Resource Allocation for Multicarrier DSL Systems

Saswat Panigrahi

Department of Electrical and Computer Engineering
McGill University
Montreal, Canada

June 2005

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Engineering.

Your file Votre référence
ISBN: 978-0-494-22660-5
Our file Notre référence
ISBN: 978-0-494-22660-5

This thesis is dedicated to my parents and my teachers

Abstract

In this thesis, resource allocation problems relevant to multicarrier DSL environments are considered. The resources in a multicarrier DSL environment are bandwidth (i.e., subcarriers), the power allocated, the bits loaded and possibly the coding scheme used on each subcarrier. The major impairments in DSL environments are *crosstalk* and *external noise*. *Crosstalk* (predictable & controllable) is the interference caused by other DSL users sharing the same medium. *External Noise* (unpredictable and uncontrollable) consists of relatively high energy bursts due to electromagnetic interference from physical phenomena, electrical switches, motors and home appliances which are invariably present in the close vicinity of DSL modems.

In current DSL applications, *Crosstalk* is kept in control by specifying a fixed peak-power constraint (based on the worst-case) which each user has to obey – a technique known as Static Spectrum Management (SSM). This simplifies the resource allocation to a single-user optimization problem. But the worst-case peak power constraint is overly restrictive and results in poorer rates. Recently Dynamic Spectrum Management (DSM) techniques which dynamically vary the multiuser power allocation for crosstalk control instead of using fixed peak power constraint have proven to provide much better *rates*. This improvement however comes at the cost of dealing with a more complicated multi-user optimization problem.

On the other hand, due to its unpredictability, e*xternal noise* (and *RFI* pickup) has been combated with the use of a suitable safety *margin*.

Based on application and scenario, Resource Allocation is either performed with the objective of *Rate* maximization (at a fixed *margin*) or *Margin* maximization (at a fixed *rate*) since each of the quantities indirectly imposes a cap on the other.

But in both *rate* maximizing and *margin* maximizing resource allocation algorithms available in literature the bits loaded were always constrained to integers because most scalable modulation schemes such as QAM or PSK support integer bit/symbol. It was initially believed that most (not all) of the granularity losses (due to the integer bit constraint) could be recovered through 'bit-rounding' and 'energy re-scaling'. But this was

observed only for the total power constrained case. With the advent of peak power constraint, we show that the room for optimization in the energy domain is severely restricted and granularity losses constitute a significant percentage of the achievable data rate. To recover these losses, we propose the Adaptive Reed Solomon aided Fine Granularity Loading (ARSFGL) scheme - a scheme that jointly optimizes the power, bit and code allocation. For achieving near-continuous rate adaptation, the family of Reed-Solomon (RS) codes has been used for their low redundancy, high flexibility in correction capability and highly programmable architecture. Simulation results with realistic VDSL-DMT systems with the currently standardized SSM framework show more than 20% improvement in rate achieved for most cases.

The extension of ARSFGL technique to multiuser (DSM) scenarios results in a purely *distributed* scheme which provides *rates* better or equal to the *rates* achieved by the *centralized*, optimal (much more complicated) multiuser integer-bit loading scheme.

Further in this work a multiuser *margin* maximization algorithm is developed. Near-continuous rates provided by the ARSFGL scheme allow continuous bit loading assumption for simplicity. Prior to this work, no multiuser *margin* maximization algorithm existed, even though the importance of *margin* maximization is well-recognized. Most existing single-user *margin* maximization algorithms rely on a fixed crosstalk assumption. But in multiuser (DSM) scenarios each user's power allocation dynamically determines other user's crosstalk. With direct extension of single-user algorithms in multiuser (DSM) scenarios, one user's *margin* maximization can lead to the failure of other users in meeting their target *rates*.

We begin by exploring the favorable monotonicity and fairness properties that multiuser *margin* exhibits over a multiuser rate region and use them to formulate a box-constrained non-linear least squares (NLSQ) problem that can be solved by using a scaled gradient trust region approach with Broyden Jacobian update. This algorithm efficiently converges to a solution providing the best common equal *margin* to all users while explicitly guaranteeing that each user's target rate requirement is satisfied. The algorithm requires only minimal coordination among users, which makes it suitable for practical implementation.

Sommaire

Dans ce mémoire nous allons considérer la problématique de l'allocation de ressources dans l'environnent DSL à multi-porteuse. Les ressources à considérer dans cet environnent sont la largeur de bande (onde sous-porteuse), l'assignement de puissance, l'allocation de bits et possiblement le système de codage utilisé sur chaque onde sous-porteuse. Les obstacles majeurs présents dans l'environnement DSL sont la diaphonie et le bruit externe. La diaphonie (prévisible et contrôlable) représente l'interférence causée par d'autres utilisateurs DSL partageant le même médium. Le bruit externe (non-prévisible et non contrôlable) est le résultat de décharge énergétique provenant de l'interférence électromagnétique créée par des phénomènes physiques, interrupteurs électriques, moteurs et appareils ménagers à proximité des modems DSL.

Dans les systèmes DSL actuels, la diaphonie est contrôlée en spécifiant une contrainte que doit obéir chaque utilisateur, et déterminée par le sommet maximum de puissance (basée sure un scénario du pire cas). Cette technique, communément appelée Aménagement Statique de Ressources (ASR), simplifie le problème d'allocation de ressources à un problème d'optimisation à un utilisateur unique. Cependant, elle est très restrictive et résulte en de faibles taux de transmission. Récemment, l'utilisation de techniques d'Aménagement Dynamique de Ressources (ADR), qui assigne dynamiquement la puissance multi-utilisateur avec comme but de contrôler la diaphonie, ont démontré la possibilité de fournir de meilleurs taux de transmission. Cette amélioration crée cependant un problème d'optimisation à multi-utilsateurs plus compliqué.

D'autre part, grace à son caractère non-prévisible, le bruit externe est combattu en utilisant une *marge* de sécurité.

Basée sur les besoins de l'application particulière, l'allocation de ressources est implémentée soit avec l'objectif de maximiser le taux de transmission (avec une *marge* fixe), soit avec l'objectif de maximiser la *marge* (avec un taux de transmission fixe) car chacune de ces deux mesures impose indirectement une limite sur l'autre.

Pour les deux algorithmes d'allocation de ressources, la maximisation du *taux* et la maximisation de la *marge*, les bits chargés ont jusqu'à présent toujours été un nombre entier, car la majorité de méthodes de modulation, comme le QAM ou le PSK, supportent que des nombres entiers de bits par symbole. Il était originalement cru que la majorité des pertes granulaires (causées par la contrainte de nombre de bits entiers) pourraient être compensés avec l'arrondissement de bits ou le ré-échelonnage énergétique. Ceci n'a été cependant observé que pour le cas de contrainte de puissance totale. Avec l'arrivée des techniques basées sur le sommet maximum de puissance, nous démontrons que la possibilité d'optimiser dans le domaine énergétique est sévèrement réduite et les pertes granulaires constituent un pourcentage significatif du taux de transmission réalisable. Pour combattre ces pertes, nous proposons la méthode de Support de Granulité Fine utilisant les codes Adaptifs Reed-Solomon (SGFARS), une méthode qui optimise conjointement l'allocation de puissance, de bits et de codes. Pour obtenir une adaptation de taux quasi-continue, la famille de codes Reed-Solomon (RS) est utilisée à cause de leur faible redondance, haute flexibilité en capacité de correction et architecture hautement programmable. Des résultats de simulation avec des systèmes VDSL-DMT réalistes, utilisant le standard SSM, démontrent des augmentations de taux de plus de 20% dans la majorité des cas.

L'extension de la méthode SGFARS au scénario multi-utilisateur résulte en un système purement *distribué*, avec des taux de transmission meilleurs ou égaux à ceux obtenus par la méthode *centralisée* et optimale de chargement de bits entiers (beaucoup plus compliquée).

