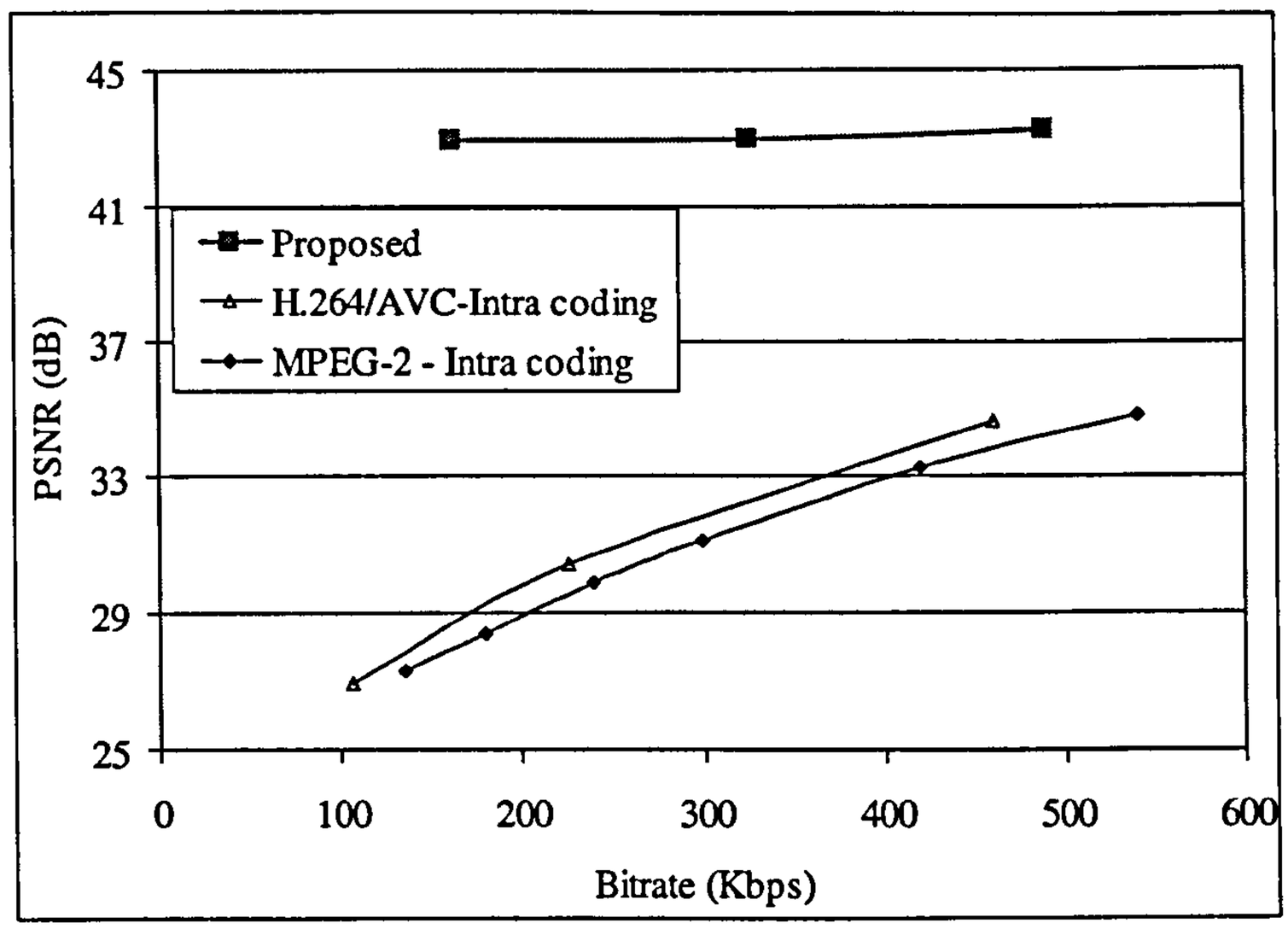



Figure 4-17 Rate distortion analysis for the *Salesman* sequence

Figure 4-18 and Figure 4-19 illustrate the effect of iterative refinement proposed in this paper, in terms of the decoded bit error rate.

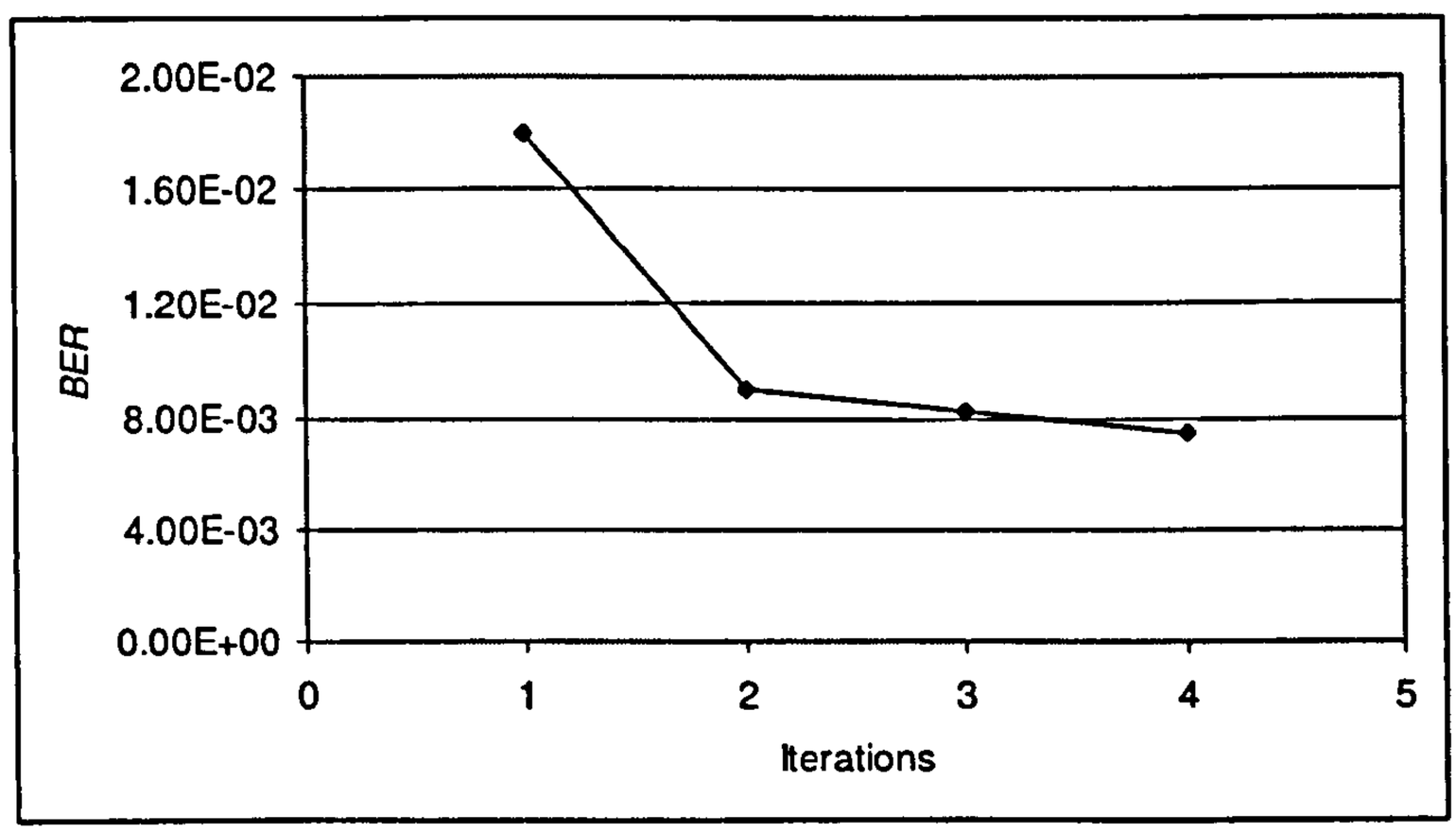



Figure 4-18 Effect of iterative refinement (*Foreman* sequence)

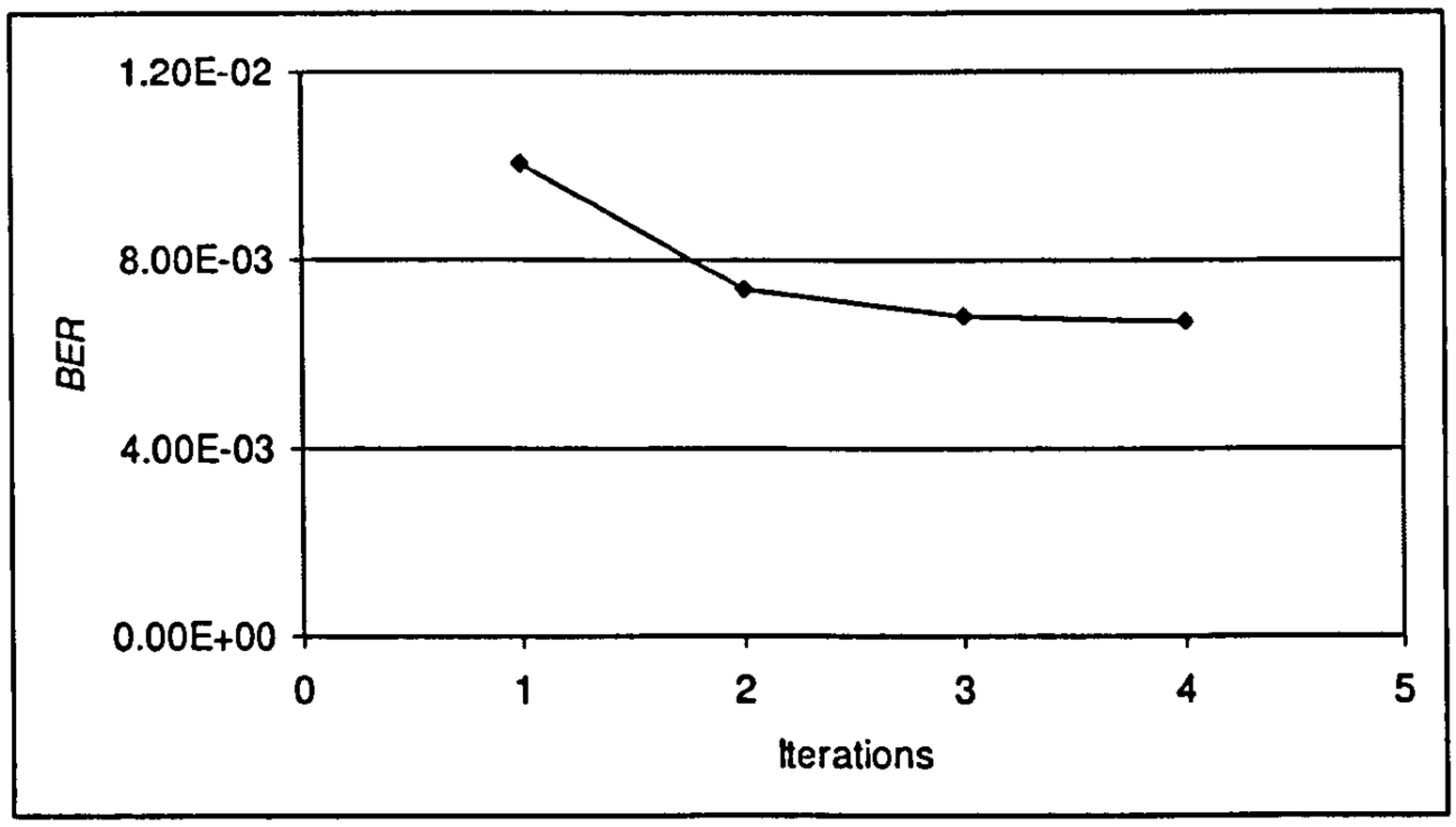



Figure 4-19 Effect of iterative refinement (*Salesman* sequence)

It is evident that the improvement starts in dramatic steps and gradually saturates over a few iterations influenced by the closeness of the reference frame to the original Wyner-Ziv frame in the case of temporal prediction and the amount of spatial correlations in spatial prediction. The effect of the block length in the encoder is illustrated in Figure 4-20. The output PSNR at a bit rate of 163 Kbps is plotted against the encoder block length. It is evident that the overall codec performance is superior at low block lengths. This shows that the contribution from the flagging and prediction operation, which induces a low block length, is dominant over the effect of turbo coding performance, which promotes a longer block length.

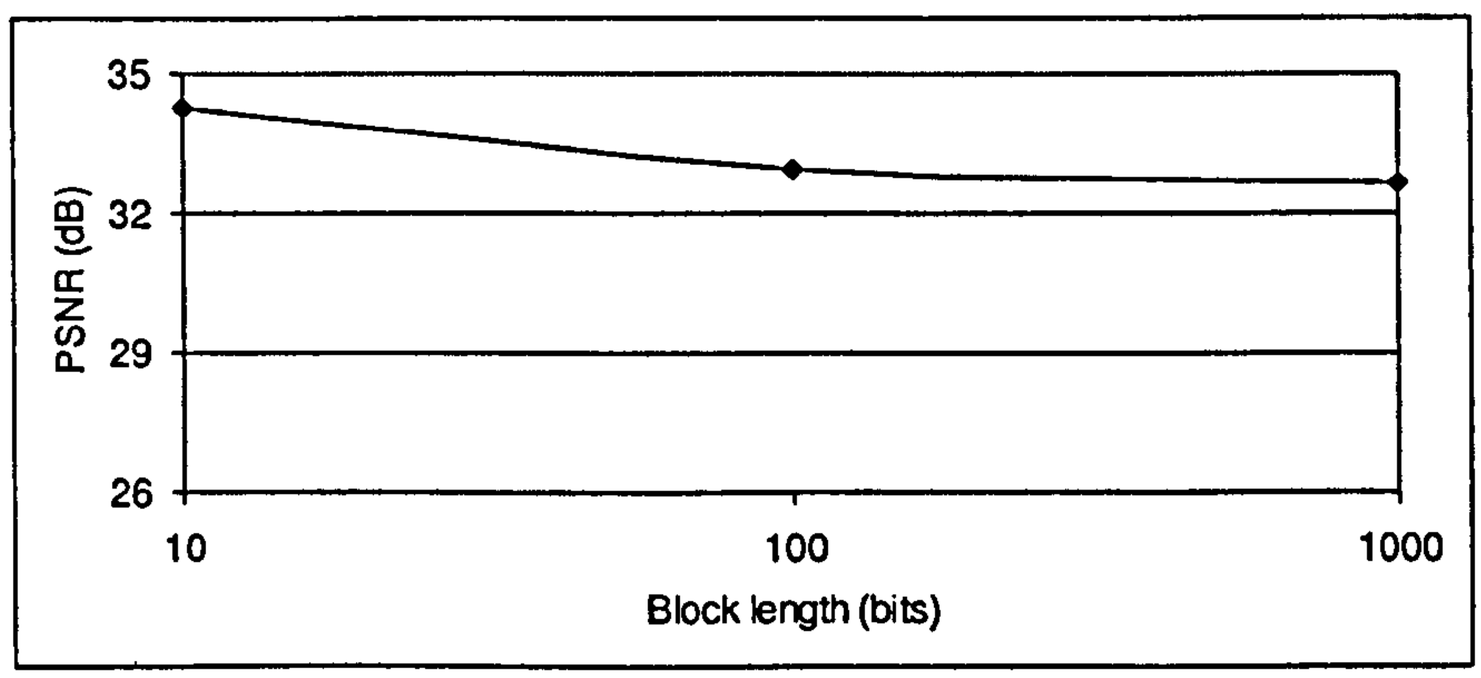



Figure 4-20 Effect of the block length selection at a bit rate of 163 Kbps (*Foreman* sequence)

From the codec R-D performance analysis presented in this section, it is clear that the use of unidirectional DVC with the refinement algorithm discussed above will enhance the picture quality of the decoded video sequence compared to the MPEG-2-Intra and H.264-Intra coding based video cameras. This is in addition to the significantly lower complexity of the DVC encoder which helps in reducing the cost of the video camera fabrication.

### Computational Complexity Discussion

It is understood that modifications proposed to the DVC codec in this paper adds considerable computational complexity to the decoder even though the widely appreciated extremely low encoder complexity remains unaffected. However, this additional cost could be justified by the fact that the DVC decoders would be heavily shared in the many-to-one type application domains discussed here. Potential beneficiaries of the low complex encoder structure include the low cost surveillance cameras and the increasingly popular low cost handheld video cameras including disposable video cameras and a number of other similar applications. The major computationally expensive tasks, including exploiting source redundancies, are taken over by the DVC decoder. The additional cost of the decoder is justified by the typically evident very high sharing factors particularly noted in the use in surveillance systems and disposable video cameras. In surveillance systems, the decoding and viewing of the data streams, which are captured and stored in the field, takes place only in limited locations. With disposable cameras, the decoding takes place in the shared decoding facilities generally located in sales outlets on the high street. They are in limited numbers compared to the millions of predicted camera use, thus the decoder cost is believed to be justified.

## 4.4 Conclusion

The reverse channel incorporated in the common Stanford DVC framework is identified as a practical burden in the targeted application scenarios, mainly due to the associated processing delay. However, this feedback loop plays a critical role in determining the optimum bit rate for the Wyner-Ziv bit stream by facilitating the parity puncturing process. In this chapter, a number of solutions were proposed to address the problems associated with suppressing the reverse channel. Encoder rate control tools are presented to facilitate the bit rate allocation at the encoder and an iterative refinement algorithm is designed to improve the quality of decoded pictures that are degraded due to sub-optimal parity allocation. In the latter, spatial and temporal prediction techniques are deployed to iteratively exploit the correlations in the decoded information.

The performances of the presented algorithms were tested for standard sequences. The simulation results were compared for both objective and subjective video quality and compression efficiency with alternate approaches with no reverse channels. Further technical analyses were provided to illustrate the effectiveness of the individual algorithms.

The next chapter looks into a number of modular elements of the DVC codec in order to improve the coding performance, focusing on both feedback-based and unidirectional frameworks. This includes a non-linear quantiser, enhanced reconstruction algorithm and a side information refinement algorithm.

*Chapter 5*

# DVC PERFORMANCE IMPROVEMENT AT MODULAR LEVEL

As discussed earlier in the Literature Review, the DVC codec based on the Stanford framework has evolved over the past six years to some degree of maturity. Practical solutions have been designed to a number of hypothetical models and assumptions in the initial proposals, including, among others, correlation noise modelling [15] and request stopping criteria [19] in the feedback-based DVC systems. Enhanced algorithms have been designed for the key components in the coding framework, dominated by the extensive work on side information estimation (e.g. [28][55]). However, it is noted that the rate-distortion performance of the DVC codec is still lagging behind the state of the art in video coding, namely H.264/AVC, although the low encoder complexity feature in DVC is more favourable in the considered application domains.

This chapter focuses on a number of enhancements to the codec considering various modular elements, including the quantisation, reconstruction, as well as the side information generation. The motivation for each of the proposed technical solution and the implementation algorithms are presented in the following sections, followed by the results of the simulation testing in comparison to the state of the art in DVC codecs.

## 5.1 Non-Linear Quantisation

The signal quantization plays a significant role in the DVC codec, as in any other form of signal coding and communication. In the considered pixel domain coding framework, quantization defines the number of discrete levels taken by the pixel and the associated step size between each discrete level. Historically, a linear quantiser has been utilized in the DVC codecs. On this backdrop, this section presents a novel non-linear quantization algorithm for DVC in view of improving the compression efficiency of the codec. The background for this proposal is initially discussed, followed by the implementation strategy and the simulation results.

### *5.1.1 The pdf of Residual frames*

In developing the hypothesis for the non-linear quantization concept, a Wyner-Ziv residual frame is introduced. The residual frame is computed in the considered system by considering the current Wyner-Ziv frame and the adjacent key frames. The simplest form is to evaluate the frame difference with the previous key frame at pixel level as shown in Figure 5-1.

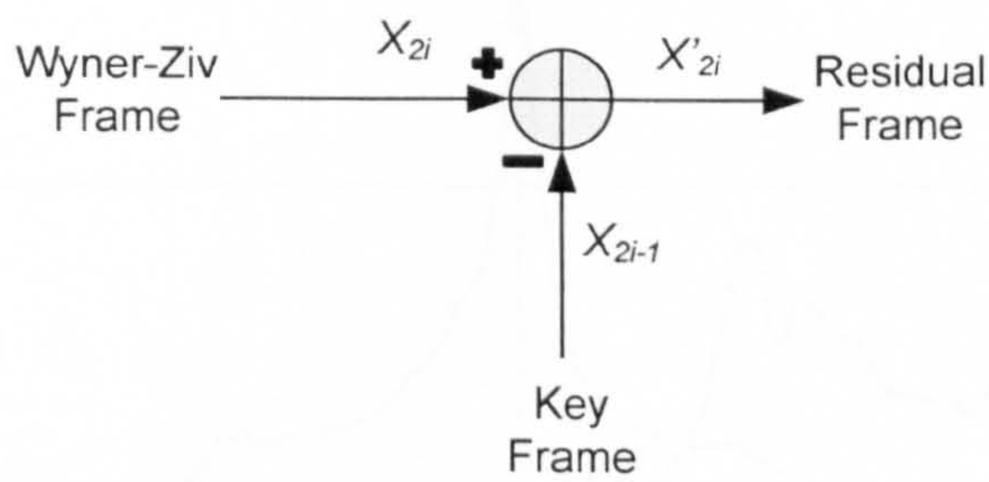



Figure 5-1 The Wyner-Ziv residual frame formation

A sample residual frame is shown in Figure 5-2.

Figure 5-2 Residual frame

To build the hypothesis for the proposed concept of non linear quantisation, the residual frame generated using the Wyner-Ziv frame and the adjacent reference frames (key frames) is obtained and the probability density function (pdf) of the statistical distribution of the pixels in

the original and residual Wyner-Ziv frames are plotted as is illustrated in Figure 5-3. The key frames are considered to be lossy-coded using H.264/AVC intra coding.

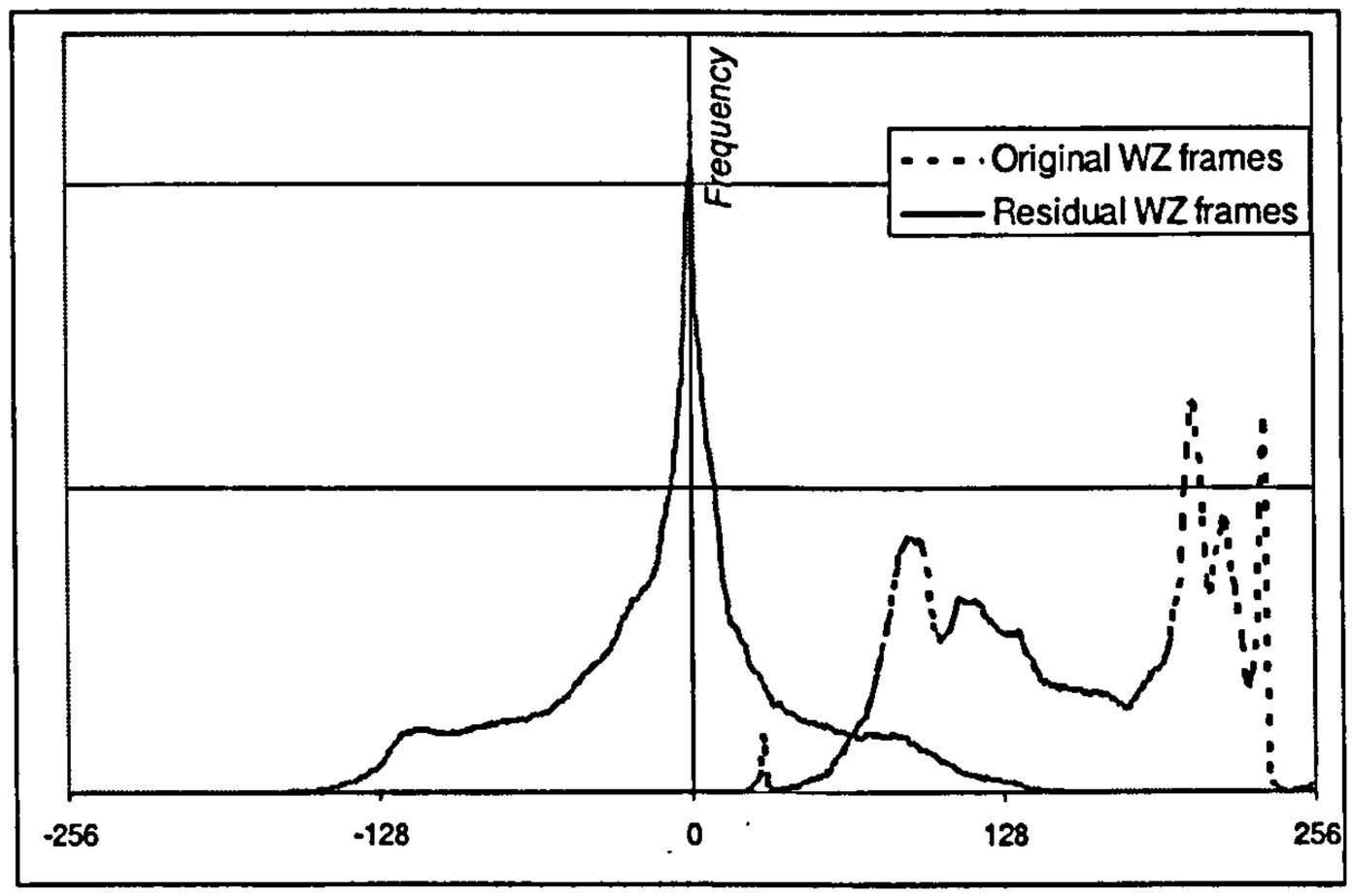



Figure 5-3 The pdf comparison of original and residual Wyner-Ziv frames

In Figure 5-3, the dotted line depicts the pdf of the pixel distribution of the original Wyner-Ziv frame sequence, and the solid line represents the same for the residual frame sequence, considering a GOP size of 2. It is noted that when the residual frame sequence is considered, pixels demonstrate a pdf with a contrastingly high density in the vicinity of near-zero mean, a notable deviation from the original pdf, which depicts a random distribution. Vast majority of the elements are concentrated around this peak distribution point. The motivation for the non-linear quantization approach discussed in this paper is to exploit this probability distribution so that small values, which make a major portion of the input residual signal, are represented with a higher precision so that a better picture quality and in effect improved compression efficiency can be achieved.

This scenario could be further explained considering the total quantization noise for the residual signal in the cases of linear and non-linear quantisation. The use of linear scaling in determining the quantization step size would introduce a large quantization noise and hence a loss of information due to the large amount of elements falling into the zero-bin as illustrated in Figure 5-4.

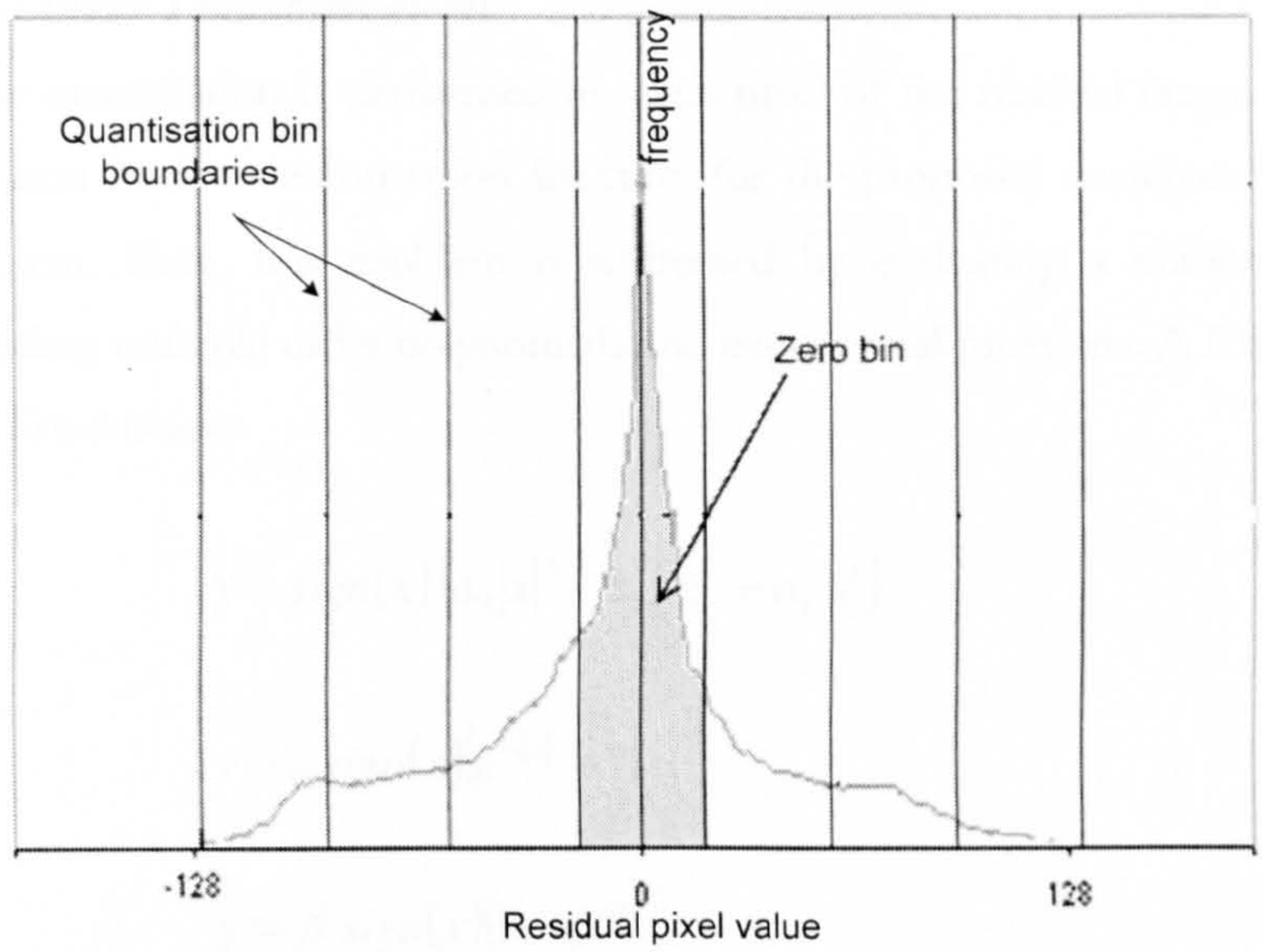



Figure 5-4 Linear Quantization bin formation (8 levels)

Alternately, when the proposed non-linear quantization is utilized, smaller step sizes defined near the zero, so that total quantization noise is significantly reduced. This scenario is illustrated in Figure 5-5.

