

行政院國家科學委員會補助專題研究計畫成果報告
基於正交分頻多重進接之無線多媒體傳收機研究及設計—總計畫(I)
OFDM-Based Mobile Wireless Multimedia Transceiver Research and
Design

計畫編號：NSC 91-2219-E-009-007

執行期限：91年8月1日至92年7月31日

主持人：林大衛 交通大學電子工程學系 教授

共同主持人：張文鐘 交通大學電信工程學系 教授

陳紹基、杭學鳴 交通大學電子工程學系 教授

摘要

本整合型計畫係以 IEEE 802.16a 標準為基礎，研究正交分頻多重進接(OFDMA)傳輸系統在作無線行動通訊服務之用時的傳收系統技術。本計畫擬為期三年，第一年(91/8-92/7)獲核定四個子計畫，第二及第三年(92/8-94/7)獲核定五個子計畫。本報告係針對第一年之研究，其中我們探討了聲訊壓縮編碼、濾波器設計、OFDM 接收器架構與 FFT 處理器架構、傳輸訊號之時間同步與頻率同步、通道估計、通道碼解碼、及通道品質預測與傳輸率控制法。在第二年以後我們將加入視訊編解碼之研究。此外，在第二年以後我們除將繼續進行各分項傳輸技術之研究外，亦擬以數位訊號處理器(DSP)為主要平台，將各傳收器組件作軟體或硬體實現並予以連結。

關鍵詞：正交分頻多重進接、正交分頻多工調變、聲訊編碼、視訊編碼、同步、通道估計、通道編碼、通道品質預測、媒體存取控制、資料流率控制

Abstract

This integrated project bases on the IEEE 802.16a standard to research into transceiving system technologies for wireless mobile communication under OFDMA (orthogonal frequency-division multiple access). The project is intended for 3 years, where for the first year (2002/8-2003/7) we have been awarded 4 subprojects and for the second and the third years (2003/8-2005/7) 5 subprojects. This report is concerned with the research of the first year, wherein we looked into audio coding, filter design, OFDM receiver architecture and FFT processor architecture, time and frequency synchronization of the transmitted signal, channel estimation, channel decoding, and channel quality prediction and transmission rate control. From the second year on, we will include research in video coding and decoding. In addition, from the second year on, besides continued research in individual transmission technologies, we also plan to use digital signal processors (DSPs) as the main platform to conduct software or hardware implementation of the transceiver components and to connect them together.

Keywords: Orthogonal Frequency-Division Multiple Access (OFDMA), Orthogonal Frequency-Division Multiplexing (OFDM), Audio Coding, Video Coding, Synchronization, Channel Estimation, Channel Coding, Channel Quality Estimation, Medium Access Control, Data Flow Control

目錄 Table of Contents

一、計畫緣由與目的	1
二、結果與討論	2
A. 總體成果摘要	2
B. 無線網路串流聲訊研究及聲視訊子系統整合(子計畫四)	3
C. 寬頻正交分頻調變傳輸系統之傳輸信號處理技術研究及傳輸子系統 整合(子計畫三)	4
D. 無線正交分頻多重進接頻道使用技術研究及全系統整合(子計畫一)	4
E. 根據預測的無線通道情況控制即時資料的傳輸速率(子計畫二)	5
三、參考文獻	6
四、圖表	7
五、計畫成果自評	8
六、可供推廣之研發成果(請見各子計畫報告)	8
七、附錄	9

一、計畫緣由與目的

第三代行動通訊系統已逐漸由研發階段步入商業運轉的階段。許多人也預期約將在十年後進入第四代。以傳輸速率而言，預期第四代約將為第三代的 10 至 100 倍。在調變方法方面，則似以正交分頻多工(OFDM)及與其相關之技術最為矚目。本計畫基於最近制定之 IEEE 802.16a 標準[1]中，關於正交分頻多重進接(OFDMA)之規範，研究高速無線行動傳輸技術。IEEE 802.16a 標準原是為了固定式寬頻無線通訊而制定，但經過約半年的醞釀之後，在 IEEE 802.16 工作團(Working Group)之下又已於去(2002)年最後一季成立了 802.16e 任務團(Task Group)，盼只將 802.16 系列標準做小幅延伸，即能符合行動傳輸之所需，且可成為第四代行動通訊系統的候選技術。目前該任務團擬於今(2003)年 9 月完成其第一版標準草案(D1)。故本計畫之執行時間，恰與其同時。有趣的是，除了 802.16e，IEEE 之內另有一個新成立(2002 年 12 月)的 802.20 工作團亦在探討行動寬頻無線傳輸。

本計畫原擬為期三年，主要目標有二：

1. 以 IEEE 802.16a 標準為基礎，研究 OFDMA 傳輸系統在作無線行動通訊服務之用時的傳收系統技術及傳輸效能。
2. 將各傳收系統組件以數位訊號處理器(DSP)或 FPGA 作軟體或硬體實現，並予以連接。

本計畫第一年(91/8-92/7)共獲核定四個子計畫，第二及第三年(92/8-94/7)獲核定五個子計畫。本報告所述之成果僅含第一年之四個子計畫，但以下關於計畫規劃之描述則包括三年。

本計畫研究標的系統之概略架構及各子計畫間的分工如圖一所示。圖左的 Video Encoder/Decoder and Error Resilience，亦即標記為 subproject x 的兩個方塊，即為第二及第三年獲核定之第五個子計畫。各子計畫除了須互相配合以達成以上述主要目標，亦可依其負責之技術項目作衍伸性之研究，以求在學術與技術方面之更多創新與進步。以下我們簡釋圖一之內容。

IEEE 802.16a 標準容許分頻雙工(FDD)與分時雙工(TDD)。鑒於未來的高速無線通訊很可能具有雙向不對稱的傳輸率，我們決定考慮 TDD 雙工，並兼研究上行與下行兩方向之傳輸(兩方向之傳輸信號結構不同)。我們並擬試使最後之 DSP(或 FPGA)實現具可程式性，以方便彈性選擇欲採用之選項參數。在 MAC 及 Link 層方面，我們主要考慮的是通道的有效使用與優良服務品質的獲得。

如圖一所示，PHY 層之各項信號處理功能，即自圖上中之 FEC (FEC encoder 之簡稱)至下中之 FED (FEC decoder 之簡稱)，係分別由子計畫四、一、三進行研究。其中且將建構一個模擬的無線通道(圖一右中之 Simulated Channel)，使系統測試時可不必建構實體 RF 通道，並使其測試能更方便而具彈性。MAC 及 Link 層之研究由子計畫二執行(見圖中央部分)，其中研究如何依據通道狀況來改變個別用戶的調變與資料頻寬以求優良的服務品質。IEEE 802.16a 標準，主要是規範 MAC 層及 PHY 層之運作。但自 2.5G 行動通訊系統開始發展以來，多媒體行動通訊日形重要，也成為第三代行動通訊系統設計上的重要考慮因素。其實所謂寬頻無線進接的一項主要功能，也就是要

提供多媒體聲視訊傳輸服務。本計畫在整個信號傳輸系統的研究上，亦考慮到適用於無線傳輸之聲視訊編解碼，以及相關傳輸誤差之處理。如圖一最左方所示，我們係於子計畫四及 x 中分別探討聲訊及視訊之編解碼與傳輸誤差之處理。子計畫二之一主要研究題目，即聲視訊資料頻寬之控制，亦與此相關。

圖一中亦顯示本計畫在系統整合方面之規劃，計分三部分，如虛線框架所示，分別由子計畫三、四、一執行。我們擬使用個人電腦及 DSP/FPGA 插板為實現平台。預期在整合完成的系統中，傳送端(含模擬之無線通道)及接收端各需要數個插板，兩端擬各架構在一台個人電腦上。

總體而言，三年的工作規劃大略如下：

1. 第一年：各分項技術之演算法研究。
2. 第二年：繼續各分項技術之演算法研究。開始 DSP/FPGA 實現工作。
3. 第三年：完成 DSP/FPGA 實現工作。進行系統整合。亦繼續各分項技術之進一步研究。

在第一年的研究中，我們希望參與者都對 IEEE 802.16a 標準有相當程度的了解。各子計畫也不宜囿於圖一所示的責任範圍，而對其責任範圍以外的方塊無所了解，以致對不同子計畫間的溝通與未來的系統整合工作造成不良的影響。在第一年中，總計畫透過隔週一次的例會，促進各子計畫之間的溝通。我們也不禁止各子計畫(特別是要負責子系統或系統整合的子計畫)去了解與研究其他子計畫負責的系統方塊之相關技術。這樣做的效果在下節的子計畫成果簡述中可以得見。

二、結果與討論

以下簡單討論本計畫的成果，其中首先摘要敘述總體成果，然後分述各子計畫的成果。後者則係依圖一中傳輸信號處理之先後順序排列，而非依國科會核定之子計畫號數，且大體上係先討論 PHY 層而後 MAC 及 Link 層，以期條理能更清楚。

A. 總體成果摘要

除人才培育外，本計畫在第一年中的主要成果可歸納為三大類：

1. 對 IEEE 802.16a 標準的規範及其中所使用的技術的了解：前者如該標準中的訊框(frame)結構與嚮導載波(pilot carriers)的時頻域位置，後者如該標準所使用的通道編碼的特性。由於該標準是在計畫進行的過程中才定稿，所以我們是不斷追蹤該標準的發展，以了解其最新版的草稿與前一版之間的差異。我們也在持續注意 IEEE 802.16e 及 802.20 標準的發展，以視是否能應用其相關資訊以幫助本計畫之進行。
2. 符合 IEEE 802.16a 或其他 OFDM 傳輸方式下的行動無線多媒體信號傳輸相關演算法設計：依圖一中傳輸信號處理之先後順序排列，包括聲訊壓縮編碼法(子計畫四)、濾波器設計法(子計畫三)、時間同步法(子計畫一)、頻率同步法(子計畫一、二)、通道估計法(子計畫二、三)、通道碼解碼法(子計畫一)、及通道品質預測與

傳輸率控制法(子計畫二)。

3. 符合 IEEE 802.16a 或其他 OFDM 傳輸方式下的信號處理硬體架構設計：包括 OFDM 接收器(receiver)架構與 FFT 處理器架構(子計畫三)。

B. 無線網路串流聲訊研究及聲視訊子系統整合(子計畫四)

本子計畫幾項研究課題為：(1)進階音訊編碼(Advanced Audio Coding, AAC)編碼法位元率控制研究，及(2)切片式算術編碼(Bit-Sliced Arithmetic Coding, BSAC)研究。茲分別敘述於下。

關於 AAC 編碼法位元率控制之研究，我們知道多媒體壓縮標準如聲訊壓縮，只界定解碼端。編碼端許多參數選擇，是設計工程師的職責。好的參數選擇產生較好的效果。參數中最重要的一類是控制位元率，以達到較佳聲訊品質與較高壓縮比。我們嘗試了兩種不同想法，分述於下。

1. 有效位元分配方式：在聲訊壓縮中傳統位元分配方法是，觀察各頻帶的 NMR (noise-to-masking ratio)值，將位元優先配給 NMR 值最大者。但如從整體效率角度，這未必是成效最好的，例如下例：

Band	NMR (dB)	NMR-Gain/Bit
A	3.5	0.5
B	3	1.0

Band A 的 NMR 值較大，但給它 1 bit 只降低 0.5 dB。若給 Band B 則降低 1 dB。如果聽覺上，不論哪一個 band，降低 1 dB 都會比降低 0.5 dB 好，則選 Band B 較有效。此即為我們提出之原則 “give bits to the band with maximum NMR-gain/bit” 或 “retrieve bits from the band with maximum bits/NMR-loss”。實驗結果此法與窮舉搜尋法達到類似效果，但計算量為百分之一。

2. 快速位元分配參數搜尋法：窮舉搜尋法可達到最佳效果，但其計算量非常大，不合於實際應用。前人觀察到編碼參數前後間之重複關係，提出 trellis-based search，其計算量遠較窮舉搜尋法為低，但仍過高。因此我們將演算法切割為段落。但分開計算雖可以大量減少運算量，卻也可能大幅降低品質。其關鍵點在找出適當的位元編碼模型(virtual Huffman codebook model)。此外，我們從統計資料中找出篩減候選參數數目的規則性。實驗數據顯示，我們可降低計算量達 trellis-based search 的百分之一，而壓縮效果無甚差別。

關於 BSAC 之研究，我們知道 MPEG-4 第二版的聲訊壓縮標準提供了編碼率精細可調式的新工具。每個可調間距大約為 1 kbits/s/ch。這功能對一些頻寬容易變動的通訊系統，例如網際網路或行動通訊來說，是非常有用的。我們首先研究切片式算術編碼的聲質效能及其對於傳輸錯誤的敏感度。接著，我們提出兩種方法試圖改善切片式算術編碼的編碼效率。比較切片式算術編碼和進階聲訊編碼的聲質效能之後，我們對實驗結果進行分析，並提出造成兩者效能差異的可能原因。因為算術編碼是一種對傳輸錯誤很敏感的編碼方式，所以我們也研究了切片式算術編碼中的錯誤傳遞問題。在改善編碼效率方面，我們研究了在切片式算術編碼過程中中會用到的機率模型。我們

也設計並測試經由實際聲訊號產生的機率模型。另一個改善編碼效能的方法是改變每個可調層分配到的位元數。主要觀念在於分配更多的位元數給較低頻的可調層。這個方法將可以看到比較明顯的效能改善。

本子計畫本年度發表二篇會議論文，見附錄 A 及 B。

C. 寬頻正交分頻調變傳輸系統之傳輸信號處理技術研究及傳輸子系統整合(子計畫三)

本子計畫幾項研究課題為：(1) OFDM 接收器架構與硬體設計研究，(2) FFT 處理器設計研究，(3) 傳輸濾波器設計研究，及(4) 通道估計法研究。

在 OFDM 接收器架構與硬體設計研究方面，本子計畫除針對 802.16a 進行研究外，亦致力於探討寬頻 OFDM 型態的傳收機之高效能的、統一的(unified)數位訊號處理技術。這是指可程式及可調整的設計，能有效率的使用一個共通的 OFDM 核心硬體架構，而能彈性調整以適用於不同的通訊系統規格。也就是能有效的達到所謂多模式(multi-mode)或多標準(multi-standard)的運作。我們已提出一個多模式與多標準的 OFDM 內接收器架構。此架構具有低硬體成本之特性，亦可彈性調整以適用於不同之 OFDM 通訊系統，如 802.11a, 802.16a, DAB, 及 DVB。並已用 Verilog 硬體設計語言完成其合成與驗證。

在 FFT 處理器設計研究方面，我們提出一個可變長度的 FFT 處理器架構，可處理 256 至 8192 個子載波的範圍。此架構具小面積及高速率的優點，並已用 Verilog 硬體設計語言合成與驗證。此處理器中包括一個新穎的可變長度資料位址產生架構、一個新穎的 twiddle factor 位址產生架構、及一個新穎的 twiddle factor 產生架構。

在傳輸濾波器設計研究方面，我們提出一個新穎的有限脈衝響應長度(FIR)濾波器合成法。該法可用以產生低複雜度、高訊雜比的 cascade 型式的定點(fixed-point) FIR 濾波器。

在通道估計法研究方面，我們探討了一些既有的通道估計法的效能(含均方差、位元錯誤率、及運算複雜度)及其在 802.16a 環境下的應用。結果顯示基於 DFT 的通道估計及二維 Wiener 內插具有較佳的結果，只是運算複雜度較高。我們又提出二種基於 DCT 的通道估計法，結果顯示它們比基於 DFT 的方法為佳。我們又考慮在快速淡化(fast fading)通道環境下的通道估計，並提出二個方法，可有效降低快速淡化狀況下的道際干擾(ICI)。

本子計畫本年度發表二篇會議論文，見附錄 C 及 D。

D. 無線正交分頻多重進接頻道使用技術研究及全系統整合(子計畫一)

本子計畫幾項研究課題為：(1) IEEE 802.16a OFDMA 之同步技術研究，(2) IEEE 802.16a OFDMA 通道編解碼研究，及(3)最佳多載波調變方法研究。

在 IEEE 802.16a OFDMA 之同步技術研究方面，我們首先了解此標準之信號結構。由此我們知道：因為下行與上行的信號結構不同，運作狀況亦有異，所以須有不同的同步設計。例如：用戶台在開機時需要抓取下行訊號的訊框與精確的載波頻率，但上

行訊號則須根據基地台所規定的時間傳送，並須將其載波頻率控制在某一誤差之內；又如：下行與上行的 pilot 信號結構與頻域位置不同；再如：下行的訊號是由一個基地台傳來，但上行的訊號則可能來自不只一個用戶台，且其間可能有相當大的時間差異。不論下行或上行傳輸都需要作時間同步以偵測信號到達時間。如果估測錯誤會降低 guard interval (即 cyclic prefix) 用來防止多重路徑延遲造成符際干擾(ISI)的能力。此外，OFDM 系統對載波頻率偏移非常敏感，些許偏移即可能造成次載波之間的正交性喪失，因而在下行傳輸中特須對頻率偏移作同步處理。