Nous développons aussi dans ce travail un algorithme de maximisation de la *marge* multi-utilisateur. Les taux quasi-continus permis par la méthode SGFARS permettent d'assumer le chargement continu de bits. Avant ce travail, il n'existait pas d'algorithme de maximisation de la *marge* multi-utilisateur, même si l'importance de la maximisation de la marge est un fait communément reconnu. La majorité des algorithmes existants de la maximisation de la marge pour un utilisateur unique assument une diaphonie fixe. Cependant, dans le cas multi-utilisateur, la distribution dynamique de puissance pour chaque utilisateur détermine dynamiquement la diaphonie perçue par les autres utilisateurs. Avec une extension directe des algorithmes mono-utilisateurs à mutli-

utilisateurs, l'augmentation de la *marge* d'un utilisateur pourrait empêcher un autre utilisateur d'atteindre son *taux* désiré.

Nous débuterons par explorer les propriétés de monotonicité favorable et de fidélité que la marge multi-utilisateur démontre sur une région multi-utilisateur, et nous les utiliserons pour formuler un problème non-linéaire de moindres-carrées (NLMC). Celui-ci sera résolu avec une approche à région de gradient proportionné avec une mise à jour de Jacobiens Broyden. Cet algorithme converge vers une solution avec une meilleure marge commune et égale pour tous les utilisateurs, et simultanément garantit que tout le taux désiré pour tous les utilisateurs est obtenu. Cet algorithme requiert seulement une coordination minimale entre utilisateurs, ce qui le rend approprié pour les applications pratiques.

Acknowledgements

First I would like to thank my supervisor, Dr. Tho Le-Ngoc for his invaluable guidance and support throughout my graduate studies at McGill University. I also acknowledge a NSERC CRD Grant with Laboratoires Universitaires Bell for partial financial support.

I am deeply grateful to Dr. Fabrice Labeau who has constantly encouraged me to perform better. I thank Dr. Ioannis Psaromiligkos, who patiently answered my questions on everything from Random Processes to CDMA each time I knocked his door without appointment. I also thank Dr. Harry Leib (*course*-Information Theory), Dr. Jan Bajcsy (*course*-Digital Communications II) and Dr. Sanjo Zlobec (*course*-Optimization) for many insightful discussions during and after class hours.

I thank Matrino and Nestor for indispensable help in DSL simulator construction. Martino, Hamid and Mohammad further showed me guiding examples of hard work, dedication and perseverance and their opinion both technical and otherwise helped me on many occasions. The greatest amount of cooperation that I have enjoyed is with Yang. We worked together in multiuser channel model construction. Our daily discussions helped me understand many subtle concepts of Optimization and Game Theory which led to conception of some ideas in this thesis. Further I am grateful to Yang, Ping and Navid for providing unconditional support and encouragement to me in every positive and negative situation that I have faced. I thank Li, Bharatram and Mohammad for proof-reading my thesis and Eugene and Nestor for the French translation of my abstract.

I thank my parents, my family and all my friends from McGill University, SGI Montreal and those back in India for their steady stream of love, support and faith.

Lastly the poem "If" by Rudyard Kipling has encouraged me and provided me direction in many situations, a few lines of which are presented below.

"If you can dream--and not make dreams your master,
If you can think--and not make thoughts your aim;
If you can meet with Triumph and Disaster
and treat those two impostors just the same;
.....
If you can fill the unforgiving minute
With sixty seconds' worth of distance run,
Yours is the Earth and everything that's in it,
And--which is more--you'll be a Man, my son!"

Table of Contents

List of Figures

List of Tables

List of Symbols and Notation

Mathematical Notation

$\{x_j\}_{j=1}^{N}$	Vector consisting of the N scalar elements $x_1, x_2, ..., x_N$
$\lfloor x \rfloor$	Floor of x = Largest integer not greater than x
$\lceil x \rceil$	Ceiling of x = Smallest integer not less than x
$round(x)$	$\lceil x \rceil$ if $\lvert x - \lceil x \rceil \rvert \leq 0.5$; $\lfloor x \rfloor$ otherwise.
N_Ω	Cardinality or the number of elements in a set Ω
$x \sim U[a,b]$	X is a random variable uniformly distributed between $[a,b]$
$\max(x,y)$	maximum of x and y
$\min(x,y)$	minimum of x and y
$\lceil x \rceil^a$	$\min(x,a)$
$\lfloor x \rfloor_b$	$\max(x,b)$
$[x]_b^a$	$\max(b,\min(x,a))$
$[x]^+$	$\max(x,0)$
$\mathrm{diag}(\mathbf{x})$	diagonal matrix with vector $\mathbf{x}$ as the diagonal
$(\mathbf{x})_i$	i^{th} element of vector $\mathbf{x}$
$\lvert \bullet \rvert$	Absolute value of a scalar
$\lVert \bullet \rVert$	Euclidean or L2 norm of a vector
$\min_{\substack{f^1(\mathbf{x})\leq 0 \\ f^2(\mathbf{x})\leq 0}} f(\mathbf{x})$	Minimize the objective function $f(\mathbf{x})$ over the design variable(s) $\mathbf{x}$ which satisfy the constraints $f^1(\mathbf{x})\leq 0$ and $f^2(\mathbf{x})\leq 0$.

Fixed Symbols

ε_j^i	Power Spectral Density used by User i on Subcarrier j
$\boldsymbol{\varepsilon}$	$K \times N$ matrix whose $(i,j)^{\text{th}}$ element is ε_j^i
H_j^{ii}	Direct channel coefficient for User i on Subcarrier j
H_j^{il}	Crosstalk coefficient from User i to User l on subcarrier j
ρ_j^i	(Direct Squared Channel gain) to (Interference+Noise Ratio) for User i on Subcarrier j
σ_j^i	$=\varepsilon_j^i \rho_j^i$;Received Signal-to-(Interference+Noise) ratio on User i, Subcarrier j
K	Number of Subcarriers

N	Number of users
R^i	Total Rate (summed across subcarriers) achieved by User i
$\mathbf{R}$	Total Rate Vector-N length vector whose i^{th} element is R^i
T^i	Target or demanded rate of User i.
$\mathbf{T}$	Target Rate Vector-N length vector whose i^{th} element is T^i
P^i	Total Power (summed across subcarriers) utilized by User i
$\mathbf{P}$	Total Power Vector- N length vector whose i^{th} element is T^i
$P^i_{\max}$	Total Power Constraint for User i. $P^i \leq P^i_{\max}$
$\mathbf{P}_{\max}$	Total Power Constraint Vector - N length vector whose i^{th} element is $P^i_{\max}$
γ^i_M	Performance or SNR Margin of User i
$\varepsilon_j^{\max}$	Spectral Mask or PSD constraint on subcarrier j
Δf	Inter-carrier spacing in the multicarrier system
r_j	code rate(number of information bits per coded bit) on subcarrier j.
Γ_j	SNR-Gap to Shannon capacity on subcarrier j.
n	Code word length in a RS(n,k)
k	Number of information words in a RS(n,k) code
$\mathrm{Pr}_{e,ch}$	Probability of bit-error after demodulation(before or without error correction decoding)
$\mathrm{Pr}_{e,dec}$	Probability of bit-error after error correction decoding
∂b_j^G	Granularity Loss in bits on subcarrier j
∂b^G	Total Granularity loss(summed across subcarriers)
$b(\sigma)$	Rate function
$\Re_{ALG}$	Rate Region - Set of achievable rate combinations by an algorithm ALG
$\overline{\Re_{ALG}}$	Rate Region Boundary – outermost set of achievable rate combinations by an algorithm ALG.

- Unless otherwise specified, single letter (lower case) subscripts refer to subcarrier number, and single letter superscript refers to User number
- In the above context, when distinguishing among users is not required the super-script is dropped.
- In the above context, when distinguishing among subcarriers is not required the subscript is dropped.