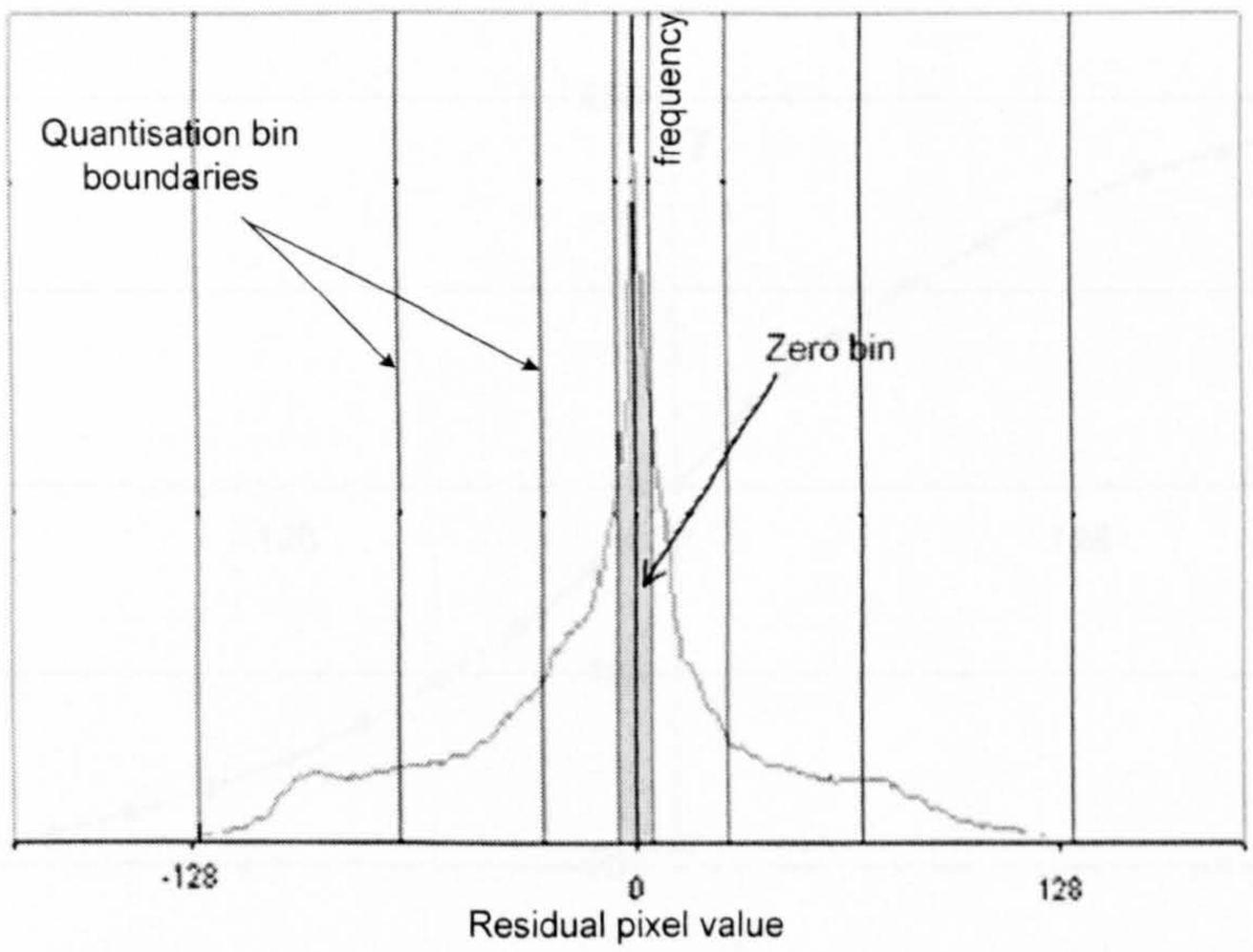



Figure 5-5 Non-Linear Quantization bin formation (8 levels)

### 5.1.2 *Non-Linear Transformation*

The non-linear quantization is performed on each pixel of the residual image. The design of the optimum non linear transformation function for the proposed quantizer is an important research problem. Here, this problem is addressed by evaluating a number of candidate formulas including multiple order polynomials and exponential functions. A few candidates for the above are listed below:

$$y = sign(x)\left(a_3|x|^3 + a_2|x|^2 + a_1|x|\right) \qquad 5\text{-}1$$

$$y = a\, sign(x)\left(e^{b.|x|} - 1\right) \qquad 5\text{-}2$$

$$y = \beta\, sign(x)\left(1 - e^{-\alpha|x|}\right) \qquad 5\text{-}3$$

where, x and y represent the input and output of the non-linear transformation respectively, and $a_1$, $a_2$, $a_3$, $a$, $b$, $\alpha$ and $\beta$ are constants that represent the level of non-linearity. Experimental evaluations have proved equation 5-3 to be the best out of the considered candidates for this purpose, which produced the highest rate-distortion performance in the codec. A graphical illustration of the transformation function depicted in equation 5-3 is shown in Figure 5-6.

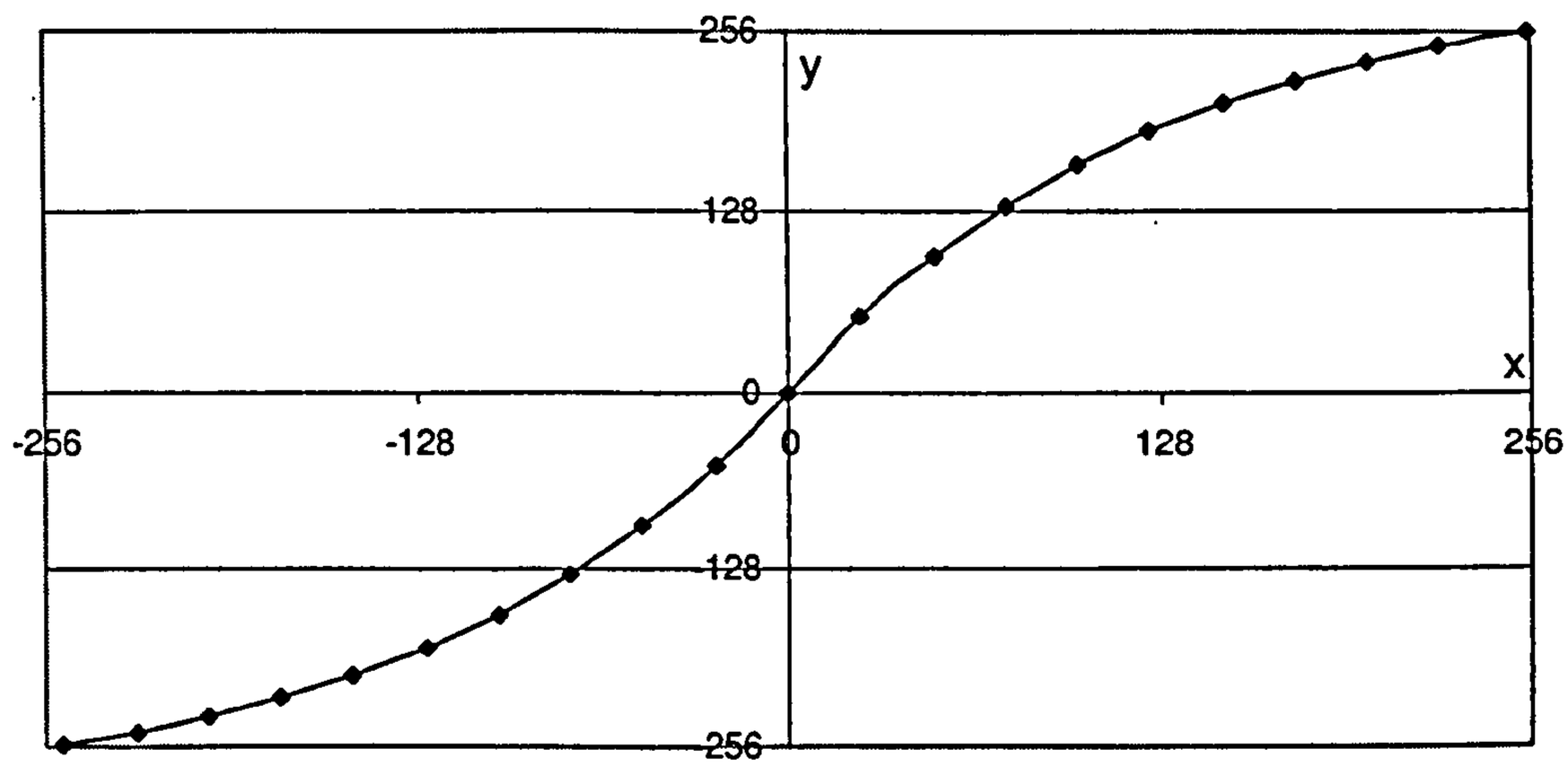



Figure 5-6 The non-linear transformation function

The Reconstruction function is also modified to incorporate the inverse-non-linear-transformation as defined by:

$$\bar{x} = -\frac{1}{\alpha} sign(y) \ln\left(1 - \frac{|y|}{\beta}\right) \qquad 5\text{-}4$$

The proposed quantization algorithm is incorporated to the pixel domain implementation of the existing DVC framework described earlier in section 2.5.3. The modified codec architecture is shown in Figure 5-7.

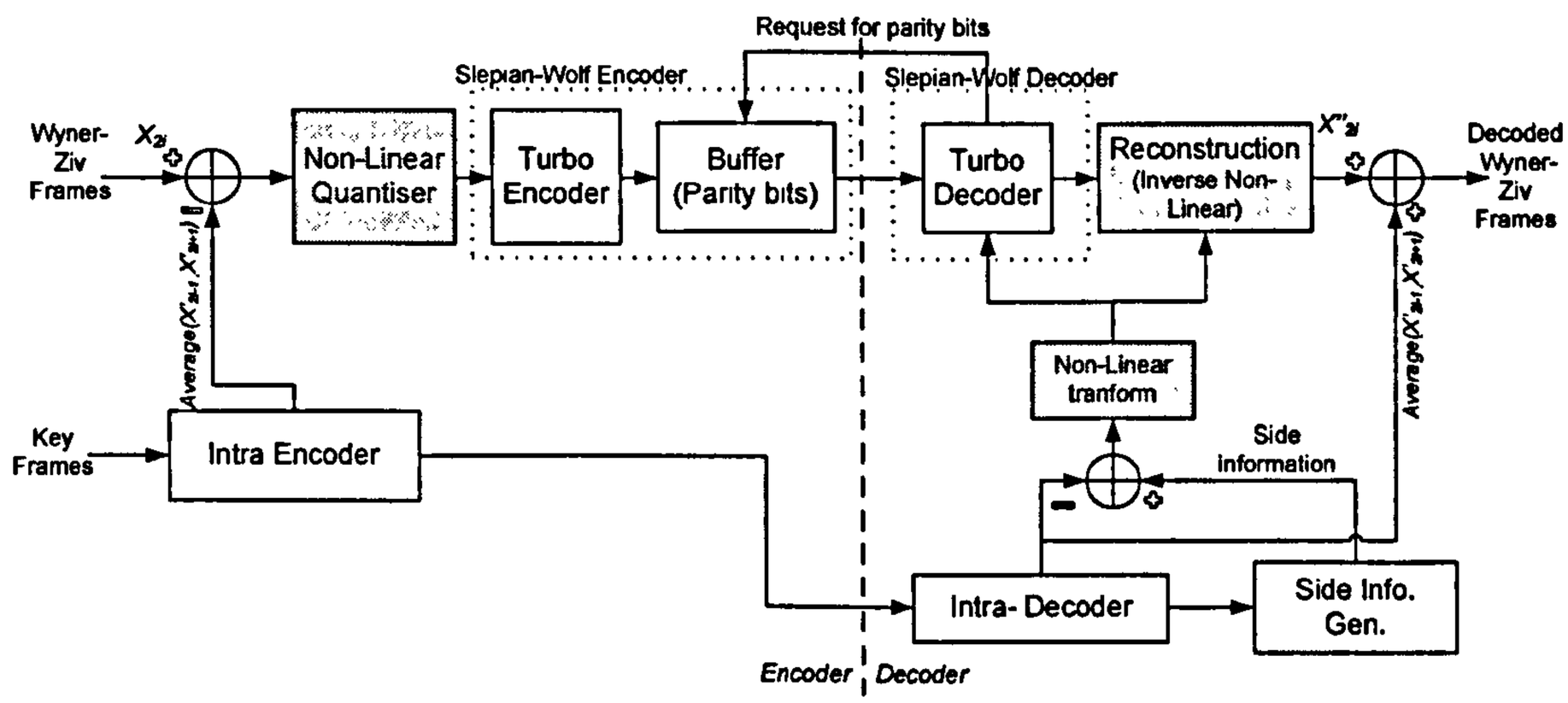



Figure 5-7 DVC codec with Non-linear quantiser

First, residual image computation is performed to derive the residual frames both at the encoder and at the decoder. At the encoder, the difference between the current Wyner-Ziv frame ($X_{2i}$) and the average of adjacent key frames ($X_{2i-1}$ and $X_{2i+1}$) is taken. Considering multiple reference frames in this case helps reducing the propagation of decoding errors in the key frames. Then, the non-linear quantization is performed for each pixel and then bit-plane-extracted for turbo encoding. At the decoder, side information is obtained by taking the difference of the motion interpolated frame ($Y_{2i}$) and the average of two intra coded key frames ($X_{2i-1}$ and $X_{2i+1}$). Similar to the process performed at the encoder, each pixel is non-linear-quantized. This bit stream is fed into turbo decoder and the reconstruction block. After the decoding and reconstruction, inverse non-linear transform is performed. Finally, the decoded

residual signal is added to the average of the adjacent key frames to reconstruct decoded frame $X'_{2r}$.

### *5.1.3 Simulation Results*

The rate-distortion (RD) performance of the proposed non-linear quantization in pixel domain DVC is tested for standard video sequences using the test conditions listed below.

**Test Conditions**

- Video sequences: *Foreman* (149 frames), *Coastguard* (149 frames), *Hall Monitor* (163 frames); QCIF (176×144); YUV 4:2:0; Frame rate: 15fps; GOP size:2
- Bit rate and PSNR are calculated and displayed considering both Wyner-Ziv frames and key frames; averaged over the sequence.
- PSNR is calculated for the luminance signals only.
- The rate-distortion points are achieved by varying the pixel domain quantiser parameter: $M$, where $2^M = \{2,4,8\}$ defines the number of quantization levels.
- Key frames are H.264/AVC intra coded. The quantization parameters for the H.264/AVC intra frame codec were selected so that the average PSNR of the decoded key frames and Wyner-Ziv frames is equivalent.
- The Laplacian parameter for the side information was optimized for each sequence.
- The $\alpha$ and $\beta$ parameters for the non-linear transformation function were set to 0.004 and 398 respectively based on experimental optimizations. The parameters were optimised by varying $\alpha$ in steps of 0.0005. It is noted that the $\beta$ parameter is dependant on $\alpha$ when the dynamic range of the residual pixel values is constrained within the range from -255 to 255 after the non linear transformation.

The above sequences are selected to represent different motion levels. The performance of the proposed algorithm is compared with the pixel domain implementation of the existing DVC codec which uses a linear quantiser. Figure 5-8, Figure 5-9 and Figure 5-10 illustrate the

performance comparison of the proposed codec for the *Foreman, Coastguard* and *Hall Monitor* sequences respectively.

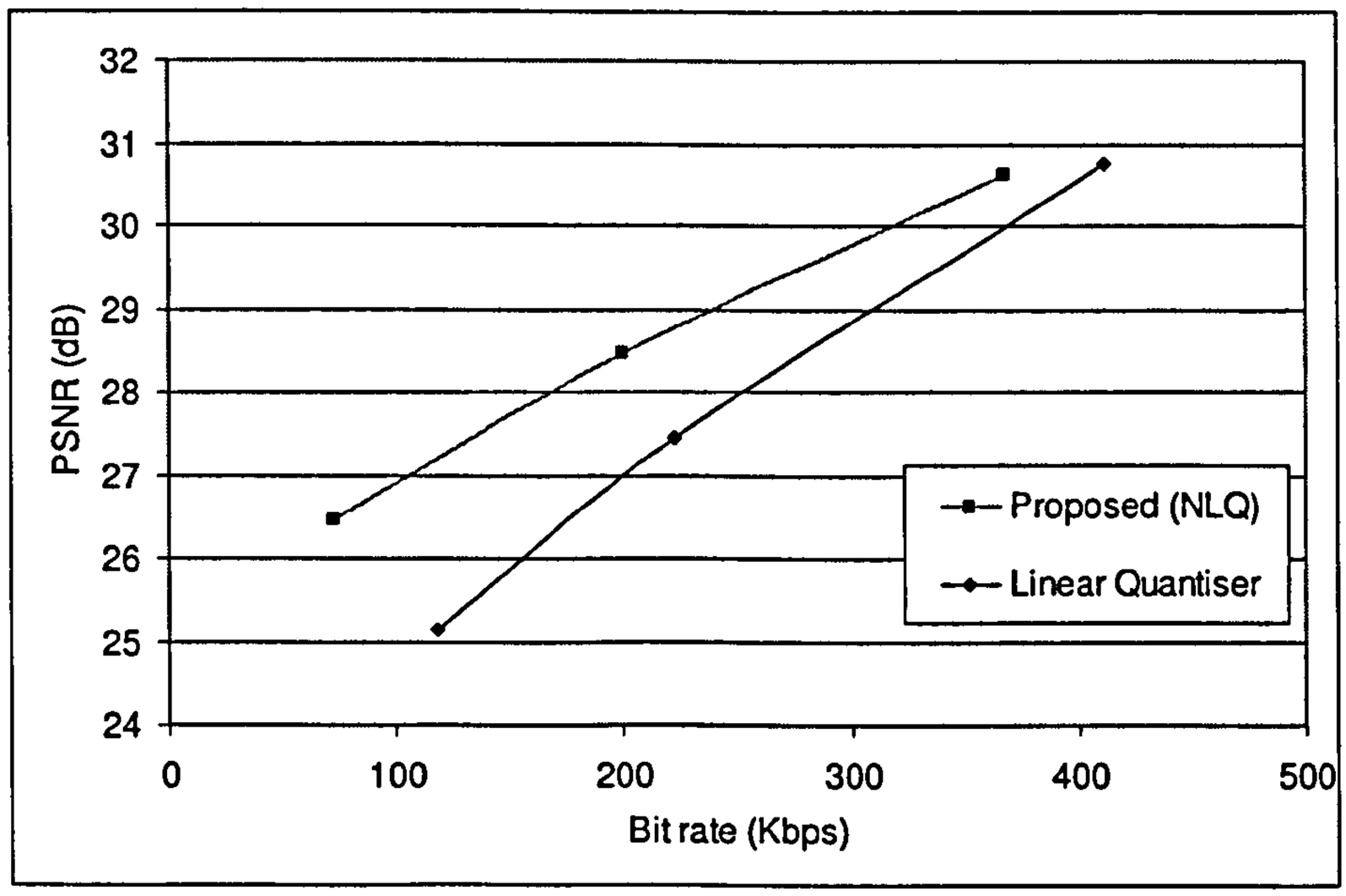



Figure 5-8 Performance comparison on non-linear quantiser (*Foreman* sequence)

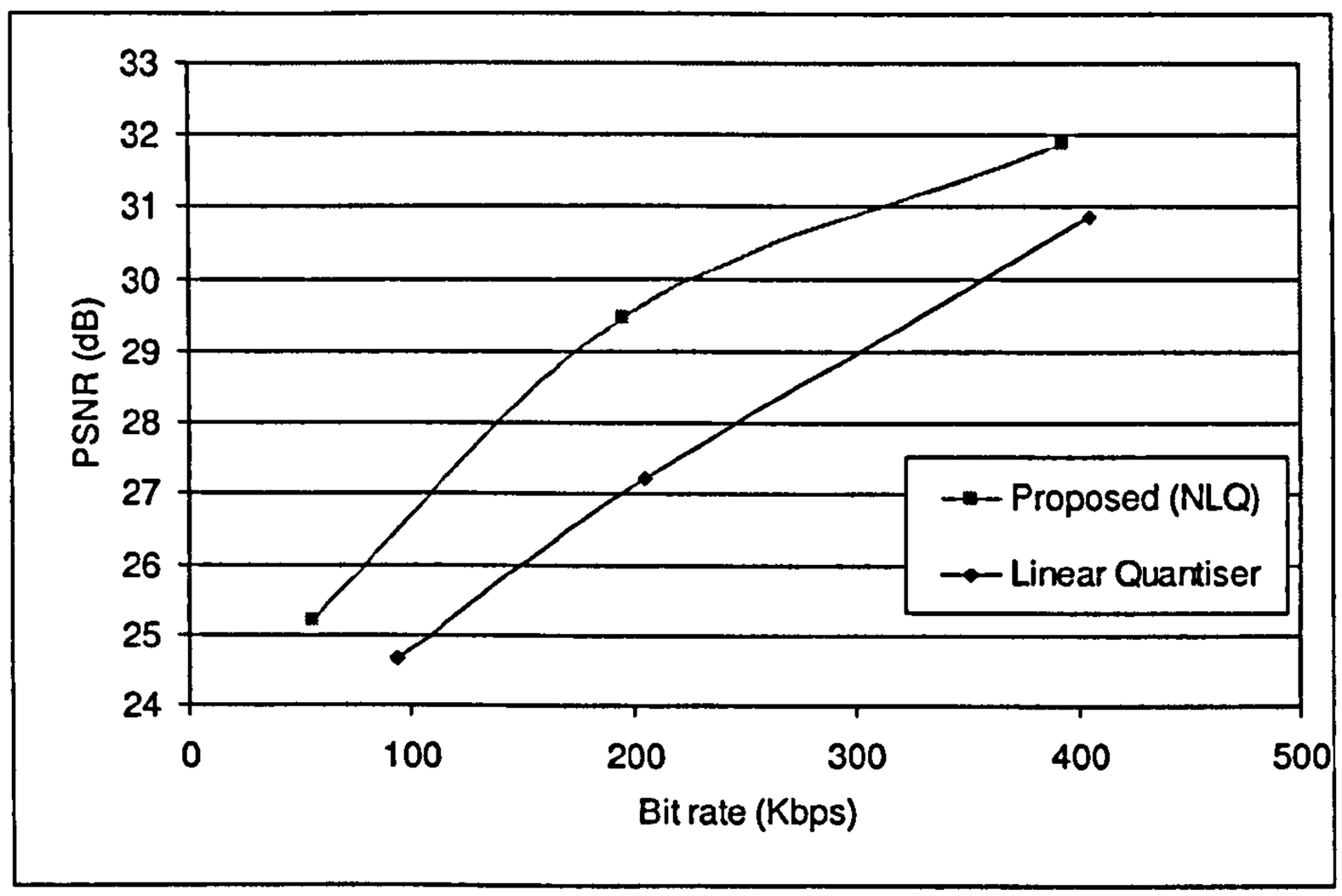



Figure 5-9 Performance comparison on non-linear quantiser (*Coastguard* sequence)

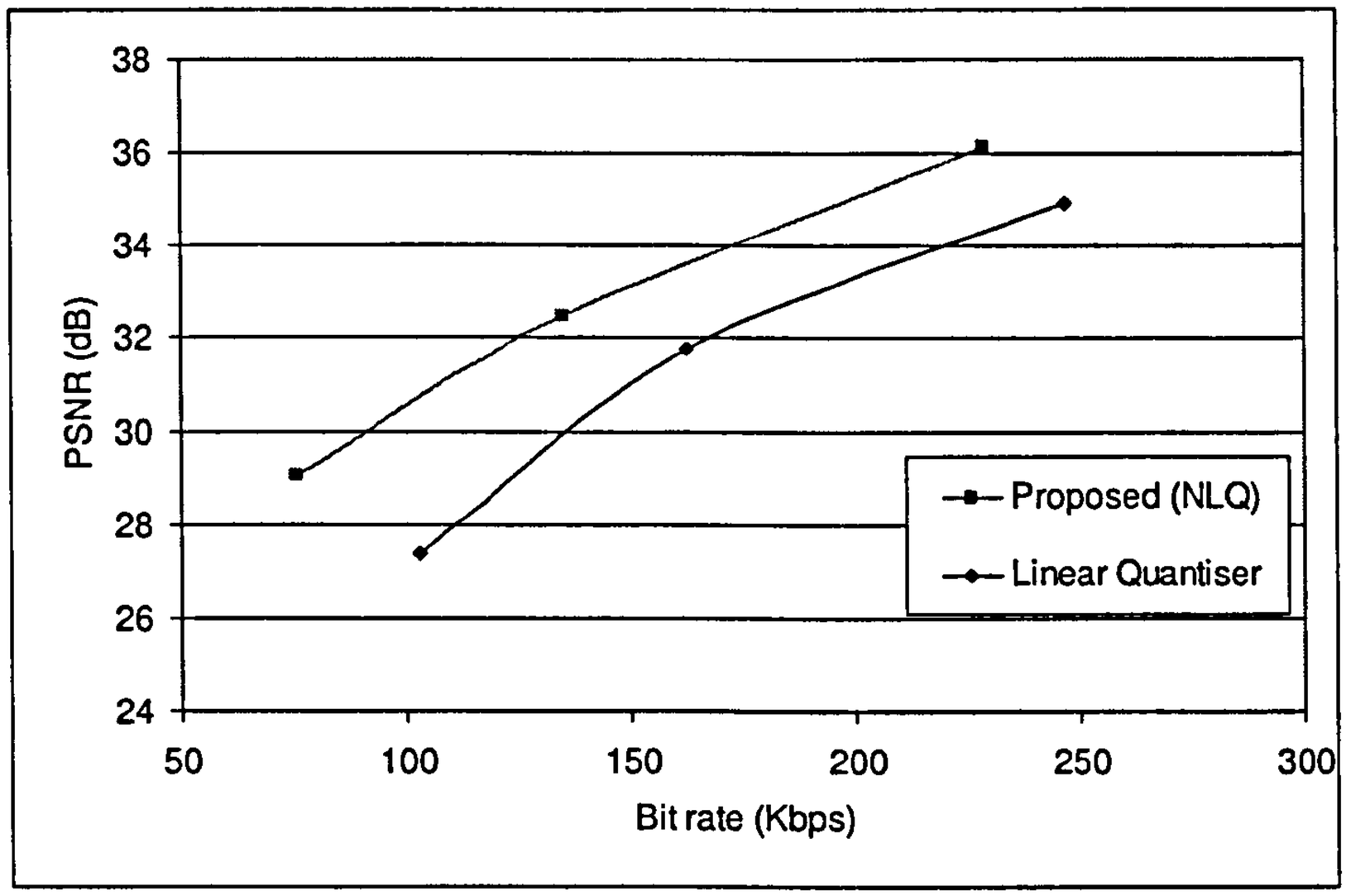



Figure 5-10 Performance comparison on non-linear quantiser (*Hall Monitor* sequence)

It can be observed that the proposed algorithm helps to improve the performance of the DVC codec by a significant margin consistently over different bit rates and different motion levels in the input video sequence. Over 3dB PSNR gain is evident at some points for a given bit rate. This gain could be attributed to the reduced quantisation noise resulting from reducing the quantisation step size around zero-bin. This factor has overridden the usually higher bit rate consumption caused by the increasing correlation noise when the step size is reduced. Thus effectively, manipulating the quantisation step sizes is a trade-off of quantisation noise and correlation noise and this research has determined an optimum operating point by introducing non linear quantisation steps. This is a very significant achievement, also considering that a heavy computational burden is also not incurred. It is further noted that the non linearity parameters $\alpha$ and $\beta$ would have distinct optimum values for each sequence depending on the frame statistics. However, a universal set of parameter values were utilised in the simulations for practicality.

There is an apparent drawback in considering the residual signal, though, in the sense that error propagation could result in if the key frames used in the computation of the residual signal are erroneously decoded. In this proposal, this problem is minimised by using the average of two

key frames for the residual calculation. Use of more reference frames would further strengthen the case. There is scepticism of the residual calculation in the DVC framework; if it's within the Slepian-Wolf guidelines for distributed source coding. Anyhow, considering the evident significant performance gain at very little computational cost to the encoder, this should be considered as an incremental approach towards the applications targeted by DVC.

## 5.2 Enhanced Reconstruction algorithm

In the Stanford DVC framework, the *reconstruction* module plays a significant role somewhat resembling an inverse quantization process. When the pixel domain DVC architecture is concerned, while quantization effectively means limiting the number of bit planes transmitted per pixel, reconstruction is the inverse process that increases the precision of the decoded quantized pixels without an additional bit rate burden. The reconstruction process utilizes the *side information* generated at the DVC decoder. In transform domain DVC, the quantization takes place in two stages; selecting the significant transform coefficients to be encoded and then limiting the number of bit planes for each coefficient band similar to the pixels in earlier case. Accordingly, the transform domain reconstruction is adapted.

The simple reconstruction model proposed by Aaron et al. [4], discussed earlier in section 2.5.3.5, has been widely used in subsequent DVC proposals. This approach has a few notable drawbacks:

- Each pixel is reconstructed independently with no consideration of the statistical properties of the neighbouring pixels, or the whole frame in general.
- There is no consideration of the bit error probability of the decoded bit stream or the side information. This could yield severe consequences when the decoded bit stream has a high BER as these errors will be reflected as irritating artefacts in the reconstructed pictures

Addressing some of the problem above, enhancements to the reconstruction algorithm were proposed in [50] and [51] which concentrated on the calculation of the expected value for the reconstructed pixel (or the DCT coefficient in transform domain coding) based on the Laplacian noise model of the residual error in the side information with respect to the original

video signal. However, it is observed that these enhanced techniques are optimized for a DVC framework with a reverse channel, which formulates a closed loop feedback mechanism for determining the optimum parity bit rate, thus yielding a controlled very low bit error rate (BER) in the Wyner-Ziv decoded bit stream.

Having considered the practical constraints in involving a reverse feedback channel in video communication systems, increasing attention is focused on unidirectional DVC architectures; with no reverse channel [17][77]. An extensive analysis of unidirectional DVC was presented in Chapter 4. It is note that the unidirectional coding framework tends to results in a higher, possibly uncontrolled, BER in the decoded bit stream depending on the employed encoder rate control algorithm. In this case, the reconstruction process could be better modelled as a statistical process involving the finite BER of the decoded bit stream as well as the side information stream. The objective of this section is to discuss an enhanced reconstruction algorithm optimized for this scenario. Both pixel domain and transform domain DVC frameworks are considered in the discussion.