經理論及模擬研究，我們將下行同步分為四級。第一及第二級利用 guard intervals 的循環相關性來估測 OFDM symbol 開始的時間及分數頻率偏移(fractional frequency offset)。第三級利用 guard bands 及部分 pilot carriers 來判斷整數頻率偏移。最後一級藉由下行傳輸的前置資訊(preamble)來判斷訊框開始的時間。上行同步則嘗試使用兩種方式做時間同步，分別是在時域及頻域將收到的訊號與上行傳輸的前置資訊作相關度分析，以找到相關度最大的時間。

在 IEEE 802.16a OFDMA 通道編解碼研究方面，IEEE 802.16a 的通道編碼採用串接碼(concatenated code)，外層是經過縮短(shorten)和穿孔(puncture)的里德-索羅門碼(Reed-Solomon code)，內層是咬尾(tail-biting)和穿孔的迴旋碼(convolutional code)。此外，在串接碼後接了一個位元交錯器(bit interleaver)和 M 階正交振幅調變器(M-ary QAM)。我們為整個通道編碼設計了解碼演算法。我們在加成性白色高斯通道下和淡化通道下分別模擬了里德-索羅門碼，迴旋碼，和串接碼，並把模擬之結果與一些分析的結果做比較，包含在加成性白色高斯通道下 IEEE 802.16a 的編碼增益需求值，用 Shannon 極限求出的編碼極限值，及用最短碼字間距求出的增益值。在加成性白色高斯通道下，里德-索羅門碼和迴旋碼的編碼增益值幾乎達到理論值，但整體的編碼增益值離理論值或 IEEE 802.16a 的編碼增益需求值卻有很大的差距。這與能找到的 802.16a Task Group 之文獻資料相吻合。我們預測在更高的號雜比下，實際的編碼增益值會比較接近理論計算出的編碼增益值。

在最佳多載波調變方法研究方面，目前一般實用的多載波調變係採用 DFT，但文獻上不乏討論在理論上更佳的方法，如使用 SVD (singular value decomposition)。然而 SVD 的運算量極大，且須隨通道變化而調整。不過，隨著數位電路的飛速發展，也許在不久的將來確可考慮使用 SVD。本子計畫對基於 SVD 的多載波調變在多路徑淡化通道中的性能及其與基於 DFT 的方法之間的比較，獲得了一些模擬分析結果。

本子計畫有數篇會議論文在投稿中。

E. 根據預測的無線通道情況控制即時資料的傳輸速率(子計畫二)

本子計畫幾項研究課題為：(1)淡化(fading)通道預測之研究，(2)載波頻率偏移估計之研究，及(3)通道估計方法研究。

關於淡化通道預測之研究，我們知道無線通道常具淡化效應。在通道情況不佳時傳送訊號，會因資訊錯誤而造成必須重傳。若有許多資訊需要重傳，就會對訊源編碼器及通道傳輸器的功能都有不利的影響。若能預估通道的好壞，就可彈性調整訊源編碼器或是通道傳輸器的輸出位元率，減少重傳的需要。具好壞兩個狀態的馬可夫鏈，所

謂 Gilbert 通道模型，常被用來作淡化通道的模型。我們探討如何由瑞利淡化通道建構 Gilbert 通道模型，並據此以預測未來的通道狀況，進而預測未來的位元錯誤率及須重傳的資料量。這個資料量就可以從原有的傳輸量中減去(例如降低訊源編碼器的輸出位元率)，以防止如緩衝器溢滿等問題。

關於載波頻率偏移估計之研究，現行的 OFDM 系統中，為了避免頻帶間的干擾，不會將所有的次載波都拿來傳送資料。這些沒有傳送資料的次載波，我們稱之為空載波。我們利用這些空載波與另外有傳送資料的次載波間的正交特性衍生出一套估計載波頻率偏移的方法。也就是說，在沒有載波頻率偏移的情形下(不考慮雜訊)，我們用這些空載波與收到的訊號作內積 結果將會是零。一旦有載波頻率偏移發生時，我們只要將這些空載波修正到正確的載波頻率偏移，再與接收訊號作內積，得到的結果也會是零。因此，根據收到訊號在修正過的空載波上的投影量，我們可以正確的決定載波頻率偏移。在考慮雜訊的情況下，投影量最小所對應的載波頻率偏移即是我們要的值。我們根據此原則設計了一個載波頻率偏移估計法，其中使用平行與疊代搜尋以降低運算量，結果顯示此方法有很好的效能。

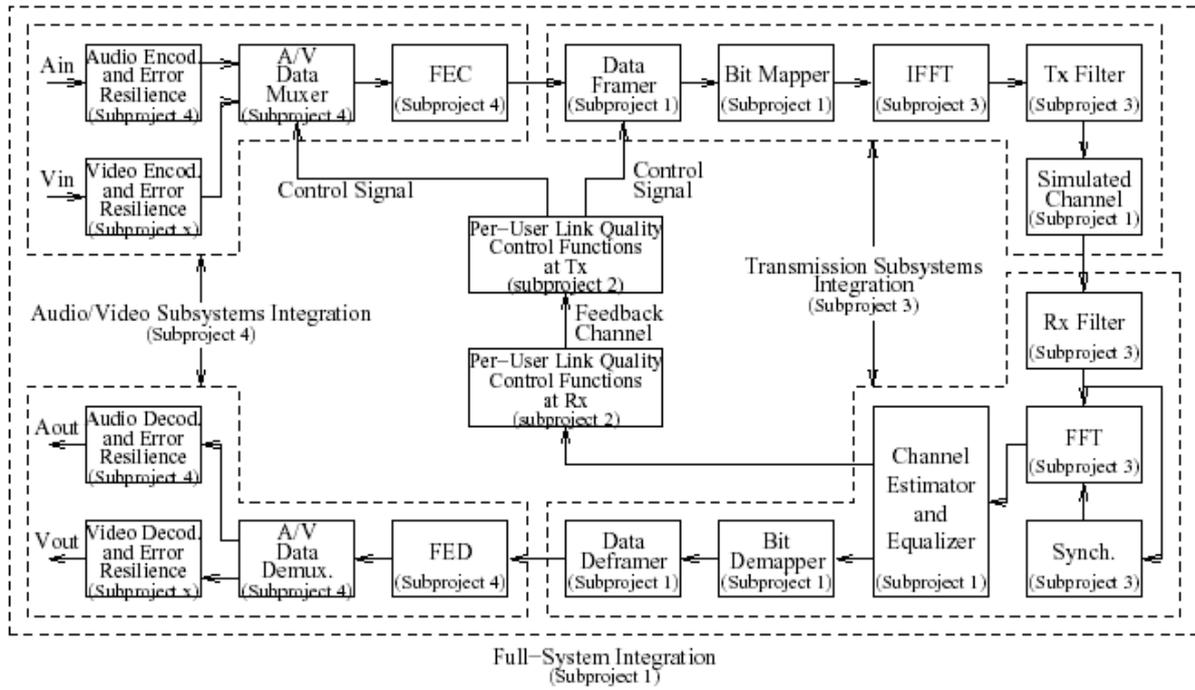
關於通道估計方法研究，OFDM 系統通常會傳送嚮導載波(pilot carriers)。這些 pilots 可用來作通道估計之用。文獻中曾有人提出 comb type 和 block type 兩種 pilots 配置的方法。我們透過模擬驗證 comb type 較不受都卜勒偏移的影響(故適合快速淡化的環境)，但對頻率選擇性淡化卻較敏感，而 block type 則相反。我們也利用適應性訊號處理的技巧以提昇估計效能。我們亦考慮 IEEE 802.16a 下鏈傳輸的通道估計，透過計算機模擬來看不同的延遲擴散(delay spread)和都卜勒偏移之下的通道估計表現。

本子計畫本年度發表一篇會議論文，見附錄 E。

三、參考文獻

- [1] IEEE Std 802.16a-2003 (Amendment to IEEE Std 802.16-2001), *IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Amendment 2: Media Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz*. New York: IEEE, Apr. 1, 2003.

四、圖表



圖一：計畫架構

五、計畫成果自評

研究內容與原計畫相符程度：達成第一年之主要目標，即對 IEEE 802.16a 標準規範及其中所使用之技術的了解，以及在 OFDMA 傳輸方式下之行動無線多媒體信號傳輸相關演算法設計。

達成預期目標情況：總計畫之主要成就，除設備之採購、計畫進行之協調等外，係透過其協調工作，促成各子計畫創新之發現、理論之推導、技術水準之提昇、計算機模擬軟體之建立、及人才培育等成果。

成果之學術與應用價值等：總計畫本身之價值，其中很重要的一點是在計畫進行中所累積的經驗與所建構的實驗環境，可為後續研究工作之用。各子計畫成果的學術價值高，迄現在已發表數篇學術論文，並有其他論文在陸續投稿中。應用價值方面，各子計畫的所設計的演算法與硬體架構，以及其相關模擬結果，可為 IEEE 802.16a 及相關無線多媒體 OFDM 傳輸系統開發之參考。

綜合評估：本總計畫及各子計畫獲得不少具有學術與應用價值的成果，並達人才培育之效。自評為「佳」。

六、可供推廣之研發成果(請見各子計畫報告)

七、附錄

本附錄共含五篇會議論文，如下列：

- A. C.-H. Yang and H.-M. Hang, “Efficient bit assignment strategy for perceptual audio coding,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp. V-405—V-408, April 2003 (4 pages).
- B. C.-H. Yang and H.-M. Hang, “Cascaded trellis-based optimization for MPEG-4 Advanced Audio Coding,” in *Audio Engineering Society 115th Convention*, New York, Oct. 2003 (9 pages).
- C. Y.-H. Yeh and S.-G. Chen, “Efficient channel estimation based on discrete cosine transform,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp. IV-676—IV-679, April 2003 (4 pages).
- D. C.-K. Chang, C.-P. Hung, and S.-G. Chen, “An efficient memory-based FFT architecture,” in *Proc. IEEE Int. Symp. Circuits Syst.*, pp. II-129—II-132, May 2003 (4 pages).
- E. W.-T. Chang, “Rate control for real time media based on predictive wireless channel condition,” poster presentation at 2003 年消息理論及通訊秋季講習會暨國科會成果發表會 (2003 Fall Workshop on Information Theory & Communication), 花蓮縣兆豐休閒農場, Aug. 26-27, 2003 (4 pages).

以上 A 及 B 為子計畫四所發表，C 及 D 為子計畫三所發表，E 為子計畫二所發表。

EFFICIENT BIT ASSIGNMET STRATEGY FOR PERCEPTUAL AUDIO CODING

*Cheng-Han Yang and Hsueh-Ming Hang***

Department of Electronics Engineering
National Chiao Tung University Hsinchu, Taiwan, R.O.C.

**h nhang@cc.nctu.edu.tw; Fax: (886)-3-5723283

ABSTRACT

For the purpose of efficient audio coding at low rates, a new bit allocation strategy is proposed in this paper. The basic idea behind this approach is “Give bits to the band with the maximum NMR-Gain/bit” or “Retrieve bits from the band with the maximum bits/NMR-Loss”. The notion of “bit-use efficiency” is suggested and it can be employed to construct a bit assignment algorithm operated at band-level as compared to the traditional frame-level bit assignment methods. Based on this strategy a new bit assignment scheme, called Max-BNLR, is designed for the MPEG-4 AAC. Simulation results show that the performance of the Max-BNLR scheme is significantly better than that of the MPEG-4 AAC Verification Model (VM) and is close to that of TB-ANMR [3], which is the (nearly) optimal solution. Moreover, the Max-BNLR scheme has the advantages of low computational complexity comparing to TB-ANMR.

1. INTRODUCTION

Many highly efficient and high quality audio coding schemes have been developed and proposed to meet the growing demand of multimedia applications. The MPEG-4 Advanced Audio Coding (AAC) is one of the most recent audio coder specified by the ISO/IEC MPEG standards committee [1]. It is a very efficient audio compression algorithm aiming at a wide variety of applications, such as Internet, wireless, and digital broadcast arenas [2]. For the applications where the bandwidth is very limited, the low rate audio coding with good quality becomes essential.

The procedure of bit assignment is one of the most important elements in audio coding. Particularly, when bits are scarce, how to make the best use of the limited bits is critical in producing the best quality audio. Up to now, the popular strategies on bit assignment are as follows ([2][3][5]).

1. “Give bits to the band which has the largest value of NMR (perceptual distortion).”
2. “Give bits to the bands of which the distortion is larger than the masking threshold”.

In these strategies, the bit-use (giving away bits) is considered at frame-level and only the value of distortion is taken into consideration at band-level. Hence, it is hard to control the bit-use efficiency (the NMR improvement due to adding one bit) at band level and thus results in a less efficient compression scheme.

In this paper, we suggest the notion of *bit-use efficiency* and propose a new strategy to improve the bit-use efficiency, which can be evaluated at band-level. Moreover, a new bit assignment scheme based on this new strategy is proposed for MPEG-4 AAC.

The organization of the paper is as follows. Section 2 describes the aforementioned new strategy. A new AAC bit assignment scheme is delineated in section 3. Finally, the complexity analysis and the simulation results are presented in section 4.

2. EFFICIENT BIT-USE STRATEGY

How to make use of the bits more efficiently is always the key issue in audio coding. The traditional strategies, “Giving bits to the band with the largest NMR” or “Giving bits to the bands of which the distortion is larger than masking threshold”, do not necessarily provide the best bit-use efficiency. For example, there are two candidate bands, A and B, and their NMR characteristics are listed in the table below. Which band should the first available bit be assigned to? In this table, NMR-Gain/bit means the gain in NMR by allocating one bit to this particular band. A more precise definition of NMR-Gain/bit will be given in section 3.

Band	NMR (dB)	NMR-Gain/bit
A	3.5	0.5
B	3	1.5

Following the traditional strategy, we would assign this one bit to band A; however, considering the bit-use efficiency, this one bit should be assigned to band B so that the overall NMR reduction is maximized. The essence of this new strategy can be summarized by the following statements.

“Give bits to the band with the maximum NMR-Gain/bit” or “Retrieve bits from the band with the

maximum bits/NMR-Loss”, where bits/NMR-loss is the number bits we save if we give away one unit of NMR.

3. MAX BITS/NMR-LOSS BIT ASSIGNMENT SCHEME

In this section, a new bit assignment scheme designed for MPEG-4 AAC based our new strategy is described. First, we define NMR-Gain/bit and bits/NMR-Loss by the following equations.

$$NMR - Gain / bit = \frac{(NMR_{ref} - NMR_{new})}{(bits_{new} - bits_{ref})} \quad (1)$$

$$\text{and } bits / NMR - Loss = \frac{(bits_{new} - bits_{ref})}{(NMR_{new} - NMR_{ref})}. \quad (2)$$

Figure 1 is the block diagram of the Max bits/NMR-Loss based bit assignment scheme. Each step in Figure 1 will be elaborated in the following sub-sections.

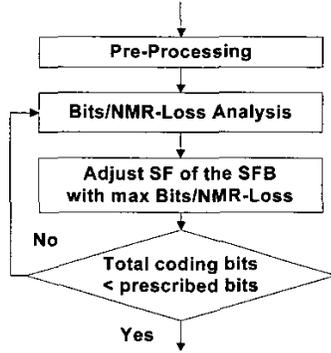


Figure 1. Max bits/NMR-Loss bit assignment scheme

3.1. Pre-Processing

The pre-processing step is to initialize two of the major parameters in the *bits/NMR-Loss* analysis: the reference NMR and the reference bits. There are no particular values associated with these parameters and thus the design of the pre-processing is case-dependent. In our implementation, we set the reference $NMR=1$ (0dB) for all the scale factor bands (SFB) at the beginning of processing a frame. After that, the reference scale factor (SF) for each SFB and the reference bits are calculated based on the input audio data.

3.2. Bits/NMR-Loss Analysis and SF Adjustment

In this scheme, only one SF value (of one SFB) is adjusted in one adjustment iteration. The detailed process is described below.

1. Initialization. Get the reference bits (B_{ref}), and the reference SFs (sf_{ref}) and NMRs (NMR_{ref}) for all SFBs (N_SFB SFB in total) from the pre-processing step.

Start the max *bits/NMR-Loss* analysis from the first SFB and thus set the SFB index $i=1$.

2. Find the local max *bits/NMR-Loss* ratio of the i th SFB, $BNLR_i$, by computing

$$BNLR_i = \max_{sf} \{ (B_{ref} - B_{sf}) / (NMR_{sf,i} - NMR_{ref,i}) \}$$

$$\forall sf \text{ and } sf_{ref,i} < sf \leq sf_{max,i}$$

The B_{sf} is the new value of the total coding bits in the current frame if the SF value (of the i th SFB) is changed from $sf_{ref,i}$ to $sf_{sf,i}$. The $sf_{max,i}$ is the SF value that quantizes all the spectral coefficients in the i th SFB to zero. The local optimal SF (of the i th SFB), $sf_{opt,i}$, is the SF with the maximum $BNLR$. The local optimal coding bits of the i th SFB, $B_{opt,i} = B_{sf_{opt,i}}$, is also recorded.

3. If $i < N_SFB$, update i to $i+1$ and go to step 2.

4. Find the global maximum *bits/NMR-Loss* ratio, $BNLR_{global}$ by computing

$$BNLR_{global} = \max_i \{ BNLR_i \} \quad \forall i, 0 \leq i < N_SFB$$

The global optimal SFB, $sf_{b_{global}}$, is the SFB that has the $BNLR_{global}$. Then, the global optimal SF, sf_{global} , is the local optimal SF of the $sf_{b_{global}}$ -th SFB. Similarly, the global optimal coding bits, B_{global} , is the coding bits of the $sf_{b_{global}}$ -th SFB.