Acronyms and Abbreviations

ADSL	Asymmetric Digital Subscriber Loops
ANSI	American National Standards Institute
ARSFGL	Adaptive Reed Solomon aided Fine Granularity Loading
AWG	American Wire Gauge
AWGN	Additive White Gaussian Noise
DMT	Discrete Multi-tone
DS	Downstream
DSL	Digital Subscriber Loop
DSLAM	Digital Subscriber Line Access Multiplexer
DSM	Dynamic Spectrum Management
EMM	Enlightened Margin Maximization
FEQ	Frequency Domain Equalizer
FEXT	Far-End Crosstalk
FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform
ISI	Inter Symbol Interference
IWF	Iterated Water Filling
MA	Margin Adaptive (Loading Algorithm)
NEXT	Near-End Crosstalk
OFDM	Orthogonal Frequency Division Multiplexing
OSM	Optimum Spectrum Management
OTBI	Optimum Transmit Bandwidth Identification
PPO	Peak Power (constraint) Only
PSD	Power Spectral Density
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
RA	Rate Adaptive (Loading Algorithm)
RFI	Radio Frequency Interference
RS	Reed-Solomon
SCM	Single Carrier Modulation
SMC	Spectrum Management Center
SNR	Signal to interference + noise ratio
SP	Service Provider
SSM	Static Spectrum Management
TPO	Total Power (constraint) Only – cases when peak power constraint is absent.
TPP	Total and Peak Power (constraint)
US	Upstream
VDSL	Very high-speed Digital Subscriber Line

Chapter 1

Introduction

The increase in demand for higher rates and reliability in data transmission during the last three to four decades, coupled with the development of more efficient integrated circuits, has led to the development of increasingly sophisticated and high performance communication systems. Since late 1980's Multicarrier Modulation (MCM) has attracted a lot of interest as a modulation choice for communication systems. This is because it has the ability to provide variable data services, more efficient usage of resources and simpler receiver structures to deal with Inter-Symbol Interference (ISI) than Single Carrier Modulation (SCM). The two most common forms of MCM are Orthogonal Frequency Division Multiplexing (OFDM) and Discrete Multi-tone Modulation (DMT). OFDM has been used in applications such as Digital Audio Broadcasting (DAB) and accepted for Wireless Local Area Network (WLAN) standards such as IEEE802.11a. DMT has been selected for Asymmetric Digital Subscriber Loop (ADSL) [1] and is currently being considered for Very high speed DSL (VDSL) [3].

Following the terminology in [4][5], OFDM and DMT are distinguished as follows. OFDM assigns the same number of bits to each sub-channel. However DMT goes one step further in exploiting the degrees of freedom inherent in MCM by varying the number of bits on each sub-channel. For this kind of differential assignment the transmitter needs the channel state information. This requirement makes DMT best suited for almost time-invariant channels such as Digital Subscriber Loop (DSL) channels over twisted pair copper wirings.

1.1 Megabits over Copper Wires - DSL

Twisted copper wires connecting homes for plain old telephone service since a very long time had been a mammoth existing infrastructure. But when used for analog speech, only up to 3 kHz of the bandwidth was used. It was in late 1980's that the effort was made for the first time to exploit bandwidths higher beyond 3 kHz for data transmission in twisted copper pairs. This effort has led to tremendous research activity into and also

generated huge revenues from DSL technologies which use the copper wires to provide broadband access to residences and businesses by harnessing the complete bandwidth usable in the copper wires. Over the last decade, this 'usable' bandwidth has increased with advancements in the fields of digital communications and signal processing, so much so that currently standardized VDSL-DMT systems use all the way up to 12MHz bandwidth and VDSL2 systems propose to go even further. The success has been demonstrated through the massive deployment of ADSL for example which can provide up to 5Mbps through a single copper pair. VDSL sets its aim much higher.

A simplified but typical DSL loop layout scenario is presented in Figure 1-1. For each user (or subscriber), a 'DSL modem' is installed the customer premises (e.g., The houses in 'Residential Neighborhood' in Figure 1-1) and another corresponding 'DSL modem' is installed at the central office which in turn is connected to a backbone network. Notice that the copper wires connecting these modems are already present in most cases due to plain old telephone service and the only additional investment is the modems and the network infrastructure at the central office. However the layout also shows the many challenging problems that megabit transmission in the DSL scenario presents.

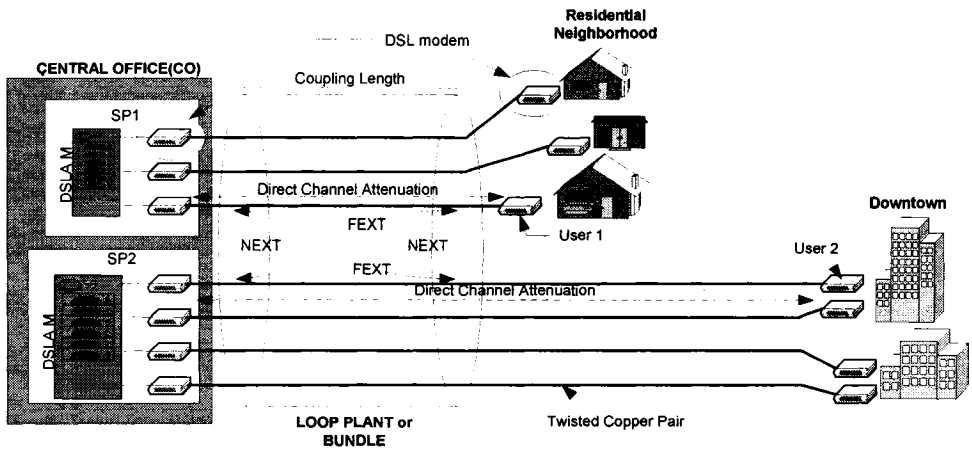


Figure 1-1: Typical DSL Loop Layout Scenario

Firstly each copper wire individually presents a highly frequency selective channel. In other words the 'Direct Channel Attenuation' in Figure 1-1 increases rapidly with fre-

quency. This high frequency selectivity corresponds to severe ISI in time domain at reasonably high signaling rates. Secondly since a number of users share the same 'Loop Plant or Bundle' for a certain 'Coupling Length' the transmission on one copper wire causes interference into another user's copper wires. This interference is commonly referred to as crosstalk (XT). The crosstalk as displayed in Figure 1-1 is of two types – Near End Crosstalk (NEXT) and Far-End Crosstalk (FEXT). Thirdly the DSL bundles as well as the modems themselves exist in close vicinity of other external sources of interference[1]. The noise (in the broad sense, i.e., including interference) in DSL environments is classified into 2 types – *capacity limiting* and *performance limiting* [22]. NEXT, FEXT, background noise and external interferences can be classified into these categories. This is discussed in Section 1.3 and 1.4.

Thus not only is the direct channel highly frequency selective, the noise also is colored, i.e., frequency dependent. Due to these reasons SCM implementations of DSL modems involve a very expensive equalizer (for dealing with ISI) and possibly noise-whitening filters. A much more economical design to deal with the above impairments is Discrete Multitone (DMT) modulation. Further, as we shall see, the degrees of freedom in terms of design provided by DMT enable us to suit transmission parameters to keep crosstalk in control and maintain robustness against external interference while meeting different users' rates (in Mbps) and performance (quality of service in terms of probability of error) based on channel state information.

1.2 Discrete Multitone Modulation

A general DMT transmitter/receiver structure is presented in Figure 1-2. In this section we provide a brief overview of the DMT structure. A more detailed introduction can be found in pp. 3-8 in [5]. The entire available bandwidth say F is divided into K parallel subchannels. Even if the channel as well as the noise might have been frequency selective over the original bandwidth F, K is chosen large enough so that over the intercarrier spacing $\Delta f=F/K$, the channel and noise experienced by each of K subcarriers appears nearly flat (or frequency non-selective).