### *5.2.1 Reconstruction model*

The proposed modification to the reconstruction algorithm is built upon the hypothesis that, when the BER of the decoded output increases, the relative significance of the side information increases in the reconstruction process. Thus the proposed algorithm is designed with particular focus on unidirectional architectures, unlike the previous proposals. This effect could be quantified considering the noise distributions of side information and the decoded output. Let's consider the case where the decoded quantized pixel falls outside the quantized side information bin. In this case, based on the above hypothesis, it is considered that the reconstructed pixel value can be modelled as a function of the side information value and the corresponding quantization bin boundaries as:

$$\hat{x} = \begin{cases} z^- + \tan(k)(y - z^-) & , y < z^- \\ z^+ + \tan(k)(y - z^+) & , y > z^+ \end{cases} \qquad 5\text{-}5$$

where $k$ is an index; defined in the proposed model as the *recon angle*. This parameter reflects the noise distributions of the decoded bit stream and the side information. A gross illustration of the proposed reconstruction model is shown in Figure 5-11.

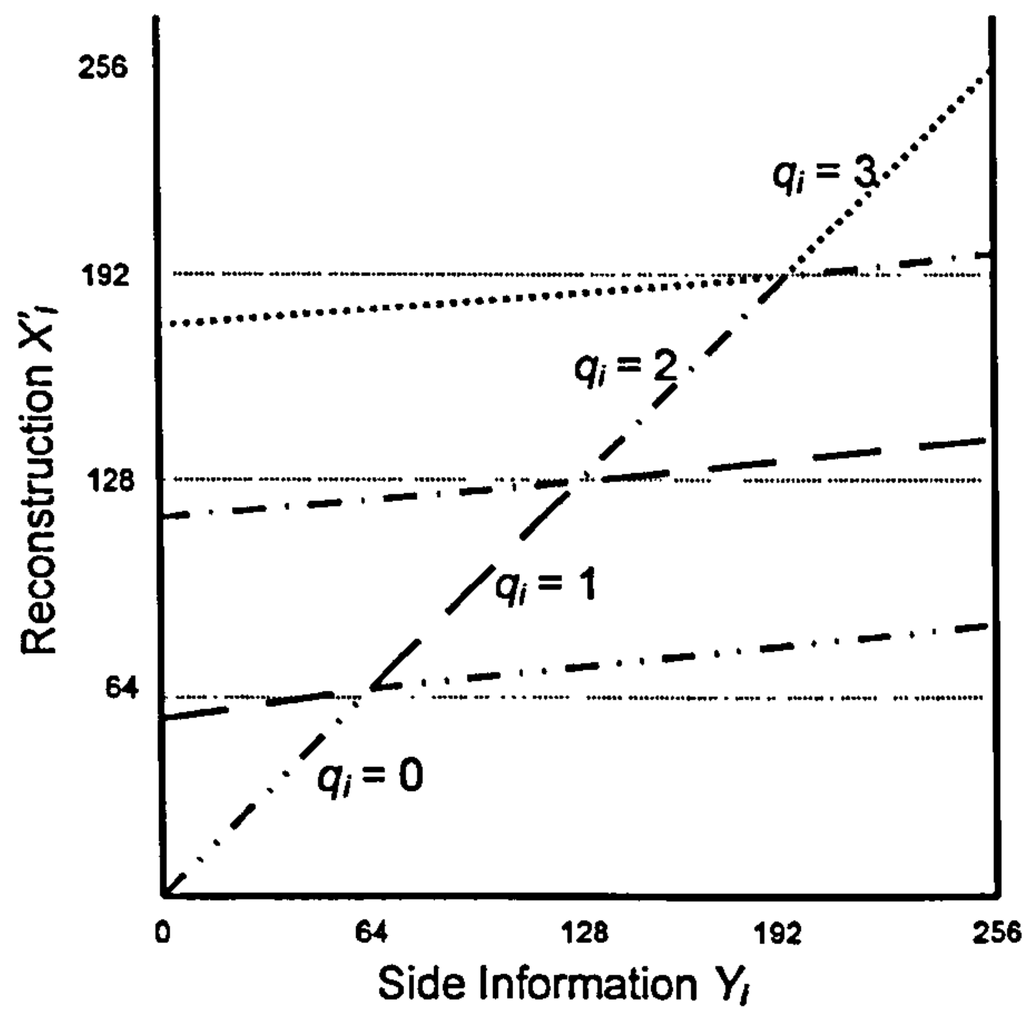



Figure 5-11 Proposed reconstruction model for 4 quantization levels

Here, the gradient of the reconstruction lines in the graphs resembles the recon angle. The derivation of the optimized recon angle formulates an important research problem. This problem could be addressed by considering the statistical properties of the decoded and side information bit streams. It is further noted that these statistical properties varies significantly within each frame based on the regional motion activity. To exploit this, the recon angle is optimized for blocks of pixels in the pixel domain implementation. In the transform domain implementation, this parameter is optimized for each coefficient band.

In the case the decoded quantized pixel falls within the quantized side information bin, the reconstructed pixel is derived as,

$$\hat{x} = y \quad , \qquad z^- < y < z^+ \qquad 5\text{-}6$$

### *5.2.2 Pixel domain implementation*

#### *5.2.2.1 Codec Integration*

The pixel domain unidirectional DVC architecture incorporating the modified reconstruction function is shown in Figure 5-12. The computationally very low complex ERC algorithm discussed earlier in section 4.2.2 is utilized for the rate control in view of maintaining the widely appreciated low complexity feature of the DVC encoder. This involves the estimation of a

form of side information at the encoder using average interpolation of two key frames. Then, for each macro block on the Wyner-Ziv frames, a BER is computed for the estimated frame with reference to the original frame. This BER is used to derive the rate decisions.

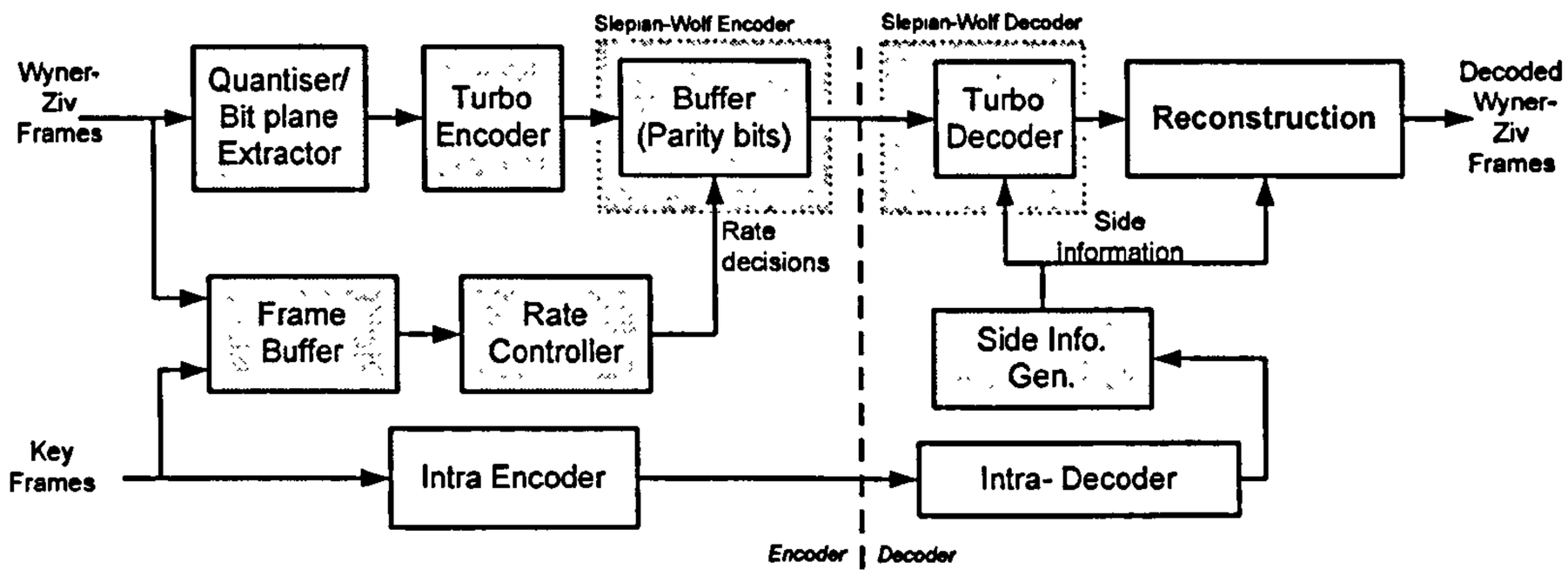



Figure 5-12 Pixel domain unidirectional DVC codec with modified reconstruction block

The reconstruction block implements the modified algorithm discussed earlier in this section, with the empirically optimised parameters.

In the proposed reconstruction model, the optimization of the recon angle formulates a critical research problem. This problem is addressed by statistical means through a training process to determine the recon angle that yields the optimum picture quality with minimum distortion. As per the common practice in this thesis, the objective picture quality, measured in terms of PSNR, is used in the optimization.

#### *5.2.2.2 Optimization of Recon Angle: Training process*

This section presents the training sequence performed for experimentally determining the optimum recon angle for different BER levels of the decoded bit stream. The sequential steps of the training process deployed for the macroblock based pixel domain coding framework are described below. The codec incorporates the same ERC algorithm during the training process as well as the subsequent simulations. The test conditions used in the training process are listed below.

**Test conditions for training process**

- Video sequences: *Soccer*, *Foreman*, *Coastguard*; QCIF (176×144); YUV 4:2:0; Frame rate: 15fps; GOP size:2

- PSNR (luminance only) and BER of the decoded bit stream are calculated for each macroblock separately.

- The macroblock size is set to 352 pixels. Therefore, in the bit plane level coding structure, the Slepian-Wolf coder operates on blocks of 352 bits each.

- The recon angle is varied in the range 0 – 45 degrees.

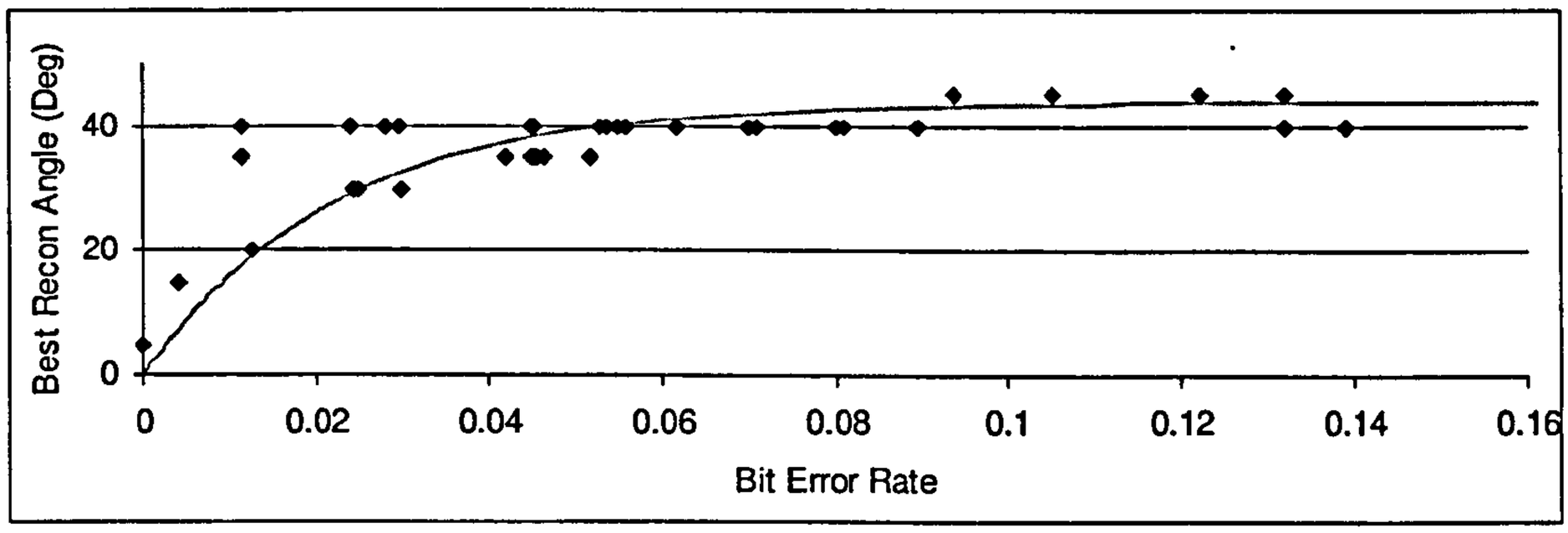
- Key frames are H.264/AVC Intra coded.

**The stepwise training process**

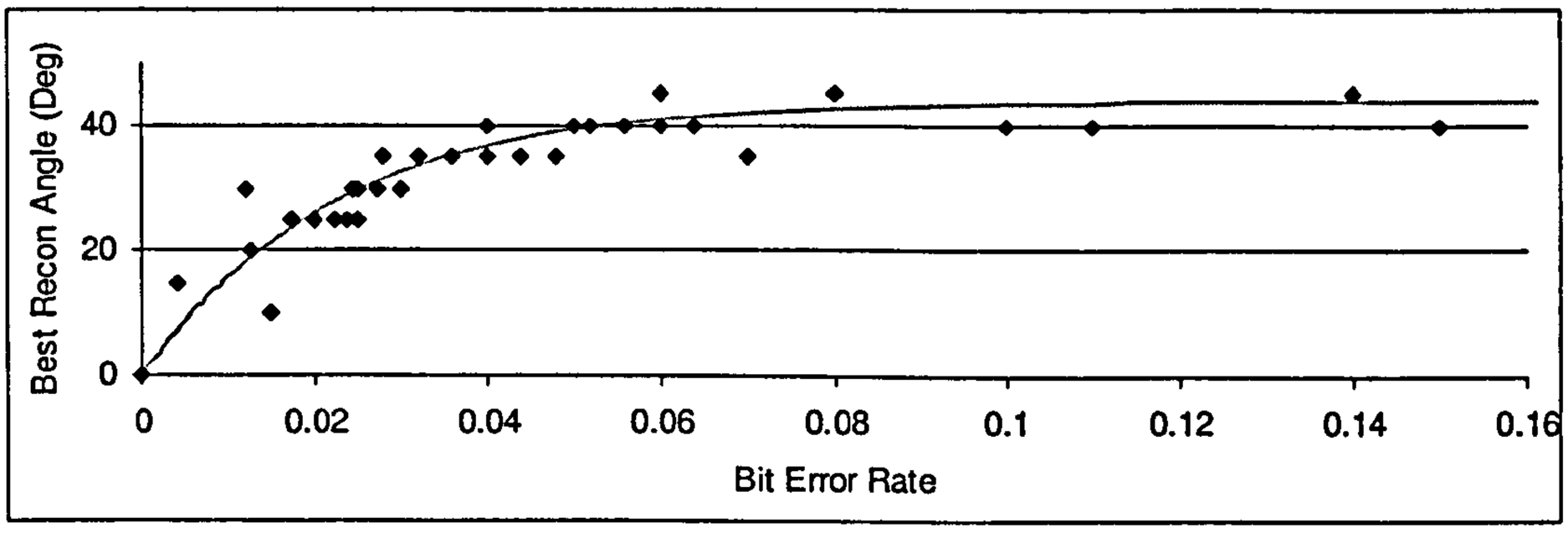
1. The codec is run using the above test conditions and recon angle is initially set to 0 degrees. Once the decoding is completed, calculate the PSNR and BER for each macro block. The BER calculation accounts for all decoded bit planes (the number of bit planes is determined by the pixel domain quantization parameter in use) of the macroblock.

2. The above step is repeated increasing the recon angle value up to 45 degrees in small steps, noting down the macroblock-wise PSNR for each increment in recon angle. Note that the BER for each macroblock is unaffected by the change of recon angle parameter.

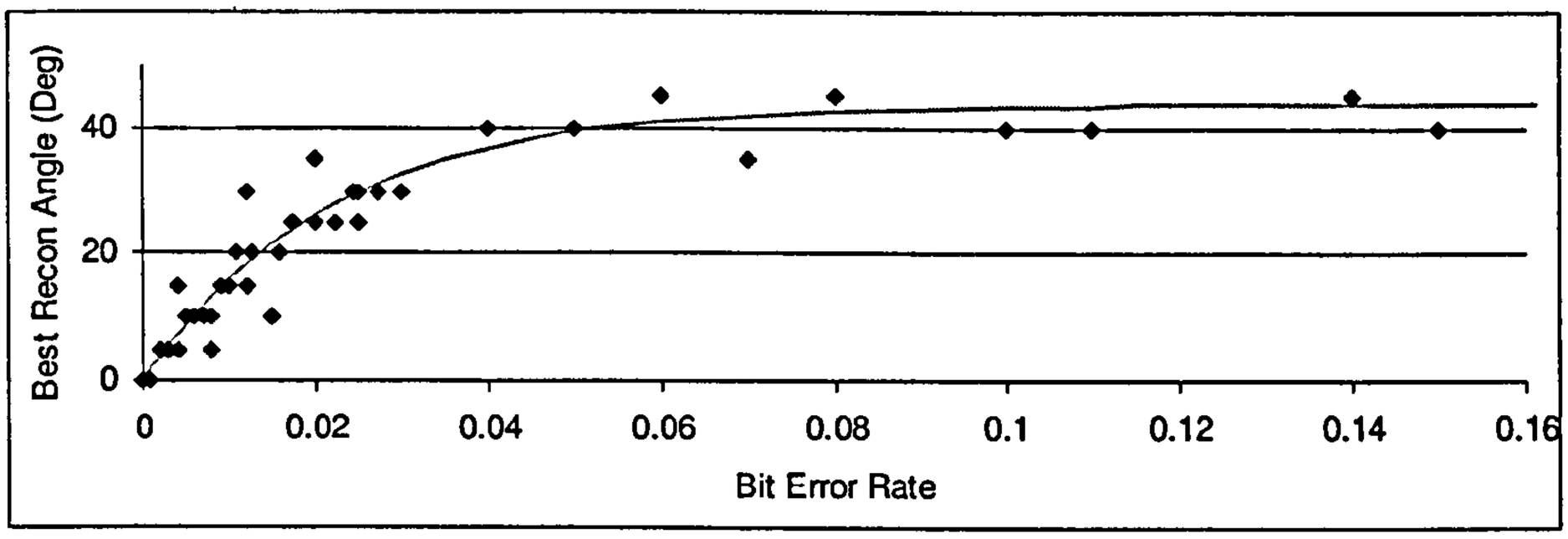
3. For each macroblock, evaluate the recon angle that yield the optimum picture quality, measured in terms of PSNR. Now these optimum recon angles are plotted on a scatter graph as illustrated in Figure 5-13. Note that each marker correspond to the optimum (in terms of objective picture quality) combination for one macroblock.




(a) *Soccer* sequence




(b) *Foreman* sequence




(c) *Coastguard* sequence

Figure 5-13 Scatter plot for the optimum recon angle for each block and the corresponding BER of the decoded bit stream

4. The regression line is drawn on the plots. This produces an approximation to the best-fit recon angle for the BER values for the macro blocks. It is observed that the regression model is consistent for all sequences irrespective of the contained motion levels, albeit differences in the density of distribution. When the motion level in the

sequence is higher, the average BER tends to increase resulting in a more dense distribution of operating points towards right on the BER axis.

The training process yields the optimum recon angle to be used for each value of BER per block in the decoded bit stream. The fact that the regression model is consistent for each sequence as observed in Figure 5-13 produces proof of a universal reconstruction model, subject to the test conditions listed above.

#### *5.2.2.3 Simulations and results*

The proposed reconstruction algorithm incorporated in to the pixel domain unidirectional DVC codec is tested for a number of test video sequences. The optimized recon angle parameter values are derived from the training process discussed above. The test conditions adopted in the simulations are listed below.

**Test conditions**

- Video sequences: *Foreman* (149 frames), *Hall Monitor* (149 frames); QCIF (176×144); YUV 4:2:0; Frame rate: 15fps; GOP size:2

- The recon angle is varied for each macroblock depending on the BER of decoded bit stream, following the regression line derived in the training process.

- Key frames are H.264/AVC Intra coded. The key frame quality is maintained identical for the corresponding test cases: with and without the proposed reconstruction algorithm.

- The PSNR is calculated for each Wyner-Ziv frame of the sequence, and plotted against the frame number.

- The block size is set to 352 bits.

- The performance is compared with the reconstruction model proposed in [4] keeping all other coding parameters identical.

**Results and analysis**

Figure 5-14 and Figure 5-15 depict the comparative results for *Foreman* and *Hall Monitor* sequences, respectively.

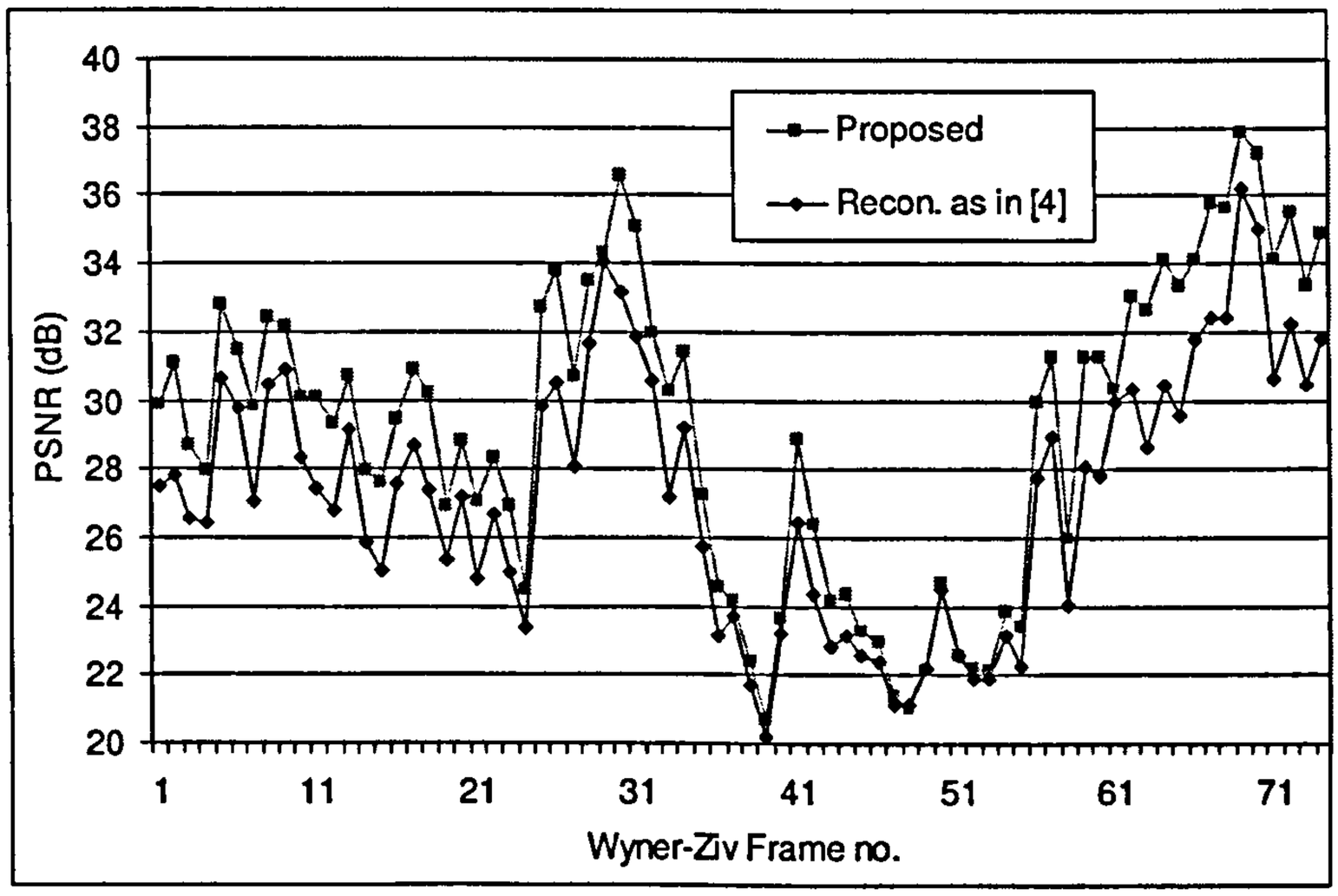



Figure 5-14 Performance comparison of the proposed algorithm (Pixel domain coding, *Foreman*)

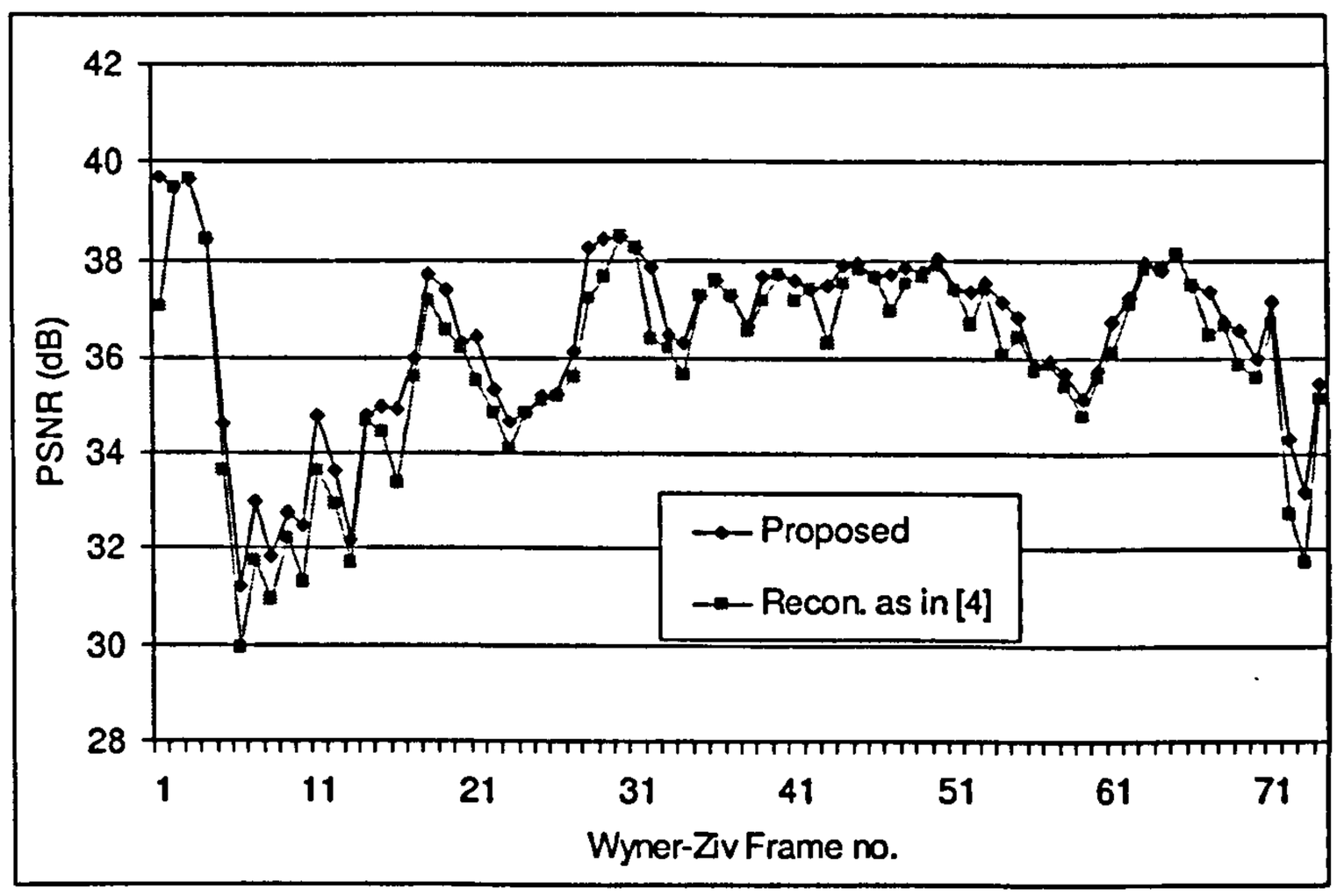



Figure 5-15 Performance comparison of the proposed algorithm (Pixel domain coding, *Hall Monitor*)

The simulation results show a very promising and consistent improvement in average PSNR of 1.9 dB for the *Foreman* and 0.5 dB for the *Hall Monitor* sequences. Figure 5-16 gives an illustration of the perceptual quality improvement produces by the proposed reconstruction algorithm.