5. Set the SF of the $sf_{b_{global}}$ -th SFB to sf_{global} . Update parameters for the $sf_{b_{global}}$ -th SFB; that is, $sf_{ref, sf_{b_{global}}} = sf_{global}$ and $NMR_{ref, sf_{b_{global}}} = NMR_{sf_{global}, sf_{b_{global}}}$.

6. Compare B_{global} to the prescribed rate, B . If $B_{global} > B$, update B_{ref} to B_{global} and go to step 2.

Note that, in performing the local maximum *bits/NMR-Loss* ratio analysis in step 2, only the SF of one SFB that is being examined is modified. The SF of the other SFBs are kept unchanged.

3.3. Trellis-Based Optimization on HCB

Total coding bits calculation in step 2 in the **Bits/NMR-Loss Analysis** (in sub-section 3.2) is one of the most computational-intensive processes. When the SF for each SFB is determined, the quantized spectral coefficients are also fixed. Before calculating the total coding bits, the HCB for each SFB has to be chosen first. The MPEG-4 AAC Verification Model (VM) has a simple algorithm for this purpose; however, a more efficient algorithm is needed for HCB decision. Thus, we adopt the Viterbi-based approach in this paper.

The problem for finding the optimal HCB can be reformulated as minimizing the following cost function:

$$C_{HCB} = \sum_i b_i + R(h_{i-1}, h_i), \quad (3)$$

where b_i is the coding bits of the quantized spectral coefficients for the i th SFB, h_i is the HCB for the i th SFB, and R is the run-length coding function (bits needed) for coding HCB. We find that the contribution of h_i to C_{HCB} depends only on the *previous* choice, h_{i-1} . Therefore, the minimization of C_{HCB} can be achieved by finding the optimal path through the trellis using the Viterbi algorithm.

A trellis is thus constructed for minimizing C_{HCB} . Each stage in the trellis corresponds to an SFB and each state at the i th stage represents a HCB candidate for this scale factor band. In other words, for the i th stage, if a path passes through the m th state, the m th HCB is employed for encoding the i th SFB. The Viterbi search procedure is outlined below.

The k th state at the i th stage is denoted by $S_{k,i}$ and the minimum accumulative-partial cost ending at $S_{k,i}$ is denoted by $C_{k,i}$. The transition cost from $S_{n,i-1}$ to $S_{m,i}$ is $R(h_{n,i-1}, h_{m,i})$.

1. Initialize $C_{m,0} = 0, \forall m$. Initialize $i=1$.
2. *Search.* $\forall m$, the best path ending at $S_{m,i}$ is found by computing

$$C_{m,i} = \min_n \{C_{n,i-1} + b_{m,i} + R(h_{n,i-1}, h_{m,i})\}$$

3. If $i < N$, set $i = i+1$ (SFB) and go to step 2.

3.4. Fast algorithm for Bits/NMR-Loss Analysis

The most time-consuming computation in this bit assignment scheme is the trellis-based HCB optimization for coding bits calculation in step 2 (Search). For each SF modification in step 2, the new value of total coding bits needs to be recalculated. Therefore, for one SF adjustment iteration, we need to perform $(sf_{\max,i} - sf_{\text{ref},i})$ times trellis-based HCB optimization processes for the local *bits/NMR-Loss* analysis. Hence, the total number of calculations for finding the global maximum *bits/NMR-Loss* is

$$\sum_{i=1}^{N_{SFB}} (sf_{\max,i} - sf_{\text{ref},i}). \quad (4)$$

There are at least two ways to reduce computations. One is to reduce the complexity of the trellis-based HCB optimization; the other is to reduce the number of trellis-based HCB optimization.

By analyzing the local optimal parameters, $sf_{\text{opt},i}$ and $BNLR_i$, we find some interesting properties.

1. The average value of the difference between the local optimal SFs of the m th and the $(m+1)$ th iterations, $sf_{\text{diff},ave}$, is often close to zero.

$$sf_{\text{diff},ave} = \frac{1}{(N_{SFB} - 3)} \times \sum_{i \in S} \text{abs}(sf_{\text{opt},i}^{m+1} - sf_{\text{opt},i}^m),$$

where $S = \{sf_{\text{global}}^m - 1, sf_{\text{global}}^m, sf_{\text{global}}^m + 1\}$ and sf_{global}^m is the global optimal SFB of the m th SF adjustment iteration. 2. The average value of the difference between the local max *bits/NMR-Loss* ratio of the m th and the $(m+1)$ th iteration, $BNLR_{\text{diff},ave}$, is typically quite small.

$$BNLR_{\text{diff},ave} = \frac{1}{(N_{SFB} - 3)} \times \sum_{i \in S} \text{abs}(BNLR_i^{m+1} - BNLR_i^m)$$

Using these two properties, we can drastically reduce the number of *bits/NMR-Loss* analyses (trellis-based HCB optimizations). We only need to perform the *bits/NMR-Loss* analysis on three SFBs after the first SF adjustment iteration.

4. SIMULATION RESULTS

The computational complexity and objective quality based on our simulations are summarized in this section. The bits assignment schemes used in comparison are as follows.

- (1) The MPEG-4 VM of AAC (VM-TLS) without modification.
- (2) The modified MPEG-4 VM of AAC (VM-TLS-M), in which the HCB decision algorithm is replaced by the TB-HCB optimization procedure described in section 3.3.
- (3) The trellis-based ANMR optimization (TB-ANMR) and the MNMR optimization (TB-MNMR), which are implemented as described in [3] and [4].
- (4) The normal and fast max *bits/NMR-Loss* schemes (max-BNLR).

Ten audio files with sampling rate 44.1K are used as test sequences. Two of them are extracted from MPEG SQAM [6], and the others are from EBU [7].

4.1. Computational complexity

The storage and computational complexity of one iteration in various schemes are summarized in Table 1.

Table 1. Complexity Analysis

	Search complexity	Storage
VM-TLS	1	--
VM-TLS-M	$12^2 \times N_{SFB}$	$12 \times N_{SFB}$
TB-ANMR TB-MNMR	$(60 \times 2)^2 \times 12^2 \times N_{SFB}$	$60 \times 2 \times 12 \times N_{SFB}$
Max-BNLR	$N_{SFB} \times \text{Ave}_{SF} \times 12^2 \times N_{SFB}$	$12 \times N_{SFB}$
Fast Max-BNLR	$\times (a) (N_{SFB} \times \text{Ave}_{SF} \times 12^2 \times N_{SFB})$ $(b) 3 \times \text{Ave}_{SF} \times 12^2 \times N_{SFB}$	$12 \times N_{SFB}$
※ (a) is only for the first iteration; all the rest are using (b)		

In this table, Ave. SF is the average number of SF tested for the max BNLR analysis for each SFB and its typical value is around 17 or so. Table 2 is the statistics collected from the simulations on audio sequences. It is clear that in terms of computational requirement:

Fast Max-BNLR << Max-BNLR << TB-ANMR(MNMR)

Table 2. Statistics on Computational Complexity

	Average iteration /frame	Average TB HCB optimizations/frame	Average TB HCB optimizations/iteration	Complexity ratio
TB-ANMR (MNMR)	12	14400*12	14400	1
Max-BNLR	50	10103	10103/12 = 842	1/17
Fast Max-BNLR	50	1153	1153/12 = 96	1/150

4.2. Objective results

Two common objective quality measurements, average noise to mask ratio (ANMR) and maximum noise to mask ratio (MNMR) [5], are adopted in the performance comparison. Note that, in evaluating distortion, the NMR is set to 0 dB if the original NMR value is less than 0 dB. The rate-distortion curves of six bit assignment schemes are shown in Figures 2 and 3. (Note: TB-ANMR and TB-MNMR are similar algorithms aiming at two different target NMRs.) We can find that the ANMR performance of the Max-BNLR scheme is almost as good as that of TB-ANMR. There is almost no loss of ANMR performance in using the fast algorithm for Max-BNLR either. The MNMR values of TB-ANMR, Max-BNLR and Fast Max-BNLR are also similar. The characteristic of the proposed Max-BNLR scheme is closer to that of TB-ANMR as compared to TB-MNMR. Again, TB-ANMR and TB-MNMR are the optimal solutions tuned for their target cost functions, ANMR and MNMR, respectively [3][4].

4. CONCLUSIONS

In this paper, we propose a new concept, bit-use efficiency, for improving audio coding performance. Furthermore, a new bits assignment scheme based on this new concept (strategy) is proposed for MPEG-4 AAC, named Max-BNLR. Simulation results show that the Max-BNLR scheme has a performance close to TB-ANMR and is much better than the MPEG VM. In addition, its computational complexity is much lower than that of TB-ANMR.

5. REFERENCES

- [1] ISO/IEC JTC1/SC29, "Information technology – vary low bitrate audio-visual coding," *ISO/IEC IS-14496 (Part 3, Audio)*, 1998
- [2] M. Bosi, et al., "ISO/IEC MPEG-2 advanced audio coding," *Journal of Audio Engineering Society*, vol. 45, pp. 789-814, October 1997.
- [3] A. Aggarwal, S.L. Regunathan, K. Rose, "Trellis-based optimization of MPEG-4 advanced audio coding," *Proc. IEEE Workshop on Speech Coding*, pp. 142-4 2000.
- [4] A. Aggarwal, S.L. Regunathan, K. Rose, "Near-optimal selection of encoding parameters for audio coding," *Proc. of ICASSP*, vol. 5, pp. 3269-3272, Jun 2001.
- [5] H. Najafzadeh and P. Kabal, "Perceptual bit allocation for low rate coding of narrowband audio," *Proc. of ICASSP*, vol. 2, pp. 893-896, 2000.
- [6] "The MPEG audio web page." <http://www.tnt.uni-hannover.de/project/mpeg/audio>.
- [7] European Broadcasting Union, *Sound Quality Assessment Material: Recordings for Subjective Tests*, Brussels, Belgium, Apr. 1988.

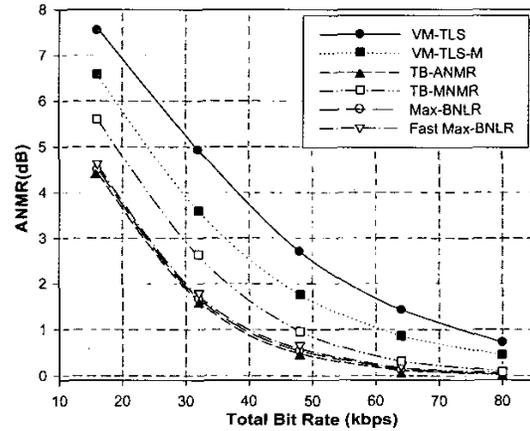


Figure 2. ANMR rate-distortion analyses

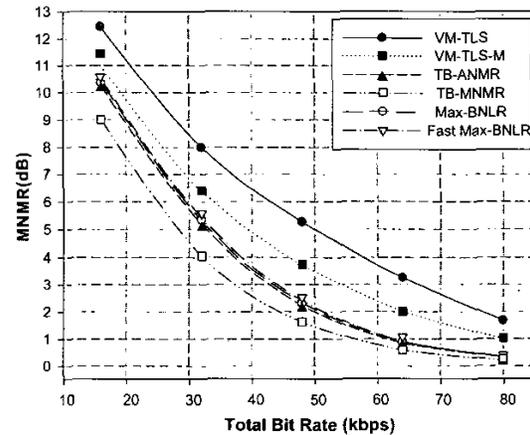


Figure 3. MNMR rate-distortion analyses



Audio Engineering Society Convention Paper

Presented at the 115th Convention
2003 October 10–13 New York, New York

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Cascaded Trellis-Based Optimization For MPEG-4 Advanced Audio Coding

Cheng-Han Yang¹, Hsueh-Ming Hang¹

¹Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

hmhang@mail.nctu.edu.tw; Fax: (886)-3-5723283
u8911831@cc.nctu.edu.tw; Fax: (886) -3-5731791

ABSTRACT

A low complexity and high performance scheme for choosing MPEG-4 Advanced Audio Coding (AAC) parameters is proposed. One key element in producing good quality compressed audio at low rates in particular is selecting proper coding parameter values. A joint trellis-based optimization approach has thus been previously proposed. It leads to a near-optimal selection of parameters at the cost of extremely high computational complexity. It is, therefore, very desirable to achieve a similar coding performance (audio quality) at a much lower complexity. Simulation results indicate that our proposed cascaded trellis-based optimization scheme has a coding performance close to that of the joint trellis-based scheme, and it requires only 1/70 in computation.

1. INTRODUCTION

To meet the demand of various multimedia applications, many high-efficient audio coding schemes have been developed. The MPEG-4 Advanced Audio Coding (AAC) is one of the most recent-generation audio coders specified by the ISO/IEC MPEG standards committee [1]. It is a very efficient audio compression algorithm aiming at a wide variety of different applications, such as Internet, wireless, and digital broadcast arenas [2].

One key element in an AAC coder is selecting two sets of coding parameters properly, the scale factor

(SF) and Huffman codebook (HCB) in the rate-distortion (R-D) loop. Because encoding these parameters is inter-band dependent, i.e., the coded bits produced for the second band depend on the choice of the first band, the choice of their proper values so as to minimize the objective quality becomes fairly difficult. As discussed in [3][4], the poor choice of parameters for rate control is one shortcoming of the current MPEG-4 AAC Verification Model (VM) and therefore its compression efficiency is not as expected at low bit rates.

Some methods such as vector quantizers rather than scalar quantizers have been suggested to reduce the side information [5][6]. They would alter the syntax of the standards. In this paper, we focus on finding the parameters in the existing AAC standard that produce the (nearly) optimal compressed audio quality for a given bit rate.

In [3] and [4], a joint optimization scheme, which takes the inter-band dependence into account, is proposed for choosing the encoding parameters for all the frequency bands. This joint optimization is formulated as a trellis search and is, therefore, called *trellis-based optimization*. Although the complexity of this joint trellis-based optimization scheme can be reduced by adopting the Viterbi algorithm, its search complexity is still extremely high and is thus not suitable for practical applications.

In this paper, we propose a *cascaded trellis-based* (CTB) optimization scheme for selecting the proper encoding parameters. Our scheme retains the good audio quality offered by the joint trellis-based (JTB) optimization while its search complexity is drastically decreased.

The organization of this paper is as follows. In section 2, a brief overview of MPE-4 AAC is provided. The proposed CTB scheme with several variations for choosing the optimal coding parameters is described in sections 3, 4 and 5. The algorithm complexity analysis and the simulation results are summarized in section 6.

2. OVERVIEW OF AAC ENCODER

The basic structure of the MPEG-4 AAC encoder is shown in Figure 1. The time domain signals are first converted into the frequency domain (spectral coefficients) by the modified discrete cosine transform (MDCT). For tying in with the human auditory system, these spectral coefficients are grouped into a number of bands, called scale factor bands (SFB). The pre-process modules, which are the optional functions, can help removing the time/frequency domain redundancies of the original signals. The psychoacoustic model calculates the spectral coefficient masking threshold, which is the base for deciding coding parameters in the R-D loop. The R-D loop, our focus in this paper, is to determine two critical parameters, SF and HCB for each SFB so as to optimize the selected criterion under the given bit rate constraint. The SF is related to the step size of the quantizer, which determines the quantization noise-to-masking ratio (NMR) in each band. The quantized coefficients are then entropy-coded by one of the twelve pre-designed HCBs. In addition, the

indices of SFs and HCBs are coded using differential and run-length codes respectively and are transmitted as side information.

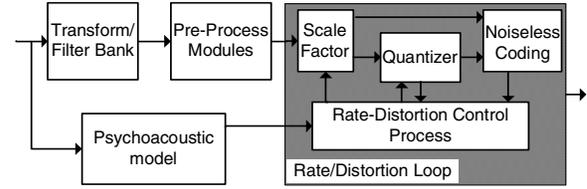


Fig. 1. Basic structure of the MPEG-4 AAC encoder

3. CASCADED TRELLIS-BASED OPTIMIZATION

The JTB optimization approach can substantially enhance the coding performance at low bit rates [3][4]. However, this approach also results in a very high computational complexity. The coding parameters in the JTB scheme, SF and HCB, are optimized simultaneously by using the trellis search. The states at the i th stage in the trellis for the JTB scheme represent all combinations of SF and HCB for the i th SFB. Different from the JTB scheme, our scheme, so-called *cascaded trellis-based scheme* (CTB), finds the proper coding parameters, SF and HCB, in two consecutive steps. The search complexity can thus be drastically reduced, while the advantage of trellis-based optimality is mostly retained.

The way that the trellis search performs depends on the optimization criterion it adopts. There are two frequently used criteria, the average noise-to-mask ratio (ANMR) and the maximum noise-to-mask ratio (MNMR) [7]. Both criteria will be used in this paper.

3.1. Trellis-Based ANMR Optimization on SF

The constrained optimization problem for the ANMR criterion is formulated as below.

$$\min \sum_i w_i d_i \quad \text{s.t.}$$

$$\sum_i (b_i + D(sf_i - sf_{i-1}) + R(h_{i-1}, h_i)) \leq B$$

, where w_i is the inverse of the masking threshold and d_i is the quantization distortion. Under this criterion, we minimize the sum of the perceptually weighted distortion. The coding parameters, SF and HCB, for the i th SFB is denoted by sf_i and h_i . Symbol D is the differential coding function performed on SF and symbol R is the run-length coding function performed on HCB. The returned function values in both cases are bits to encode the arguments. Parameter b_i is the bits for coding the quantized

spectral coefficients (QSCs) and the parameter B is the prescribed bit rate for a frame.