[1] A rigorous treatment and characterization of DSL channels is presented in Chapter 2.

The input bit stream is first passed though a channel encoder. The encoded bits are then buffered and modulated onto the K sub-carriers. The modulation operation on each sub-channel is independent of other subchannels and is determined based on the channel state information. The 'Mod1' to 'Mod$K$' blocks perform this operation and output complex numbers corresponding to a baseband signal point in a 2-dimensional constellation which is generally chosen to be Quadrature Amplitude Modulation (QAM).

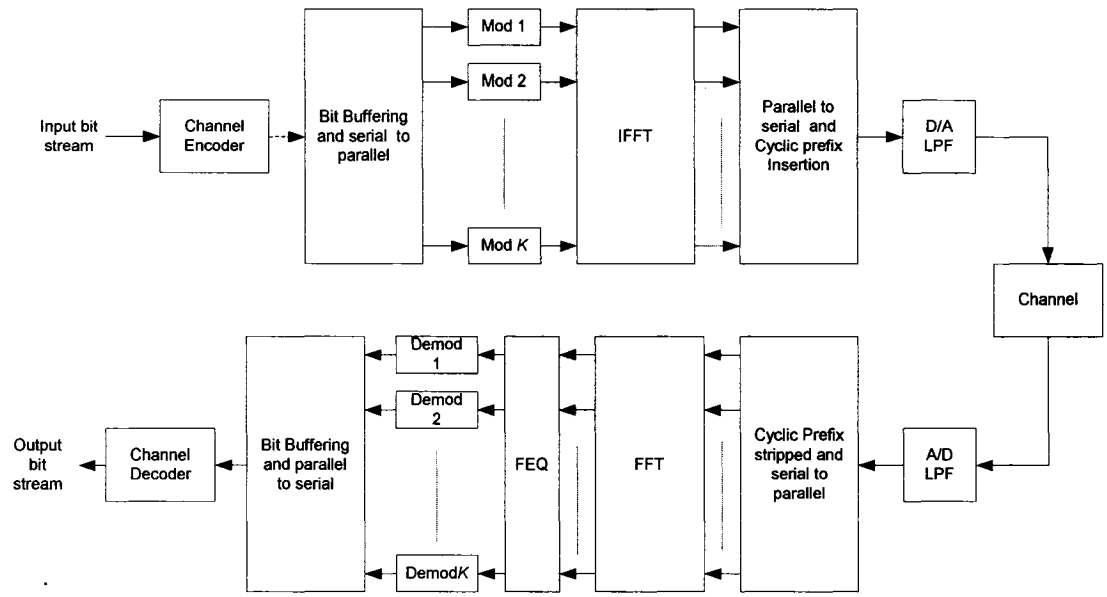


Figure 1-2: Generic DMT system model

A K-point IFFT operation is performed to generate the discrete time domain version of a single DMT symbol. Following this a guard interval, of length greater than the impulse response of the channel (see [5]) in the form of a cyclic prefix is introduced. Stripping this guard interval at the receiver ensures zero ISI between adjacent DMT symbols and that the linear convolution of the channel can be treated as a circular convolution, which in turn ensures a one tap frequency domain equalizer (FEQ) at the receiver. Other receiver blocks are inverses of operations in the transmitter.

Thus there are 3 key design parameters which can be varied based on channel state information and can be viewed as degrees of freedom offered by MCM.

- **Energy Allocation** – Each of the subcarriers can be allocated a portion of the total transmit power budget. This energy allocation might be further constrained due to spectral masks or peak power constraints used to control crosstalk among users. The energy allocated together with the channel state information on a particular subcarrier decides the final SNR which in turn indirectly decides the rate and performance. The energy allocation of one user also affects the interference profile to other users due the crosstalk coupling among copper wires sharing the bundle.
- **Bit Allocation** - Based on the channel state information and the results from the energy allocation problem, the number of bits (fractional or integer) to be allocated to each sub-carrier is decided. In reference to Figure 1-2, this amounts to choosing the QAM size in 'Mod1'-'Mod$K$' blocks. This bit-allocation summed across the subcarrier gives the overall rate of a user. Further the particular bit allocation also decides the overall BER of the system and hence must be conducted with a minimum BER or performance constraint in mind.
- **Coding Scheme** - The coding scheme is generally chosen to translate the BER from the vicinity of 10^{-3} to 10^{-7} (which is standard BER requirement for DSL applications). The coding gain that the code provides can be viewed as an increase in the energy allocated to each subcarrier which in turn allows us to load more bits and hence increase the information rate. As shown in Figure 1-2, in most schemes in practice and in literature today the coding scheme is a singular structure independent of the sub-carrier energy, bit allocation and channel state information. Hence it has been treated a one-time design decision rather than a degree of freedom. However throughout this thesis we shall exploit this degree of freedom as well and show its benefits.

The family of algorithms which varies these available degrees of freedom based on channel state information while respecting power constraints, medium constraints and system design constraints towards the objective of maximizing the rate at a specific performance (quality of service) or the objective of maximizing performance or robustness while satisfying a target demanded rate is known as loading algorithms. The channel state information needed for this purpose involves knowing, not only the direct channel attenuation,

but also the noise (and interference) affecting the channel. The noise in DSL environments is classified into 2 types – *capacity* limiting and *performance* limiting [22].

1.3 Capacity Limiting Noise

Thermal noise and crosstalk fall into the category of *capacity* limiting noise [22] because these limit the best achievable rate by the system. The existence of 20-50 twisted copper cables within the same bundle, cause significant crosstalk into each other's channels. Crosstalk is typically 10-20 dB higher than the background thermal noise [27].

Loading algorithms which have the objective of maximizing the rate achieved at a specific performance requirement are known as rate adaptive loading algorithms. Given a static noise and interference profile the capacity achieving (or rate maximizing) energy allocation has long been known to be the 'water-filling' solution (see pp. 383-390 in [7]). However this solution tacitly assumes infinite granularity in bit allocation, e.g., it assumes that something like 3.1415 bits/symbol can actually be sent on a carrier. This is very difficult to implement in practice because most of the known modulation schemes (e.g., PSK or QAM) support integer bit constellations and hence is a system-design constraint. It was initially observed in [30] and [9] that most of the integer-bit granularity losses could be recovered by rounding off the bit-allocation to integers and scaling energies accordingly after starting with a water-filling [9] or flat [30] energy distribution. However the freedom for such re-scaling is considerably reduced in the presence of peak-power constraint [2][38]. Peak-power constraint (maximum power that can be allocated to a subcarrier) arises from spectrum compatibility requirements to enable co-existence among diverse services. When the peak-power constraint is far stricter than the total-power constraint, as is often the case in VDSL-DMT, there is hardly any room left for maneuverability (or rescaling) in the energy domain (to compensate for unavailability of non-integer choices in bit-domain) and significant losses (known as granularity losses) in achievable data rates of integer-bit algorithms are observed. But we note that among the degrees of freedom stated in Section 1.2, the coding scheme has not yet been exploited. We shall show in Chapter 3 that through the exploitation of these additional degrees of freedom, most of the granularity losses can be recovered.