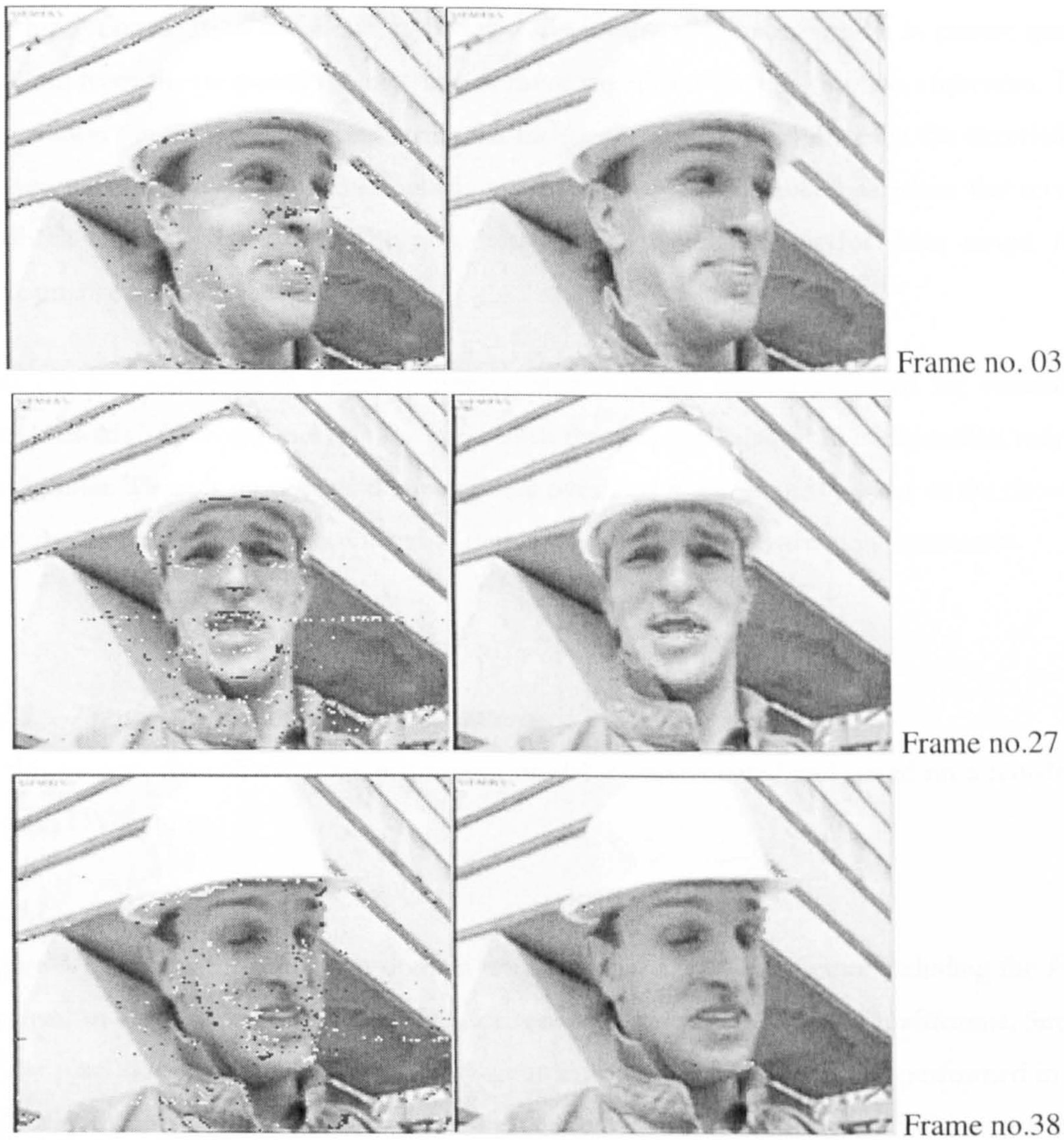



Figure 5-16 Performance comparison of the proposed algorithm (Pixel domain coding, *Foreman* sequence: frame no.3)

(a) without proposed model

(b) with proposed model

It is noted that, in the unidirectional architecture, the decoded bit stream has residual errors that yield a notable degradation of video quality as seen in Figure 5-16 (a). The proposed algorithm helps to solve these artefacts by biasing the pixels with higher error probability towards the side information.

It is further noted from the shown results that the comparative improvement in picture quality resulting from the proposed modification is more significant for high motion sequences. This observation can be explained considering the higher probability of bit errors in the decoded bit stream in high motion sequences, which may not be finely controlled as when the reverse feedback channel is available. The proposed algorithm compensates for these errors, thus performance gain is higher.

This can be considered as a very promising feature which can be exploited for concealing errors resulting from channel noise as well, since the proposed algorithm is dependant only on the decoder. There is no notable computational overhead to either the encoder or the decoder, once the training process is completed to derive the universal reconstruction parameters.

### *5.2.3 Transform domain implementation*

In this section, the enhanced reconstruction model is implemented and tested on a transform domain DVC codec.

#### *5.2.3.1 Codec Integration*

Figure 5-17 shows the transform domain unidirectional DVC architecture including the ERC mechanism and the proposed algorithm for reconstructing the transform coefficients. Similar to the pixel domain implementation, average interpolation of key frames is performed in the ERC algorithm and DCT is taken for the estimated frame. Then, the rate decisions for each coefficient band are taken in the rate controller, based on the correlation noise between the DCT coefficients of the interpolated frame and the original Wyner-Ziv frame.

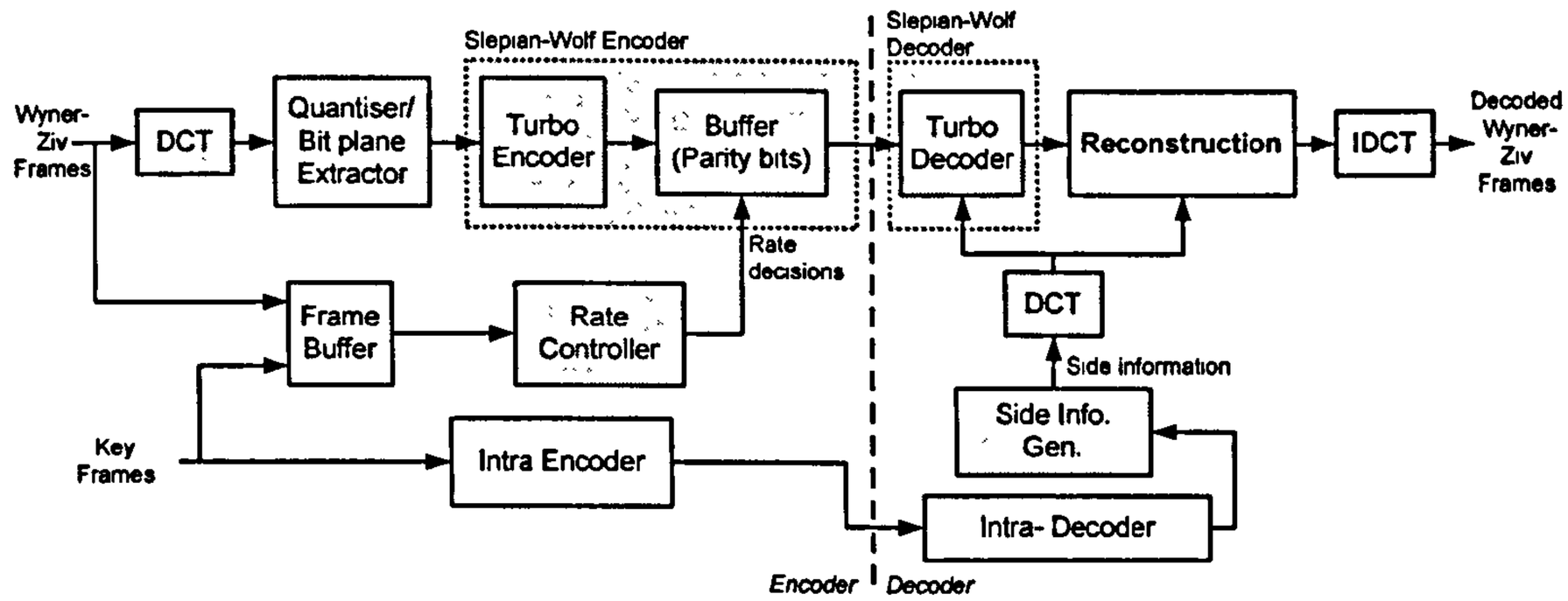



Figure 5-17 Transform domain unidirectional DVC codec with modified reconstruction block

A training process is performed for the transform domain codec similar to that discussed earlier in this chapter for the pixel domain case, to determine the optimized recon angle values for each coefficient band.

#### *5.2.3.2 Simulations and results*

This section presents the simulation results of the proposed reconstruction algorithm implemented on the transform domain unidirectional DVC codec. The test conditions adopted in the simulations are listed below.

**Test conditions**

- Video sequences: *Foreman* (149 frames), *Coastguard* (149 frames); QCIF (176×144); YUV 4:2:0; Frame rate: 15fps; GOP size:2

- The recon angle is varied for each DCT coefficient band depending on the BER of decoded bit stream, following the regression line derived in the training process.

- Key frames are H.264/AVC Intra coded. The key frame quality is maintained identical for the corresponding test cases: with and without the proposed reconstruction algorithm.

- The PSNR and bit rate are calculated for all Wyner-Ziv and key frames of the sequence.

- One block in the Slepian-Wolf coder correspond to one bit plane of a DCT coefficient band; thus the block size is 1584 bits.

- The performance is compared with the reconstruction model proposed in [4] keeping all other coding parameters identical.

Figure 5-18 and Figure 5-19 depict the comparative results for *Foreman* and *Coastguard* sequences, respectively. A PSNR gain of up to 2dB is observed for the *Foreman* and up to 1dB for the *Coastguard* by utilizing the proposed reconstruction technique. Similar to the pixel domain case discussed earlier in section 5.2.2, the performance gain is attributed to the proportionate biasing of the coefficient blocks with high bit error probability towards the side information, which then is the more reliable information.

Figure 5-20 graphically illustrates the perceptual quality improvement using selected frames from *Foreman* and *Coastguard* sequences. The performance improvement observed here is consistent with the above results. Note here that the block artefacts caused by the bit errors in the transform domain coding scenario are contrastingly different to the *dots* in pixel domain coding due to the 4X4 spatial blocks considered in the DCT transform.

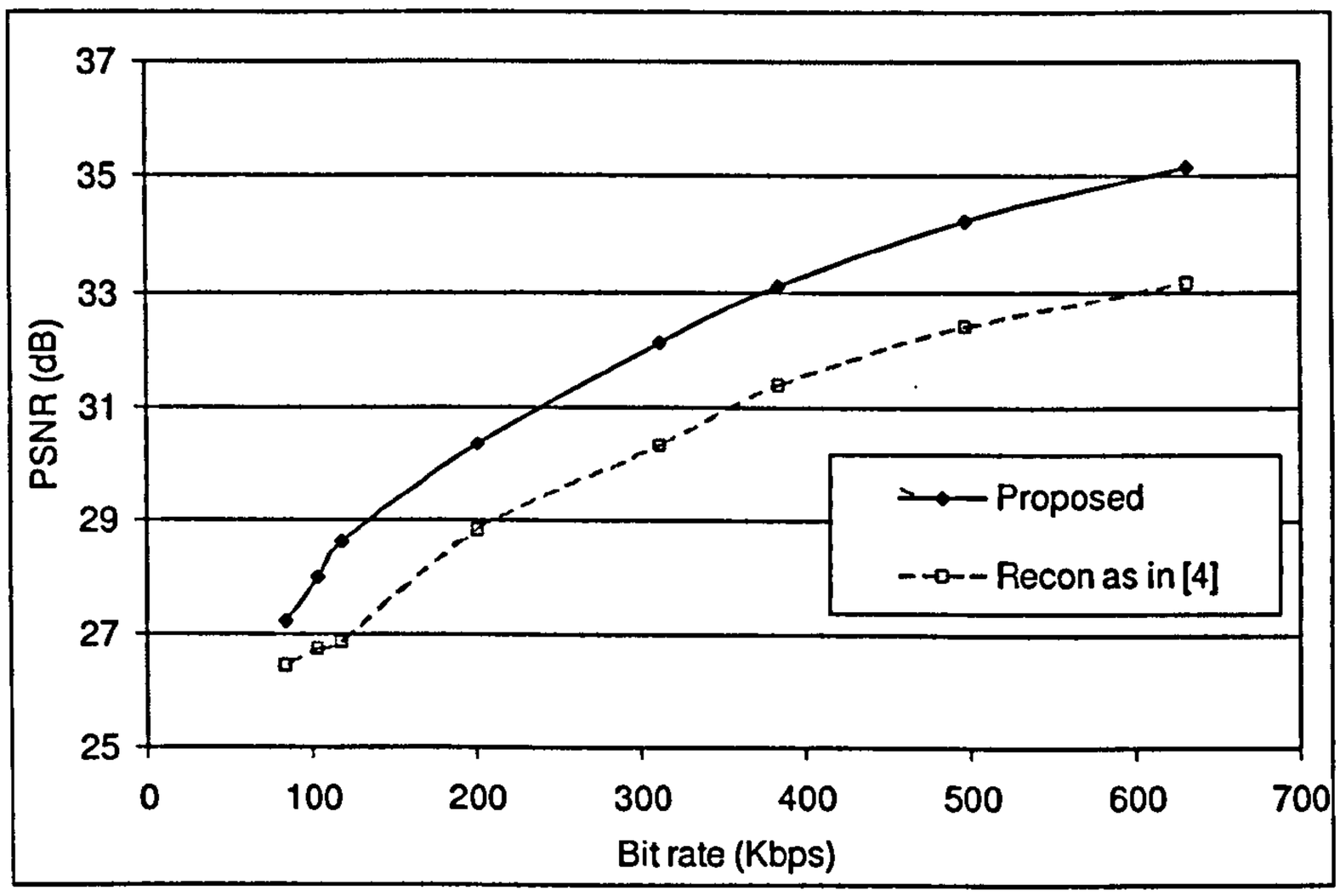



Figure 5-18 Performance comparison of the proposed algorithm (Transform domain coding, *Foreman* sequence)

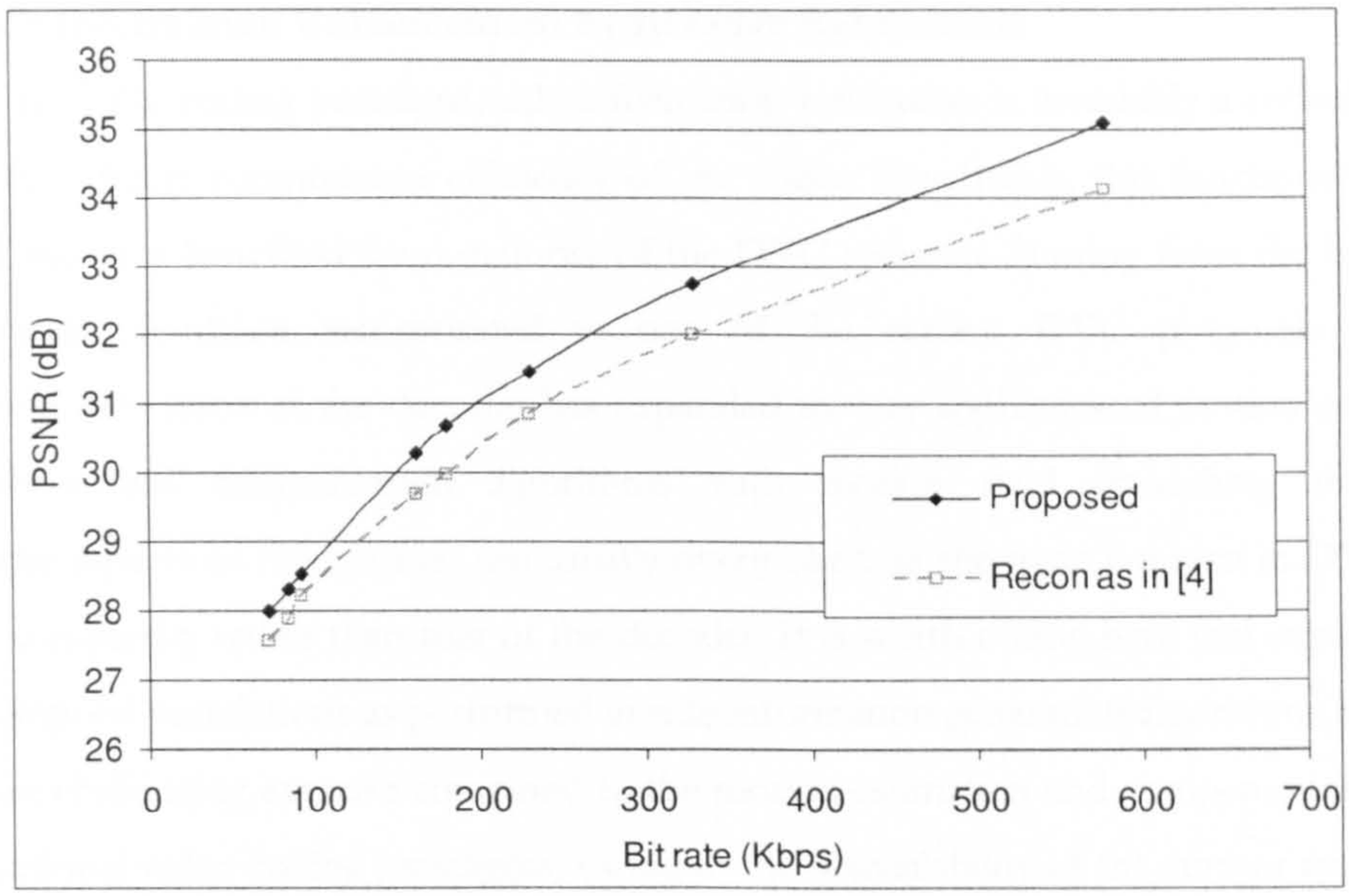
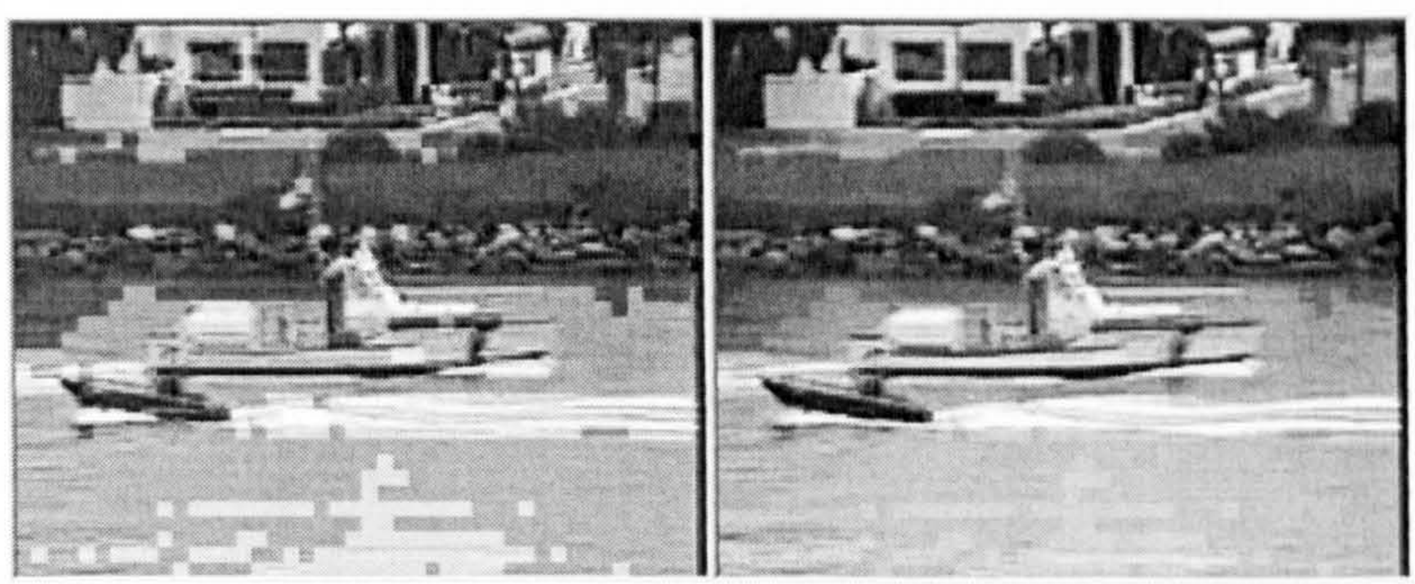



Figure 5-19 Performance comparison of the proposed algorithm (Transform domain coding, *Coastguard* sequence)

*Foreman*



*Coastguard*

(a) (b)

Figure 5-20 Performance comparison of the proposed algorithm (Transform domain coding)

(a) without proposed model

(b) with proposed model

## 5.3 Side Information Enhancement by Iterative Refinement

In the Wyner-Ziv coding paradigm, side information estimation is invariably a critical process that directly affects compression efficiency of the codec. Resultantly, this functional area has attracted and thus benefited from majority of the DVC research. Starting from the basic pixel interpolations algorithm incorporated in one of the earliest DVC proposals [4], side information estimation at the decoder has expanded to very sophisticated motion estimation, interpolation and compensation algorithms with motion field smoothing techniques. Complexity aspects of this process are usually overlooked, as the main concern in DVC is the encoder complexity rather than that of the decoder. It is worth noting here that exploiting the source temporal correlations as performed in side information generation algorithms in DVC is a far more challenging exercise compared to the motion estimation and compensation process in conventional video coding paradigms, owing to the unavailability of the current target frame in the process. This problem was elaborated in detail, earlier in section 2.5.3.3 of this thesis.

In addition to the side information generation algorithms, there has been some efforts in recent research to dynamically refine the side information over the decoding process [29][79]. The concept of side information refinement has been feasible due to the incremental nature of the encoding and decoding processes. When the pixel domain coding framework is concerned, the Wyner-Ziv frames are usually coded in a bit plane-wise structure, usually starting from the most significant bit plane. Thus, once a bit plane is decoded, it can be used as *prior information* to refine the side information for the subsequent bit planes.

The objective of this section is to present an approach for the dynamic refinement of side information by iterative decoding, and thereby improve the rate-distortion performance of the codec. In this proposal, the spatial and temporal correlations of the frame sequence are iteratively exploited. The proposed technique is used in conjunction with the DVC decoding algorithm proposed in [78] which is a combined enhanced version of the *3-D motion refinement* [29] and *multiple side information* [79] techniques

### *5.3.1 The refinement algorithm*

The efficiency of a side information generation algorithm is dependant on the optimally exploiting of the spatial and temporal correlations in the adjacent reference key frames in the sequence. It is further noted that the partially decoded information in the current Wyner-Ziv frame also can be used to enhance this process. The proposed design is intended at utilizing an

iterative processing algorithm to maximize the exploitation of the source redundancies. The block diagram of the proposed DVC codec incorporating the enhanced side information generation technique is shown in Figure 5-21.

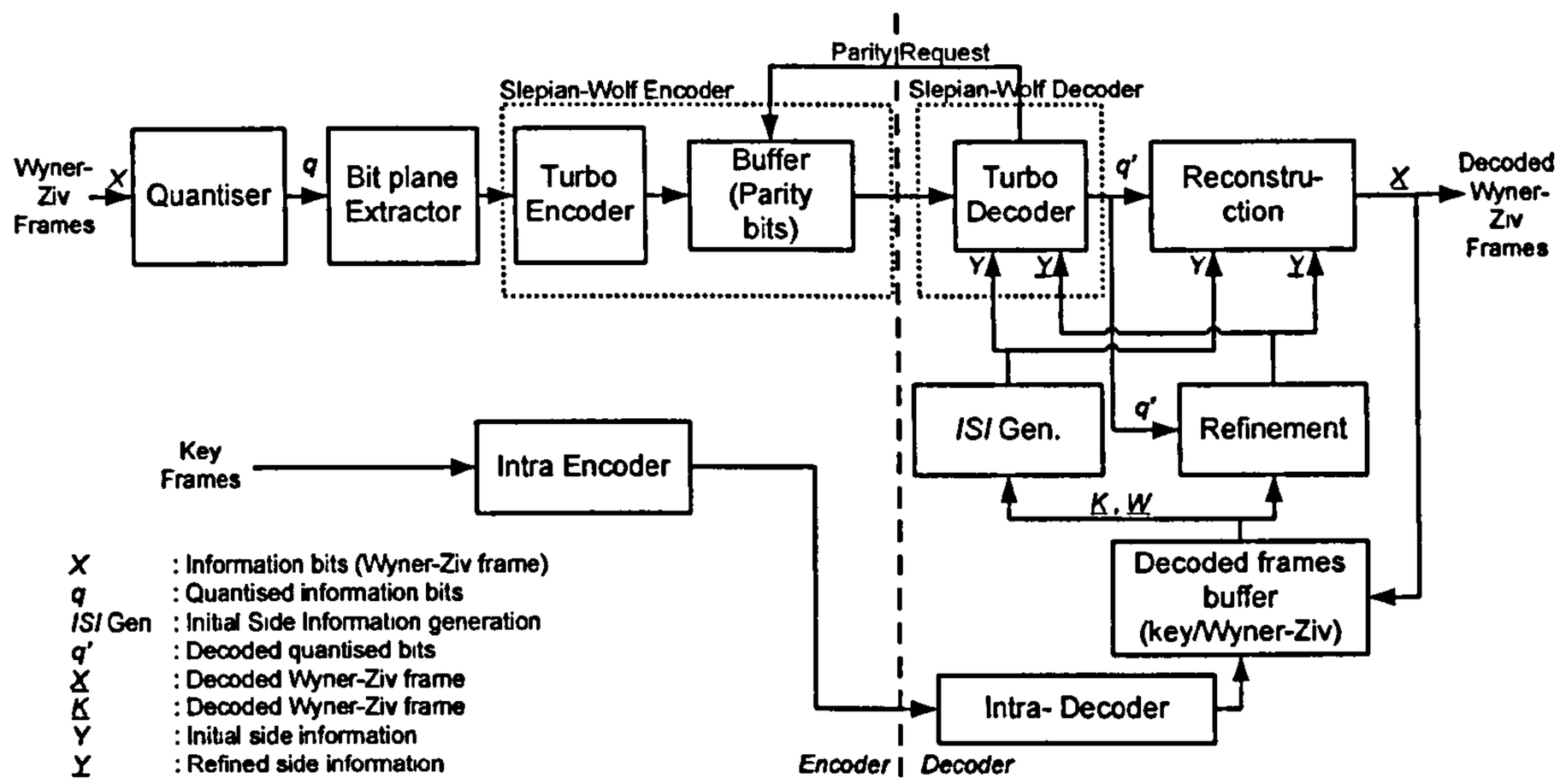



Figure 5-21 DVC codec with proposed side information refinement

The side information refinement technique is illustrated in Figure 5-22, and described below.

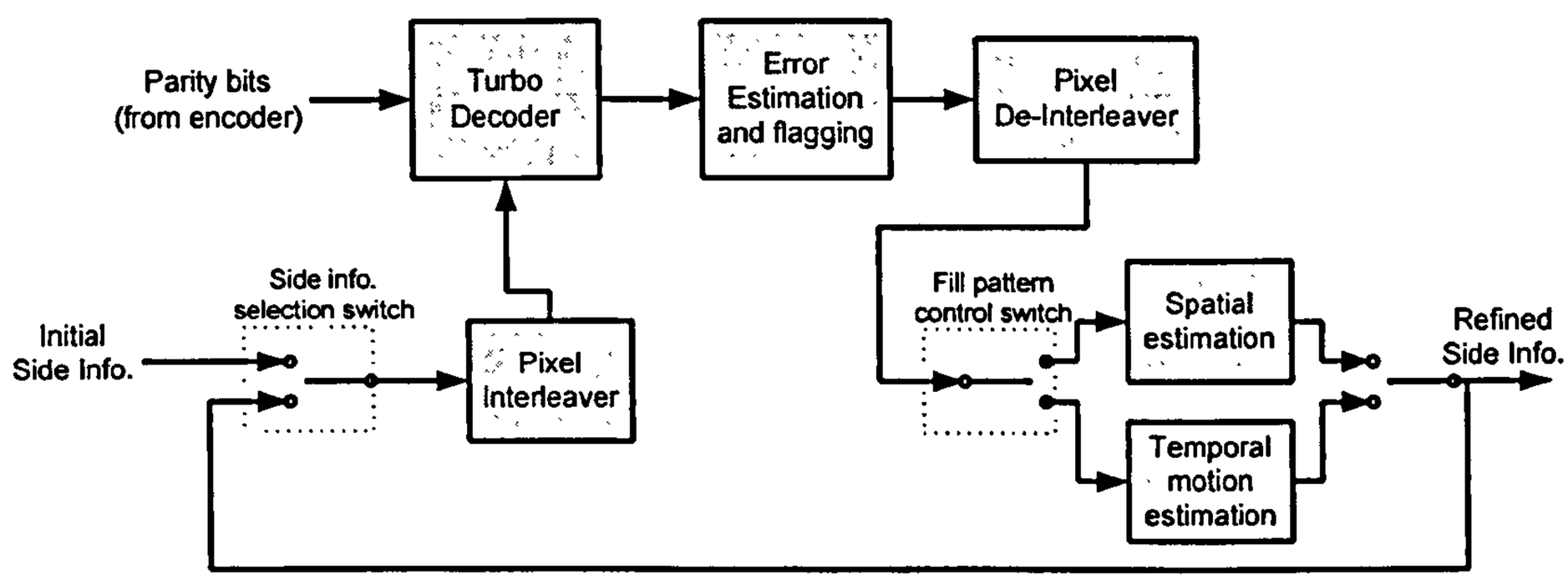



Figure 5-22 Side information refinement algorithm

The process starts by producing *initial side information* using an existing side information generation algorithm. For the simulations discussed later in this section, a motion extrapolation

[26] based algorithm utilizing the previously decoded key frames is used. In the existing DVC architecture, as discussed earlier in chapter 2, this side information is used in turbo decoding in conjunction with the parity bit stream received from encoder and then the decoded bit stream is subjected to an error estimation process. The parity bit rate starts at an arbitrarily small value; and dynamically increase based on the outcome of the error estimation process. The proposed scheme starts similarly with the initial side information and arbitrarily small parity rate and performs the error estimation on the decoded bit stream. However, unlike in the conventional algorithm, the decoded bit stream is forwarded to the new refinement module if an above-threshold BER occurs in the error estimation.

The codec incorporates a pixel interleaver similar to that proposed in section 4.2.1. The interleaving helps to suppress the adverse effects of burst errors typically evident in the side information stream by scattering the bit sequence. The bit stream is split to linear blocks for this process and optimum block size (significantly smaller than a full bit plane size) was determined experimentally by repeatedly simulating the codec varying the block size. The step size in varying the block size in the above optimisation was initially set to 100 and then fine tuned using steps of 10. If the post decoding error estimation for the block yields a BER above the pre-set threshold, all bits in the block are flagged similar to that in the iterative refinement algorithm proposed for unidirectional DVC in section 4.3.2 above. Next, the decoded bit plane is de-interleaved resulting in the flagged bit bursts to scatter. A prediction algorithm is then utilized to fill the flagged holes in the bit plane. The flag-filled bit plane is then interleaved again and enters the next iteration of the turbo decoder with the same parity bit stream. The flagging and spatio-temporal prediction based flag-filling cycle continues for either a predetermined number of iterations, or a predetermined BER, is reached. It was experimentally determined that alternative use of spatial and temporal prediction techniques, with spatial prediction as initial step, in the above process yields optimum results. The maximum iteration count was set to 4 for the simulations as a trade-off of computational complexity and performance.

If the above refinement loop was terminated by the iteration count, rather than achieving the desired BER, then the parity bit stream is increased dynamically by sending a parity request message to the encoder buffer. The parity request loop would then continue until the pre-set BER threshold is reached as in existing feedback based DVC codecs. Once the BER for each bit plane of the frame is decreased below the threshold, the bit stream is forwarded to the reconstruction module.

### *5.3.2 Simulation Results*

The proposed DVC codec was tested with a number of test video sequences using the test conditions listed below.

**Test Conditions**

- Video sequence: *Foreman* (100 frames); QCIF (176×144); YUV 4:2:0; Wyner-Ziv frame rate: 15fps; GOP size:2

- Bit rate and PSNR are calculated and displayed for Wyner-Ziv frames, averaged over the sequence.