As described in [3], the ANMR optimization problem can be reformulated as minimizing the unconstrained cost functions, C_{ANMR} , with the Lagrangian multiplier λ .

$$C_{ANMR} = \sum_i w_i d_i + \lambda \cdot (b_i + D(sf_i - sf_{i-1}) + R(h_{i-1}, h_i)) \quad (1)$$

Different from that in the original JTB scheme, the optimization problem in our CTB scheme is reformulated as minimizing two unconstrained cost functions, C_{SF_ANMR} and C_{HCB} , as follows.

$$C_{SF_ANMR} = \sum_i w_i d_i + \lambda \cdot (b_i + D(sf_i - sf_{i-1})) \quad (2)$$

$$C_{HCB} = \sum_i b_i + R(h_{i-1}, h_i) \quad (3)$$

The minimization of C_{SF_ANMR} is described in this sub-section, and the minimization of C_{HCB} will be described in section 3.3.

Similar to the approach in the JTB scheme, the goal for finding proper SFs that minimize C_{SF_ANMR} can be achieved by finding the optimal path through the trellis. Each stage in the trellis corresponds to an SFB. (There are N_SFB stages in total.) However, different from JTB, each state at the i th stage in our scheme only represents a SF candidate for the i th SFB. In other words, at the i th stage, if a path passes through the m th state, it means that the m th SF candidate is employed to encode the i th SFB. For a given value of λ , the Viterbi search procedure described in [3] is modified as stated below.

The k th state at the i th stage is denoted by $S_{k,i}$ and the minimum accumulative-partial cost ending at $S_{k,i}$ is denoted by $C_{k,i}$. The state-transition cost, $T_{l,i-1 \rightarrow k,i}$, from $S_{l,i-1}$ to $S_{k,i}$ is $\lambda \cdot D(sf_{k,i} - sf_{l,i-1})$.

1. Initialize $C_{k,0} = 0, \forall k$ and $i=1$.
2. Search for, $\forall k$, the best path ending at $S_{k,i}$ by computing

$$C_{k,i} = \min_l \{ C_{l,i-1} + w_i d_{k,i} + \lambda \cdot b_{k,i} + T_{l,i-1 \rightarrow k,i} \} \quad (4)$$

3. If $i < N_SFB$, set $i = i+1$ and go to step 2.

3.2. Trellis-Based MNMR Optimization on SF

The constrained optimization problem for the MNMR criterion is formulated below.

$$\min \left(\max_i w_i d_i \right) \quad \text{s.t.} \\ \sum_i (b_i + D(sf_i - sf_{i-1}) + R(h_{i-1}, h_i)) \leq B$$

, where $\max_i w_i d_i$ is the maximum NMR in a frame.

Again, using the unconstrained format, the cost function in the JTB scheme [4] becomes

$$C_{MNMR} = \sum_i b_i + D(sf_i - sf_{i-1}) + R(h_{i-1}, h_i) \quad (5)$$

Different from the cost function in the JTB scheme [4], the MNMR optimization problem in our CTB scheme is reformulated as the minimization of two cost functions, C_{SF_MNMR} and C_{HCB} (Eqn.(3)), under the constraint: $w_i d_i \leq \lambda, \forall i$, for some constant value of λ ,

$$C_{SF_MNMR} = \sum_i b_i + D(sf_i - sf_{i-1}) \quad (6)$$

Similar to the trellis-based ANMR optimization on selecting SF, a trellis is constructed for minimizing C_{SF_MNMR} and each state at the i th stage only represents a SF candidate for the i th SFB. The Viterbi search procedure described in [4] is modified as stated below. The state-transition cost, $T_{l,i-1 \rightarrow k,i}$, from $S_{l,i-1}$ to $S_{k,i}$ is $D(sf_{k,i} - sf_{l,i-1})$.

1. Initialize $C_{k,0} = 0, \forall k$ and $i=1$.
 2. Find the valid states for the i th stage, $S_{k,i}, \forall k$. A state is valid if the NMR ($w_i d_{k,i}$) corresponding to that state parameter is $\leq \lambda$.
 3. Search for, $\forall k$, the best path ending at the valid state $S_{k,i}$ by computing
- $$C_{k,i} = \min_l \{ C_{l,i-1} + b_{k,i} + T_{l,i-1 \rightarrow k,i} \} \quad (7)$$
4. If $i < N_SFB$, set $i = i+1$ and go to step 2.

As pointed in [3][4], in the trellis for selecting the optimal SF (for both ANMR and MNMR), each state is further split into two consecutive states. In the first state, the spectral coefficients are quantized using the assigned valid SF, and in the second state, all quantized values of spectral coefficients are set to zero.

3.3. Trellis-Based Optimization on HCB

The HCB optimization is performed under the condition that the SF for each SFB has already been determined. With a determined SF, QSCs (quantized spectral coefficients) for each SFB are fixed and thus the b_i term in the cost function C_{HCB} (Eqn.(3)) only depends on HCB. The optimization procedure here is to find the HCBs that minimize the cost function C_{HCB} and this can be achieved again by finding the optimal path through the trellis.

A trellis is thus constructed for minimizing C_{HCB} . Each stage in this trellis corresponds to an SFB (There are N_{SFB} stages in total.) and each state at the i th stage represents a HCB candidate for the i th SFB. In other words, at the i th stage, if a path passes through the m th state, the m th HCB candidate is employed for encoding the i th SFB. The state-transition cost, $T_{n,i-1 \rightarrow m,i}$, from $S_{n,i-1}$ to $S_{m,i}$ is $R(h_{n,i-1}, h_{m,i})$. The Viterbi search procedure for finding optimal HCBs is as follows.

1. Initialize $C_{m,0} = 0, \forall m$. Initialize $i=1$.
2. Search for, $\forall m$, the best path ending at $S_{m,i}$ by computing

$$C_{m,i} = \min_n \{ C_{n,i-1} + b_{m,i} + T_{n,i-1 \rightarrow m,i} \} \quad (8)$$
3. If $i < N_{SFB}$, set $i = i+1$ and go to step 2.

3.4. Cascaded Trellis-Based Optimization

The block diagram of the CTB optimization scheme is shown in Figure 2 and the processing steps are described below.

1. Initialize λ .
2. For a given λ , a set of optimal SF, sf_{opt} , is determined by the trellis-based SF optimization procedure using the Virtual-HCB Mode.
3. For the given sf_{opt} obtained from step 2, a set of optimal HCB, hcb_{opt} , is determined by the trellis-based HCB optimization procedure.
4. For the given hcb_{opt} obtained from step 3 and λ , a set of recalculated optimal SF, sf'_{opt} , is obtained from the Fixed-HCB Mode trellis-based SF optimization procedure.
5. *Adjust rate.* For the given optimal sf'_{opt} (or sf_{opt}) and hcb_{opt} , the total coding bit rate is calculated and compared to the prescribed bit rate (B). Adjust λ and go to step 2 if the constraint is not met.

In the preceding procedure, the trellis-based ANMR (MNMR) optimization on SF is applied to steps 2 and 4.

As described in the preceding procedure, the trellis-based optimization procedure on SF is operated in two different modes, Virtual-HCB Mode and Fixed-HCB Mode. In these two modes, the value of $b_{k,i}$ in Eqn.(4) and Eqn.(7) is determined in different ways, and these will be described in section 4.

The preceding procedure is called the *full optimization mode* (or *Two-Loop mode*), because the

optimization procedure on SF is done twice. The second optimization procedure on SF (step 4) can help in recovering some improper SF values determined in step 2. Furthermore, the CTB optimization can also be operated in a simpler optimization mode (so-called *One-Loop mode*), in which step 4 is not included.

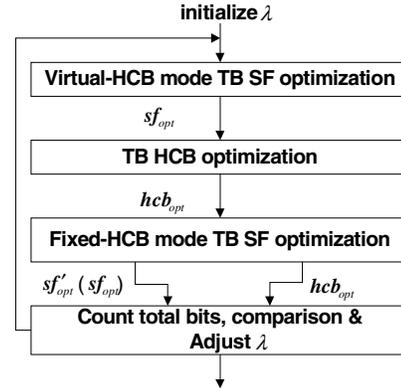


Fig. 2. Cascaded trellis-based optimization scheme

4. FIXED AND VIRTUAL HCB MODE FOR SF OPTIMIZATION

For an identified $C_{k,i}$ in Eqn.(4) and Eqn.(7) in sections 3.1 and 3.2, $w_i d_{k,i}$ or $D(sf_{k,i} - sf_{l,i-1})$ is unique for a given state or state transition. However, the value of $b_{k,i}$ depends not only on the state parameter $sf_{k,i}$; it also depends on the choice of HCB. In the JTB optimization scheme, for each candidate value of SF, all possible b values, corresponding to 12 pre-designed HCBs, are evaluated. But in our SF optimization scheme, we have to determine one proper value of $b_{k,i}$ for the state $S_{k,i}$. According to our implementation, the trellis-based ANMR or MNMR SF optimization scheme can operate in two modes, the Fixed-HCB mode and the Virtual-HCB mode, for determining the value of $b_{k,i}$.

In the Fixed-HCB mode, a set of fixed HCBs, $[h_1^f, h_2^f, \dots, h_{N_{SFB}}^f]$, is determined beforehand. For all the states at the i th stage, the QSCs, $q_{k,i}$, are encoded using h_i^f ; thus, $\forall k, b_{k,i} = h_i^f(q_{k,i})$.

In the Virtual-HCB mode, a Virtual HCB, $h_{k,i}^v$, is employed for state $S_{k,i}$. Thus, $h_{k,i}^v$ needs to be pre-constructed to help us in determining $b_{k,i}$ and it can be constructed in several ways. For example, $h_{k,i}^v$

may be one of the 12 pre-designed HCBs or a compound codebook. Consequentially, the more accurate of the $b_{k,i}$ and $h_{k,i}^v$ we can estimate, the higher accuracy of SF optimization we can achieve. In order to improve the accuracy of the estimated values of $b_{k,i}$ and $h_{k,i}^v$, we did some analysis on the JTB optimization scheme.

For a given value of λ , by applying the JTB scheme, we can find a set of optimal parameters, sf_{opt}^{JTB} , hcb_{opt}^{JTB} and b_{opt}^{JTB} that minimize the cost function C_{ANMR} (Eqn.(1)) or C_{MNMR} (Eqn.(5)). For comparison purpose, we also construct an ideal set of bits for coding QSCs, b_{min}^{JTB} . For the i th SFB, $b_{min,i}^{JTB}$ is the minimum value of bits for coding $q_{opt,i}^{JTB}$ using 12 pre-designed HCBs and is formulated as:

$$b_{min,i}^{JTB} = \min_m \{ h_m(q_{opt,i}^{JTB}) \}$$

, where $q_{opt,i}^{JTB}$ is QSCs quantized by $sf_{opt,i}^{JTB}$.

The histogram of the differences between b_{opt}^{JTB} and b_{min}^{JTB} , denoted by $b_{opt-min}^{JTB}$, is shown in Figure 3. We can find that over 91% of $b_{opt-min}^{JTB}$ is less than 3 for both ANMR and MNMR criteria. From the other viewpoint, we tend to choose the HCB that results in nearly the minimum QSCs bits.

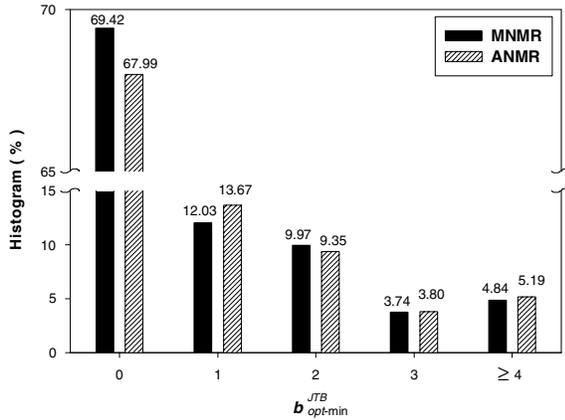


Fig. 3. Comparison against $b_{opt-min}^{JTB}$

Observing this characteristics of b_{opt}^{JTB} , we derive a rule in determining $h_{k,i}^v$ and $b_{k,i}$. For state $S_{k,i}$, the candidate $h_{k,i}^v$ values are the set of HCB that satisfies the proposed rule in Eqn.(9); namely,

$$h(q_{k,i}) \leq \min_m \{ h_m(q_{k,i}) \} + \delta \quad (9)$$

Then, $b_{k,i}$ is formulated as:

$$b_{k,i} = \frac{1}{|h_{k,i}^v|} \sum_{n \in h_{k,i}^v} h_n(q_{k,i}) + \alpha \cdot R_v(h_{l,i-1}^v, h_{k,i}^v) \quad (10)$$

R_v is the run-length coding function performed on the Virtual HCB and it is similar to R in our implementation.

$$R_v(h_{l,i-1}^v, h_{k,i}^v) = \begin{cases} 0, & \text{if } (h_{l,i-1}^v \cap h_{k,i}^v) \neq \emptyset \\ 9, & \text{else} \end{cases} \quad (11)$$

, where α is a weight for including $R_v(h_{l,i-1}^v, h_{k,i}^v)$ in $b_{k,i}$. Finally, we have to determine suitable values for δ and α . The simulation results of the normalized difference, $(C_{ANMR}^{JTB} - C_{ANMR}^{CTB})$ or $(C_{MNMR}^{JTB} - C_{MNMR}^{CTB})$, versus different values of δ and α are shown in Figure 4 and 5.

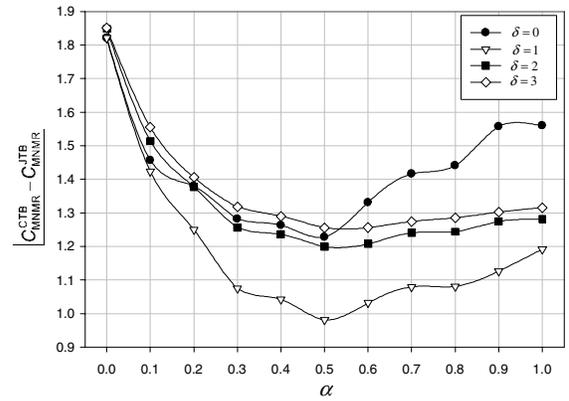


Fig. 4. $(C_{MNMR}^{JTB} - C_{MNMR}^{CTB})$ v.s (δ, α)

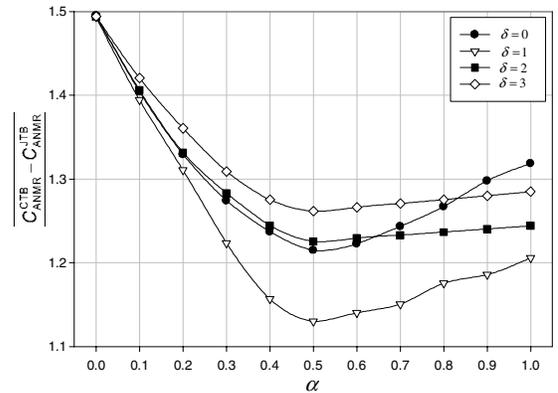


Fig. 5. $(C_{ANMR}^{JTB} - C_{ANMR}^{CTB})$ v.s (δ, α)

In this notation, C_{ANMR}^{JTB} (or C_{MNMR}^{JTB}) is the minimal C_{ANMR} (or C_{MNMR}) derived from the JTB scheme and C_{ANMR}^{CTB} (or C_{MNMR}^{CTB}) is the minimal C_{ANMR} (or C_{MNMR}) derived from the CTB scheme. We find that for a

wide range of δ values, we can achieve better performance when $R_v(h_{l,i-1}^v, h_{k,i}^v)$ is included in $b_{k,i}$ ($\alpha > 0$). As shown in Figure 4, the CTB scheme can achieve the nearly best performance when $\delta=1$ and $\alpha=0.5$. Therefore, we choose 1 for δ and 0.5 for α in our implementation.

5. FAST SEARCHING ALGORITHM

The computational complexity of the trellis-based optimization scheme depends on the searching range (number of states) of each stage in the trellis. Hence, reducing the candidate states at each stage is an effective way in reducing the complexity. Base on this idea, we propose fast searching algorithms for the trellis-based optimization schemes on SF and HCB.

5.1. Fast Searching Algorithm for HCB Optimization

In MPEG-4 AAC, SFs are differentially coded and HCBs are coded by run-length coding. Run-length coding can be viewed as a special case of differential coding; therefore, the procedure of trellis-based optimization on HCB is similar to that on SF. However, the output of run-length coding has only two possible values, either 0 or 9. In looking for $C_{k,i}$ in Eqn.(8), the cost of run-length coding is as follows.

$$R(h_{n,i-1}, h_{m,i}) = \begin{cases} 0, & \text{if } n = m \\ 9, & \text{else} \end{cases} \quad (12)$$

In HCB optimization, each state at the i th stage represents a HCB candidate. As shown in Figure 6(a), for finding the optimal path ending at $S_{m,i}$, all the HCB candidates at $(i-1)$ th stage have to be taken into consideration. In MPEG-4 AAC, there are 12 pre-designed HCBs, so the searching complexity for finding all the optimal paths ending at the i th stage is 12×12 .