Crosstalk being the major impairment, spectrum management is essential to enable co-existence among multiple users and diverse services sharing the bundle. For example in Figure 1-1, a typical near-far scenario is visible. The 'Downtown' user's signal will be severely attenuated by the time it reaches the central office due to the greater attenuation of the direct channel. However since the residential users are closer to the central office, they will encounter less attenuation and also the FEXT they induce into the copper wires corresponding to the 'Downtown' users will be higher. Thus the downtown users will always be in a lose-lose situation due to the extremely strong 'Residential' users and will be able to afford much lower rates or performance than them. This scenario is definitely not desirable. Customers would like to receive rates in accordance with the service they pay for and not their geographical location. This necessitates spectrum management. In the currently standardized and implemented forms of DSL, admissible spectral masks (i.e., peak power constraint discussed above) [2] have been specified for each user (This way the damage the stronger 'Residential' users do to 'Downtown' users is controlled). These spectral masks however being based on worst case crosstalk scenario resulted in the aforementioned Static Spectrum Management (SSM) techniques to be unduly restrictive and thereby led to conservative rates. This realization has recently motivated significant research activity in Dynamic Spectrum Management (DSM) [23] – which seeks to jointly optimize transmit spectra towards minimizing crosstalk and has led to very rewarding improvements in achievable rates. Based on level of centralized control and coordination necessary, DSM schemes are graded from Level 1 to Level 3 [24]. Level 1 represents the most distributed control which is desirable in a practical situation because more than one service providers (e.g., SP1 and SP2 in Figure 1-1) often share the same bundle. Expecting a high level of coordination among diverse service providers is unrealistic.

In DSM techniques, the peak power constraint is either not present or much less strict when present, than in SSM schemes. However the rate demanded by a certain user indirectly restricts the maximum interference that other users can cause to it, thus acting as implicit peak power constraint. Hence DSM techniques also suffer from granularity losses. As we shall see in Chapter 3 an adaptive Reed Solomon scheme can be used to exploit the degrees of freedom in coding scheme described in Section 1.2 to recover these

losses. We end this section with 2 reminders. 1. Though DSM techniques are fast gaining popularity SSM is still the standardized and implemented version. 2. DSM techniques (and rate adaptive algorithms in general) present a promising solution to *capacity* limiting noise:

1.4 Performance Limiting Noise

The other major form of noise that affects DSL systems is *performance* limiting noise. *Performance* limiting noise is constituted by impulse noise and Radio Frequency Interference (RFI) which are non-stationary, geographically variable and unpredictable. Impulse noise consists of relatively high energy bursts due to electromagnetic interference from physical phenomena, electrical switches, motors and home appliances which are invariably present in the close vicinity of DSL modems. In addition to interleaving, which has the disadvantage of being the primary contributor of end-to-end delay undesirable for interactive applications [37]; the primary defense against this form of noise is performance *margin* of the system. With VDSL system bandwidth going up to 12MHz (and higher in VDSL2) and increased antenna efficiency of network cables at these frequencies RFI pickup becomes an issue [22] and amateur radio interference appearing outside reserved RF bands (due to circuit nonlinearities, imperfect filtering [37]) can make interference duration well above what interleaving and error correction can handle [37]; thus leaving *margin* as the only defense. Further since most DSL applications are constant bit rate applications; maximizing the *margin* (which is equivalent to minimizing the system probability of error as proven in lemma 4.1 in [30]) while satisfying a certain bit-rate demand is desirable [4], [30]–[36]. Further an adequate *margin* allows error free transitions (when certain modems change their mode from passive to active).

The importance of *margin* maximization due to the above reasons, has led to the development a number of *margin* adaptive loading algorithms [30]–[36]. One of the first solutions [31] to the *margin* maximization problem was based on reaching the solution by iterating over the *margin* and using the SNR-gap rate function (defined in Section 2.2.1). In [32], the authors approached the problem directly in terms of minimizing probability of error by maximizing the SNR profile leading to slightly improved results. In [34], the equivalence of the *margin* maximization problem and the *energy* minimization problem

was established. This equivalent reformulation being convex (unlike its original counterpart) was used to develop optimal *margin* maximization algorithms in [35] and [36]. The energy saved by solving the *energy* minimization problem could be directly transferred into gain in *margin* by multiplying all subcarrier energies with the ratio of the total energy constraint to the total utilized energy, i.e., optimal value of *energy* minimization problem and this ratio itself was the final *margin* of the system.

However all of these algorithms having been developed in the SSM era, have an inherent fixed crosstalk and noise profile assumption and hence are essentially single-user algorithms. Thus no direct extensions to the new DSM scenario are possible, where each user's power allocation decides other users' crosstalk profiles. Furthermore the duality between the *margin* maximization problem and *energy* minimization problem is more involved in a DSM scenario. The energy saved by solving the *energy* minimization problem cannot be converted into gain in *margin* by any simple multiplication because any such multiplication would destroy the stationarity or equilibria that exist between the crosstalk profiles of different users due to the DSM algorithm, e.g., IWF; and this in turn can result in other users loosing rate or *margin* or both[2]. Hence the multiuser *margin* maximization problem deserves an independent study instead of extensions from single user *margin*.

In chapter 4, we show that *margin* possesses certain favorable fairness properties over a multiuser rate region which can be exploited to design a Level 1 coordination multiuser *margin* maximization algorithm to ensure the best *margin* in the min-max sense to all users while guaranteeing their target (demanded) rate satisfaction.

1.5 Contributions and Thesis Outline

Chapter 2 begins with a characterization of the DSL direct and interference channels. Subsequently a near exhaustive review of various optimization problems[3] that have been solved in literature in the context of loading algorithms, solution properties and practical implementations is presented. The review begins with the simplest theoretical water-

[2] Due to this reason there is some skepticism about modems working in Margin Adaptive (MA) in DSM scenario [29] which used to be very popular during SSM phase [31]–[36].

[3] The concept of Karush Kuhn Tucker conditions from classical optimization theory are utilized in this chapter. A brief review of these conditions is presented in Appendix B.

filling problems and progressively addresses problems of increasing difficulty with additional constraints until the true problem encountered by loading algorithms in practical DSL modems. The various problems are classified based on their objective function (the cost they intend to minimize), the constraint they address, complexity of implementation and optimality of result achieved. The motivations towards a fine-granularity scheme and the absence of a multiuser *margin* maximization algorithm are outlined.

Chapter 3 begins with a theoretical quantification of granularity loss. This further justifies the claim that granularity losses due to integer bit constraint are substantial in current DSL scenarios[4]. To recover these losses the Adaptive Reed Solomon aided Fine Granularity Loading (ARSFGL) scheme is proposed[4,5]. The performance of this scheme is compared with integer bit algorithms in both SSM and DSM cases with phenomenal results. In SSM cases ARSFGL schemes result in 20% improvement in rate achieved in most scenarios. The extension of ARSFGL scheme to DSM cases results in a purely distributed scheme which can give better or equal rate as compared to the centralized optimal integer-bit scheme[6].

In the process of comparing the ARSFGL scheme developed in Chapter 3 with various integer-bit algorithms in use today for various DSL scenarios, we discovered that bit-rounding algorithms [30][9] which are very popular loading algorithms load bits aggressively leading to minimum performance requirement violation. We found the causes of these violations and suggested repairs[7]. This is presented in Appendix A. In Chapter 3, we compare the ARSFGL scheme with these repaired versions of the bit-rounding algorithms described in Appendix A along with other algorithms described in Chapter 2.

---

[4] Saswat Panigrahi and Tho Le-Ngoc, "Fine-Granularity Loading Schemes using Adaptive Reed-Solomon Coding for xDSL-DMT Systems," *EURASIP – Journal on Applied Signal Processing*, to appear 4th quarter 2005.

[5] Saswat Panigrahi and Tho Le-Ngoc, "Fine-Granularity Loading Schemes using Adaptive Reed-Solomon Coding for Discrete Multitone Modulation Systems", *IEEE International Conference on Communications* (ICC), May 2005. (5 pp. - proceedings to be published)

[6] Saswat Panigrahi, Yang Xu and Tho Le-Ngoc, "Enhanced Multiuser Resource Allocation using near-continuous loading algorithms with Adaptive coding for Digital Subscriber Loops", *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 132-135, May 2005.