- PSNR is calculated for the luminance signals only.

- Each rate-distortion point is achieved for pixel domain quantiser parameter: $M$, where $2^M = \{2,4,8,16\}$ defines the number of quantization levels.

- The iteration count in the side information refinement was set to 4.

The simulation results for the side information refinement algorithm for the *Foreman* sequence are illustrated in Figure 5-23. The results are compared with a number of existing pixel domain DVC codecs [78][29][79][26].

It is evident from the illustrated results that the proposed enhancement has resulted in a significant improvement in rate-distortion performance of the pixel domain DVC codec. The PSNR gain is up to 3 dB for a similar bit rate in the reference results. With the iterative refinement, the side information gets statistically much closely correlated to the original Wyner-Ziv frame. Thus the Slepian-Wolf decoder demands less number of parity bits from the encoder in the feedback based parity-on-demand architecture, to achieve the pre-determined bit error rate threshold in the decoded output bit stream. Thus the resultant bit rate for a given quantiser value is reduced. Further, the improved quality of the side information yields an improved PSNR value for the output frames sequence during the reconstruction process.

The tests carried out with other video sequences derived similar results

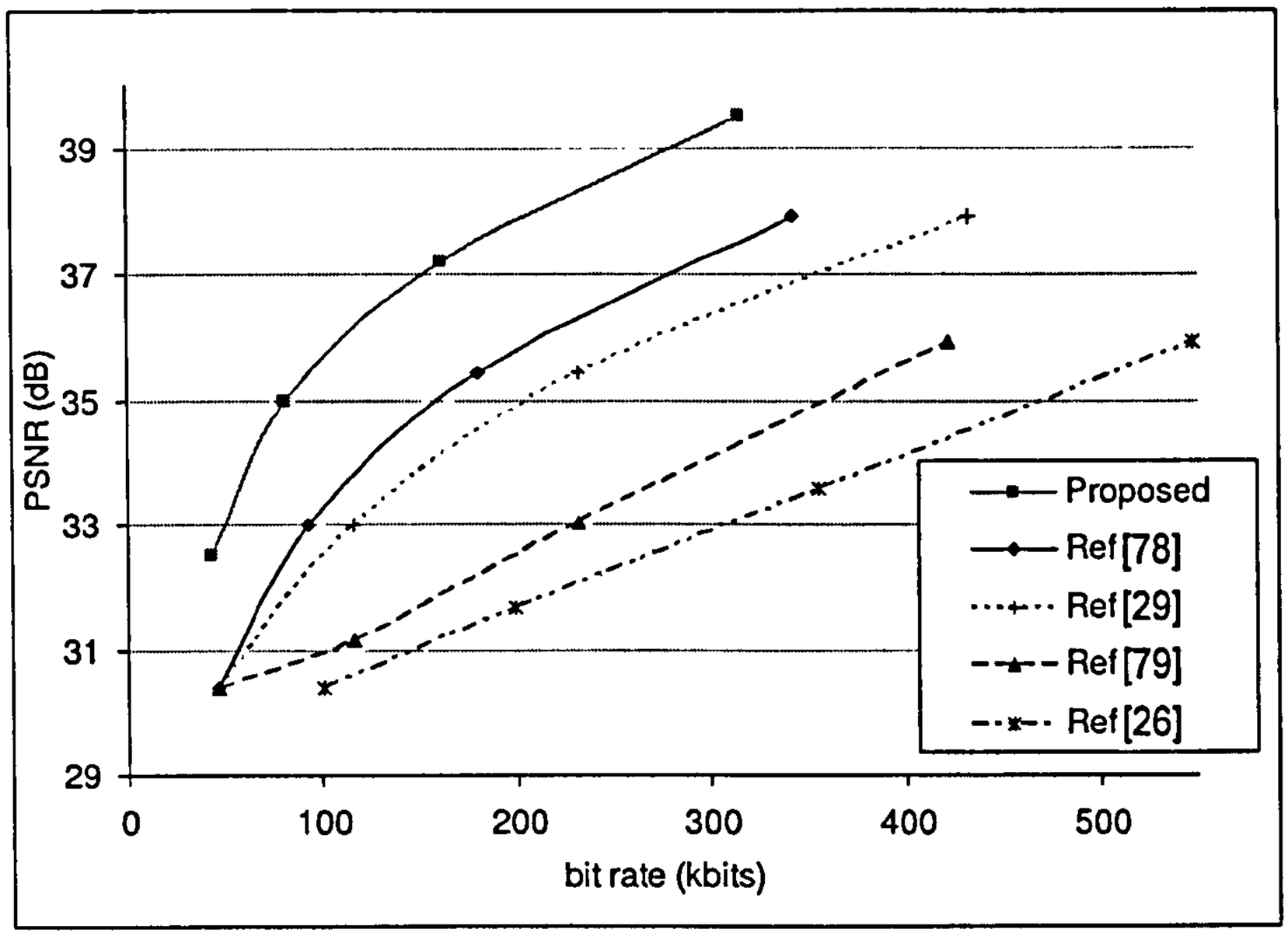



Figure 5-23 Performance Comparison of refinement algorithm for *Foreman* Sequence

**Computational Complexity Analysis**

It is understood the proposed algorithm introduces additional computational complexity to the DVC decoder. Table 5-1 shows the results of a quantitative analysis of the computational complexity of the proposed decoder. The complexity is measured in terms of average processing delay. The simulations were run (repeatedly) on a personal computer with the following specifications: Intel(R) Core (TM)2 CPU 6400 @ 2.13 GHz, 3.00GB of RAM.

The proposed refinement is a computer intensive task that involves the following major tasks carried out in each iteration (4 iterations are performed in the simulations):

- One turbo decoding cycle for flagging
- Temporal block search based flag-filling *or* Spatial flag-filling

Table 5-1 Complexity analysis: Average decoding time per frame

| Sequence | Proposed decoder | | Without side information refinement [78] | |
|---|---|---|---|---|
| | Average decoding time per frame (s) | Average bit rate (kb per frame) | Average decoding time per frame (s) | Average bit rate (kb per frame) |
| *Foreman* | 8.33 | 2.91 | 4.83 | 3.10 |
| *Coastguard* | 11.25 | 3.01 | 6.43 | 3.19 |

At the end of 4 iterations of the side information refinement, the proposed DVC decoder incorporates a full turbo decoding process that would normally include multiple turbo decoding cycles based on the closed loop feedback process in the DVC codec. It is noted here that even though this full turbo decoding process is identical to that in the reference algorithm used here for comparison of performance, the number of parity request iterations is less in the proposed design since the correlation noise in the refined side information is significantly low, forcing the feedback loop to terminate faster. This gives a complexity saving to the proposed algorithm. The simulation results depicted in Table 5-1 show that the overall complexity has risen by approximately 73% for a PSNR gain of approximately 2dB. This is believed to be a beneficial trade-off.

It is further noted here that processing delay figures in Table 5-1 do not reflect a real-time coding scenario. This is mainly due to the un-optimised software coding of the algorithms. Code optimisation and hardware implementations are expected to produce faster real-time processing.

## 5.4 Conclusion

This chapter focused on a number of modular elements of DVC, and proposed innovative algorithms to enhance the performance. First, a non-linear quantisation algorithm was presented in contrast to the linear scalar quantiser traditionally utilised in the DVC designs. The proposed algorithm minimises the high quantisation noise present in the Wyner-Ziv residual coding scenario when linear quantisation is used. The performance gain of the proposed non-linear quantisation algorithm was evaluated against the state-of-the-art pixel domain video codec that uses a linear quantiser and a significant rate-distortion performance gain was evident.

Then an enhanced reconstruction algorithm was presented focusing on the unidirectional DVC framework discussed earlier in Chapter 4. When the unidirectional architecture is considered, the decoded bit stream tends to contain residual bit errors due to the possibly sub-optimal parity puncturing process. A solution to this problem is proposed by tracking the error probability of each bit in the decoded bit stream and accordingly modifying the reconstruction function by biasing the reconstructed pixels towards side information. A *recon angle* parameter is defined and a systematic training process is introduced to derive the optimum recon angle corresponding to the bit error rate in the decoded bit stream. Simulations were carried out for both pixel domain and transform domain coding frameworks and a significant PSNR improvement was observed at no additional bit rate burden to the codec.

Considering the significance of side information in the DVC codec, a side information refinement algorithm was also presented in this chapter. The algorithm starts with an *initial side information* generated by any existing algorithm, then keep on refining it iteratively by exploiting the spatial and temporal correlations in the current Wyner-Ziv frame and its neighbouring frames.

The side information refinement algorithm, built into the existing reverse channel feedback based pixel domain DVC framework, was tested for standard video sequences and a significant PSNR gain is evident when compared with the existing pixel domain codec that uses the same initial side information. An analysis of the additional computational complexity involved in the proposed iterative refinement algorithm was also presented.

The next chapter focuses on the redesign of the noise models incorporated in turbo decoding in the DVC decoder in order to improve the coding efficiency.

*Chapter 6*

# DVC CODEC OPTIMISATION FOR WIRELESS CHANNELS

Since the initial proposals on distributed video coding a few years back, the DVC researchers have been focusing on improving the compression efficiency of the codec assuming a virtually error free availability of the Wyner-Ziv encoded bit stream at the decoder. In the Wyner-Ziv coding scenario, the encoded bit stream is constituted of the parity bit stream that is punctured either dynamically or statically depending on the availability of the reverse channel in the coding framework.

DVC has been widely proposed for a class of applications that involve a network of remote video sensors that capture the video information and upload to centralized decoding nodes. The common application domain is video surveillance. In these deployments, wireless networks are the widely available access networks of choice, and internet could be largely utilized as a low-cost transport network. Since these are highly error prone media, this communication scenario demands for a significant consideration on the error resilience properties of the coding technologies. This is a relatively untouched research area, possibly since DVC is still a developing technology. Analysis of the performance of the currently available DVC codec solutions over noisy channels has been the focus of some recent research publications, notably for packet based networks [80]. However, it is noted that a number of potential applications of DVC involve wireless sensor networks with multipath fading effects in the transmission media. The resultant performance degradation due to the channel fading has hardly been discussed in literature.

In the light of above, this chapter focuses on a study of the performance of the DVC codec over noisy channel conditions arisen in wireless communications. Further, novel noise models will be proposed for the relevant bit streams involved, and accordingly the decoding algorithms will be modified to enhance the error resilience of the codec.

## 6.1 Dual Channel System

In this discussion, the wireless channel noise scenario in DVC is modelled as a dual channel system. This scenario is illustrated in Figure 6-1; a description follows.

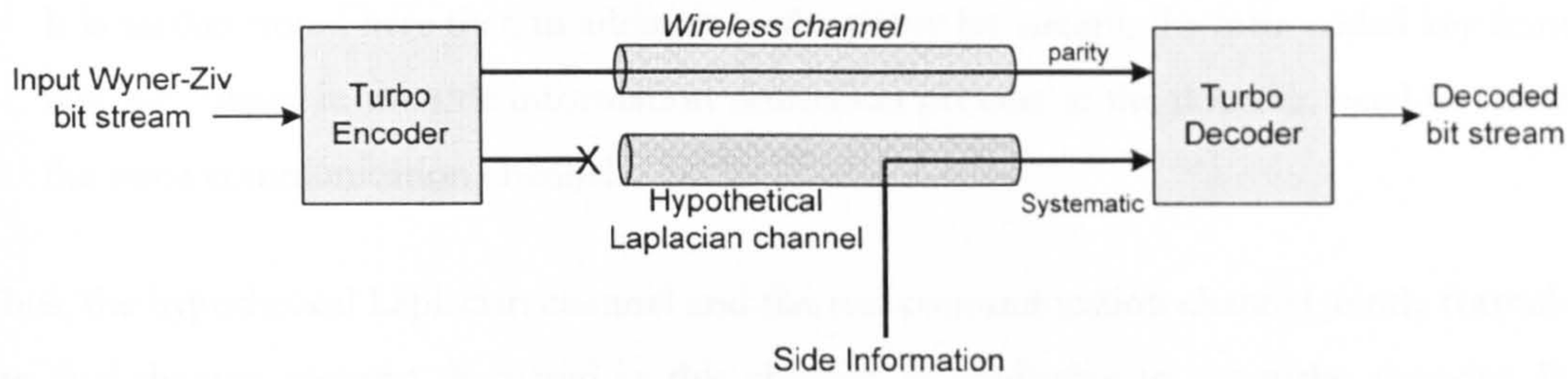



Figure 6-1 The dual-channel model

The Wyner-Ziv frame is sent through a turbo encoder, which is constituted of two parallel recursive systematic convolutional (RSC) encoders. The output parity bit stream of the turbo encoder, after puncturing, formulates the Wyner-Ziv encoded bit stream, whereas the systematic bit stream is discarded. At the Wyner-Ziv decoder, the turbo decoder awaits the systematic and parity bit streams that are transmitted from the encoder in a typical communication scenario. However, in Wyner-Ziv coding only the parity bit stream is received from the encoder and the systematic bit stream is sourced from the locally generated side information sequence. The side information is usually modeled using a Laplacian noise process. In this setup, there are clearly two identifiable channels as relevant to the turbo decoder:

- Hypothetical Laplacian channel

  This hypothetical channel is modeled between the original systematic bit stream, which is used at the encoder to generate the parity bit stream (and subsequently dropped), and the side information bit stream that is fed into the turbo decoder. The correlation noise modeling process performed on the side information, to determine the statistical properties of this hypothetical Laplacian noise was briefly discussed earlier in section 2.5.3.4.

- Real communication channel

  This is mainly applied to the punctured parity bit stream transmitted from the encoder. The statistical noise model is dependant on the utilized transmission media. An additive white

Gaussian noise (AWGN) channel is commonly resembled. A wireless channel would usually incur time varying and frequency selective multipath fading of the modulated signals.

It is further noted here that, in addition to the parity bit stream, the intra coded key frames that are utilized in the side information estimation process at the decoder, need to traverse the same communication channel.

Thus, the hypothetical Laplacian channel and the real communication channel jointly formulate the dual-channel scenario discussed in this chapter, as applicable to the turbo decoder. It is emphasized here that this channel model differs from a conventional turbo coding scenario in a communications system. In the latter, the systematic and parity bit streams are jointly transmitted over the communication channel and thus these two streams assume similar noise models at the decoder. The maximum a-posteriori (MAP) decoding algorithm [81] utilized in turbo decoding is thus designed for this latter scenario and makes itself suboptimal for use in the DVC decoding over error prone wireless channels.

In this chapter, appropriate noise models are developed for the above dual channel system and thereby the necessary modifications to the MAP algorithm are discussed to enhance the performance of the DVC codec with noise and fading effects on the communication channel. AWGN and W-CDMA wireless channel models are considered for simulations.

## 6.2 Noise Modelling in the Dual Channel System

This section covers the mathematical modelling of the proposed dual channel system. In turbo decoding, the soft outputs are typically represented by the Log Likelihood Ratio (LLR) of which the polarity determines the output bit. The LLR of a data bit $u_k$ is defined by:

$$L(u_k) = \ln\left(\frac{P(u_k = +1)}{P(u_k = -1)}\right) \qquad 6\text{-}1$$

Given the LLR L(uk), it is possible to calculate the probability that $u_k = +1$ or $u_k = -1$ as follows.

Noting that, $P(u_k = -1) = 1 - P(u_k = +1)$ and taking the exponent of both sides in Equation 6-1, we can write:

$$e^{L(u_k)} = \frac{P(u_k = +1)}{1 - P(u_k = +1)} \qquad 6\text{-}2$$

$$P(u_k = +1) = \frac{e^{L(u_k)}}{1 + e^{L(u_k)}} \qquad 6\text{-}3$$

$$= \frac{1}{1 + e^{-L(u_k)}}$$

Similarly,

$$P(u_k = -1) = \frac{1}{1 + e^{+L(u_k)}} \qquad 6\text{-}4$$

$$= \frac{e^{-L(u_k)}}{1 + e^{-L(u_k)}}$$

And hence we can write the probability of the data bit value as,

$$P(u_k = \pm 1) = \left( \frac{e^{-L(u_k)/2}}{1 + e^{-L(u_k)}} \right) . e^{\pm L(u_k)/2} \qquad 6\text{-}5$$

Where, the bracketed term is the same for $u_k = +1$ or $u_k = -1$.

The conditional LLR $L(u_k \mid y_k)$ is defined as:

$$L(u_k \mid y_k) = \ln \left( \frac{P(u_k = +1) \mid y_k}{P(u_k = -1) \mid y_k} \right) \qquad 6\text{-}6$$

where $y_k$ is the received parity information.

Incorporating the code's trellis, this may be written as,

$$L(u_k) = \ln\left(\frac{\sum_{s+} p(s_{k-1} = s', s_k = s, y) / p(y)}{\sum_{s-} p(s_{k-1} = s', s_k = s, y) / p(y)}\right) \quad 6\text{-}7$$

where $s_k \in S$ is the state of the encoder at time instance $k$, $S^+$ is the set of ordered pairs $(s', s)$ corresponding to all state transitions $(s_{k-1} = s') \rightarrow (s_k = s)$ caused by data input $u_k = +1$ and $S^-$ is similarly defined for $u_k = -1$ [82].

From the BCJR (named after its inventors: Bahl, Cocke, Jelinek and Raviv) algorithm, $p(s', s, y) = p(s_{k-1} = s', s_k = s, y)$, is computed as,

$$p(s', s, y) = \alpha_{k-1}(s') \cdot \gamma_k(s', s) \cdot \beta_k(s) \quad 6\text{-}8$$

where, $\alpha_k(s)$ and $\beta_k(s)$ are the forward and backward trellis state probabilities respectively, and $\gamma_k(s', s)$ is the state transition probability, defined as follows:

$$\alpha_k(s) = \alpha_{k-1}(v_1)\gamma_k(v_1, s) + \alpha_{k-1}(v_2)\gamma_k(v_2, s) \quad 6\text{-}9$$

$$\beta_k(s) = \beta_{k+1}(w_1)\gamma_{k+1}(w_1, s) + \beta_{k+1}(w_2)\gamma_{k+1}(w_2, s) \quad 6\text{-}10$$

$$\gamma_k(s', s) = P(s \mid s')p(y_k \mid s', s)$$

$$= P(u_k)p(y_k \mid u_k) \quad 6\text{-}11$$

where $v_1$ and $v_2$ denote valid previous states and $w_1$ and $w_2$ denote valid next states for the current state.

It is recalled here that in DVC, the parity stream is transmitted over a noisy wireless channel and the systematic stream is taken from the locally generated side information frame which could be modelled using a hypothetical additive white noise (AWN) channel. In a practical deployment, the side information also would inherit some properties of the wireless transmission channel, inherited through the intra-decoded key frames used in the side

information estimation process. However, in this discussion, this noise contribution is assumed to be suppressed by a channel coder incorporated to the H.264/AVC intra coding based key frame transmission system. This matter is open for further research. Then, the noise properties in parity and systematic components of the input to the turbo decoder could be considered as independent, thus forming the *dual channel* model. In order to cater for this independent channel scenario, $p(y_k \mid u_k)$ in equation 6-11 can be written as:

$$p(y_k \mid u_k) = \left( \frac{P(y_k^p \mid u_k^p = +1).P(y_k^s \mid u_k^s = +1)}{P(y_k^p \mid u_k^p = -1).P(y_k^s \mid u_k^s = -1)} \right) \qquad 6\text{-}12$$

where $y_k = \{y_k^p, y_k^s\}, u_k = \{u_k^p, u_k^s\}$ are received and transmitted parity and systematic information respectively. The noise modelling of received parity, and systematic information are defined in the following sub-sections.

**Parity Noise Model**

For the parity bit sequence received through a multipath fading wireless channel, the probability of the received bit $y_k^p$ conditioned on the transmitted bit $u_k^p$ can be written as:

$$p(y_k^p \mid u_k^p = +1) = \frac{E_b}{\sqrt{2\pi\sigma^2}} e^{\frac{-[y_k^p - \sum_{i=1}^{N} |a_i|^2]^2}{2\sigma^2}} \qquad 6\text{-}13$$

and,

$$p(y_k^p \mid u_k^p = -1) = \frac{E_b}{\sqrt{2\pi\sigma^2}} e^{\frac{-[y_k^p + \sum_{i=1}^{N} |a_i|^2]^2}{2\sigma^2}} \qquad 6\text{-}14$$

where $i$ is the path index in the multipath fading model, $a_i$ is the complex fading coefficient for each path, $\sigma$ is the standard deviation in the Gaussian probability distribution, $E_b$ is the energy per bit, and $N$ is the total number of fading paths.

### Systematic Noise Model

The systematic bit stream at the decoder is constituted by the side information. The correlation of the side information compared to the original Wyner-Ziv bit stream is usually modelled as Laplacian distributed.

$$p\left(y_k^s \mid u_k^s = +1\right) = \frac{\alpha}{2}\exp(-\alpha \mid y_k^s - 1 \mid) \qquad 6\text{-}15$$

$$p\left(y_k^s \mid u_k^s = -1\right) = \frac{\alpha}{2}\exp(-\alpha \mid y_k^s + 1 \mid) \qquad 6\text{-}16$$

with variance of $2/\alpha^2$. Even though the channel noise effects on the intra coded key frames would necessarily have an effect on the above noise model, it is assumed that this is sufficiently suppressed by a channel coder (note however, that a channel coder is not necessary for the Wyner-Ziv bitstream as in the proposed solution) and any residual effect is absorbed into the above noise model.

### Composite Dual Channel Noise Model

Substituting the Equations 6-13 - 6-16 in Equation 6-12 and taking logarithm of both sides:

$$L(y_k \mid u_k) = \ln\left\{\left(\frac{\exp\left(-\frac{E_b}{2\sigma^2}\left(y_k^p - \sum_{i=1}^{N} \mid a_i \mid^2\right)^2\right).\exp\left(-\alpha\left(y_k^s - 1\right)\right)}{\exp\left(-\frac{E_b}{2\sigma^2}\left(y_k^p + \sum_{i=1}^{N} \mid a_i \mid^2\right)^2\right).\exp\left(-\alpha\left(y_k^s + 1\right)\right)}\right)\right\} \qquad 6\text{-}17$$

$$= \frac{E_b}{2\sigma^2} 4a.y_k^p + \alpha\left(\mid y_k^s + 1 \mid - \mid y_k^s - 1 \mid\right)$$

where, the term: $4a\frac{E_b}{2\sigma^2}$ is usually labelled as $L_c$, or the channel reliability. This depends on the channel SNR and the fading amplitude, both of which are derived from channel estimation performed on the wireless channel.

Now, the equations are ready for use in the calculation of the branch metric, $\gamma_k(s',s)$ in Equation 6-11, for the state transition $\left(s_{k-1} = s'\right) \rightarrow \left(s_k = s\right)$ on the trellis. Then the forward

and backward path probabilities $\alpha_k(s)$ and $\beta_k(s)$ are computed using Equations 6-9 and 6-10 for the current state $s$, in the MAP decoding algorithm [81].

## 6.3 Simulations and Results

The performance of the DVC codec with the proposed enhancements is tested for standard video sequences. The simulation system with the incorporated channel modulation and demodulation functions is introduced with the test conditions first, in the next sub-section. The results and a discussion will follow.

### *6.3.1 Simulation System Configuration*

The considered pixel domain DVC codec structure with the additional components for simulating the WCDMA wireless system is presented in this section. A block diagram of the system is shown in Figure 1.

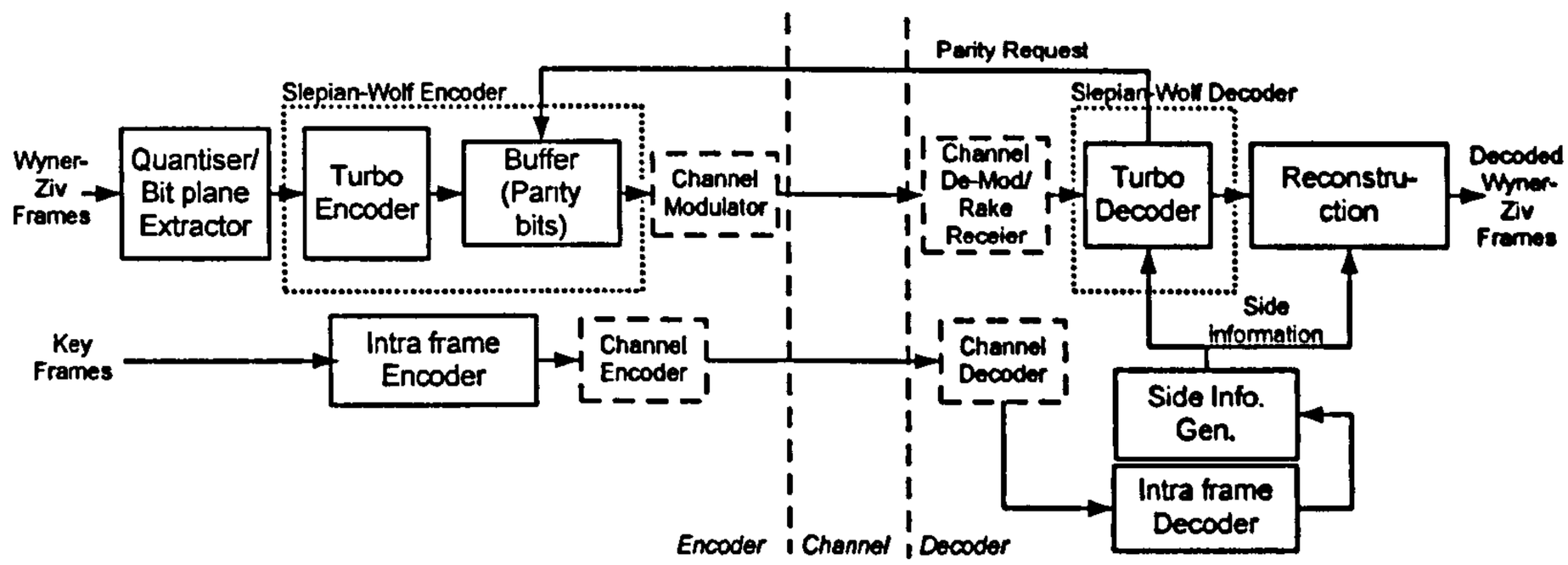



Figure 6-2 The system block diagram with the channel transceiver

The DVC encoder generates the parity bit stream and stores in the buffer as detailed earlier in chapter 2. Upon dynamic requests from the decoder, received over the reverse feedback channel, the parity bits are transmitted through the noisy channel.

**Rayleigh fading channel**

As shown in Figure 6-2, a *quadrature phase shift keying* (QPSK) modulator and a rake receiver are used in the transmitting and receiving nodes of the wireless channel, respectively. In wireless communications it is common to use a discrete time Rayleigh fading channel model consisting of multiple narrow-band channel paths separated by delay elements as illustrated in Figure 6-3.

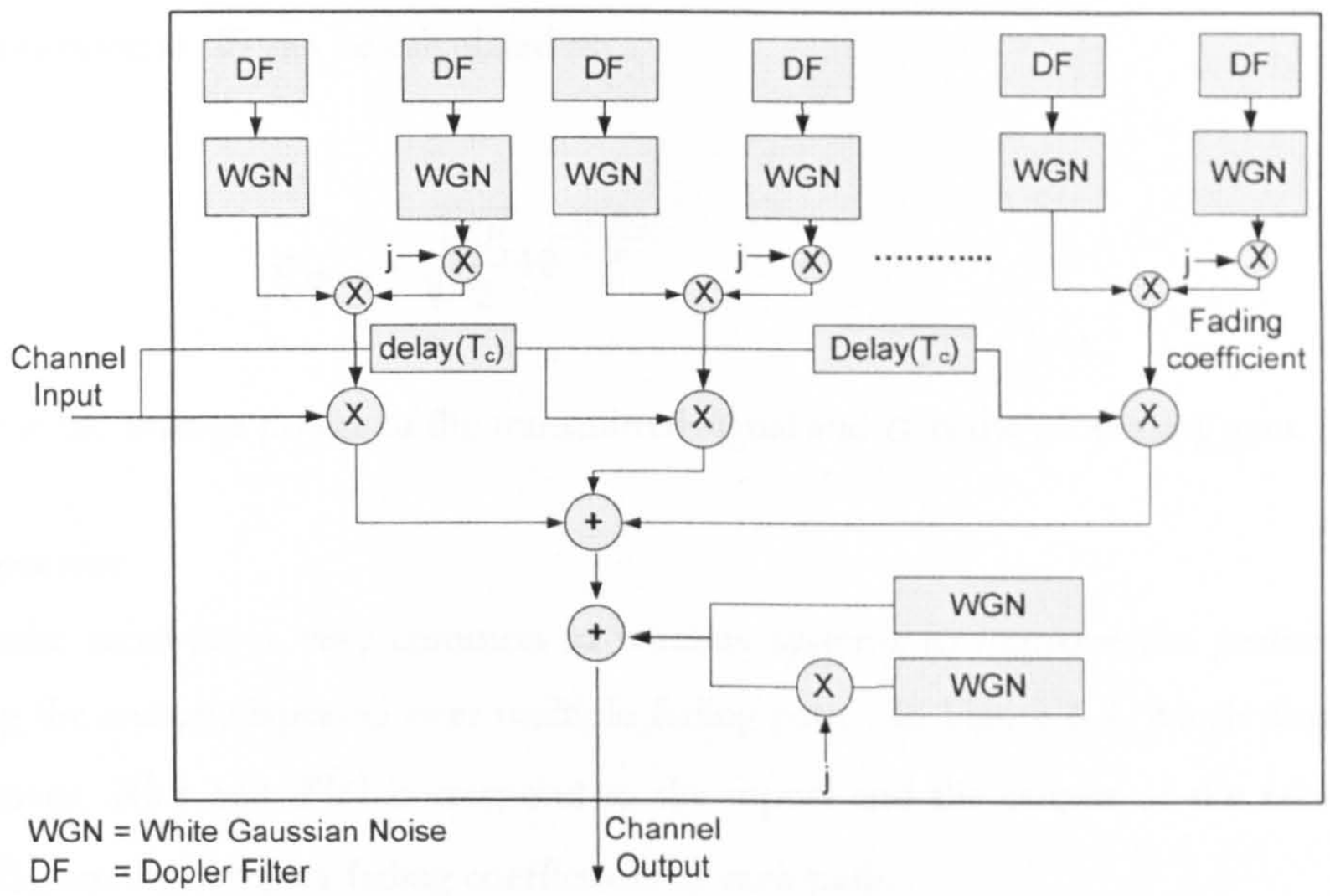



Figure 6-3 Rayleigh fading channel model

The number of taps has to be decided based on the delay spread in the radio environment. The values of variances of each Gaussian noise source in each tap are selected according to the exponential power delay profile. The average power of each path is given by,

$$P(n) = \frac{e^{-\frac{nT_c}{T_d}}}{\sum_{n=0}^{L-1} e^{-\frac{nT_c}{T_d}}} \qquad 6\text{-}18$$

where $T_d$ is the decaying factor, $T_c$ is the path delay, and $L$ is the total number of paths in that fading model. The standard deviation $\sigma_n$ of each Gaussian noise source of the $n^{th}$ tap can be found using the following expression:

$$\sigma_n = \sqrt{\frac{P(n)}{2}} \qquad n \in \{0, \cdots, L-1\} \qquad 6\text{-}19$$

The complex baseband *additive white Gaussian noise* (AWGN) samples are generated as shown in Figure 6-3, using two Gaussian noise sources each having zero mean and standard deviation $\sigma_{AWGN}$. The value of $\sigma_{AWGN}$ for specific values of the signal to noise ratio (SNR) $E_b/N_0$, measured in dB can be calculated as:

$$\sigma_{AWGN} = \sqrt{\frac{GP_t}{2}} 10^{-\frac{(E_b/N_0)}{20}} \qquad 6\text{-}20$$

where $P_t$ is the average power of the transmitted signal and $G$ is the processing gain.