The number next to the arrow in Figure 6 is the state-transition cost. Except for the path $S_{m,i-1} \rightarrow S_{m,i}$, the state-transition costs of the other paths ending at $S_{m,i}$ are all the same (equal to 9). Therefore, in calculating $C_{k,i}$ in Eqn.(8), among these 11 paths, the path with the smallest $C_{n,i-1}$ will result in the smallest $C_{m,i}$. Based on this property, a fast searching algorithm is proposed and is divided into two steps.

1. Among the 12 candidate states at $(i-1)$ th stage, the state with the minimum cost, $C_{\min,i-1}$, is chosen and treated as the virtual thirteenth state, $S_{\min,i-1}$.

$$C_{\min,i} = \min_n \{ C_{n,i-1} \}$$

2. As shown in Figure 6(b), while finding the optimal path ending at $S_{m,i}$, we only have to consider two paths, path ($S_{m,i-1} \rightarrow S_{m,i}$) and path ($S_{\min,i-1} \rightarrow S_{m,i}$). The rest of searching procedure is the same to that in section 3.3.

The searching complexity (in terms of branch metric calculation) of this fast algorithm is approximately $12 + 2 \times 12$. The first “12” term is the computational complexity for determining $S_{\min,i-1}$. Note that the performance (accuracy) of the fast searching algorithm is the same to that of the full searching algorithm.

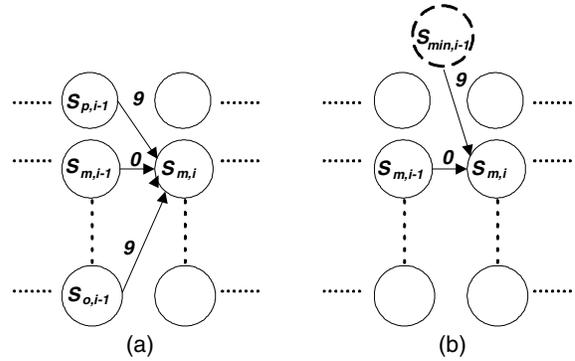


Fig. 6. Trellis representation of HCB optimization

5.2. Fast Searching Algorithm for SF Optimization

In SF optimization, each state in the trellis represents a candidate value of SF. Searching over a larger set of SF candidates can result in better performance, but at the cost of higher searching complexity.

In general, the state numbers (sn) for all the stages in the trellis are the same and the searching complexity for each stage transition in this uniform sn trellis is $sn \times sn$. In this section, we propose two non-uniform (adaptive) sn algorithms, in which the sn for each stage in the trellis can vary to reduce the searching complexity. The first one is called “global minimum reference SF restricted non-uniform trellis”, or “Gm_Nu” in short, and the second one is called “local minimum reference SF restricted non-uniform trellis”, or “Lm_Nu”. In both cases, a reference SF is first identified and then the number of candidates is reduced against this reference.

In the first step, we define the reference SF, sf_i^{ref} , for i th SFB as the largest SF among all the valid states at the i th stage. Then we can find a global

minimum reference SF, $sf_{G_Min}^{ref}$, which is the minimum SF among all the reference SFs. In the Gm_Nu algorithm, we restrict the SF candidates at the i th stage in the range of $[sf_i^{ref}, sf_{G_Min}^{ref} - \epsilon]$. Thus, the sn at the i th stage, $sn_{Gm,i}$, equals to $(sf_i^{ref} - sf_{G_Min}^{ref} + 1 + \epsilon)$.

Next, we define the n th-order local reference minimum SF at the i th stage, $sf_{L_Min,i}^{ref}$, where

$$sf_{L_Min,i}^{ref} = \min_{i-n \leq j \leq i+n} \{sf_j^{ref}\} \quad (13)$$

In the Lm_Nu algorithm, we restrict the SF candidates at the i th stage in the range of $[sf_i^{ref}, sf_{L_Min,i}^{ref} - \epsilon]$. Therefore, the sn for the i th stage, $sn_{Lm,i}$, equals to $(sf_i^{ref} - sf_{L_Min,i}^{ref} + 1 + \epsilon)$. In both cases, ϵ is a parameter to control the searching range for all stages. In the simulations in section 6, the values of n in Eqn.(13) and ϵ are both set to 1.

6. SIMULATION RESULTS

In this section, we will discuss the computational complexity and the coded audio quality in our experiments. Three types of bits allocation algorithms have been tested and compared as described below.

- (1) The MPEG-4 VM of AAC (VM-TLS).
- (2) The joint trellis-based ANMR and MNMR optimization schemes, abbreviated as JTB-ANMR and JTB-MNMR respectively, described in [3] and [4].
- (3) The cascaded trellis-based ANMR and MNMR optimization schemes, abbreviated as CTB-ANMR CTB-MNMR respectively, described in section 3.

Ten two-channel audio sequences with sampling rate 44.1kHz are tested. Two of them are extracted from MPEG SQAM [8], and the others are from EBU [9].

6.1. Complexity Analysis

The computational complexity analyses for the aforementioned several coding schemes are summarized in Table 1. The value in ‘‘Computation’’ column is the searching complexity in calculating one stage transition in the trellis in terms of branch metric computation. For the convenience of comparison, the full-search JTB is set as the reference (ratio=1) and all the other schemes are rated based this base. Also shown in Table 1 is the storage requirement. Again, it is measured in terms of one branch metric computational needs.

We can find from Table 1 that the n -Loop CTB scheme is approximately $(142/n)$ times faster than the JTB scheme. Moreover, the storage requirement for

the trellis search in the CTB scheme is much smaller than that in the JTB scheme.

For the JTB scheme, the fast HCB optimization algorithm can reduce the complexity to 1/4. Note that sn_{Gm}^{ave} and sn_{Lm}^{ave} in Table 1 are the average sn in the Gm_Nu and Lm_Nu algorithms and are calculated by Eqn.(14) and Eqn.(15).

$$sn_{Gm}^{ave} = \left(\frac{1}{NB_SFB} \sum_{i=1}^{NB_SFB} (sn_{Gm,i-1} \cdot sn_{Gm,i}) \right)^{1/2} \quad (14)$$

$$sn_{Lm}^{ave} = \left(\frac{1}{NB_SFB} \sum_{i=1}^{NB_SFB} (sn_{Lm,i-1} \cdot sn_{Lm,i}) \right)^{1/2} \quad (15)$$

The simulation results show that typical sn_{Gm}^{ave} is approximately 12 and sn_{Lm}^{ave} is about 5. Hence, the Gm_Nu algorithm can reduce the complexity to 1/25 and the Lm_Nu algorithm can reduce the complexity to 1/144.

Table 1. Complexity Analysis

	Computation	Ratio	Storage
JTB	$(60 \times 2)^2 \times 12^2$	1	$60 \times 2 \times 12$
n -Loop CTB	$n \times (60 \times 2)^2 + 12^2$	$n / 142$	60×2
JTB + Fast HCB	$(60 \times 2)^2 \times 36$	1/4	$60 \times 2 \times 12$
JTB + Fast HCB + Gm_Nu	$(sn_{Gm}^{ave} \times 2)^2 \times 36$	1/100	$60 \times 2 \times 12$
JTB + Fast HCB + Lm_Nu	$(sn_{Lm}^{ave} \times 2)^2 \times 36$	1/576	$60 \times 2 \times 12$
n -Loop CTB + Gm_Nu + Fast HCB	$n \times (sn_{Gm}^{ave} \times 2)^2 + 36$	$n / 3600$	60×2
n -Loop CTB + Lm_Nu + Fast HCB	$n \times (sn_{Lm}^{ave} \times 2)^2 + 36$	$(n+0.4) / 20736$	60×2

6.2. Objective Quality Analysis

The rate-distortion curves of these bit allocation schemes are shown in Fig. 7 and 8. Two major evaluative methodologies, ANMR and MNMR, are used for distortion. We can find that the performance of the CTB scheme is similar to that of the JTB scheme. The ANMR performance loss is less than 0.2dB for One-Loop CTB-ANMR and less than 0.1dB for Two-Loop CTB-ANMR (the lowest three curves in Fig. 7). The MNMR performance loss is less than 0.1 dB for both One- and Two-Loop CTB-MNMR (the lowest three curves in Fig. 8). Both of them are much better than the MPEG-4 VM (the top line).

The differences of performance between the fast searching algorithms and the original CTB-MNMR scheme are shown in Fig. 9 and 10. In light of the complexity analyses on Gm_Nu and Lm_Nu, and the uniform NB_SF fast algorithms, with $NB_SF=12$ and 5, are chosen for comparison. There is nearly no performance loss for the Gm_Nu algorithm (ANMR or MNMR Difference ≈ 0). The advantage of the non-uniform algorithms over the uniform algorithms at about the same complexity is clearly shown in Figs. 9 and 10.

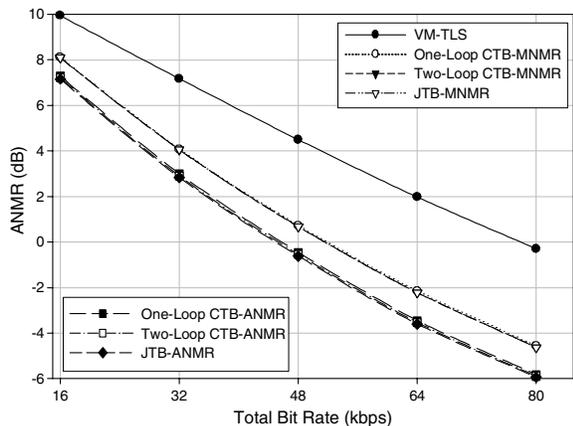


Fig. 7. ANMR Rate-Distortion Analysis

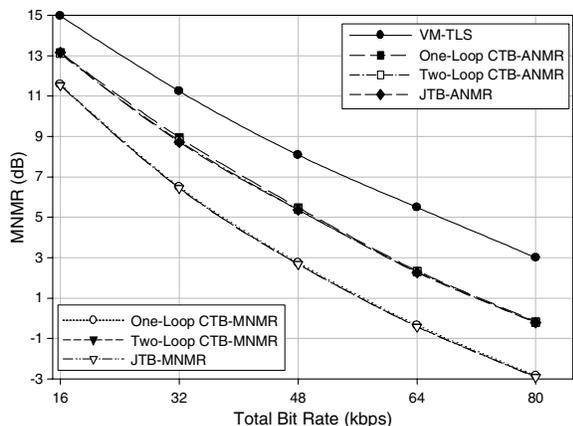


Fig. 8. MNMR Rate-Distortion Analysis

6.3. Subjective Quality Analysis

Listening test by human ears is the traditional method to subjectively evaluate the audio quality and is also the most recognized subjective quality test. However, such subjective test is expensive, time consuming, and difficult to reproduce. Informal listening tests on the aforementioned schemes show that it is hard to differentiate between JTB and various CTB schemes. In addition, a “simulated” subjective test, Objective Difference Grade (ODG), has been conducted. ODG

is a measure of quality designed to be comparable to the Subjective Difference Grade (SDG). It is calculated based on the difference between the quality rating of the reference and test (coded) signals. The ODG has a range of $[-4, 0]$, in which -4 stands for very annoying difference and 0 stands for imperceptible difference between the reference and the test signals [10][11]. The ODG results for various search schemes discussed in this paper are shown in Fig. 11 and the reference signal is the original audio sequence. According to the collected test data (Fig. 11), the difference between JTB and CTB schemes is quite small. The ODG results, which are relative to the CTB-MNMR scheme, for various fast searching algorithms are shown in Fig. 12. Again the performance of the non-uniform NB_SF algorithms is better than that of uniform NB_SF algorithms at about the same computational complexity.

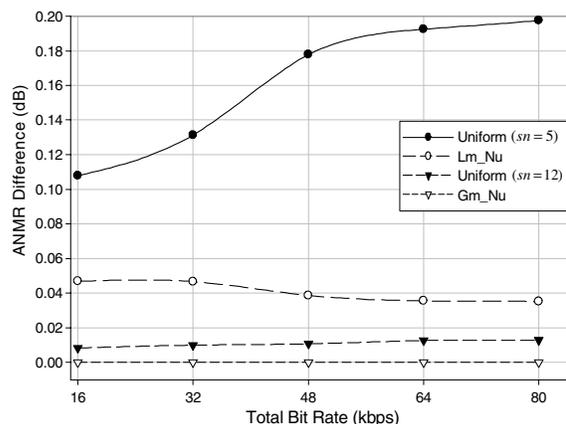


Fig. 9. ANMR Difference Analysis

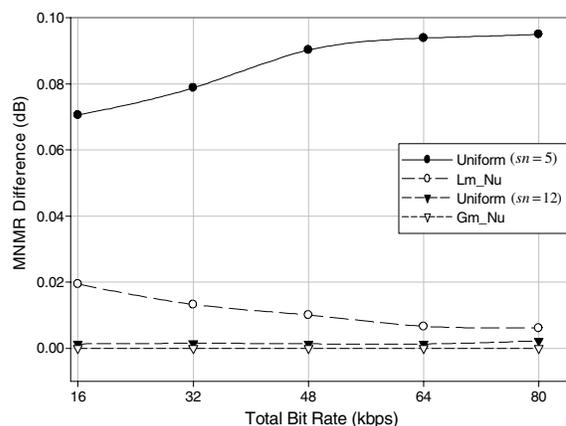


Fig. 10. MNMR Difference Analysis

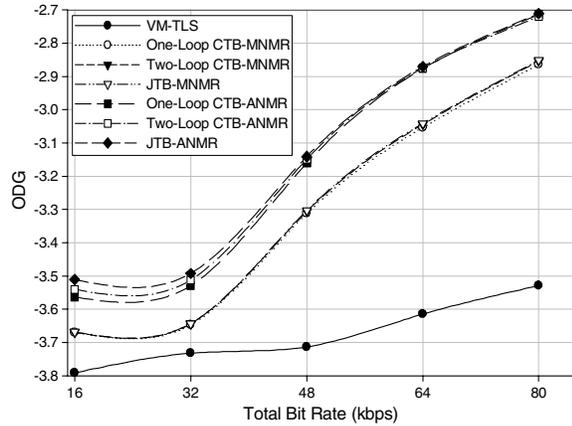


Fig. 11. ODG for VM-TLS, JTB and CTB

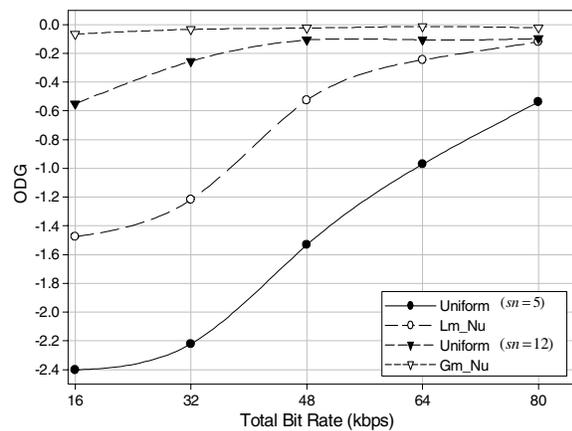


Fig. 12. ODG for Various Fast Searching Algorithms

7. CONCLUSIONS

In this paper, we propose a CTB optimization scheme for the MPEG-4 AAC coder, in which the optimization procedures for finding coding parameters, SF and HCB, are separated in two consecutive steps. Based on the complexity analysis, the proposed CTB scheme is approximately 71 to 142 times faster than the JTB scheme. Moreover, the simulation results show that both the objective and subjective quality of the CTB scheme is close to that of the JTB scheme. In addition, we also propose a lossless fast searching algorithm for trellis-based HCB optimization, which is about 4 times faster. Furthermore, two non-uniform searching algorithms, Gm_Nu and Lm_Nu, are proposed for trellis-based SF optimization. The simulation results show that the non-uniform searching algorithms can achieve better performance than uniform searching algorithms under the same complexity.

8. ACKNOWLEDGMENT

This work was supported by National Science Council, Taiwan, R.O.C., under Grant NSC-91-2219-E-009-011.