[7] Saswat Panigrahi and Tho Le-Ngoc, "Total Power Constraint Equality in Integer Bit Loading Algorithms for Multicarrier Systems", *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 116-119, May 2005. **** Received 'Student Paper Prize Award' at CCECE 2005. ****

Chapter 4 addresses the problem of multiuser *margin* maximization. Due to the near-continuous results of the ARSFGL scheme developed in Chapter 3, throughout this chapter continuous bit loading is assumed which definitely simplifies the problem. We discover certain fairness properties that multiuser *margin* exhibits over a multiuser rate region. We exploit these properties to develop a Level 1 coordination algorithm which achieves the best common equal *margin* for all users. The fairness properties discovered and demonstrated here are entirely new and the developed algorithm is the first multiuser *margin* maximization algorithm in a DSM scenario[8].

Chapter 5 summarizes the conclusions of the thesis.

[8] Saswat Panigrahi, Yang Xu and Tho Le-Ngoc, "Enlightened Margin Maximization in Multiuser Interference Digital Subscriber Line Channels," accepted at *IEEE Global Communications (GLOBECOM) Conference*, 2005.

Chapter 2

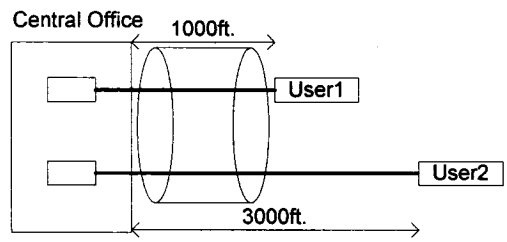
Literature Review

2.1 DSL Channels

In Figure 1-1 we presented a simplified DSL loop layout scenario and Section 1.3 gave a qualitative overview of crosstalk and direct channel attenuation. In this section we examine the crosstalk and direct channel attenuation in greater detail. Consider the general scenario of N users (twisted pairs) sharing a bundle. In Figure 1-1 e.g., N=7. In real life DSL scenarios N can vary between 20 and 50. The channel transfer function (of crosstalk) from User i to User l with $l \neq i$ is denoted by $H^{il}(f)$. Similarly the direct channel transfer function for User i is $H^{ii}(f)$. $H^{ii}(f)$ depends on the length of the loop, source impedance, load impedance, and the characteristic impedance of the loop (see Section A.2.1 in [2]). The absolute value of $H^{il}(f)$ is modeled in terms of $H^{ii}(f)$ and the coupling length[9] l_c as (see FEXT models in [2] and [3]):

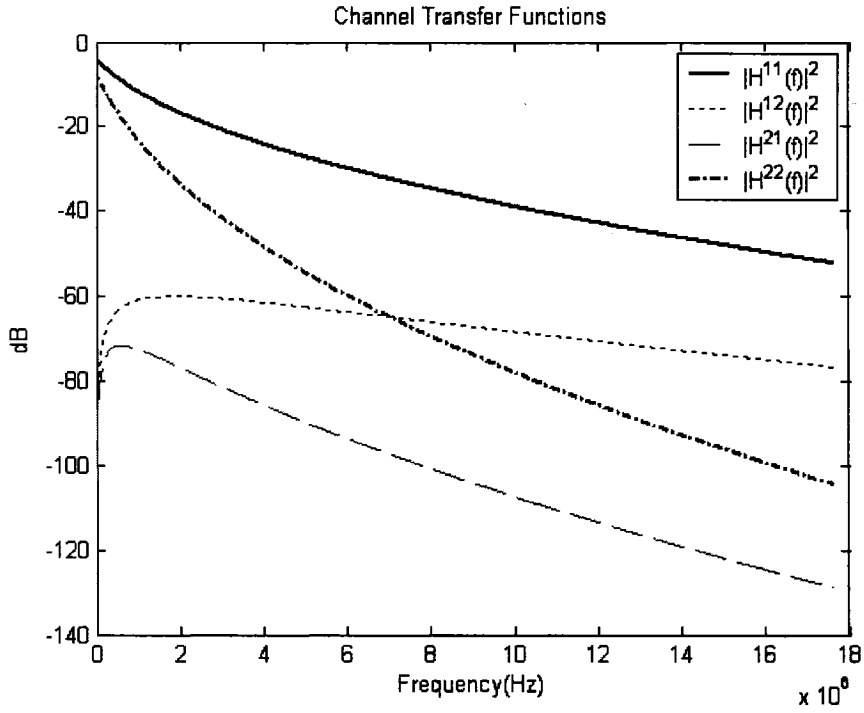
$$\left|H^{il}(f)\right|^2 = K_{FEXT} \cdot f^2 l_c \left|H^{ii}(f)\right|^2 \tag{2.1}$$

where K_{FEXT} is a constant. A simpler scenario (than Figure 1-1) of 2 loops, as shown in Figure 2-1(a), one of which is at 1000 ft. (User 1) from the central office and the other at 3000ft. (User 2) from the central office is considered. Notice that l_c=1000ft. in this case. The corresponding direct channel transfer functions $\left|H^{11}(f)\right|^2$, $\left|H^{22}(f)\right|^2$ and the crosstalk transfer functions $\left|H^{12}(f)\right|^2$ and $\left|H^{21}(f)\right|^2$ are plotted in Figure 2-1(b).



(a) Setup

[9] Coupling length is the distance for which the N twisted pairs travel within the same bundle.



(b) Transfer Functions

Figure 2-1: Sample DSL Channel Transfer Functions

The following important observations from Figure 2-1(b) will be useful throughout this thesis:

- The channel is asymmetric, i.e., $|H^{12}(f)|^2 \neq |H^{21}(f)|^2$.
- The User closer to the CO (User1 in this case) has a much better direct channel, i.e., $|H^{11}(f)|^2 >> |H^{22}(f)|^2$ and causes strong crosstalk (see $|H^{12}(f)|^2$) to User 2; which already has a poor or weak direct channel.
- If we consider either $|H^{11}(f)|^2$ or $|H^{22}(f)|^2$ individually, the direct channel of both users exhibits a similar trend. The attenuation is low at low frequencies and rapidly increases towards higher frequencies.
- DSL channels exhibit very high frequency selectivity, e.g., $|H^{22}(f)|^2$ changes by more than 100dB over the 18MHz bandwidth. Thus if SCM is used at reasonably high signaling rates, the high frequency selectivity will lead to severe inter-symbol interference (ISI). To counter such ISI, a very expensive equalizer would be required. Multicarrier methods [5] have proved to an efficient alternative to

deal with frequency selective channel. In this case, the entire bandwidth is divided into K subcarriers with a uniform inter-carrier spacing of Δf. Note that the typical value of K for ADSL is 256 and VDSL is 4096. For Δf small enough each subcarrier will experience a nearly flat or frequency non-selective channel, i.e., nearly be ISI free. Further since the lower subcarriers are expected to have much better SNRs, it would seem natural to vary the bits or QAM-size along with the energy allocation in order to achieve higher rates or better performance at fixed rate. OFDM is a widely used multicarrier scheme. In [5] and [9] an OFDM scheme that does not keep the bit allocation equal across subcarriers is called DMT and we shall use the same terminology.

The generic class of algorithms, which allocate energy and bits with the objective to maximize rate, performance or robustness given energy constraints, system-design constraints and performance requirements, are known as loading algorithms. We first discuss single user loading algorithms in Section 2.2 and then multiuser loading algorithms (or DSM) in Section 2.3.

2.2 Single User Loading Algorithms

Loading algorithms require channel state information to be available at the transmitter to optimize the energy and bit-allocation. Since DSL channels are quasi-static in time, this requirement is satisfied much more easily than say in wireless channels. The rich dividends in terms of overall system performance, obtained from loading algorithms, the number of possibilities and the practical feasibility of their implementation has led to intense research in the field over the last decade [8]–[11],[25],[27],[30]–[36],[38].