**Rake Receiver**

Use of rake receivers is very common in wireless systems to improve the performance by collecting the energy dispersed over multiple fading paths. In Figure 6-4, which illustrates the rake receiver, $R(t)$ and $Y(t)$ correspond to the inputs and the output of the rake receiver. $G_1, G_2 \ldots G_N$ are the complex fading coefficients of each path.

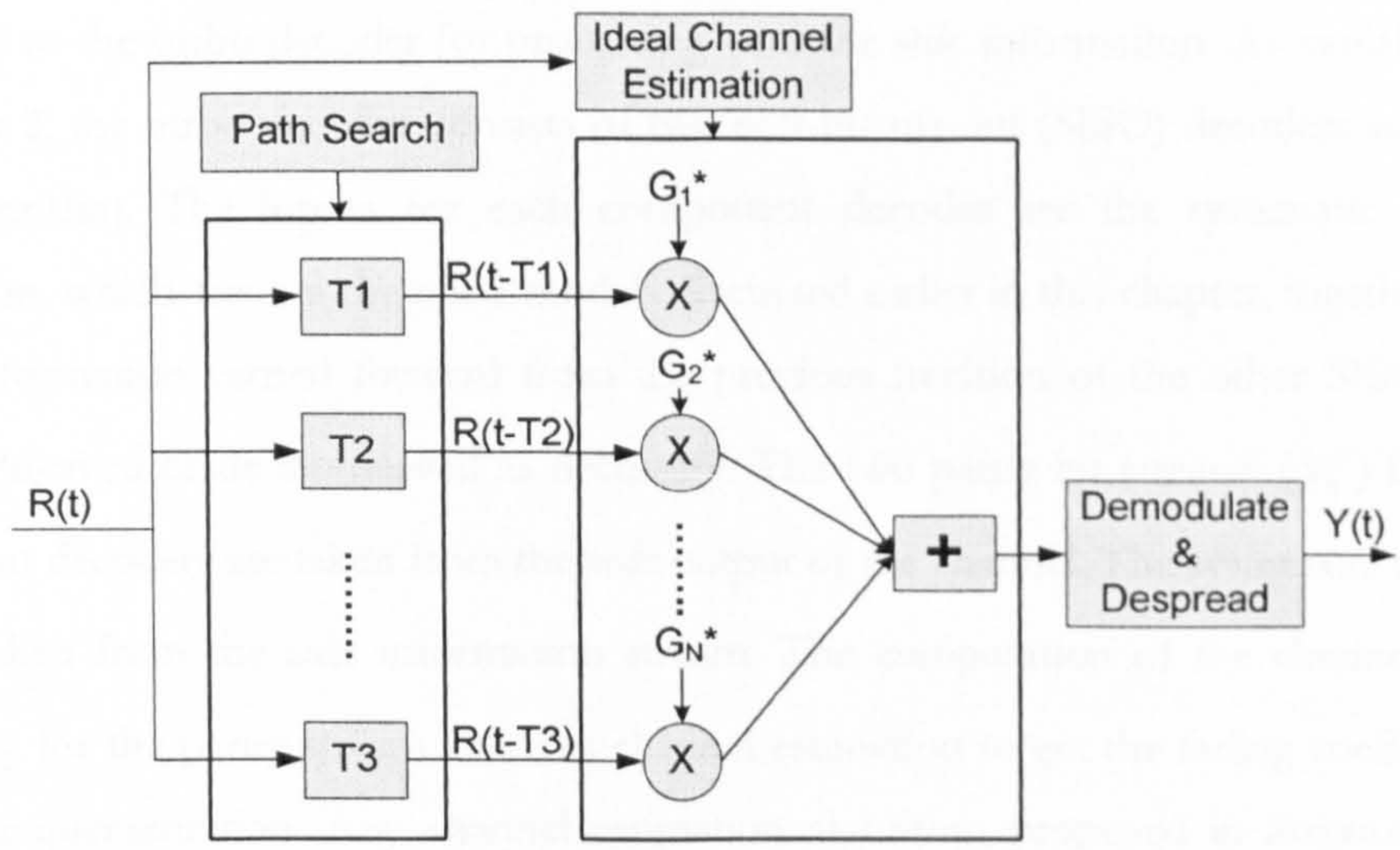



Figure 6-4 Block diagram of the rake receiver

The output of the rake receiver $Y(t)$, with respect to signals received from $N$ fading paths is computed as:

$$Y(t) = \sum_{i=1}^{N} G_i^* R_i \qquad 6\text{-}21$$

Where

$$R_i = \sum_{j=1}^{N} \left( G_j X(t+2-i-j) \right) + n(t) \qquad \forall i \in \{1,2,...,N\} \qquad 6\text{-}22$$

thus,

$$Y(t) = \left( \sum_{i=1}^{N} |G_i|^2 \right) X(t) + \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} G_i^* \{ G_j X(t+2-i-j) + n(t) \}$$

$$Y(t) = \left( \sum_{i=1}^{N} |G_i|^2 \right) X(t) \quad + \quad n'(t) \qquad 6\text{-}23$$

where $X(t)$ is the transmitted symbol, $G_i^*$ represent the complex conjugate values of the fading coefficient of path index $i \left( i \in \{1,2,..,N\} \right)$ and $n'(t)$ is the composite noise in the rake receive output.

Then, the QPSK demodulator computes the soft channel output for the parity bit stream to be forwarded to the turbo decoder for processing with the side information. As explained earlier in chapter 2, the turbo decoder consists of two soft-in-soft-out (SISO) decoders adopting the MAP algorithm. The inputs for each component decoder are the systematic and parity information, which assume the noise models discussed earlier in this chapter, together with the *a priori* information carried forward from the previous iteration of the other SISO decoder, either interleaved or de-interleaved as necessary. The two parity bit streams ($y_k^p$) for the two component decoders are taken from the soft output of the channel. The systematic component ($y_k^s$) is taken from the side information stream. The computation of the channel reliability factor ($L_c$) for the parity stream involves channel estimation to get the fading coefficients in a practical implementation. Any channel estimation algorithm proposed in literature, such as [83], can be utilized for this purpose.

**Wireless Channel Types**

Two variations of the transmission channel models are considered in this discussion: (a) Additive White Gaussian Noise (AWGN) Channel and (b) W-CDMA Wireless Channel. It is noted that, the AWGN channel model is a special case of the wireless channel model discussed in section 6.2 with no time varying multipath fading; hence unity fading amplitude (a = 1). The configuration parameters for each of above are discussed later in this section.

**Side information generation**

In the simulation system, the decoder generates the side information using the previously decoded key frames and the Wyner-Ziv frames of the sequence using a 3-dimensional motion compensation and extrapolation (MC-E) algorithm with bitplane-wise improvement proposed in [29]. Using forward frame extrapolation enables low delay real time implementation of the codec.

**Test Conditions**

- Video sequences: *Foreman* (medium/high motion), *News* (low motion); QCIF (176×144); YUV 4:2:0; 100 frames; Wyner-Ziv frame rate: 15fps

- GOP sizes: 2 and 12 used in simulations

- Bit rate and PSNR calculated and displayed only for Wyner-Ziv frames, and averaged over the sequence.

- PSNR is calculated for the luminance signals only.

- Each rate-distortion point is achieved for pixel domain quantiser parameter: $M$, where $2^M \in \{2,4,8,16..\}$ defines the number of quantization levels.

- H.264/AVC codec is simulated in IPIP GOP structure (JM10.1/Main profile) for identical channel conditions. This frame pattern is comparable with the DVC frame sequence, when only forward motion extrapolation is used in the side information estimation. *Frame Copy* error concealment algorithm is incorporated for whole frame losses [84]. Accordingly, when a frame is corrupted due to transmission noise, it is concealed by copying each pixel value from the corresponding pixel of the previous

decoded reference frame. Partial losses were concealed using the default means in the reference codec [84].

- AWGN and W-CDMA channels are used in the simulation. Typical parameter values are considered for the W-CDMA wireless channel as summarized in Table 6-1 [85]. Practical variations of these are noted to make no significant effect on the comparative performance enhancement of the proposed decoding algorithm.

Table 6-1 Configuration Parameters of the W-CDMA Channel

| Channel model | W-CDMA uplink and downlink channels (slowly-varying frequency-selective fading) |
|---|---|
| Modulation | QPSK |
| Chip rate | 3.84 Mchip/s |
| Spreading factor | 32 |
| Spreading sequence | OVSF sequences |
| Carrier frequency | 2 GHz |
| Doppler speed | 100 km/h |

- The signal to noise ratio (SNR) of the transmission channel traversed by the parity bit stream has been varied in the range 6~16 dB as appropriate.

- Channel coding has not been used for both Wyner-Ziv bit stream and H.264/AVC encoded bit stream, so that a reasonable R-D performance and error resilience analysis of the two codecs could be performed.

### *6.3.2 Analysis of Results*

The simulation results are presented and analysed in this section.

**AWGN Channel**

The results for the Foreman and News sequences are shown in Figure 6-5 and Figure 6-6, for a GOP size of 2. The performances of the two codecs are plotted for different channel conditions, where SNR$\in \{6,8,10,12\}$dB. It is observed that the DVC codec with the proposed modification to the decoding algorithm has outperformed the considered H.264/AVC codec when used with AWGN channels below 8dB SNR. It is observed that the H.264/AVC codec relies significantly on the error concealment algorithm when high packet drop occurs at low SNR levels. Notably, the proposed DVC codec does not involve an explicit error concealment algorithm, yet it demonstrates strong error resilience characteristics evident by the remarkably stable performance as shown in the figures.

There are several key observations that can be made with the simulation results. The proposed DVC codec has a consistent performance in the SNR range 6dB to 12dB; in fact the performance saturates at 10dB SNR level. The H.264/AVC coder has a superior performance at very high SNR levels, but its performance drops notably with the increased BER of the bit stream received over the channel.

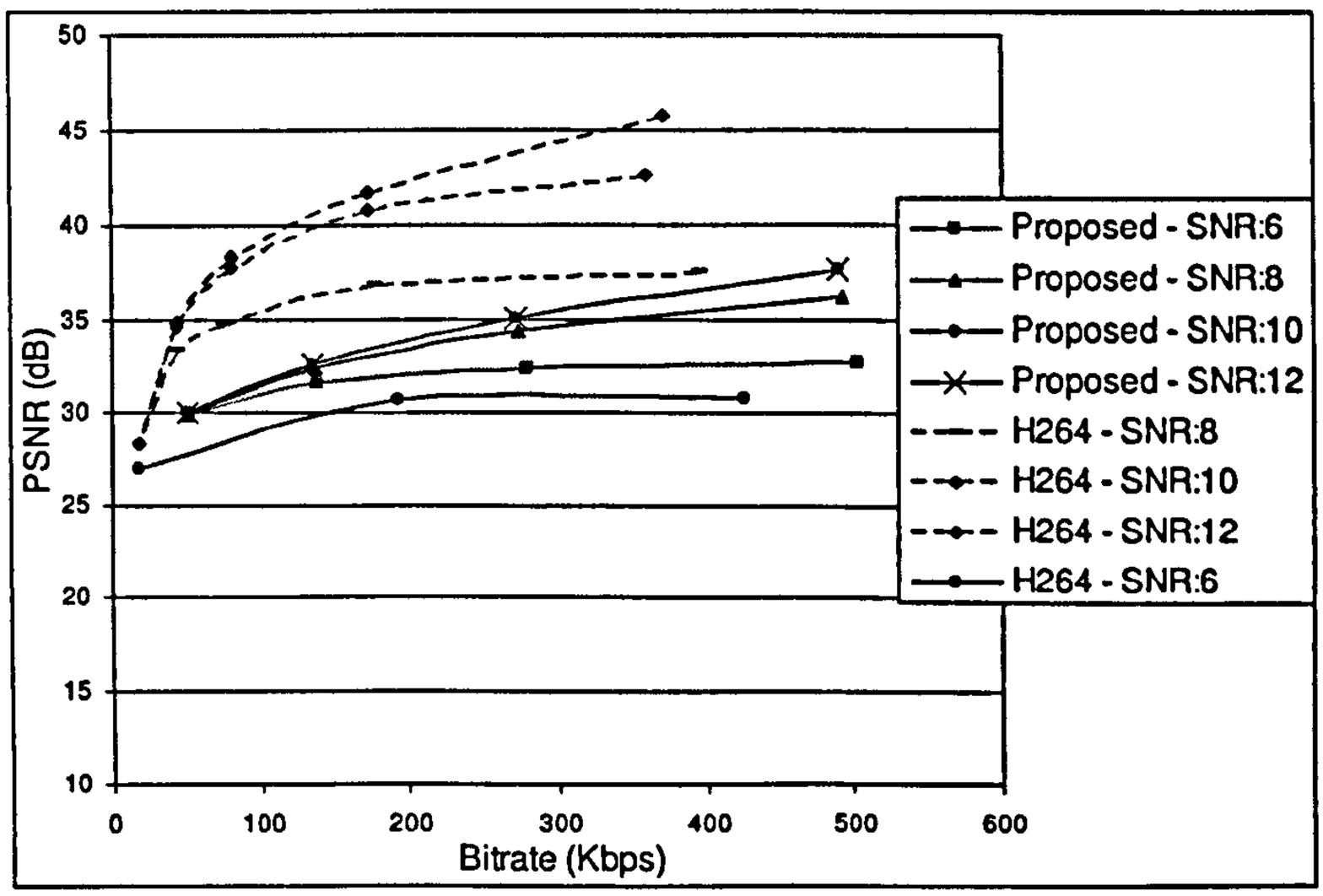



Figure 6-5 Performance comparison for the *Foreman* test sequence for the AWGN channel

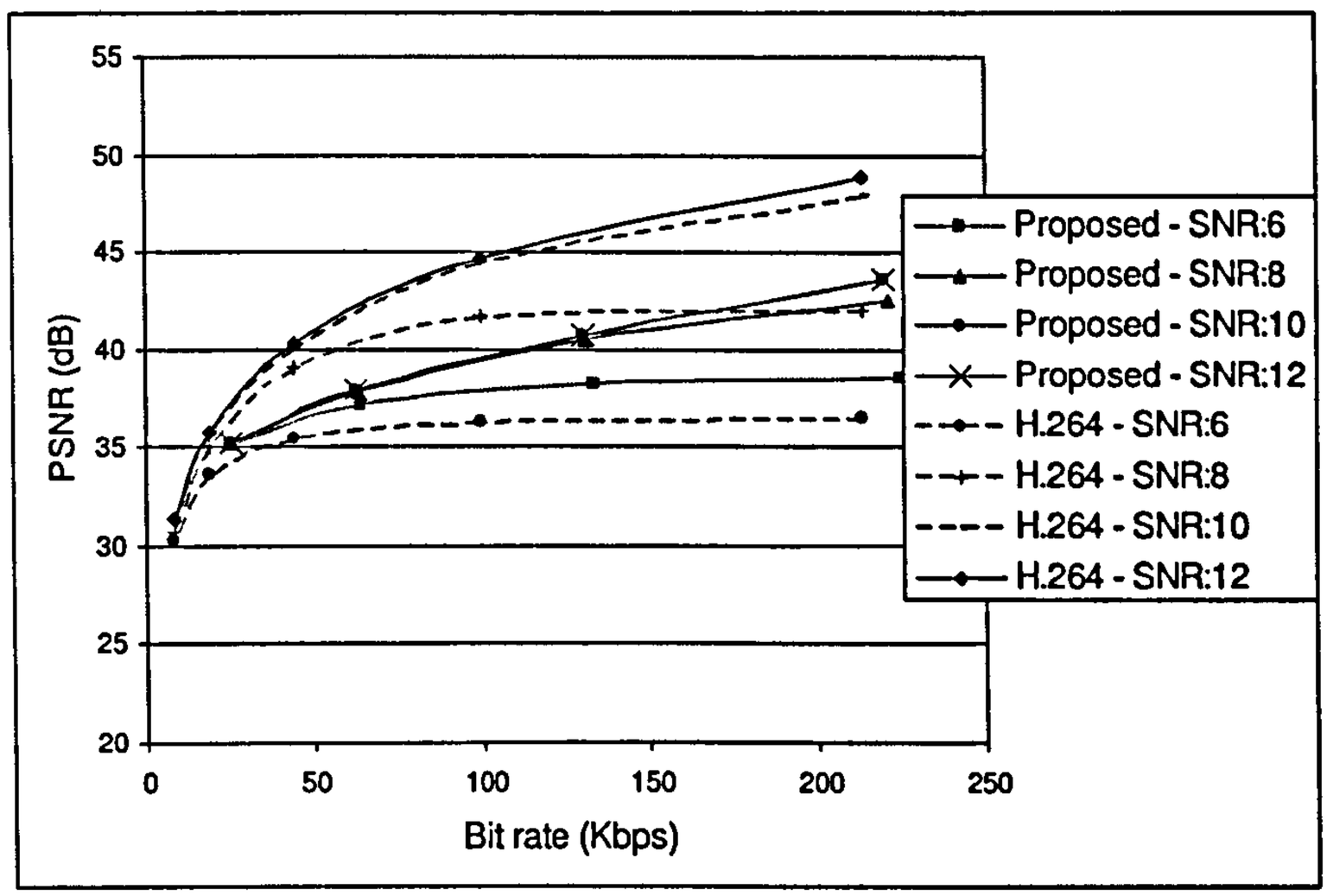



Figure 6-6 Performance comparison for the *News* test sequence for the AWGN channel

**W-CDMA Channel**

The proposed DVC codec is tested on a simulated W-CDMA channel for the same video sequences for GOP sizes of 2 and 12. The results are illustrated in figures below. Figure 6-7 and Figure 6-8 show the performance comparison for GOP size of 2. Figure 6-9 and Figure 6-10 depict the performance for a larger GOP size (12) with and without motion estimation for the H.264/AVC codec.

***GOP size: 2***

In this scenario, again, the proposed codec demonstrates very good and steady performance over the tested SNR range of 10dB to 16dB. As expected, the PSNR gradually increases with the increase of SNR for a given bit rate. The H.264 codec demonstrated superior results only at very high channel SNR levels, typically above 16dB, depending on the statistics of the frame sequence. This observation is in agreement with the prevailing H.264 – DVC performance gap in ideal noise free channel scenario [11]. However, at channel SNR of 16dB and below, the DVC performance is very stable and significantly outperform the H.264/AVC – P frames.

***GOP size: 12***

Increasing the GOP size is a preferred choice in practical video codec implementations so that the frequency of intra coded frames can be reduced as intra-coding is naturally an expensive

process. The simulation results show that even though the R-D performance of DVC codec drops significantly when the GOP size is increased, the relative error resilience performance compared to conventional video coding remains high.

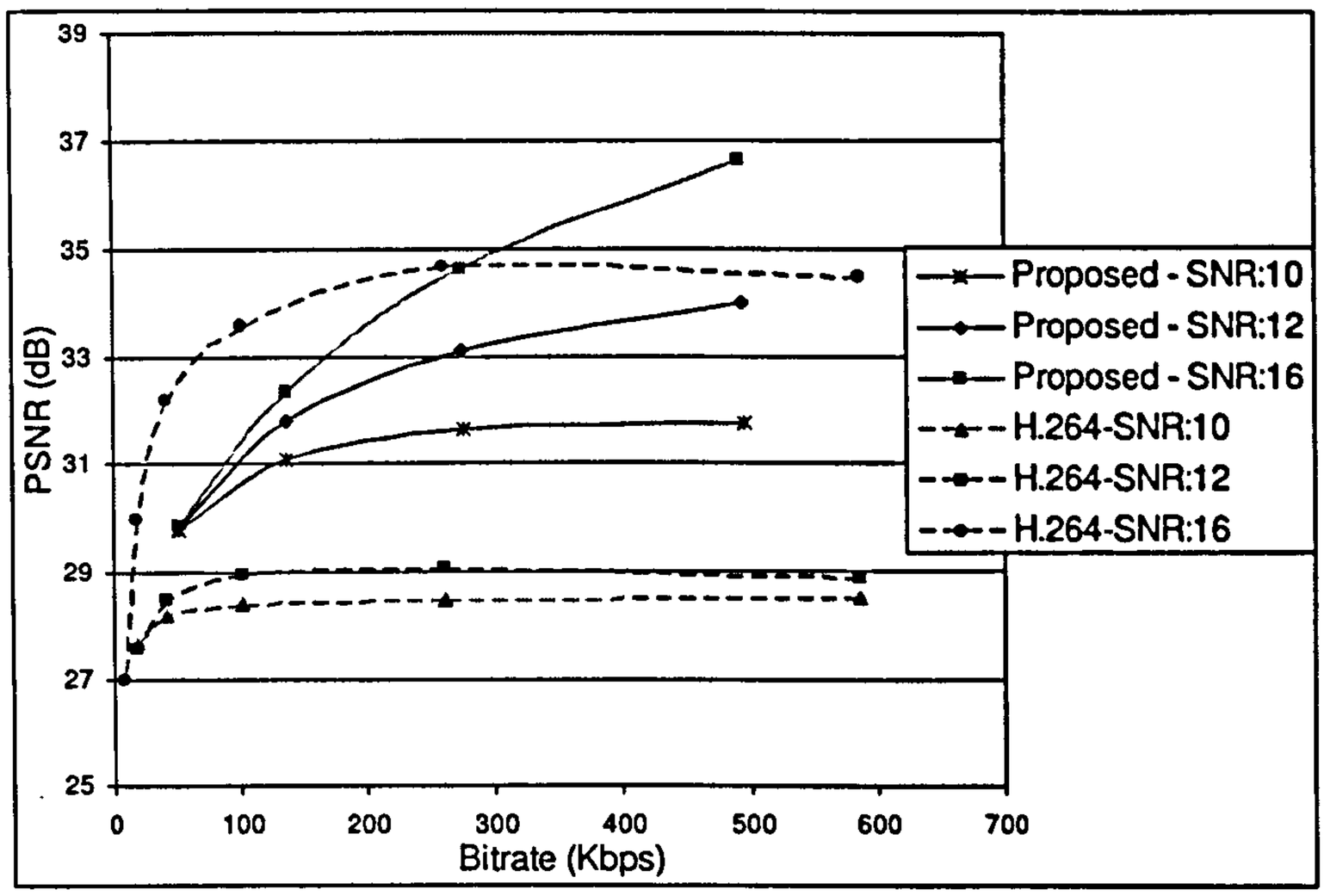



Figure 6-7 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, GOP=2

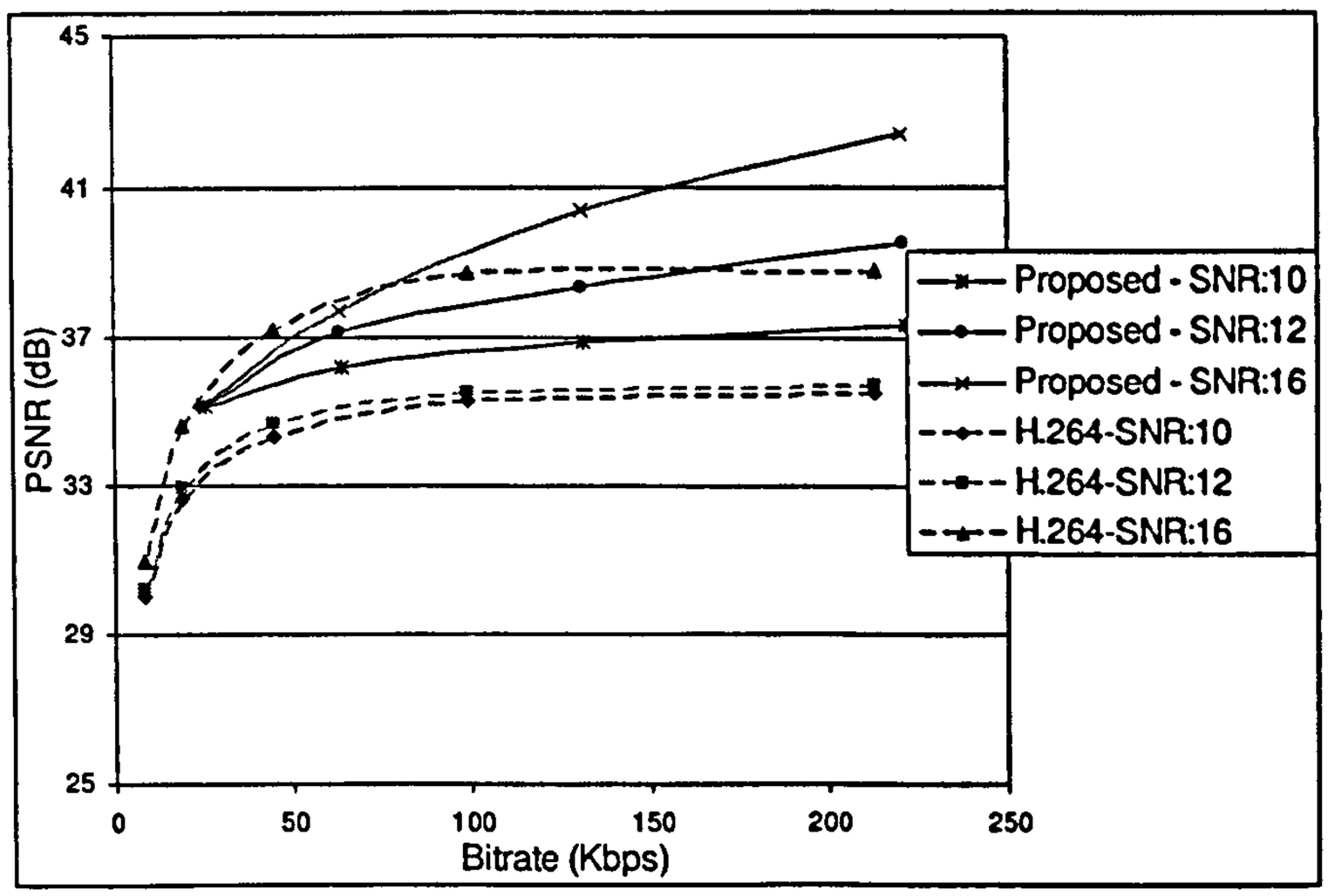



Figure 6-8 Performance comparison for the *News* test sequence for the W-CDMA channel, GOP=2