9. REFERENCES

- [1] ISO/IEC JTC1/SC29, "Information technology – vary low bitrate audio-visual coding," *ISO/IEC IS-14496 (Part 3, Audio)*, 1998.
- [2] M. Bosi, *et al.*, "ISO/IEC MPEG-2 advanced audio coding," *Journal of Audio Engineering Society*, vol. 45, pp. 789-814, October 1997.
- [3] A. Aggarwal, *et al.*, "Trellis-based optimization of MPEG-4 advanced audio coding," *Proc. IEEE Workshop on Speech Coding*, pp. 142-4 2000.
- [4] A. Aggarwal, *et al.*, "Near-optimal selection of encoding parameters for audio coding," *Proc. of ICASSP*, vol. 5, pp. 3269-3272, Jun 2001.
- [5] T. V. Sreenivas and M. Dietz, "Vector quantization of scale factors in advance audio coder (AAC)," *Proc. of ICASSP*, vol. 4, pp. 3641-3644, May 1998.
- [6] H. Najafzadeh and P. Kabal, "Improving perceptual coding of narrowband audio signals at low rates," *Proc. of ICASSP*, vol. 2, pp. 913-916, March 1999.
- [7] H. Najafzadeh and P. Kabal, "Perceptual bit allocation for low rate coding of narrowband audio," *Proc. of ICASSP*, vol. 2, pp. 893-896, 2000.
- [8] "The MPEG audio web page." <http://www.tnt.uni hannover.de/project/mpeg/audio>.
- [9] European Broadcasting Union, *Sound Quality Assessment Material: Recordings for Subjective Tests* Brussels, Belgium, Apr. 1988.
- [10] Draft ITU-T Recommendation BS.1387: "Method for objective measurements of perceived audio quality," July. 2001.
- [11] A. Lerchs, "EAQUAL software", Version 0.1.3 alpha, <http://www.mp3-tech.org/>

EFFICIENT CHANNEL ESTIMATION BASED ON DISCRETE COSINE TRANSFORM

Yen-Hui Yeh and Sau-Gee Chen

National Chiao Tung University
Department of Electronics Engineering and Institute of Electronics
1001 Ta Hsueh Rd, Hsinchu, Taiwan, ROC

ABSTRACT

Channel impairment caused by multi-path reflections can deeply degrade the transmission efficiency in wireless communication systems. Based on the property of the channel frequency response and the concept of interpolation, in this paper a DCT-based pilot-aided channel estimator for orthogonal frequency division multiplexing is proposed. This approach can mitigate the aliasing effect in the DFT-based channel estimator when there is non-sample-spaced path delay. Compared with DFT-based estimator, DCT-based estimator significantly improves the performance with a comparable complexity. In addition, a noise reduction scheme is introduced and combined with the estimator. In implementation, the DCT-based estimator has the advantages of utilizing mature fast DCT algorithms and compatible FFT algorithms, which is favorable to other matrix-based channel estimation methods.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) is a highly efficient and popular technique for high bit-rate data transmission over wireless communication channels. It has been adopted in wireless LAN and MAN standards IEEE 802.11a and 802.16, and the European digital audio broadcasting (DAB) and digital video broadcasting (DVB) standards.

In wireless communication channels, multi-path is a very common and severe problem. It causes inter-symbol interference (ISI) in the signal stream and this may degrade the transmission efficiency. OFDM can easily avoid this problem by inserting guard interval (GI). Besides ISI, multi-path also causes frequency-selective fading. If coherent demodulation is adopted, the effect of amplitude and phase fluctuation should be mitigated. One typical solution is to perform channel estimation, followed by channel equalization. Generally, there are two types of approaches for channel estimation. One is blind type of algorithms [1] and the other is pilot-aided type of algo-

rithms [2]. Although the pilot-aided algorithms waste a little more bandwidth than the blind algorithms, their performance is usually better than that of the blind case. Pilot-aided approach has been adopted in many standards such as 802.11a and many others. Therefore, in this paper we will focus on the pilot-aided case.

The optimal interpolation filtering in Minimum Mean Square Error (MMSE) sense for channel estimation [3], [4] needs the information of channel statistic and the associated computation complexity is very high. This may be hard to implement in practice. The approach of DFT-based interpolation [5] can theoretically achieve ideal lowpass interpolation, and has the advantages of low complexity by employing FFT algorithms. This technique works well when multi-path delays are integer multiples of the sampling time. However, this hardly happens in practical transmission environment. When the condition is not satisfied, performance of the DFT-based algorithm may degrade considerably. This is because the equivalent channel impulse response will be a disperse version of the original shorter one [6]. As a result, the DFT-based interpolation process will be based on the aliased data of the disperse impulse response.

In this paper, for the consideration of better channel interpolation result and lower aliasing error, we will propose a DCT-based channel estimation method, as detailed below.

2. CHANNEL ESTIMATION BASED ON DFT INTERPOLATION

2.1. OFDM System Model

We assume that an OFDM symbol contains N sub-carriers, and the OFDM symbol duration is T . Then the sampling period will be T/N and the sub-carrier spacing is $1/T$. The transmitted signal can be expressed as:

$$s(t) = \sum_{i=-\infty}^{\infty} \sum_{n=0}^{N-1} D_{i,n} \theta_{i,n}(t) \quad (1)$$

where $D_{i,n}$ is the data on the n -th sub-carrier in the i -th OFDM symbol and

$$\theta_{i,n}(t) = e^{\frac{j2\pi n}{T}(t-T_g-iT_c)} [u(t-iT_c) - u(t-(i+1)T_c)] \quad (2)$$

where T_g is the guard time interval, $T_c = T_g + T$ is the total symbol duration, and $u(t)$ is the unit step function.

A multi-path channel can be characterized as:

$$h(t, \tau) = \sum_{i=0}^{L-1} \alpha_i(t) \cdot \delta(\tau - \tau_i) \quad (3)$$

where $\alpha_i(t)$ is the time-varying gain and τ_i is the delay time for the i -th path. L is the total number of the paths. Usually, the magnitude of $\alpha_i(t)$ is modeled Rayleigh distributed, and the variation is associated with Doppler frequency f_d , $f_d = f_c v / c$, where f_c is the carrier frequency, v is the vehicle speed and c is the velocity of light.

The received OFDM signal passing through the AWGN time-varying multi-path channel can be expressed as

$$r(t) = \sum_{i=0}^{L-1} \alpha_i(t) \cdot s(t - \tau_i) + n(t) \quad (4)$$

where $n(t)$ is the white Gaussian noise. After sampling the signal and removing guard interval, the equivalent channel frequency response is (assuming $\alpha_i(t)$ is constant over one OFDM symbol)

$$H_{s,k} = \sum_{i=0}^{L-1} \alpha_{s,i} \cdot e^{\frac{j2\pi\tau_i k}{T}} \quad (5)$$

where $H_{s,k}$ is channel frequency response corresponding to the k -th sub-carrier of the s -th symbols, and $\alpha_{s,i}$ is the gain of the i -th path during the s -th symbol period. The received signal on the k -th sub-carrier of the s -th symbol can be expressed as

$$Y_{s,k} = D_{s,k} \cdot H_{s,k} + N_{s,k} \quad (6)$$

The corresponding impulse response is [6]

$$h_{s,n} = \frac{1}{\sqrt{N}} \sum_{i=1}^{L-1} \alpha_{s,i} e^{-\frac{j\pi(n+(N-1)\lambda_i)}{N}} \frac{\sin(\pi\lambda_i)}{\sin(\pi(\lambda_i - n)/N)} \quad (7)$$

where $h_{s,n}$ is the n -th tap of channel impulse response during the s -th symbol and $\lambda_i = \tau_i / T_s$, where T_s is the sampling period. By this equation, when non-integer λ_i exists, the power will leak to all taps $h_{s,n}$, as shown in Fig. 1.

2.2 DFT-based channel estimation [7]

Assume that M pilots are evenly assigned to M sub-carriers out of total N sub-carriers at a spacing of N/M sub-carriers, where N/M is an integer. The DFT-based channel estimation algorithm begins with the least-square (LS) estimation of the pilot sub-carriers.

$$\hat{H}_{p,m} = Y_{p,m} / p_m \quad (8)$$

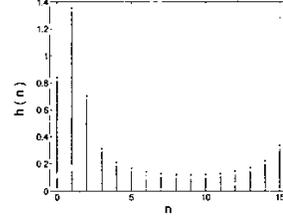


Fig.1. The equivalent impulse response for the continuous channel $h(t) = \delta(t) + \delta(t - 0.5Ts) + \delta(t - 1.4Ts)$

where $Y_{p,m}$ is the received signal at the m -th pilot sub-carrier and p_m is the pre-assigned pilot value for the m -th pilot sub-carrier. Then $\hat{H}_{p,m}$ is multiplied by some linear-phase shift as shown below

$$\hat{H}_{p',m} = \hat{H}_{p,m} \cdot e^{j\pi \frac{m\beta}{MT}} \quad (9)$$

where β is the minimum integer greater than all the path delays. The operation amounts to a corresponding time shift of the impulse response. It would make the power of the impulse response much more concentrate around $t=0$, while the impulse response values in the middle time positions would be smaller. This will facilitate zero insertion in those positions, and lead to a more effective up-sampling result of the channel frequency response, than the case without phase adjustment, as detailed below.

First the M -point impulse response is obtained by

$$\{\hat{h}_p\} = \text{IFFT}\{\hat{H}_p\} \quad (10)$$

Next the zero-insertion impulse response is formed by inserting $(N-M)$ zeros in the middle time indices:

$$\hat{h}_{N,n} = \begin{cases} \hat{h}_{p,n} & n \leq M/2 - 1 \\ 0 & M/2 \leq n \leq N - M/2 - 1 \\ \hat{h}_{p,n-N+M} & \text{otherwise} \end{cases} \quad (11)$$

Then the interpolated channel frequency response is solved after performing FFT on \hat{h}_N .

$$\{\hat{H}_{sh}\} = \text{FFT}\{\hat{h}_N\} \quad (12)$$

Finally, the actual estimated channel frequency response is obtained by canceling the phase shift operations performed in the beginning stage of the algorithm:

$$\hat{H}_n = \hat{H}_{sh,n} \cdot e^{-j\pi \frac{n\beta}{NT}} \quad (13)$$

3. THE PROPOSED ALGORITHM

As mentioned before, there will be leakage in channel impulse response, when the path delays are non-integer multiples of the sampling period. It is obvious that DFT-

based interpolation is not suitable for channel estimation under this condition. This is because the leakage will cause severe aliasing, when the mentioned DFT-based method is used. [7] proposed a windowed DFT-based approach to improve the performance. However, this approach must sacrifice some bandwidth. Next, we will propose a DCT-based interpolation algorithm to mitigate the aliasing problem. DCT is a well-known technique extensively used in image processing. DCT can reduce the high frequency component in the transform domain compared with DFT. The reason is that when given a sequence of N -point data, DFT conceptually treats it as a periodic signal with a period of N points. Hence, there is a tendency of noticeable high-frequency components, due to signal discontinuity in between consecutive period boundary. In contrast, DCT conceptually extends the original N -point data sequence to $2N$ -point sequence by doing mirror-extension of the N -point data sequence. As a result, the waveform will be smoother and more continuous in the boundary between consecutive periods. Correspondingly, high frequency components will be reduced. This benefits interpolation process. The proposed DCT-based channel estimation algorithm is detailed below.

3.1 The new DCT-based channel estimation algorithm

First, we also use LS estimation to get the channel frequency response on the pilot sub-carriers. After that, we perform DCT

$$\hat{h}_{c,k} = w_k \sum_{m=0}^{M-1} H_{p,m} \cos \frac{\pi(2m+1)k}{2M}, \quad k=0, \dots, M-1 \quad (14)$$

$$w_k = \frac{1}{\sqrt{M}}, \quad k=0; \quad w_k = \sqrt{\frac{2}{M}}, \quad k \neq 0$$

The next step inserts zeros in the DCT domain. However, different from DFT-based interpolation, zeros must be inserted at the end of \hat{h}_c as

$$\hat{h}_{N,k} = \begin{cases} \hat{h}_{c,k} & k \leq M-1 \\ 0 & \text{otherwise} \end{cases} \quad k=0, \dots, N-1 \quad (15)$$

Here IDCT can't be directly performed on \hat{h}_N to get the channel frequency response due to the following reason. Compared with DFT, DCT has a shift in the time domain data. Due to this characteristic, the value of the original low-rate data would not remain the same after interpolation process by employing IDCT. Therefore, the interpolation result would be poor. The solution is to use extendible IDCT (EIDCT) [8]. Based on EIDCT, we can get the interpolated channel estimation as

$$\hat{H}_n = \sum_{k=0}^{M-1} w_k \hat{h}_{N,k} \cos\left(\left(\frac{n}{N} + \frac{1}{2M}\right)\pi k\right) \quad n=0, \dots, N-1 \quad (16)$$

Alternatively, since the transform is derived from the concept of DFT, we can get the same result by first doing mirror-duplication to get doubled-length data and then applying the DFT-based interpolation.

One may argue that we can exchange the DCT and IDCT processes in the interpolation, then the time shift problem will not occur. Indeed, this is true. However, if we adopt this approach, another problem similar to DFT-based interpolation will be introduced. In the M -point DCT transform (14), its value is always zero at $k=M$. Therefore, if we treat the original data as DCT transform domain signal, the estimated channel frequency response after interpolation will decay to zero outside the last pilot sub-carrier. As it turns out, this would lead to degradation of performance at the edge of spectrum.

3.2 Combining a noise reduction scheme

When the delay time of each path is close to zero, the white Gaussian noise can be effectively reduced in the DCT domain. If the path delays are all small, the channel frequency response will be smoother (with less high frequency components). As such, in the DCT domain, the power in the high frequency region can be viewed as noise, and we can eliminate it by setting the value of high frequency to zero. The method works better in the DCT domain than in the DFT domain [5]. Especially, it is most effective when the pilot power is not much larger than the noise power. When the pilot power is limited to a lower level, for low-power consideration, this method can improve performance. The whole operations are detailed below.

After DCT operations, the accumulated power counting from the first index can be calculated. The value is compared with a threshold to determine the region occupied mostly by noises. One way to define the threshold is using percentage of total power, e.g. 90% of total power. After the index is determined, all the impulse response values after this index are set to zero as

$$\hat{h}_{cc,k} = \begin{cases} \hat{h}_{c,k} & 0 \leq k \leq b \\ 0 & b < k \leq M-1 \end{cases} \quad (17)$$

where b is the index of threshold.

Note that regardless of DCT-based or DFT-based approaches, the delay spread must be smaller than $(M \cdot T_s)$. Otherwise, the estimation will be error prone. This can be explained by the concept of down sampling. The frequency responses at the pilot sub-carrier frequencies are the down sampling version of the complete channel frequency response at all N sub-carrier frequencies. Hence, if the delay spread is equal to or larger than $(M \cdot T_s)$, then the aliasing of channel impulse will occur. There is no way to recover the aliased impulse response.

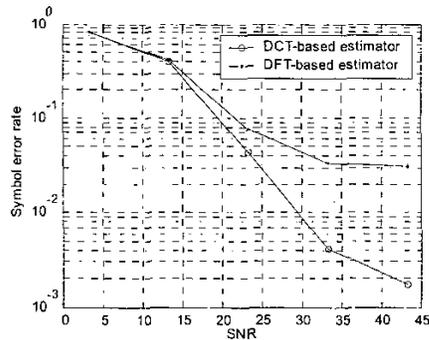


Fig.2. The SER performance with DCT-based estimator compared with DFT-based estimator

4. SIMULATION RESULT

In this section, we present the simulation result of the DCT-based estimator and compared it with DFT-based approach. The multi-path Rayleigh fading channel is simulated by Jakes' model. And each path gain follows the exponential power delay profile.

$$E[|\alpha_r(t)|^2] = e^{-\mu r} \quad (18)$$

We assume the channel has 4 paths and the set of delay spread is $\{0, 3.5T_s, 7.3T_s, 10.4T_s\}$. Meanwhile, we choose μ such that the average power of last path will be 20dB less than first path.

The number of total sub-carriers is 1024. 32 pilots are evenly inserted into the sub-carriers, and the first pilot is put on the first sub-carrier. Assume the transmission bandwidth is 5MHz. Then the sub-carrier spacing is 4.883KHz, and the sampling period is $0.2 \mu s$. The Doppler spread is fixed at 50Hz, such that $f_d T \approx 0.01$. The modulation scheme on each sub-carrier is 16QAM. The guard time interval is 32 sample periods. As for the value assigned to pilot, the outmost constellation point in 16QAM is chosen. Fig.2 shows the simulation result. It is obvious that DCT-based approach noticeably has higher performance especially in high SNR, than the DFT counter part.

We also simulate the case when the proposed new algorithm method combines with a noise reduction scheme as mentioned before. In this case, the set of delay spread is assumed $\{0, 0.5T_s, 2.2T_s, 3.1T_s\}$. As explained in section 4, the delay values cannot be too far away from zero. Also we change the pilot value from the outmost constellation point in 16QAM to the innermost point to reduce the pilot power. The threshold is set to 90% of the total power. Fig. 3 depicts the simulation result.

5. CONCLUSION

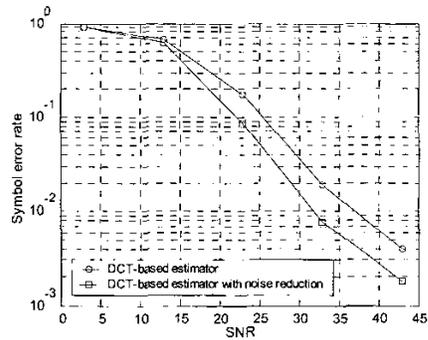


Fig.3. The SER performance of DCT-based estimator with noise reduction

A DCT-based pilot-aided channel estimator of OFDM system in the multi-path fading channel with non-integer sample-spaced path delay has been proposed in this paper. It achieves significant improvement over the DFT-based approach. It can be realized by the mature, low-complexity fast DCT algorithms in the literature. It is much lower than many other well-known matrix-based estimators. For the case of small path delay spreads and pilots with low power level, we also propose an effective noise reduction method to improve the performance.