Let us consider a multicarrier system of K subcarriers. Let ε_j be the controllable transmit Power Spectral Density (PSD) and ρ_j be the normalized received SNR when ε_j=1 over the j^{th} sub-carrier. The inter-carrier spacing Δf is assumed to be small enough and the cyclic prefix is assumed to be larger than the delay spread of the channel so that each subcarrier can be treated as an independent AWGN channel. Thus the channel of the j^{th} subcarrier can be completely characterized by ρ_j and the received SNR is $\sigma_j = \varepsilon_j \rho_j$.

The rate-function $b(\sigma)$ is defined as the maximum information rate in bits that can be supported at SNR of σ while keeping the conceded error probability lower than or equal to a specified target error probability [38]. Clearly, the rate function contains the summary of the behavior of the modulation and coding scheme in use.

Loading algorithms with the objective of maximizing rate are known as Rate Adaptive (RA) loading algorithms and those with the objective of maximizing the performance *margin* are known as Margin Adaptive (MA) loading algorithms. We discuss single user RA algorithms in Section 2.2.1 and single user MA algorithms in Section 2.2.2.

2.2.1 Rate Adaptive Loading Algorithms

The simplest rate-maximizing optimization problem is stated as follows. The objective function is the total supported rate,

$$R = \sum_{j=1}^{K} b\left(\varepsilon_j \rho_j\right). \tag{2.2}$$

The traditional total power constraint and the constraint of non-negative power distribution can be expressed as

$$\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j \le P_{\max} \quad \text{and} \quad \varepsilon_j \ge 0, \quad 1 \le j \le K. \tag{2.3}$$

2.2.1.1 Rate Adaptive Water-Filling

The first solution for this type of total power constrained is the Orthodox Water-Filling solution for the capacity of parallel AWGN channels by Gallager (see pp.383-390 in [7]). However in [7] $b(\sigma)$ was assumed to be the Shannon capacity, i.e., $\log_2(1+\sigma)$ because the objective was to determine capacity. But in reality we never truly operate at Shannon capacity. To account for this non-ideality the concept of SNR-gap Γ was introduced [5][4]. Γ is a single parameter characterization of the distance from Shannon capacity in terms of SNR[10]. Thus the SNR-gap rate function is $b_{GAP}(\sigma) = \log_2\left(1+\sigma/\Gamma\right)$.

[10] The value of Γ depends on the target symbol error probability. For example, for uncoded QAM, and a target SER of 10^{-7}, Γ=9.8dB [5][31] and while for a target symbol error probability of 10^{-3}, Γ=5.48dB [11].

Thus the SNR-gap based rate maximization[11] problem is

$$\min_{\substack{\Delta f\sum_{j=1}^{K}\varepsilon_j \le P_{\max} \\ \varepsilon_j \ge 0}} \left(-\sum_{j=1}^{K} \log_2\left(1+\frac{\varepsilon_j \rho_j}{\Gamma}\right)\right). \tag{2.4}$$

In the above problem, the objective function is a sum of negative of logarithms which is easily proven to be convex. The constraints are also affine (i.e., linear and hence convex). Thus we have a convex optimization problem. The Lagrangian corresponding to the above problem is:

$$L\left(\{\varepsilon_j\}_{j=1}^{K};\lambda,\{\mu_j\}_{j=1}^{K}\right) = -\sum_{j=1}^{K}\log_2\left(1+\frac{\varepsilon_j\rho_j}{\Gamma}\right)+\lambda\left(\sum_{j=1}^{K}\varepsilon_j-\frac{P_{\max}}{\Delta f}\right)-\sum_{j=1}^{K}\mu_j\varepsilon_j\,. \tag{2.5}$$

For a convex optimization problem with differentiable objective and constraints, we know that the KKT conditions (see Appendix B) are necessary and sufficient conditions of optimality [41]. Thus the optimal power allocation $\{\varepsilon_j^*\}_{j=1}^{K}$ must satisfy the KKT conditions corresponding to the above Lagrangian are stated in (2.6)-(2.8) for some $\lambda \ge 0$ and $\mu_j \ge 0; 1 \le j \le K$.

$$\frac{\partial L(.)}{\partial \varepsilon_j} = \frac{-1}{\ln 2}\cdot\frac{1}{\varepsilon_j^* + \Gamma/\rho_j} + \lambda - \mu_j = 0;\;\; 1 \le j \le K \tag{2.6}$$

$$\mu_j \varepsilon_j^* = 0;\;\; 1 \le j \le K \tag{2.7}$$

$$\lambda\left(\sum_{j=1}^{K}\varepsilon_j^* - P_{\max}/\Delta f\right) = 0 \tag{2.8}$$

Note that between (2.6)-(2.8), there are $2K+1$ unknowns and $2K+1$ equations. For the constraints we know that $\varepsilon_j^* \ge 0;\; 1 \le j \le K$. For all l such that $\varepsilon_l^* > 0$, from (2.7) we have $\mu_l = 0$. If $\mu_l = 0$, (2.6) for $j=l$ simplifies to $\varepsilon_l^* = 1/\lambda \ln 2 - \Gamma/\rho_l$. Now using the assumption $\rho_j > 0;\; 1 \le j \le K$, we obtain the following 'water-filling' conditions for the SNR-gap rate function with total power constraint.

---

[11] In this thesis all maximization problems will be changed into minimization by adding a negative sign to the objective function which is a common practice in most optimization texts. This is because there is no loss of generality in this process and in optimization theory most theorems and algorithms are stated for minimization problems and hence treating maximization problems as minimizations problems enables more direct usage of these available tools.

$$\varepsilon_j^* = \left[\frac{1}{\lambda \ln 2} - \frac{\Gamma}{\rho_j}\right]^+ \tag{2.9}$$

where $[x]^+ = \max(x, 0)$ and λ is obtained by solving

$$\Delta f \cdot \sum_{j=1}^{K} \left[\frac{1}{\lambda \ln 2} - \frac{\Gamma}{\rho_j}\right]^+ = P_{\max}. \tag{2.10}$$

Note the appearance of Γ in (2.9) as compared to the orthodox water-filling. The water-filling level is $1/\lambda \ln 2$ and is illustrated in Figure 2-2.

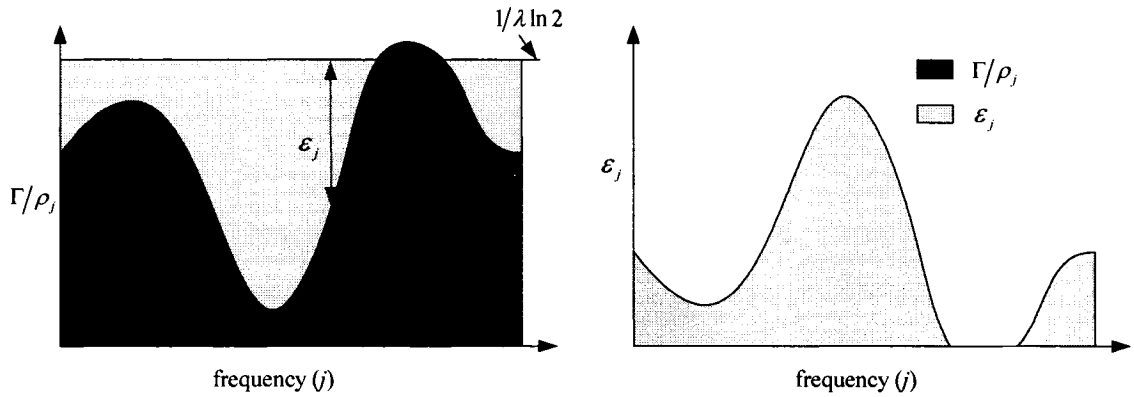


Figure 2-2: Illustration of Rate-Adaptive Water-Filling

Understanding of water-filling is vital to form an intuitive understanding of loading. But notice that after we obtain the optimal energy allocation $\{\varepsilon_j^*\}_{j=1}^K$ from (2.9), the final bit allocation on the j^{th} subcarrier will be $\log_2(1+\varepsilon_j^*\rho_j/\Gamma)$ which is a *real* number. But most practical modulation schemes such as M-PSK or M-PAM or M-QAM can support only integer bits/symbol. Thus water-filling though theoretically attractive and intuitively useful, assumes continuous bit loading and hence cannot be implemented practically.