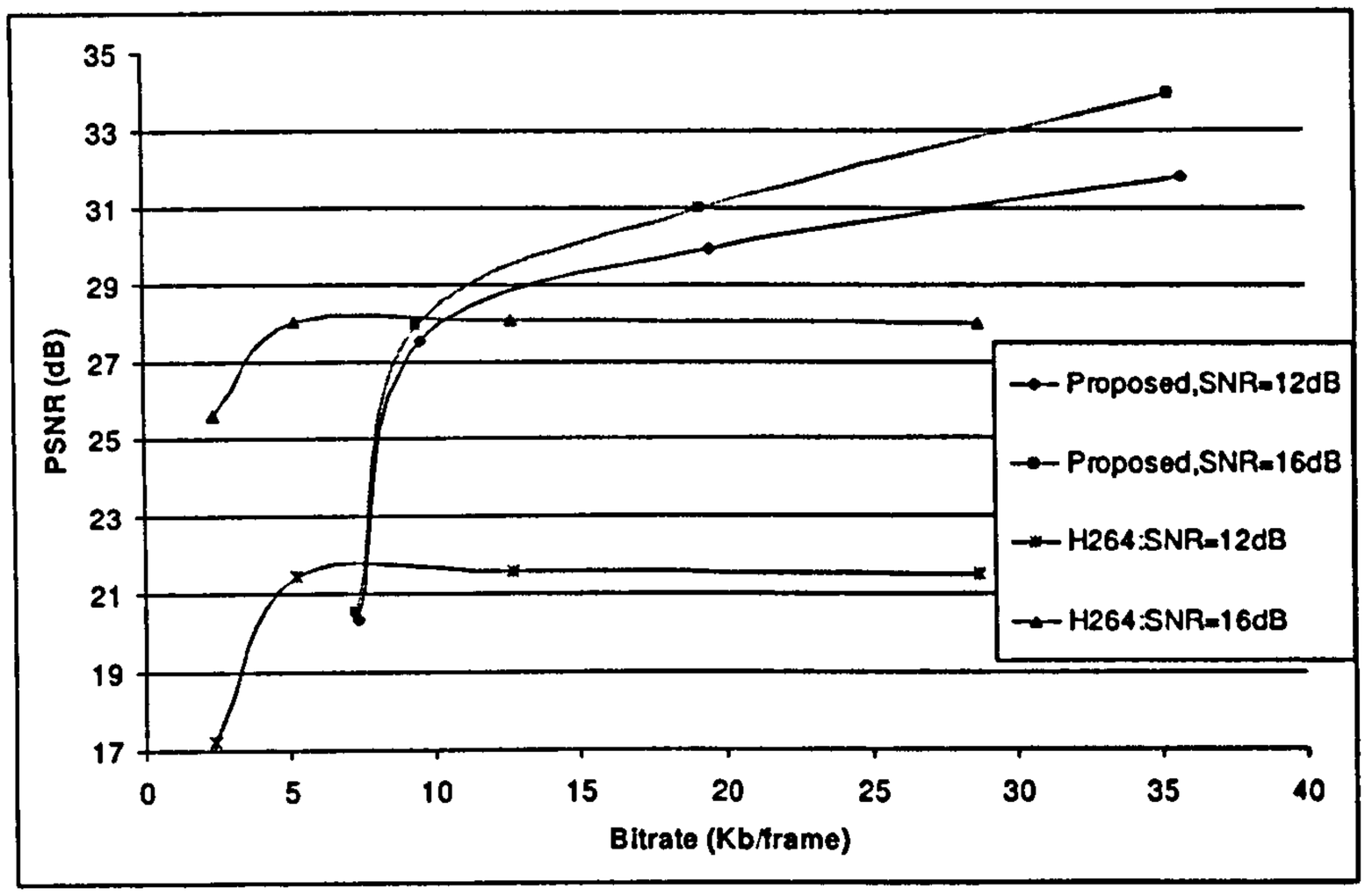



Figure 6-9 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, GOP=12

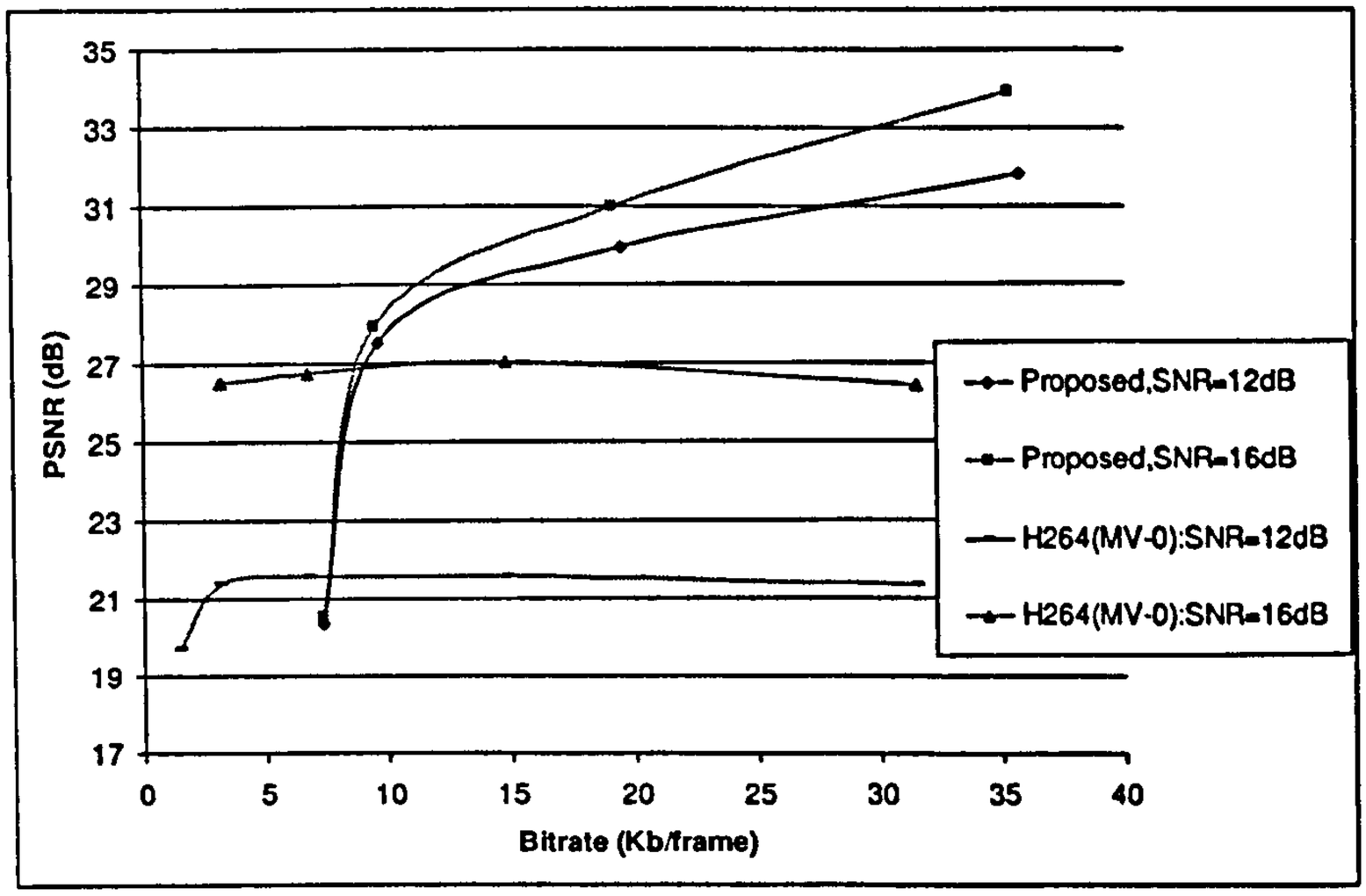



Figure 6-10 Performance comparison for the *Foreman* test sequence for the W-CDMA channel, with motion search disabled for H.264/AVC, GOP=12

**Comparison with the existing decoding algorithm**

In this section, the performance of the proposed algorithm is compared with the existing decoding algorithm [29] operating over a noisy wireless channel. Figure 6-11 depicts the comparison for the Foreman sequence. It is evident that the proposed algorithm produces an improvement in RD performance by reducing the bit rate for a given PSNR. A change in PSNR is not evident since the decoding uses the same side information and the Slepian-Wolf decoder runs against the identical error threshold in the feedback loop, which are the determining factors for the PSNR. Also note that the evident reduction in bit rate is more significant at lower SNR values in the channel. This observation is explained by the higher noise level in the soft output of the channel, which creates a larger performance drop to be lifted back by the proposed algorithm. At high SNR levels, typically above 16dB in this case, the channel output has a very low noise level so there is not much scope for the proposed algorithm, thus the performance is saturated.

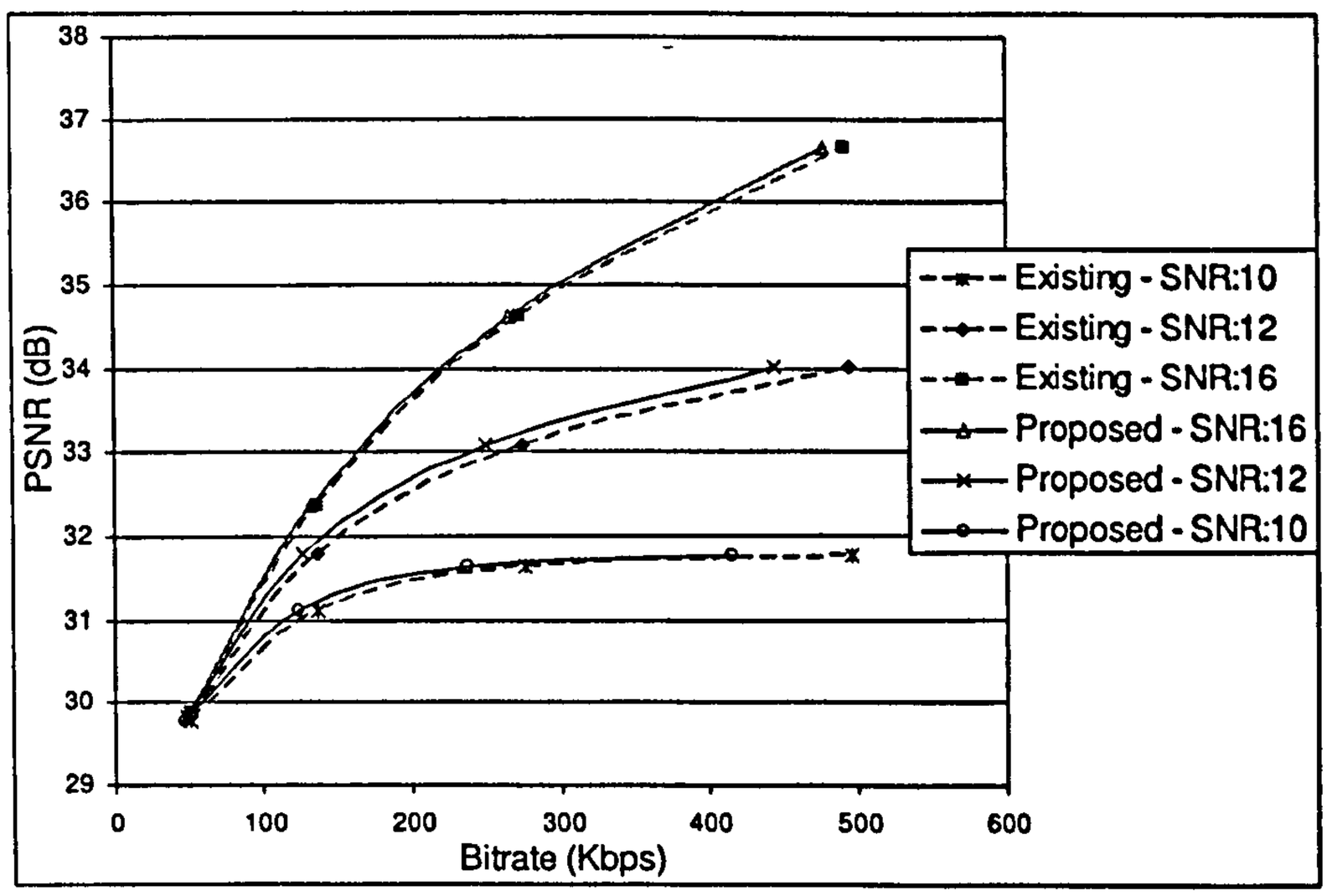



Figure 6-11 Performance comparison of the proposed technique with the existing decoding algorithm (*Foreman*, W-CDMA channel , GOP=2)

**Complexity Analysis**

It is noted that the performance comparison between the proposed DVC codec and H.264/AVC codec is carried out in the light of a significant difference in the computational complexities in the two encoders. The considered H.264/AVC encoder demands higher

computational power due to the motion estimation process involved therein. Suppressing the motion search range enables a more reasonable comparison in terms of the encoder complexity. However, suppressing the motion search affects to further deteriorate the H.264/AVC performance as evident from Figure 6-10.

The DVC codec is optimized for a set of applications which desperately demands for a low complex encoder. However, due to the lack of DVC codecs tested for noisy channels available in literature for comparison, state-of-the-art H.264/AVC was used for comparison. This proves that, with the proposed modifications, DVC codec demonstrates strong error resilience characteristics to be utilized in video communications over noisy wireless channels.

## 6.4 Conclusion

DVC is a technology that is largely proposed for video surveillance and like applications which are usually deployed over wireless transport networks. Therefore, the performance of the codec when the compressed bit stream is transmitted over noisy wireless channels becomes a significant consideration in the deployment of video solutions. Accordingly, this chapter presented an investigation of the performance of DVC codec over wireless channels. It was noted that the simulation system with a pixel domain DVC codec and a wireless channel resembled a unique noise scenario when the hypothetical channel model within the DVC codec is also collectively considered. To closely represent this scenario, a novel dual-channel noise model was derived. Then the turbo decoding algorithm within the Slepian-Wolf decoder was modified incorporating the novel noise model.

It was observed that the proposed noise model and the associated decoder modification improve the rate distortion performance of the codec when compared with a similar system without the modifications over a wireless channel with different channel conditions. Further, the modified DVC codec proved to even outperform the state-of-the-art H.264/AVC codec when tested over harsh channel conditions when an explicit channel coder (which is an additional bandwidth and complexity burden) is not incorporated.

*Chapter 7*

# CONCLUSIONS

DVC is an emerging video coding technology that is built upon the concept of distributed source coding. The objective of the thesis is to bring forward the DVC paradigm towards a practical realization, ready for standardizations and commercial developments. In order to achieve this, the scope of this research is focused on a number of aspects of the DVC coding framework:

- Enhancement of the performance and robustness of the DVC codec
- Design a unidirectional coding framework for DVC to suppress the dependence on a reverse channel
- Optimise the performance of the codec for utilization over error prone wireless channels

The specific technical contributions presented in this thesis in order to achieve the above objectives are summarized below.

## 7.1 Technical contributions of the thesis

### *7.1.1 Improving the Slepian-Wolf coding performance with TTCM*

The existing DVC proposals are dominantly based on turbo coding in the Slepian-Wolf codec. In this regard, a novel approach was proposed in Chapter 3 of this thesis by introducing the TTCM concept into DVC. This proposal is motivated by the high coding gain of TTCM. In designing the TTCM based Slepian-Wolf codec, notably, the TTCM encoder was arranged so that the parity bit stream generated at the encoder is punctured and transmitted as a bit stream to the Slepian-Wolf decoder for mapping into the complex symbols together with the side information. A set partitioning technique was utilised in mapping to the signal constellation for optimising the decoding performance.

The performance of the proposed TTCM based codec is compared with the more conventional turbo coding based approach. The simulation results depict a significant rate distortion performance gain in using the proposed approach in the codec.

### *7.1.2 Unidirectional coding framework development*

It is understood that the involvement of a reverse feedback channel in the common Stanford DVC framework could incur a practical burden in the targeted application scenarios as discussed in detail in Chapter 4. However, this feedback loop plays a critical role in determining the optimum bit rate for the Wyner-Ziv bit stream by facilitating the parity puncturing process. With the removal of reverse channel, the main challenge is in determining the optimum parity patterns. On this backdrop, two approaches were considered in the Chapter 4:

1. Encoder rate control: perform rate decisions based on correlation noise estimation at the encoder. However, this introduces an additional complexity burden to the DVC encoder, which is otherwise extremely low complex.

2. Refinement of decoded frames: improve the picture quality of the decoded Wyner-Ziv frames using iteratively decoding at no additional encoder complexity or additional bit rate burden.

Innovative algorithms for both approaches above were presented in this thesis. A pixel interleaving technique and a very low complex correlation noise estimation algorithm were proposed for ERC. An iterative side information refinement algorithm was proposed for the decoded frame enhancement by exploiting the spatial and temporal correlation in the decoded frame sequence.

The performances of the presented algorithms were tested for standard sequences. The simulation results were compared for both objective and subjective video quality and compression efficiency with alternate approaches with no reverse channels. Further technical analyses were provided to illustrate the effectiveness of the individual algorithms.

### *7.1.3 Enhanced functional modules for the DVC codec*

A review of the related literature proves that the compression efficiency of the existing DVC implementations lags behind the state of the art in video coding, namely H.264/AVC, by a notable margin. Even though DVC is not proposed as a cross-platform competitor to the

conventional video coding technologies; but a complementary technology targeted at specialised applications, improving the compression efficiency is vital in pushing DVC paradigm forward towards commercial developments and standardisations. In the light of above, Chapter 5 considers a number of aspects of DVC and proposes innovative algorithms to enhance the performance.

**Non-linear quantisation**

In this section, a non-linear quantisation algorithm was presented for DVC in contrast to the linear scalar quantiser traditionally utilised in the DVC designs. The concept was motivated by the contrastingly high concentration of information around the zero-bin when the Wyner-Ziv residual signal is considered. In this regard, the proposed algorithm minimises the quantisation noise as discussed in section 5.2 and thus helps in improving the rate-distortion performance of the codec. The optimum non-linear transformation function was experimentally determined by evaluating a number of candidate formulas.

The performance gain of the non-linear quantisation algorithm was evaluated against the state-of-the-art pixel domain video codec that uses a linear quantiser. Accordingly, a very significant PSNR gain was evident at a given bit rate.

**Enhanced reconstruction**

In section 5.2, an enhanced reconstruction algorithm was presented focussing on the unidirectional DVC framework discussed earlier in Chapter 4. It was noticed that, when the unidirectional architecture is considered, the decoded bit stream tends to contain residual bit errors due to the possibly sub-optimal parity puncturing. These errors are reflected in the reconstructed frame as irritating artefacts. A solution to this problem is proposed by evaluating the error probability of each bit in the decoded bit stream and accordingly modifying the reconstruction function by biasing the reconstructed pixels towards side information. A *recon angle* parameter in defined and a systematic training process is introduced to derive the optimum recon angle corresponding to the bit error rate in the decoded bit stream.

The enhanced reconstruction algorithm was simulated for a number of video sequences and the test results were compared with a similar UDVC codec, but involving the existing reconstruction algorithm proposed by Aaron et al. [4] in the initial Stanford DVC proposal. In this case, both pixel domain and transform domain coding frameworks were considered for the

simulations and in all test cases a significant PSNR improvement was observed at no additional bit rate burden to the codec.

**Side information refinement**

Side information is arguably one of the most important components in the DVC codec. Accordingly, the quality of side information, defined in terms of the level of correlation with the original Wyner-Ziv frame (higher the better), has a direct impact on the rate distortion performance of the codec. Considering this significance, a side information refinement algorithm was proposed in section 5.3. The algorithm starts with an *initial side information* generated by any existing algorithm, then keep on refining it iteratively by exploiting the spatial and temporal correlations in the current Wyner-Ziv frame and its neighbour frames.

The side information refinement algorithm, built into the existing reverse channel feedback based pixel domain DVC framework, was tested for standard video sequences and a significant PSNR gain is evident when compared with the existing pixel domain codec that uses the same initial side information. The additional computational complexity involved in the DVC decoder in utilising the iterative refinement algorithm is also quantitatively analysed. It is believed that the increase in the computational cost is justified considering the significant performance improvement, particularly in the context of the target applications of DVC that involve largely shared centralised decoders.

### *7.1.4 Optimising DVC performance over wireless channels*

DVC is a technology that is largely proposed for video surveillance and like applications which are most likely deployed over wireless transport networks. Therefore, the performance of the codec when the compressed bit stream is transmitted over noisy wireless channels becomes a significant consideration in the deployment of video solutions. Motivated by this scenario, an investigation of the performance of DVC codec over wireless channels was presented in 5.4. It was noted that the simulation system with a pixel domain DVC codec and a wireless channel resembled a unique noise scenario when the hypothetical channel model within the DVC codec is also collectively considered. Accordingly, a novel dual-channel noise model was derived to closely represent this video communications system. Then the turbo decoding algorithm within the Slepian-Wolf decoder was modified incorporating the novel noise model.

It was observed that the proposed noise model and the associated decoder modification improve the rate distortion performance of the codec when compared with a similar system without the modifications over a wireless channel with different channel conditions. Further, the modified DVC codec proved to even outperform the state-of-the-art H.264/AVC codec when tested over harsh channel conditions when an explicit channel coder (which is an additional bandwidth and complexity burden) is not incorporated.

## 7.2 General Conclusions

The research pertaining to this thesis covers a time span of three years. It was started based on an ambitious DVC framework that was initially proposed three years earlier in 2002 [4]. The state of the art in DVC was a premature system involving a number of hypothetical models and assumptions in relation to the correlation noise and decoded error rate modelling, key frame coding etc. together with basic side information estimation algorithms. The compression efficiency compared to the conventional video coding was significantly low, albeit simulated on a theoretical base; key frame transmission, for example. On this backdrop, the goal of this research was to push the DVC concept forward towards a realistic practical implementation through increasing the robustness of the codec and improving the compression efficiency exploiting the guidelines set by Slepian-Wolf [2] and Wyner-Ziv [3].

In line with the above goal, there have been significant contributions made in this research as descriptively discussed in the previous section. The effectiveness of the proposed algorithms was verified through simulations. Also, 35 research publications were made in international refereed journals and conference proceedings over the past three years in relation to this activity. As a result of this, and other parallel research in the community, the state of the art in DVC has been enhanced significantly in terms of the practicality of the codec framework and the compression efficiency. Albeit with the promising developments, some further effort is due in realising DVC as a commercially viable technology. The next section presents some possible future work suggested in achieving this goal.

## 7.3 Suggested Future Work

Based on the understanding achieved through this research and review of peer research, there are a number of aspects of DVC identified as having potential for future research. Some of these are discussed below with notes on possible approaches to be considered. This includes

possible extensions to the algorithms proposed in this thesis as well as other significant research areas.

1. UDVC: Unidirectional DVC is likely to be the future path of DVC since the practical hindrances of the feedback framework are increasingly appreciated. Encoder rate control algorithms are already been studied, yet a significant effort is vital in devising a robust coding framework. Further, the decoder based refinement algorithm presented in this thesis could be developed extended to the transform domain and thus could make it play vital role in a future robust UDVC solution.

2. Side information:

    a. Side information generation has always been playing a key role in determining the coding efficiency and will continue to do so. Since the motion estimation and compensation in DVC is itself a challenging task, less straightforward compared to that in conventional coding paradigms, there is always room for improvement. Particularly when the motion level in the video sequence is high, the existing motion interpolation techniques do not perform well. In these cases, spatial prediction exploitation based refinement techniques should be handy tools.

    b. Even though, there are number of literature regarding the transform domain DVC, it is not evident if they employ side information refinement techniques, which should be incorporated in order to increase the rate-distortion performance of transform domain DVC.

    c. In current DVC designs GOP size is mostly set to 2. This is mainly because at higher GOP lengths, side information generation process deteriorates dramatically due to the high temporal distance between reference (key) frames resulting in lack of sufficient information for motion estimation. Therefore, innovative approaches in this aspect is a must to improve performance at higher GOP sizes.

3. Key frames: It is noted that, in the DVC encoder, majority of the computational complexity (at least 90%) is usually attributed to the intra frame codec (commonly

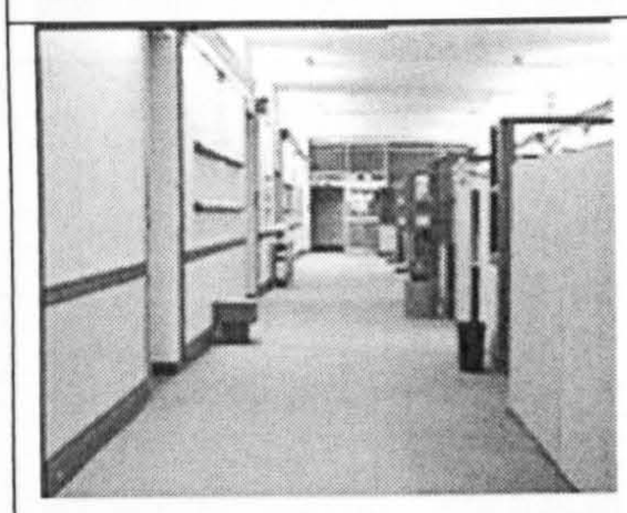
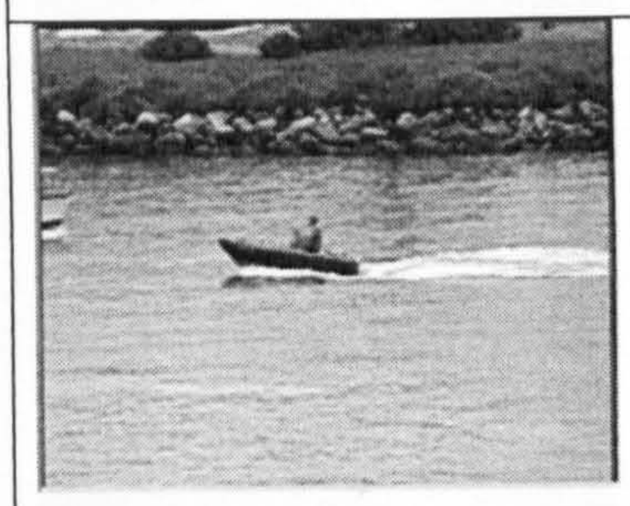
borrowed from H.264/AVC standard) used for key frame coding. In a situation where the encoder complexity is an important consideration in DVC, the above appears as a problem. An initial effort to solve this problem was taken in [14]; however, more attention needs to be paid to this aspect of DVC.

Above suggested are some of the possible future considerations in developing a stable and robust DVC codec leading to standardization and commercial deployments.

# APPENDIX A

## Test Video Sequences

The simulations carried out within this research utilize a number of test video sequences. Some basic details of each of these sequences are listed below. These sequences are selected in the simulations following those in the reference literature so that the performance of proposed algorithms could be easy compared.

| Frame #0 | Sequence name | Frame format | Temporal Resolution (fps) | Motion activity level |
|---|---|---|---|---|
| | Foreman | QCIF (spatial res: 176×144; sampling: 4:2:0) | 30, 15 | Medium |
|  | Hall Monitor | QCIF | 30, 15 | Low |
|  | Coastguard | QCIF | 30, 15 | Low~medium |

| | | | | |
|---|---|---|---|---|
| | Soccer | QCIF | 30, 15 | High |
| | Clair | QCIF | 30, 15 | Low |
| | Salesman | QCIF | 30, 15 | Low |
| | Carphone | QCIF | 30, 15 | Medium |
| | News | QCIF | 30, 15 | Low |

# APPENDIX B

## Presentation of simulation results

In this thesis, the results of the codec simulations with standard test video sequences are presented mainly in two methods:

- Objective quality results
- Subjective quality results

### Objective quality results

The objective quality is measured in terms of PSNR as:

$$PSNR = 10\log_{10}\left[\frac{255^2}{(1/N)\sum_i\sum_j\left(Y_{ref}(i,j)-Y_{prc}(i,j)\right)^2}\right] \qquad \text{B-1}$$

where $Y_{ref}(i,j)$ and $Y_{prc}(i,j)$ represent the pixel values of the reference and decoded pictures respectively and $N$ is the total number of pixels in the picture. The PSNR is calculated for each frame and then either plotted against frame number for a given bit rate, or averaged over the sequence and plotted against the bit rate. Example illustrations are shown below:

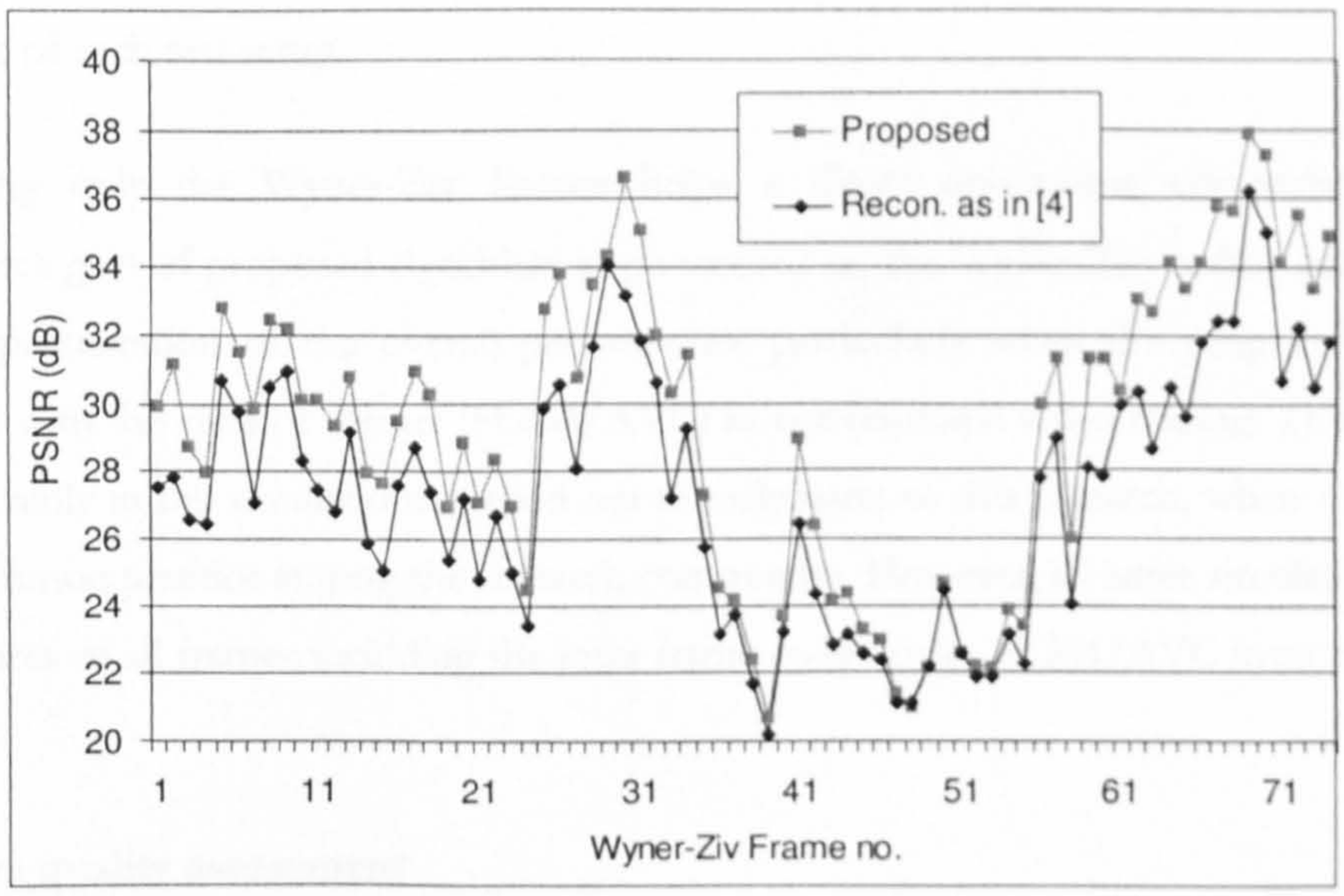



Figure B-1 PSNR vs. frame no. plot

In PSNR vs. frame number plots, the PSNR results are displayed only for the Wyner-Ziv frames of the sequence.