6. REFERENCES

- [1] H. Bolcskei, R. W. Heath, Jr., and A. J. Paulraj, "Blind Channel Identification and Equalization in OFDM-Based Multiantenna Systems," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 96-108, January 2002.
- [2] Y. Li "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp.1207-1215, July 2000.
- [3] Y. Li, L. J. Cimini, Jr., and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902-915, July 1998.
- [4] O. Ddfors, M. Sandell, J. J. van der Beek, S. K. Wilson, and P. O. Borgesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol 46, no. 7, pp. 931-939, July 1998.
- [5] Y. Zhao and A. Huang, "A Novel Channel Estimation Method for OFDM Mobile Communication Systems Based on Pilot Signals and Transform-Domain Processing," *Proc. VTC '97*, pp. 2089-2094.
- [6] J.J. van der Beek, O. Edfors, M. Sandell, S.K. Wilson, and P. O. Borgesson, "On Channel estimation in OFDM systems," *Proc. VTC '95*, pp. 815-819.
- [7] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Windowed DFT Based Pilot-symbol-Aided Channel Estimation for OFDM Systems in Multipath Fading Channels," *Proc. VTC '00-Spring*, vol. 2, pp.1480-1484.
- [8] Y. F. Hsu, Y. C. Chen, "Rational interpolation by extendible inverse discrete cosine transform," *Electronics Letters*, vol. 33, no. 21, pp. 1774-1775, 9 Oct. 1997.

AN EFFICIENT MEMORY-BASED FFT ARCHITECTURE

Chao-Kai Chang, Chung-Ping Hung, and Sau-Gee Chen

Department of Electronics Engineering & Institute of Electronics
National Chiao-Tung University
1001 Ta Hsueh Rd, Hsinchu, Taiwan, ROC
sgchen@cc.nctu.edu.tw; asuma.ee90g@nctu.edu.tw

ABSTRACT

This paper proposes an efficient memory-based radix-2 FFT architecture, which greatly improves the memory-based FFT [5], [6] by reducing 50% memory size requirement, while maintaining a simple address generator. Specifically the memory size is reduced to 1.25N words. In addition, the multiplier utilization rate is 100%.

1. INTRODUCTION

FFT is applied to virtually every field of signal processing applications. Particularly, the OFDM technique [1] based on FFT, is recognized as the most efficient broadband wireline and wireless modulation technique [2], [3] at the time being and future to come. Because of high throughput rate demand of an OFDM system, an efficient FFT processor is required for real-time operations. FFT architectures can be categorized as two types: the pipeline architectures such as [4] and memory-based architectures, [5], [6]. Pipeline architectures bring the regularity of VLSI implement into full play and get higher performance by using more processing elements (PE), than the memory-based architectures. Memory-based architectures generally use only one butterfly PE (or more than one to establish parallelism), some memory blocks to store input or intermediate data, and a sophisticated control unit to handle memory read/write and data flow direction. In general, memory-based FFT processors can satisfy most of the applications including the mentioned OFDM communication systems, based on the state-of-art technology.

The memory-based FFT processors have the advantage of simple address generator design [5], [6] and overlapped data loading and FFT operations between two different FFT computations [6], at the cost of 1.5N [5] or 2.5N [6] memory sizes. Architecture of [6] has a better timing performance and higher multiplier utilization ratio (i.e., 100%) than [5]. This paper improves the FFT architectures of [5], [6] by reducing the memory size to only 1.25N. The reduction does not affect the performance of the original design, with respect to throughput rate (the same as that of [6]), multiplier utilization (the same as that of [6]), and simplicity of address generator.

2. MEMORY-BASED FFT PROCESSOR OF [5], [6]

In [5], [6], two types of FFT processor architecture based on the same design idea were proposed, which we call them Type I [5] and Type II [6] structures for short. Both of them were designed to implement radix-2 decimation-in-frequency (DIF) FFT processing as shown below:

$$X[2r] = \sum_{n=0}^{(N/2)-1} (x[n] + x[n + (N/2)]) W_{N/2}^{nr} \quad (1)$$

$$X[2r+1] = \sum_{n=0}^{(N/2)-1} (x[n] - x[n + (N/2)]) W_N^n W_{N/2}^{nr} \quad (2)$$

$r=0,1,\dots,(N/2)-1$

In Type I FFT structure, as shown in Figure 1, N -point data is separated into two $N/2$ -point parts (i.e., the first half and the second half), and loaded into RAM-1 and RAM-2, respectively. As suggested in eq. (1) and eq. (2), in the beginning of an FFT computation, the FFT structure picks up $x[n]$ and $x[n + (N/2)]$ from RAM-1 and RAM-2, respectively, and feed them to the butterfly PE. As shown in Figure 1, the butterfly PE consists of two subblocks: one is for the addition operation, while the other is for the subtraction of the input data. As suggested by eq. (1) and eq. (2) output data from the addition subblock do not have to be multiplied by twiddle factors and can be directly written back to proper memory locations. On the other hand, output data from the subtraction subblock are stored in RAM-3 temporarily, and will be multiplied by twiddle factors later when all the addition/subtraction operations in the butterfly PE are completed. In the first stage of an N -point FFT, outputs from the addition subblock are written sequentially back to first half part of RAM-1. After completion of the butterfly operations, the intermediate data in RAM-3, produced by the subtraction operations in the butterfly PE, are fed into the multiplier subblock for twiddle factor multiplication, and then sequentially written back to the second half parts of RAM-1 and RAM-2. In the 2nd, 3rd, and the remaining FFT stages, correspondingly RAM-1 and RAM-2 are partitioned into to four subparts, eight subparts, and so on, for the writing back of the butterfly PE outputs to the RAM memories. Butterfly operations and twiddle factor multiplications are done in different phases for different FFT stages. Under this condition, the multiplier utilization ratio is 50%. This architecture has low complexity in control unit, because that the connections between memory blocks and PEs are sim-

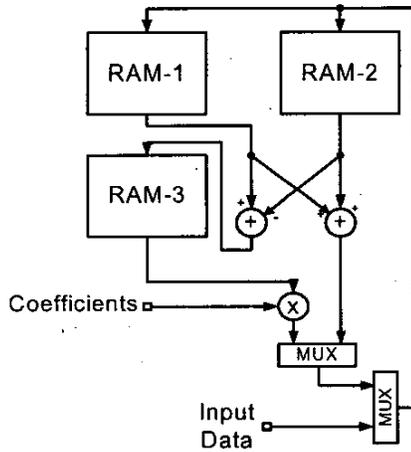


Figure 1. Architecture of Type I FFT processor [5].

ple, but at the cost of memory size (due to non-in-place computation). It needs a memory size of $1.5N$ words [5].

The Type II structure [7], as shown in Figure 2, is an improved version of Type I structure. Here, RAM-1 and RAM-2 in Type I structure are duplicated that facilitates the loading of the next N FFT data points into the other set of RAM-1 and RAM-2 memories in parallel with the current N -point FFT operations. Basically, Type II structure is the same as the Type I structure, except that two N -point FFT computations are executed simultaneously. The structure achieves 100% multiplier utilization, by interleaving the butterfly PE operations and twiddle-factor multiplications for the first and second FFT computations. However, in this case the total memory size increases to $2.5N$ words, which is much higher than the Type I structure. In order to reduce the memory size and simultaneously maintains the merits of Type-I and Type-II structures, the next section proposes a new memory-based structure that is much more efficient.

3. THE NEW IMPROVED FFT STRUCTURE

Without loss of generality, let's take a radix-2 8-point FFT for example. As shown in Figure 3, after the first FFT stage, the output data is split into two independent parts, i.e., the upper and lower parts. Therefore, we can process the first-stage FFT operations by following the configuration of Type-I structure. On the other hand, for the rest of the FFT stages, we can treat the computations as two independent $N/2$ -point FFT's, which are to be executed simultaneously. Figure 4 depicts the new improved FFT structure. In this case, for the 1st-stage FFT computation, RAM-1 and RAM-2 (each of $N/4$ words) are combined as a single memory block like RAM-1 of the Type I

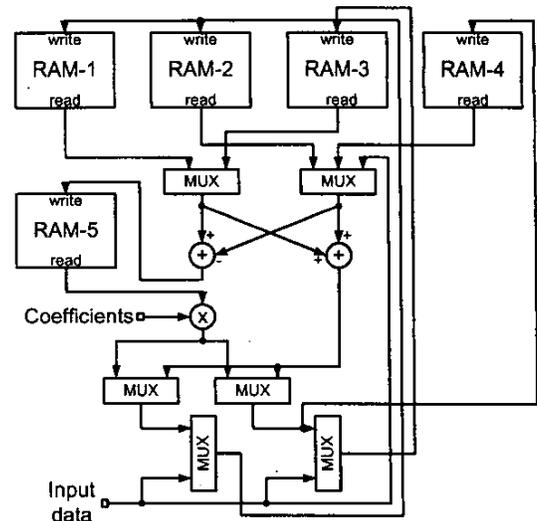


Figure 2. Architecture of Type II FFT processor [6].

structure, while RAM-3 and RAM-4 (each of $N/4$ words), are combined as a single memory block like RAM-2 of the Type I structure. For the rest of the FFT stages, similar configuration to Type II structure is adopted, except that the data path and control circuit are a little different from the Type II structure.

The detailed mechanism of the new improved FFT architecture is summarized in the following execution steps:

Step 1. Load half of input data;

Load $x[0], x[1], \dots, x[N/4-1]$ to RAM-1;
Then, load $x[N/4], x[N/4+1], \dots, x[N/2-1]$ to RAM-2.

Step 2. Process the first stage butterflies, including add, subtract, and coefficient multiply;

Read $x[k]$ from address k of RAM-1, and $x[N/2+k]$ from external input buffer and send them to butterfly PE, and perform $a[k]=x[k]+x[N/2+k]$,
 $b[k]=(x[k]-x[N/2+k]) \times W_N^k$, $k=0,1,2,\dots,N/4-1$.

Store $a[r]$ in address r of RAM-1, and

$b[r]$ in address r of RAM-3,

$r=0,1,2,\dots,N/4-1$.

Then, read $x[k]$ from address $k-N/4$ of RAM-2, and $x[N/2+k]$ from external input buffer and send them to butterfly PE, and perform

$a[k]=x[k]+x[N/2+k]$, $b[k]=(x[k]-x[N/2+k]) \times W_N^k$.
 $k=N/4,N/4+1,N/4+2,\dots,N/2-1$.

Store $a[r]$ in address $r-N/4$ of RAM-2, and

$b[r]$ in address $r-N/4$ of RAM-4,

$r=N/4,N/4+1,\dots,N/2-1$.

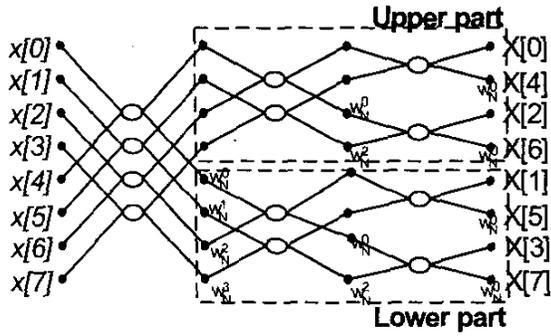


Figure 3. Signal-flow graph of an 8-point FFT. After the first FFT stage, data flow is split into two independent parts.

Step 3. Add and subtract data of upper-part butterflies;

Read $a'[k]$ from address k of RAM-1 and $b'[k]$ from address k of RAM-2 and send them to butterfly PE, and perform $a[k] = a'[k] + b'[k]$, $b[k] = a'[k] - b'[k]$, $k=0,1,2,\dots,N/4-1$.

Store $a'[r]$ in address r of RAM-1, $r=0,1,2,\dots,N/8-1$.

Store $a'[r+N/8]$ in address r of RAM-2, $r=0,1,2,\dots,N/8-1$.

Store $b'[r]$ in address r of RAM-5, $r=0,1,2,\dots,N/4-1$.

Step 4. Complete upper-part butterflies with coefficient multiplications and simultaneously add and subtract data of lower-part butterflies;

$$c[k] = b[k] \times W_{N/2}^k, k=0,1,2,\dots,N/4-1.$$

Store $c[r]$ in address r of RAM-1, $r=0,1,2,\dots,N/8-1$.

Store $c[r+N/8]$ in address r of RAM-2, $r=0,1,2,\dots,N/8-1$.

Read $a'[k]$ from address k of RAM-3 and $b'[k]$ from address k of RAM-4 and send them to butterfly PE, and perform $a[k] = a'[k] + b'[k]$, $b[k] = a'[k] - b'[k]$, $k=0,1,2,\dots,N/4-1$.

Store $a[r]$ in address r of RAM-3, $r=0,1,2,\dots,N/8-1$.

Store $a[r+N/8]$ in address r of RAM-4, $r=0,1,2,\dots,N/8-1$.

Store $b[r]$ in address r of RAM-5, $r=0,1,2,\dots,N/4-1$.

Step 5. Complete coefficient multiplications of lower-part butterflies of one stage and simultaneously add and subtract data of upper-part butterflies of the next stage;

$$c[k] = b[k] \times W_{N/2^{t-1}}^k, k=0,1,2,\dots,N/4-1, t=3,4,\dots,\log(N).$$

Store $c[r]$ in address r of RAM-3, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=0\}, t=3,4,\dots,\log(N)$.

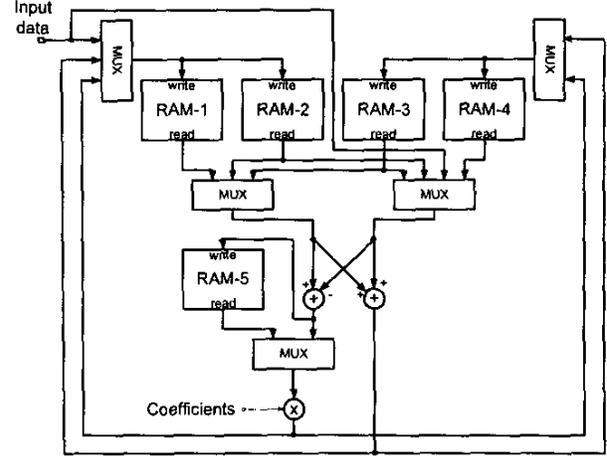


Figure 4. The new improved architecture (omitting control signal).

Store $c[r]$ in address $r-N/2^t$ of RAM-4, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=1\}, t=3,4,\dots,\log(N)$.

Read $a'[k]$ from address k of RAM-1 and $b'[k]$ from address k of RAM-2 to butterfly PE, and perform $a[k] = a'[k] + b'[k]$, $b[k] = a'[k] - b'[k]$, $k=0,1,2,\dots,N/4-1$.

Store $a[r]$ in address r of RAM-1, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=0\}, t=4,5,\dots,\log(N)$.

Store $a[r]$ in address $r-N/2^t$ of RAM-2, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=1\}, t=4,5,\dots,\log(N)$.

Store $b[r]$ in address r of RAM-5, $r=0,1,2,\dots,N/4-1$.

Step 6. Complete coefficient multiplying of upper part butterflies of one stage and simultaneously add and subtract data of lower part butterflies of the same stage;

$$c[k] = b[k] \times W_{N/2^{t-1}}^k, k=0,1,2,\dots,N/4-1, t=4,5,\dots,\log(N).$$

Store $c[r]$ in address r of RAM-1, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=0\}, t=4,5,\dots,\log(N)$.

Store $c[r]$ in address $r-N/2^t$ of RAM-2, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=1\}, t=4,5,\dots,\log(N)$.

Read $a'[k]$ from address k of RAM-3 and $b'[k]$ from address k of RAM-4 to butterfly PE, and perform $a[k] = a'[k] + b'[k]$, $b[k] = a'[k] - b'[k]$, $k=0,1,2,\dots,N/4-1$.

Store $a[r]$ in address r of RAM-3, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=0\}, t=4,5,\dots,\log(N)$.

Store $a[r]$ in address $r-N/2^t$ of RAM-4, $r=\{y\lfloor y/(N/2^t) \rfloor \bmod(2)=1\}, t=4,5,\dots,\log(N)$.

Store $b[r]$ to address r of RAM-5, $r=0,1,2,\dots,N/4-1$.

Table 1. Performance comparison of the new design and the existing designs.

	Memory size	Average throughput rate	Multiplier Utilization
Improved	1.25N	$\frac{2}{N \log(N) + N}$	100%
Type I [5]	1.5N	$\frac{1}{N \log(N) + N/2}$	50%
Type II [6]	2.5N	$\frac{2}{N \log(N) + N}$	100%

Step 7. Repeat Step 5 and Step 6 for the following stages until the last stage.

In the processing of the first-stage FFT, output data from the butterfly subtraction subblock need not to be stored in RAM-5. Instead, they are directly multiplied by the twiddle factors and written back to memory. After processing the first-stage FFT, the FFT structure proceeds to process the upper part of the intermediate output data stored in from RAM-1 and RAM-2, and simultaneously process the lower part of the intermediate output data stored in RAM-3 and RAM-4, using the similar operation configuration of Type II structure. Whenever, there is one part of intermediate data is fed into butterfly PE, the other part of intermediate data is multiplied by twiddle factor. Since RAM-5 always sends one intermediate data point to multiplier and receives one intermediate data point from butterfly PE to the same address, no data hazard is caused in RAM-5.

This architecture has 5 memory blocks, each of size $N/4$ words. As such, the total memory size is $1.25N$, which is only half that of Type II architecture.

4. PERFORMANCE EVALUATION

According to the discussed operation steps, we evaluate and compare the architectural performances of the new structure, Type I and Type II structures, based on the required total number of machine cycles, throughput rate, multiplier utilization ratio and memory size. For the evaluation, a few assumptions are made. First, the memory units support two-phase independent accesses: read phase and write phase. Secondly, maximum period between data read and write is defined as one cycle. We can figure out that critical path of this architecture contains one butterfly PE and one multiplier, that is bounded by the Step 2 mentioned in Section 3. As a result, for an N -point FFT, the required machine cycles are contributed by:

Step1. $N/2$ cycles.

Step2. $N/2$ cycles.

Step3. $N/4$ cycles.

Step4. $N/4$ cycles.

Step5. $(N/4) * (\log(N)-2)$ cycles.

Table 2. Required FFT clock rates for some OFDM communication standards using the improved FFT processor.

Communication systems	FFT size(operating frequency in MHz)
DAB	2048(12),1024(11),512(10),256(9)
DVB-T	8192(56),2048(48)
802.11a	64(70)

Step6. $(N/4) * (\log(N)-2)$ cycles.

Hence, without considering data output overhead, total latency of an N -point FFT is $N/2 + (N/2) * \log(N)$ cycles. Note that the period of a machine cycle is bounded by the critical path containing one butterfly PE, one multiplier, and several memory access overhead, and they are the same for Type I, Type II, and the proposed structures.

The Type II structure needs $N + N * \log(N)$ cycles to compute two N -point FFT's. Therefore, in average, it has the same time performance as the proposed architecture, but it requires double memory size that of the proposed structure. Performance comparison is summarized in Table 1.

Based on the proposed structure, the required clock cycles for some OFDM communication systems such as DAB, DVB-T, and 802.11a, are summarized in Table 2.

5. CONCLUSION

The proposed new structure significantly reduces the memory size, while maintains the same speed performance, compared with its predecessors [6]. The future works will be enhancing the proposed architecture for variable length FFT's that suits for different OFDM systems, and realizing the processor for particular applications.

6. REFERENCES

- [1] Richard van Nee and Ramjee Prasad, *OFDM Wireless Multimedia Communications*, Artech House, 2000
- [2] IEEE Standard 802.16-2001 IEEE Standard for Local and Metropolitan area networks Part 16: Air Interface.
- [3] IEEE Standard 802.11a-1999 Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications.
- [4] Shousheng He and Mats Torkelson, "A New Approach to Pipeline FFT Processor," Parallel Processing Symposium, The 10th International, pp. 766-770, 1996.
- [5] Ching-Hsien Chang, Chin-Liang Wang, and Yu-Tai Chang, "A novel memory-based FFT processor for DMT/OFDM applications," Acoustics, Speech, and Signal Processing, 1999. Proceedings. 1999 IEEE International Conference on, Volume:4, pp. 1921-1924, 1999
- [6] Chin-Liang Wang and Ching-Hsien Chang, "A DHT-based FFT-IFFT processor for VDSL transceivers," Acoustics, Speech, and Signal Processing, 2001. Proceedings. 2001 IEEE international Conference on, Volume:2, pp. 1213-1216, 2001.

Rate control for real time media based on predictive wireless channel condition

根據預測的無線通道情況控制即時資料的傳輸速率

計畫編號：NSC 91-2219-E-009-009

執行期限：91 年 08 月 01 日至 92 年 07 月 31 日

主持人：張文鐘 交通大學電信系

E-mail: wtchang@cc.nctu.edu.tw

一, 中文摘要

關鍵字：(通道模型, 自動重傳, 通道預測) 兩個狀態的馬可夫鍊被用來描述衰變通道。在這個模型, 好狀態代表正確傳輸, 而壞狀態代表位元錯誤。這兩個狀態模型的轉換機率是由通道情況推導而來, 像是都卜勒頻率, 平均信號雜音比及符號長度。利用這麼個模型來代表目前通道情況的目的是要用轉換機率來推測未來的通道情況。進兒預測未來的位元錯誤率及自動重傳的量。這個自動重傳的量就可以從原先已分配好的位元量減掉, 以防止因重傳而使緩衝器擠滿。

ABSTRACT (keywords: channel model, ARQ, channel prediction): A two-state Markov chain is used to model the fading channel. In this model, good state indicates correct transmission and bad state indicates bit error. The transition probability of the two-state model is derived as a function of the channel condition such as Doppler frequency, average SNR and symbol timing. The purpose of the state model to represent the current channel condition is that the future channel condition can be predicted from the transition matrix. Based on this model, the future channel condition and bit error rate can be predicted and the amount of ARQ can be pre-determined. This amount of ARQ is thus subtracted from the pre-allocated target bit for real time media to pre-compensate for future re-transmission.

二, 緣由與目的

Most of the error in transmission is due to channel fading. The current wireless system employs ARQ protocol to deal with erroneous packets. This kind of error concealment will increase the transmission burden and pose a problem for real time media data. The direct consequence is that the effective buffer output rate will decrease due to the retransmission. To prevent overflow, one of the method is to adapt the source coding rate according to the buffer condition. The buffer fullness reflects the channel condition prior to the current encoding moment. With such a strategy, the current coding rate will be a function of the amount of ARQ that was issued before. However, if the channel condition can be predicted in advance such that during the source coding period the transmission error rate can be estimated, the source coding rate can be accordingly adapted such that ARQ will not increase the buffer fullness to prevent future buffer overflow and frame skip. To achieve this goal, a strategy that further reduces the bit rate allocated for each frame to be coded according to the channel prediction is proposed. This is used to pre-compensate the amount of future possible ARQ retransmission.

In the original model, the buffer update rule is as follows:

$$W = \max(W_{prev} + B_{prev} - R / F, 0)$$

if $W > M$, M the threshold, the next frame would be skipped. Otherwise, the frame

target for the next encoding frame is

$$B_j = R/F - \Delta, \text{ where } \Delta = \begin{cases} W/F, & W > Z.M \\ W - Z.M, & \text{otherwise} \end{cases}$$

by default, $M=R/F$, $Z=0.1$, we can see that Δ is used for buffer fullness adjustment. In the following, we discuss how the channel prediction is performed to further modify B_j with a two-state channel model.

三、研究方法及成果

The main point is to predict the possible number of error packet and subtract it from the current frame target bit rate B_j .

We use a simplified Gilbert channel [2] for channel prediction. This model describes burst-noise using a Markov chain with two states G and B. In state G, transmission is error-free. In state B, the channel has only probability h of transmitting a digit correctly. Because we use simplified Gilbert channel, we set the parameter h as 0. To simulate burst noise, the states B and G must tend to persist; i.e., the transition probabilities $b=\text{Prob}(G \rightarrow B)$ and $g=\text{Prob}(B \rightarrow G)$ will be small and the probabilities $1-b$, $1-g$ of remaining in G and B will be large. Fig. 1 is a diagram for the Markov chain.

the state probabilities at time k can be derived from the state probabilities at time t_0 recursively by

$$P^k = P^{k-1} * T = P(t_0 | \text{state}(t_0) = G) * T^{k-t_0} = [1, 0] * T^{k-t_0}$$

The average number of time units the channel is in good state and bad state are

denoted as $\bar{T}(G)$ and $\bar{T}(B)$ and are

$$\frac{1}{b}, \frac{1}{g} \text{ respectively, } \frac{1}{g} \text{ is also called as}$$

the mean burst error length or average fade duration. The stationary distribution of good state and bad state is

$$P^\infty = \begin{bmatrix} P_0, P_1 \end{bmatrix} = \begin{bmatrix} g/(b+g), b/(g+b) \end{bmatrix}$$

p_1 is the average bit error rate. To predict the future channel throughput, we use the average probability in Good state in the next k -packet interval. We define the average probability of Good state in the next k -packet intervals as $\text{avg}\{P(t | \text{state}(t_n) = \text{state})\}$ and it is the average of the following equation

$$\text{avg}\{P(t | \text{state}(t_n) = \text{state})\} = \frac{1}{k}$$

$$\sum_{t=1}^k P_0(t | \text{state}(t_n) = \text{state}), n = 0, 1, 2, \dots$$

The term $(1 - \text{avg}\{P(t | \text{state}(t_n) = G)\})$ is the average bad state probability. Multiplied by the channel throughput, the prediction of the error bits at time t_n is obtained as

$$\text{Bit}_{\text{retx}} = \frac{R}{F} * (1 - \text{avg}\{P(k | \text{state}(t_{\text{current}}) = \text{state})\}),$$

$\text{state} \in \{G, B\}$

The error bits are subtracted from the previous B_j . That is $B_j = B_j - k \text{Bit}_{\text{retx}}$.

The goal is to maximize B_j , thus it is desirable that the transmission rate R is as high as possible and the error rate is as low as possible. In this paper, we aim at the establishment of the relationship between the GE model and the Raileigh channel distribution for the purpose of channel prediction..

To derive the transition probability b and g , we assume the received envelope $\alpha = |r|$ as Rayleigh distribution, i.e.,

$$f(\alpha) = \frac{\alpha}{\sigma^2} e^{-\alpha^2/2\sigma^2} \quad \text{where } \sigma^2 \text{ is}$$

variance of Gaussian random process, the received SNR is $\gamma = \alpha^2 E_b / N_0$, its PDF is exponent distribution, i.e.,

$$f(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}}, \quad \gamma \geq 0 \quad \text{where}$$

$\bar{\gamma} = E[\alpha^2] E_b / N_0$ is the average SNR of the received signal and $E[\alpha^2]$ is the average value of α^2 . Since the GE channel consists of two states, we let γ_t

be the thresholds of the received SNR, where the channel will change state. We can calculate the stationary probabilities of the GE channel in its respective states by finding the fraction of time the Rayleigh fading channel is below and above γ_t respectively. Thus

$$P^\infty(B) = \int_0^{\gamma_t} \frac{1}{\gamma} e^{-\gamma/\bar{\gamma}} d\gamma = 1 - e^{-\gamma_t/\bar{\gamma}} = 1 - e^{-\rho^2}$$

similarly, we have

$$P^\infty(G) = \int_{\gamma_t}^{\infty} \frac{1}{\gamma} e^{-\gamma/\bar{\gamma}} d\gamma = e^{-\rho^2} \quad \text{where}$$

$\rho^2 = \gamma_t/\bar{\gamma}$. Another related parameters are the Level Crossing Rate(LCR) and Average Fade Duration(AFD). From [3] and [4],

$$L_R = \frac{\int_0^{\infty} p(R, \dot{R}) dR}{dt} = \int_0^{\infty} \dot{p}(R, \dot{R}) dR = \sqrt{2\pi} f_D \rho e^{-\rho^2}$$

$$\bar{\tau} = \frac{P_r[r \leq R]}{L_R} = \frac{1 - e^{-\rho^2}}{\sqrt{2\pi} f_D \rho e^{-\rho^2}} = \frac{e^{\rho^2} - 1}{\sqrt{2\pi} f_D \rho}$$

the transition probability can be derived as,

$$g = \frac{\rho f_D T_S \sqrt{2\pi}}{e^{\rho^2} - 1}, \quad b = \rho f_D T_S \sqrt{2\pi}$$

where $f_D = \frac{v}{\lambda}$ is the Doppler frequency, and T_S is the symbol duration.

Fig. 2 shows an example of packet error in a typical wireless channel for 160 frames. These error packets will cause retransmission, Without ARQ, the buffer fullness will not be affected. With ARQ, the effective throughput is reduced. Fig. 3 shows the buffer fullness condition when ARQ is used. When the buffer fullness is above a certain threshold, frame skip happens. With the above mentioned channel prediction and

adaptive frame layer bit allocation, the improvement is shown in Fig.4.

四、結論

The main point is to use a two-state Markov model to represent the current channel condition. The transition matrix is a close approximation of the channel memory. From the transition matrix, we estimate the future bit error rate and the possible retransmission amount. This is equivalent to the future channel prediction. From the predicted amount of ARQ, we subtracted it from the pre-allocated target bit in real time media coding. This is to leave room for the future possible bit error and ARQ such that the retransmission ARQ will not cause transmitter buffer overflow.

Reference:

- [1] Supavadee Aramvith, I-Ming Pao, and Ming-Ting Sun, "A Rate-Control Scheme for Video Transport over Wireless Channels", IEEE Trans. Video Tech. Vol.11, NO.5 May 2001.
- [2] E. N. Gilbert, "Capacity of a burst-noise channel," Bell Syst. Tech. J., vol. 39, pp.1253-1265, Sept. 1960.
- [3] T. S. Rappaport, Wireless Communications, Principal & Practice. Englewood Cliffs, NJ: Prentice Hall, 1996.
- [4] L. Wilhelmsson and L. B. Milstein, "On the effect of Imperfect Interleaving for the Gilbert-Elliott Channel," IEEE Trans. Commun, vol. 47, no. 5, May 1999.

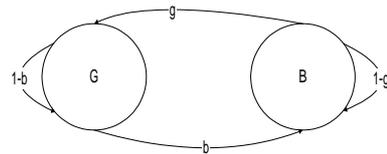


Fig.1 two state channel model

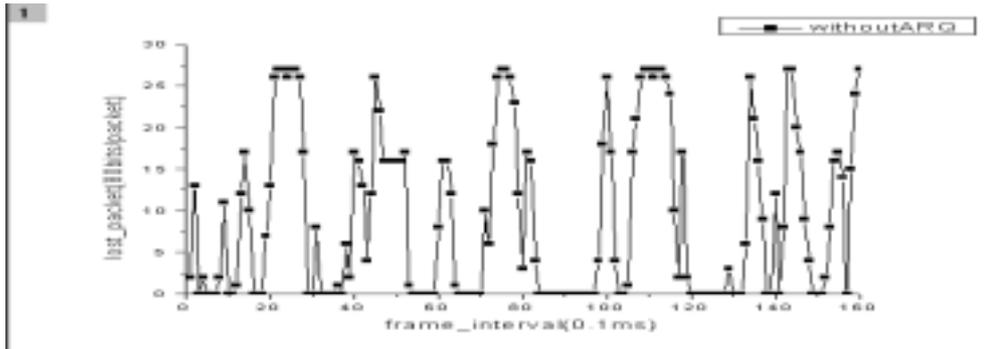


Fig 2.transmission error distribution

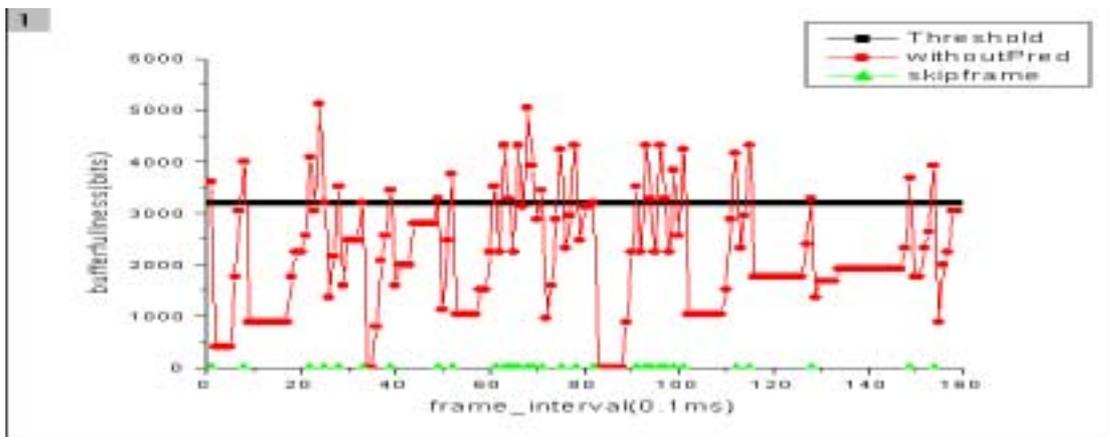


Fig 3 transmission without prediction

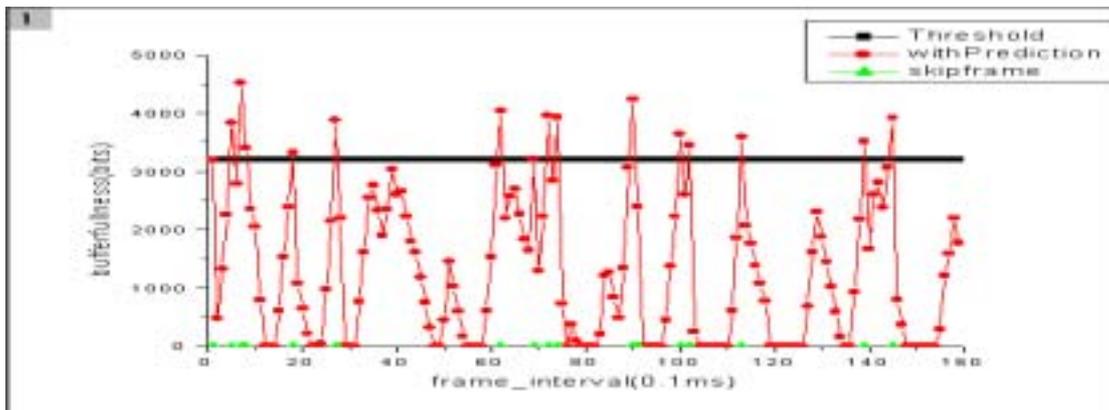


Fig 4 transmission with prediction