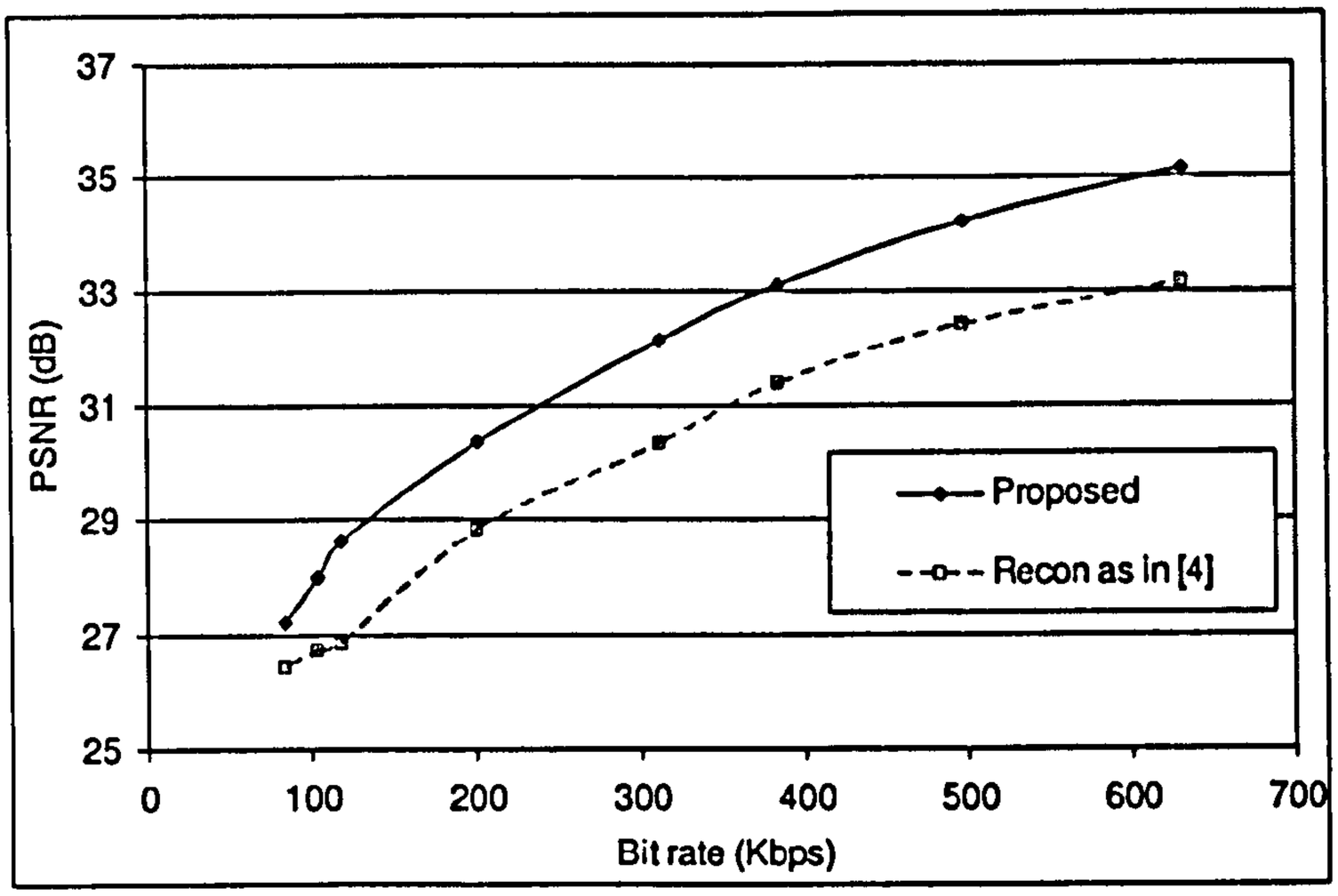



Figure B-2 PSNR vs. Bit rate plot

In PSNR vs. Bit rate plots, the average PSNR and bit rate values are calculated for the for either Wyner-Ziv frames only or all frames of the sequence including key frames. Both of the above approaches have its pros and cons and the chosen approach is specified in the test conditions of each test setup.

Considering only the Wyner-Ziv frames helps a direct one-to-one comparison of the performance gain of proposed algorithm implemented on the Wyner-Ziv coding path, yet lack a clear representation of the overall performance particularly when the proposed results is compared with the state of the art (H.264/AVC) in conventional video coding. This approach is taken mainly in the simulations carried out in early parts of this research, when that used to be the common practice among the research community. However, all latter simulations reflect the measures of all frames including the intra frame coded (uses H.264/AVC intra coding) key frames.

**Subjective quality assessment**

In this thesis, where a subjective quality assessment of a decoded frame is believed to reinforce

the improvement resulting in a given algorithm, the graphics of selected frames are illustrated in addition to the objective quality measures. An example is shown below.

Figure B-3 Graphics for subjective quality assessment

# APPENDIX C

## List of Publications

**Book Chapters:**

1. J. L. Martínez, W. A. R. J. Weerakkody, P. Cuenca, F. Quiles, W. A. C. Fernando, "Transform Domain Wyner-Ziv Codec Based on Turbo Trellis Codes Modulation," Lecture Notes in Computer Science (Advances in Multimedia Modeling), Volume 4903/2008, pp. 413 -424, Springer-Verlag 2008.

2. J. L. Martínez, W. A. C. Fernando, W. A. R. J. Weerakkody, J. Oliver, O. López, M. Martinez, M. Pérez, P. Cuenca, F. Quiles, "Low-Complexity TTCM Based Distributed Video Coding Architecture," Lecture Notes in Computer Science (Advances in Image and Video Technology), Volume 4872/2007, pp. 841–852, Springer-Verlag 2007.

**Journals:**

3. W. A. R. J. Weerakkody, W. A. C. Fernando, A. B. B. Adikari, "Unidirectional Distributed Video Coding for Low Cost Video Encoding," IEEE Trans. Consumer Electronics, vol. 53, issue 2, pp. 788 -795, May 2007.

4. W. A. R. J. Weerakkody, W. A. C. Fernando, A. B. B. Adikari and R. M. A. P. Rajatheva, "4-PSK TTCM for Wyner-Ziv frame coding in DVC," IEICE Electron. Express, Vol. 3, No. 8, pp.178-183, 2006.

5. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi, "Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames," IEE Electronics Letters, Vol. 42, Issue 7, pp 398-399, March,2006

6. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Multiple Side Information Streams for Distributed Video Coding," IEE Electronics Letters, Vol. 42, Issue 25, December 2006.

7. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Iterative Wyner-Ziv Decoding for Unidirectional Distributed Video Coding," IEE Electronics Letters, Vol. 43, Issue 2, pp. 93 – 95, January 18 2007.

8. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Independent Key Frame Coding Using Correlated Pixels In Distributed Video Coding," IEE Electronics Letters, Vol. 43, Issue 7, pp. 387 – 388, March 29 2007.

9. J.L. Martínez, W.A.R.J. Weerakkody, W.A.C. Fernando, F. Quiles, A. M. Kondoz, "Turbo Trellis Coded Modulation for Transform Domain Distributed Video Coding," IET Electronics Letters, Volume 44, Issue 15, pp. 899 - 900, July 17 2008.

10. J.L. Martinez, W.A.R.J.Weerakkody, W.A.C.Fernando, G. Fernandez-Escribano, H. Kalva, A. Garrido, "Distributed Video Coding using Turbo Trellis Coded Modulation," Accepted for the Visual Computer journal, July 2008.

11. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, A.M. Kondoz, "An Iterative Fusion Technique for Dynamic Side Information Refinement in Pixel Domain Distributed Video Coding," Accepted for IEICE Trans. on Fundamentals, Submitted April 2008.

12. M.B.Badem, W.A.C. Fernando, W.A.R.J.Weerakkody, H.K. Arachchi, A. M. Kondoz, "Transform Domain Unidirectional Distributed Video Coding Using Dynamic Parity Allocation," Accepted for IEICE Trans. on Fundamentals, Submitted April 2008.

13. M.B.Badem, W.A.R.J.Weerakkody, W.A.C.Fernando, A.M.Kondoz, "Design of a Non-Linear Quantizer for Transform Domain DVC," Accepted for IEICE Trans. on Fundamentals, Submitted April 2008.

14. S.L.P.Yasakethu, W.A.R.J.Weerakkody, W.A.C. Fernando, F.Pereira, A.M.Kondoz, "An Improved Decoding Algorithm for DVC over Multipath Error Prone Wireless Channels," Accepted for IEEE Trans. on CSVT, submitted April 2008.

**Conferences:**

15. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, " Distributed video coding of Wyner-Ziv frames using Trellis Coded Modulation," Proceedings of IEEE CCECE 2006, pp 314 – 317, Ottawa, Canada, May 2006.

16. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, Q. Fang, "Bitplane Based Wyner-Ziv Coding using Unequal Error Protection," Proceedings of IEEE CCECE 2006, pp 1192 – 1195, Ottawa, Canada, May 2006.

17. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, R.M.A.P. Rajatheva, "Distributed video coding of Wyner-Ziv frames using Turbo Trellis Coded Modulation," Proceedings of IEEE ICIP 2006, Atlanta, USA, Oct 2006.

18. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, "Distributed video coding of pixel interleaved Wyner-Ziv frames," Proceedings of IEEE ICIIS, Sri Lanka, Aug. 2006.

19. W.A.R.J. Weerakkody, W.A.C. Fernando, A.B.B. Adikari, "Distributed Video Coding of Pixel Interleaved Wyner-Ziv Frames with Iterative Decoding," Proceedings of WPMC 2006, San Diego, USA, September, 2006.

20. W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, "An Iterative Refinement Technique for Side Information Generation in DVC," Proceedings of IEEE ICME 2007, Beijing, China, July 2007.

21. W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, S.Adedoyin, "Map Decoder Design for Distributed Video Coding in Error Prone Wireless Channels," Accepted for the Proceedings of IEEE ICIIS 2007, Sri Lanka, Aug 2007.

22. W.A.R.J. Weerakkody, W.A.C. Fernando, R.M.A.P. Rajatheva, T.Kuganeswaran, K. Muraleetharan, "Distributed Video Coding With MIMO for Video Wireless Communications," Proceedings of IEEE ICSPC 2007, Dubai.

23. W.A.R.J. Weerakkody, W.A.C. Fernando, A.M.Kondoz, "An Enhanced Reconstruction Algorithm for Unidirectional Distributed Video Coding," Accepted for the Proceedings of IEEE ISCE 2007, Algarve, Portugal, April 2008.

24. A.B.B. Adikari, W.A.C. Fernando, W.A.R. Weerakkody, H. Kodikara Arachchi, " Wyner-Ziv coding with temporal and spatial correlations for motion video," Proceedings of IEEE CCECE 2006, pp 1188 – 1191, Ottawa, Canada, May 2006.

25. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi, "A Sequential Motion Estimation Based Refinement Technique for Distributed video coding of Wyner-Ziv frames," Proceedings of IEEE ICIP 2006, Atlanta, USA, Oct 2006.

26. H. Kodikara Arachchi, W.A.C. Fernando, S. Panchadcharam, W.A.R.J. Weerakkody, "Unequal Error Protection Technique for ROI Based H.264 Video Coding," Proceedings of IEEE CCECE 2006, pp 2033 - 2036, Ottawa, Canada, May 2006.

27. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H. Kodikara Arachchi, "Low Complex Key Frame Encoding With High Quality Wyner-Ziv Coding," Proceedings of IEEE ICIIS, Sri Lanka, Aug. 2006.

28. A.B.B. Adikari, W.A.C. Fernando, W.A.R. Weerakkody, H. Kodikara Arachchi, "Wyner-Ziv Intra Coding," Proceedings of WPMC 2006, San Diego, USA, September, 2006.

29. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Side Information Improvement in DVC with Two Side Information Streams and 3D Motion Refinement," Proceedings of CCECE 2007, Vancouver, Canada, 2007.

30. K.Muraleetharan, W.A.C. Fernando, W.A.R.J. Weerakkody, R.M.A.P.Rajatheva, "MIMO Wireless Channel for Video Communications using Distributed Video Coding," Proceedings of CCECE 2007, Vancouver, Canada, 2007.

31. S. Adedoyin, W.A.C. Fernando, A.Aggoun, W.A.R.J.Weerakkody, "An Es Based Efficient Motion Estimation Technique for 3d Integral Video Compression," Proceedings of IEEE ICIP 2007, San Antonio, USA, September 2007.

32. J.L. Martínez, W.A.R.J. Weerakkody, P. Cuenca, F. Quiles, and W.A.C. Fernando, "Transform Domain Wyner-Ziv Codec Based on Turbo Trellis Coded Modulation," Proceedings of MMM 2008, Kyoto Japan, January 2008.

33. J.L. Martínez, W.A.R.J. Weerakkody, P. Cuenca, F. Quiles, and W.A.C. Fernando, "Distributed Video Coding In Transform Domain," Proceedings of VIP 2007, Taiwan, Dec. 2007.

34. A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, A.M.Kondoz, "Joint Spatial And Temporal Correlation Exploitation For Wyner-Ziv Frames Coding In DVC," Accepted for the Proc. IEEE ICME, Hannover, Germany, June 2008.

35. J.L. Martínez, F.E. Gerardo, H. Kalva, A. Garrido, W.A.R.J Weerakkody, W.A.C. Fernando, "Feedback Free DVC Architecture Using Machine Learning," Accepted for the Proc. IEEE ICIP 2008, San Diego, CA, USA, October 2008.

**Under Review:**

**Journals:**

36. W.A.R.J.Weerakkody, W.A.C. Fernando, F.Pereira, A. M. Kondoz, "An Enhanced Reconstruction Algorithm for Unidirectional Distributed Video Coding," Submitted to EURASIP journal, submitted May 2008.

37. W.A.R.J. Weerakkody, W.A.C. Fernando, A.M. Kondoz , "Modifying the Turbo Decoder for DVC based Video Codecs," Submitted to IEICE Trans. on Fundamentals, Submitted June 2008.

38. W.A.R.J.Weerakkody, W.A.C. Fernando, M.B.Badem, A.M.Kondoz, "Non-Linear Quantization for Pixel Domain Distributed Video Coding," Submitted to IET Electronics Letters, submitted June 2008.

# REFERENCES


[1] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 560-576, July 2003.

[2] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Inform Theory, Vol. IT-19, pp. 471-480, July 1973.

[3] D.Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, Vol. IT-22, pp. 1-10, Jan 1976.

[4] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv coding of motion video," Proc. Asilomar Conference on Signals and Systems, Vol. 1, pp.240-244, Pacific Grove, CA, Nov 2002.

[5] R. Puri, K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," Proc. of 40th Allerton Conference on Communication, Control, and Computing, Allerton, IL, Oct 2002.

[6] C. Tonoli, M. Dalai, P. Migliorati, R. Leonardi, "Error Resilience Performance Evaluation of A Distributed Video Codec," Proc. of Picture Coding Symposium 2007, Lisbon, Portugal, November 2007.

[7] http://www.discoverdvc.org/

[8] http://www.visnet-noe.org/

[9] M. Ghanbari, "Standard Codecs: Image Compression to Advanced Video Coding," The Institution of Electrical Engineers, London, UK, 2003.

[10] Iain E. G. Richardson, "H.264 and MPEG-4 Video Compression," J.Wiley and Sons Limited, Chichester, England, 2003.

[11] A. Aaron, S. Rane, E. Setton, B. Girod, "Transform-Domain Wyner-Ziv Codec for Video," Proc. Visual Communications and Image Processing, pp.520-528, USA, January 2004.

[12] A. Aaron, S. Rane and B. Girod, "Wyner-Ziv Video Coding with Hash-Based Motion Compensation at the Receiver," Proc. of IEEE International Conference on Image Processing, Vol. 5, pp. 3097-3100, Singapore, October 2004.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," International Conference on Communications, ICC, pp 1064-1070, 1993.

[14] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Independent Key Frame Coding Using Correlated Pixels In Distributed Video Coding," IEE Electronics Letters, Vol. 43, Issue 7, pp. 387 – 388, March 29 2007.

[15] C. Brites, J. Ascenso, and F. Pereira, 'Studying temporal correlation noise modeling for pixel basedWyner-Ziv video coding," IEEE Int. Conference on Image Processing, Atlanta, CA, USA, October 2006.

[16] Meyer, Westerlaken, Gunnewiek, Lagendijk, "Distributed Source Coding of Video with Non-Stationary Side Information," Proc. SPIE VCIP, Beijing, China, June 2005.

[17] W. A. R. J. Weerakkody, W. A. C. Fernando, A. B. B. Adikari, "Unidirectional Distributed Video Coding for Low Cost Video Encoding," IEEE Trans. Consumer Electronics, vol. 53, issue 2, pp. 788 -795, May 2007.

[18] W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, "An Iterative Refinement Technique for Side Information Generation in DVC," Proceedings of IEEE ICME 2007, Beijing, China, July 2007.

[19] M. Tagliasacchi, J. Pedro, F. Pereira, S. Tubaro, "An Efficient Request Stopping Method at the Turbo Decoder in Distributed Video Coding," EURASIP European Signal Processing Conference (EUSIPCO 2007) , September 2007.

[20] C. Brites, J. Ascenso, F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance," Proceedings of IEEE ICASSP, Vol 2, pp.II-525 - II-528, Toulouse, France, May 2006.

[21] X. Guo, Y. Lu, F. Wu, W. Gao, "Distributed Video Coding Using Wavelet," Proceedings of IEEE ISCAS, May 2006.

[22] ISO/IEC International Standard 14496-10:2003, "Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding".

[23] A. Aaron, E. Setton, B. Girod, "Towards Practical Wyner-Ziv Coding Of Video," Proc. IEEE International Conference on Image Processing, ICIP-2003, Barcelona, Spain, Sept. 2003.

[24] A. Aaron, S. Rune, B. Girod, "Wyner-Ziv Video Coding with Hash-Based Motion Compensation at the Receiver" Proc. of ICIP, pp. 3097-3100, Singapore, October 2004.

[25] M. Tagliasaccchi, S. Tubaro, "A MCTF Video Coding Scheme Based on Distributed Source Coding Principles, "Visual Communication and Image Processing, Beijing, China, July 2005.

[26] L. P. Natário, C. Brites , J. Ascenso, F. Pereira, "Extrapolating Side Information For Low-Delay Pixel-Domain Distributed Video Coding," Proc International Workshop on Very Low Bitrate Video , Costa del Rey , Italy , September , 2005.

[27] J. Ascenso, C. Brites, F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding," Proc. Conference on Speech and Image Processing, Multimedia Communications and Services - EURASIP , Smolenice , Slovakia , 2005.

[28] X. Artigas, L. Torres, "Iterative Generation of Motion-Compensated Side Information for Distributed Video Coding," Proc. of ICIP'05, Genova, Italy September 2005.

[29] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi, "Sequential Motion Estimation Using Luminance And Chrominance Information For Distributed Video Coding Of Wyner-Ziv Frames," IEE Electronics Letters, Vol. 42, Issue 7, pp 396-397, March 2006.

[30] D. Kubasov, C. Guillemot, "Mesh-Based Motion-Compensated Interpolation for Side Information Extraction in Distributed Video Coding," Proc.of IEEE ICIP, pp 261 – 264, Atlanta, USA, October 2006.

[31] Toan Nguyen Dinh, GueeSang Lee, J. Y. Chang, Han-Jin Cho, "Side Information Generation Using Extra Information in Distributed Video Coding," Proc. of IEEE International Symposium on Signal Processing and Information Technology pp 138 – 143, Dec 2007.

[32] Wei-Jung Chien, L. J. Karam, G. P. Abousleman, "Distributed Video Coding With Lossy Side Information," Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, May 2006.

[33] Lei Wei, Yao Zhao, Anhong Wang, "Improved Side-Information in Distributed Video Coding," Proc. of First International Conference on Innovative Computing, Information and Control, 2006. ICICIC '06. Volume 2, 30-01 pp 209 - 212 Aug. 2006.

[34] M. Tagliasacchi, S. Tubaro, A. Sarti, "On the Modeling of Motion in Wyner-Ziv Video Coding," IEEE International Conference on Image Processing, 2006, pp 593 – 596, 8-11 Oct. 2006.

[35] M. Dalai, R. Leonardi, F. Pereira, "Improving Turbo Codec Integration In Pixel-Domain Distributed Video Coding," Proceedings of IEEE ICASSP, Vol 3, pp III-257 - III-260, Toulouse, France, May 2006.

[36] Toan Nguyen Dinh, GueeSang Lee, June-Young Chang, Han-Jin Cho, "A Novel Motion Compensated Frame Interpolation Method for Improving Side Information in Distributed Video Coding," International Symposium on Information Technology Convergence, 2007. ISITC 2007, pp 179 – 183, 23-24 Nov. 2007.

[37] Mei Guo, Yan Lu, Feng Wu, Shipeng Li, Wen Gao, "Distributed Video Coding with Spatial Correlation Exploited Only at the Decoder," IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007. pp 41 – 44, 27-30 May 2007.

[38] Yixuan Zhang, Ce Zhu, "Full Search of Side-Information in Distributed Video Coding," International Conference on Image and Graphics, 2007, ICIG 2007,pp 246 – 249, 22-24 Aug. 2007.

[39] M. Tagliasacchi, L. Frigerio, S. Tubaro, "Rate-Distortion Analysis of Motion-Compensated Interpolation at the Decoder in Distributed Video Coding," IEEE Signal Processing Letters, Volume 14, Issue 9, pp 625 – 628, Sept. 2007.

[40] Ko, Bonghyuck, Shim, Hiuk Jae, Jeon, Byeungwoo, "Wyner-Ziv coding with spatio-temporal refinement based on successive turbo decoding," IEEE International Conference on Multimedia and Expo, 2008, pp 785 – 788, June 2008.

[41] Feng, Ying, Xiao, Song, Wu, Chengke, Song, Rui, "A Refined Algorithm for the Side Information in DVC," International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP '08, pp 675 – 678, 15-17 Aug. 2008.

[42] M.B. Badem, M. Mrak and W.A.C. Fernando , "Side information refinement using motion estimation in DC domain for transform-based distributed video coding" , IET Electronics Letters, Vol. 44, No. 16, pp. 965 – 966 , July 2008.

[43] J. Ascenso, C. Brites, F. Pereira, "Content Adaptive Wyner-ZIV Video Coding Driven by Motion Activity," Proc. of 2006 IEEE International Conference on Image Processing, pp 605 – 608, 8-11 Oct. 2006.

[44] Y. Tonomura, T. Nakachi, "A New Framework for Distributed Video Coding Based on JPEG 2000," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006. Volume 3, 14-19 May 2006.

[45] Y. Vatis, S. Klomp, J. Ostermann, "Inverse Bit Plane Decoding Order for Turbo Code Based Distributed Video Coding," IEEE International Conference on Image Processing, 2007, ICIP 2007, Volume 2, pp II - 1 - II – 4, Sept. 16 2007-Oct. 19 2007.

[46] M. Ouaret, F. Dufaux, T. Ebrahimi, "Codec-Independent Scalable Distributed Video Coding," IEEE International Conference on Image Processing, 2007, ICIP 2007, Volume 3, pp III - 9 - III – 12, Sept. 16 2007-Oct. 19 2007.

[47] Wei-Jung Chien, L. J. Karam, G. P. Abousleman, "Block-Adaptive Wyner-Ziv Coding for Transform-Domain Distributed Video Coding," IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007, Volume 1, pp I-525 - I-528 15-20 April 2007.

[48] J.L Martinez, C. Holder, G. Fernandez-Escribano, H. Kalva, F. Quiles, "DVC using a half-feedback based approach," IEEE International Conference on Multimedia and Expo, 2008, pp1125 – 1128, June 23 2008-April 26 2008.

[49] R. Puri, A. Majumdar, P. Ishwar, K. Ramchandran, "Distributed Video Coding in Wireless Sensor Networks," IEEE Signal Proc. Magazine, vol. 23, no. 4, pp. 94-106, July 2006.

[50] Y. Vatis, S. Klomp, J. Ostermann, "Enhanced reconstruction of the quantized transform coefficients for Wyner-Ziv coding", Proc. of Int. Conference. on Multimedia & Expo (ICME), Beijing, China, July 2007.

[51] D. Kubasov, J. Nayak, C. Guillemot, "Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information," Proc. of Int. workshop on MMSP, Greece, October 2007.

[52] P. Ishwar, V. Prabhakaran, K. Ramchandran, "Towards a Theory for Video Coding Using Distributed Compression Principles," Proc. of International Conference on Image Processing, Vol.3, pp.687-690, Barcelona, Catalonia, Spain September 2003.

[53] B. Wu, X. Guo, D. Zhao, W. Gao, F. Wu, "An Optimal Non-Uniform Scalar Quantizer For Distributed Video Coding," IEEE International Conference on Multimedia and Expo, pp.165 – 168, Toronto, July 2006.

[54] Z. Li, L. Liu, E. J. Delp, "Wyner-Ziv Video Coding With Universal Prediction," IEEE Transaction on Circuits and Systems for Video Technology, vol. 16, no. 11, pp. 1430-1436, Nov. 2006.

[55] Z. Li, L. Liu, E. J. Delp, "Rate Distortion Analysis Of Motion Side Estimation In Wyner-Ziv Video Coding," IEEE Transaction on Image Processing, vol. 16, no. 1, pp. 98-113, Jan. 2007.

[56] D. Schonberg, S. S. Pradhan, and K. Ramchandran, "LDPC codes can approach the Slepian-wolf bound for general binary sources", presented at the Allerton Conference Communication, Control, and Computing, Champaign, IL, 2002.

[57] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, R. L. Lagendijk, "Analyzing Symbol and Bit Plane-Based LDPC in Distributed Video Coding," IEEE International Conference on Image Processing, 2007. ICIP 2007. Volume 2, pp II - 17 - II – 20. Sept. 16 2007-Oct. 19 2007.

[58] Anhong Wang, Yao Zhao, Hao Wang, "LVQ Based Distributed Video Coding with LDPC in Pixel Domain," Lecture Notes in Computer Science, Volume 4099/2006, pp 1248-1252, 2006.

[59] Hao Wang; Yao Zhao; Anhong Wang, "Performance Comparisons of Different Channel Codes in Distributed Video Coding," First International Conference on Innovative Computing, Information and Control, 2006. ICICIC '06. Volume 2, pp 225 – 228, 30-01 Aug. 2006.

[60] Fang Sheng, Liang Yong-Quan, Li Shen-Yuan, "What Affects Decoding Complexity of Distributed Video Codec Based on Turbo Code," Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Volume 2, pp 171 – 176, July 30 2007-Aug. 1 2007.

[61] Anhong Wang, Yao Zhao, Lei Wei, "Wavelet-Domain Distributed Video Coding with Motion-Compensated Refinement," IEEE International Conference on Image Processing 2006, pp 241 – 244, 8-11 Oct. 2006.

[62] Bo Wu, Xiangyang Ji, Debin Zhao, Wen Gao, "Wavelet based distributed video coding with spatial scalability," IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008. pp 3458 - 346118-21 May 2008.

[63] Y. Tonomura, D. Shirai, T. Nakachi, T. Fujii, "Optimal Bit Allocation for Wavelet-Based Distributed Video Coding," Eighth IEEE International Symposium on Multimedia, 2006. ISM'06. pp 442 – 448, Dec. 2006.

[64] M. Dalai, R. Leonardi, F. Pereira, "Improving Turbo Codec Integration in Pixel-Domain Distributed Video Coding," IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006, Volume 2, pp II - II 14-19, May 2006

[65] Lin-Bo Qing, Xiao-Hai He, Rui Lv, "Modeling non-stationary correlation noise statistics for Wyner-Ziv video coding," International Conference on Wavelet Analysis and Pattern Recognition," ICWAPR 2007. Volume 1, pp 316 - 320 2-4 Nov. 2007.

[66] Linbo Qing, Xiaohai He, Rui Lv, "Distributed Video Coding with Dynamic Virtual Channel Model Estimation," The First International Symposium on Data Privacy, and E-Commerce, 2007. ISDPE 2007. pp 170 – 1741-3 Nov. 2007.

[67] Sheng Fang, Li-Wei Zhang, Ling Huang, "An Adaptive Nested Scalar Quantization Scheme for Distributed Video Coding," IEEE Workshop on Signal Processing Systems, 2007, pp 351 – 356, 17-19 Oct. 2007.

[68] Fang Sheng, Li Xu-Jian, Zhang Li-Wei, "A Lloyd-Max-based Non-Uniform Quantization Scheme for Distributed Video Coding," Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Volume 1, pp 848 – 853, July 30 2007-Aug. 1 2007.

[69] M. Maitre, C. Guillemot, L. Morin, "3D Scene Modeling for Distributed Video Coding," IEEE International Conference on Image Processing, 2006. 8-11 Oct. 2006 pp585 – 588.

[70] X. Artigas, E. Angeli, L. Torres, "Side Information Generation for Multiview Distributed Video Coding Using a Fusion Approach," Proceedings of the 7th Nordic Signal Processing Symposium, 2006. NORSIG 2006. pp 250 – 253, June 2006.

[71]Ye Shuiming, Ouaret Mourad, Dufaux Frederic, Ebrahimi Touradj, "Hybrid spatial and temporal error concealment for distributed video coding," IEEE International Conference on Multimedia and Expo, 2008, pp 633 – 636, June 23 2008-April 26 2008.

[72] Hua Guogang, Chen Chang Wen, "Distributed video coding with zero motion skip and efficient DCT coefficient encoding," IEEE International Conference on Multimedia and Expo, 2008, pp 777 – 780, June 23 2008-April 26 2008.

[73] R. Bernardini, Rinaldo, P. Zontone, "Performance evaluation of distributed video coding schemes," IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. pp 709 – 712, March 31 2008-April 4 2008.

[74] G. Ungerboeck, "Channel Coding with Multi level/Phase Signals," IEEE Transaction Inform. Theory, IT-28, pp 55-67, January, 1982.

[75] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi, "Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames," IEE Electronics Letters, Vol. 42, Issue 7, pp 398-399, March,2006

[76] MPEG-2 video test model 5. Technical report, ISO/IEC JTCl/SC29/WGll and ITU-TS SG15 EG for ATM Video Coding, April 1993.

[77] C. Brites, F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding," IEEE International Conference on Image Processing, ICIP-2007, San Antonio, Texas, USA, Sept. 2007.

[78] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Side Information Improvement in DVC with Two Side Information Streams and 3D Motion Refinement," Proc. of IEEE CCECE, pp:32 - 35 Vancouver, Canada, April 2007.

[79] A.B.B.Adikari, W.A.C.Fernando, W.A.R.J.Weerakkody, "Multiple Side Information Streams for Distributed Video Coding," IET electronics Letters, Volume 42, Issue 25, pp:1447 – 1449, December 2006.

[80] J. Pedro. L. Ducla Soares, F. Pereira, "Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec," Picture Coding Symposium, Lisbon – Portugal, November 2007.

[81] L. R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," IEEE Transactions on Information Theory, vol. IT-20, pp. 284-287, March 1974.

[82] William E. Ryan, "A Turbo Code Tutorial," Proc. IEEE Globecom 1998. Available at, http://www.ece.arizona.edu/~ryan/publications/turbo2c.pdf

[83] G. Montalbano, D. T. M. Slock, , "Joint Common-Dedicated Pilots Based Estimation Of Time-Varying Channels For W-CDMA Receivers," Proc. of IEEE VTC, vol. 2, pp.1253 – 1257, 2003.

[84] Keng-Pang Lim, G. Sullivan , T. Wiegand , "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods ," JVT document JVT X101 Geneva, June 2007.

[85] R.N. Gajaweera, R.M.A.P. Rajatheva, K.M. Ahmed, "Low bit rate video transmission over correlated frequency selective channel," IEEE Global Telecommunications Conference (GLOBECOM) 2001, Volume 3, pp:2065 - 2069, Nov. 2001.