2.2.1.2 Integer-Bit Constraint

Integer-bit constraint arises out of the fact that most known generic M-ary modulation schemes can support only integer-bits/symbols. Thus we need that the final bit-allocation to be integer-bit over each subcarrier. This significantly complicates the problem. This is because, instead of having a convex smooth nonlinear optimization problem, we now have an integer programming problem. Based on the SNR-gap method we now

have the following step function as our rate function: $b_{GAP}^{\text{int}} = \lfloor \log_2(1+\sigma/\Gamma) \rfloor$. The floor operation follows from the definition of rate function and guarantees integer bit distribution. Thus we now have the following problem:

$$\min_{\substack{\Delta f \sum_{j=1}^{K} \varepsilon_j \le P_{\max} \\ \varepsilon_j \ge 0}} \left(-\sum_{j=1}^{K} \left\lfloor \log_2\left(1+\frac{\varepsilon_j \rho_j}{\Gamma}\right) \right\rfloor \right). \tag{2.11}$$

The above being an integer programming problem, the optimal solution is a greedy search based method [8] proposed by Hughes-Hartogs.

Table 2-1: Hughes Hartogs Algorithm

1.	$b_j = 0, \varepsilon_j = 0; 1 \le j \le K; P = 0, R = 0$
2.	*Repeat*
3.	*for* $j = 1$ to K
4.	$\Delta\varepsilon_j = \left(2^{b_j+1} - 2^{b_j}\right) \cdot \Gamma / \rho_j$
5.	*End*
6.	Determine $\hat{j}$ where $\Delta\varepsilon_j$ is minimum over $1 \le j \le K$
7.	$b_{\hat{j}} = b_{\hat{j}} + 1; \varepsilon_{\hat{j}} = \varepsilon_{\hat{j}} + \Delta\varepsilon_{\hat{j}}; R = R + 1; P = P + \Delta\varepsilon_{\hat{j}} \Delta f$
8.	*until* $\left((P \ge P_{\max}) or (R \ge T)\right)$

Hughes Hartogs Algorithm

The HH algorithm is presented in Table 2-1. At each iteration, HH algorithm generates a table of incremental energies $\Delta\varepsilon_j$ required to transmit one additional bit on each of the subcarriers (see steps 3-5). Then the subchannel with the minimum incremental energy requirement is chosen and one bit is loaded and the energy assignment ε_j and the total utilized power P are updated accordingly (see steps 6, 7). The process is repeated till the target rate[12] T is met or the total power constraint $P_{\max}$ is violated.

[12] Since in this section we are studying only RA algorithms, T can be set to a very high value so that HH algorithm stops only when the total power constraint is violated thus ensuring rate maximization.

The HH algorithm is the optimal solution for the integer-bit total power constrained problem (2.11) and hence is an important benchmark for other algorithms[13]. But note that, at each iteration, only 1 bit is added and the extensive sorting and searching makes the algorithm's complexity grow exponentially with the total number of bits loaded and the total number of subcarriers. Due to this reason, Hughes-Hartogs algorithm is impractical for usage in DSL applications in which both number of bits and number of subchannels is large. This led to slightly sub-optimal but more efficient bit-rounding algorithms.

Bit-Rounding Algorithms

Bit-rounding algorithms [4][9][30] are practically attractive because they are relatively *computationally* efficient (as opposed to greedy methods such as HH) and *implementationally* realizable because they ensure integer-bit distribution (unlike water-filling). The algorithms proposed by Chow in [30] (Sec. 4.3.4), [4] (Sec. 7.2.3.1) and by Leke in [9] start with an Optimum Transmit Bandwidth Identification (OTBI) procedure (which determines the subcarriers to be used). This identification is done through a costly iterative procedure in Chow's algorithm [4][30] which is presented in Table A-1. In Leke's algorithm [9] a non-negative energy assignment criterion was introduced, by virtue of which the identification is achieved in a single iteration over the subcarriers. The OTBI procedure of Leke's algorithm is presented in Table A-2. Furthermore, while initial energy distribution is flat in Chow's algorithm [4] [30], Leke's algorithm [9] starts with a water-filling distribution. At the end of the OTBI procedure, the total power constraint is satisfied with equality $\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j = P_{\max}$, but the bits allocated are non-integer.

The common aspect of both algorithms is that following the OTBI, the bits allocated to each of the N subcarriers, $\{b_j\}_{j=1}^{K}$ obtained from the usage of SNR-gap function, $b_{GAP}(\sigma) = \log_2(1+\sigma/\Gamma)$, are rounded to integers, i.e.,

$$\hat{b}_j = round\{b_j\};\ \ 1 \le j \le K \tag{2.12}$$

and energy allocated to each subcarrier $\{\varepsilon_j\}_{j=1}^{K}$ is adjusted to make the BER nearly equal in the sub-carriers, leading to the saw-tooth energy distribution [30], i.e.,

[13] To understand the optimality of greedy HH algorithm for (2.11), notice that the problem is a Separable Convex Discrete Resource Allocation (SCDRA) problem. See Chapter 2 in [33] for more details.

$$\varepsilon_j = \rho_j^{-1} b_{GAP}^{-1}\left(\hat{b}_j\right) = \rho_j^{-1}\left(2^{\hat{b}_j} - 1\right)\cdot\Gamma. \tag{2.13}$$

However, following this adjustment of energies in (2.13), the total utilized energy $P = \Delta f \cdot \sum_{j=1}^{K} \varepsilon_j$ is no longer equal to the total energy constraint of the system $P_{\max}$. To re-achieve this equality, in the final step of both the algorithms, $\{\varepsilon_j\}_{j=1}^{K}$ are multiplied by an energy re-scaling factor, i.e.,

$$\varepsilon_j = \varepsilon_j \times \varepsilon_{RF};\ \text{where } \varepsilon_{RF} = P_{\max}/P. \tag{2.14}$$

Leke's algorithm [9] is computationally more efficient that Chow's algorithms [4][30], due to an efficient OTBI procedure. In terms of rate achieved, Leke's algorithms are better than Chow's algorithms, but only by a very minor degree. In our simulations for DSL channels (see Section 3.4.3) we observed this improvement to be less than 1% for most cases, and a similar observation of approximately 2% percent was made in [9]. Both Chow's and Leke's algorithms are much more computationally efficient[14] than the optimal HH algorithm, but at same time they are very close in achieved rate[15] to the optimal rate achieved by HH algorithm. In our simulations for DSL channels (see Section 3.4.3) we observed this improvement to approximately 1% for all tested cases.

For reference purposes we shall refer to all the cases without peak power constraint discussed so far as Total Power Only (TPO) cases. While the algorithms discussed in this section can handle the total power constraint and the integer-bit constraint, they cannot handle the peak power constraint.

2.2.1.3 Peak Power Constraint

Peak power constraints arise out of spectral compatibility requirements which ensure a minimum rate or performance for all the users sharing a medium. Reconsider, for example the scenario discussed in Section 2.1 (see Figure 2-1). If we have only the total power constraint, then going by a water-filling type approach User 1 will utilize most of this total power budget in the lower subcarriers, because $H^{11}(f)$ is best in the lower frequencies. Now User 2, by itself has a weak direct channel $H^{22}(f)$ and hence most of

[14] Order of Complexity comparisons for HH, Chow's and Leke's algorithms is found in Table I in [11].

[15] These bit-rounding algorithms have one weakness though. The rounding procedure (2.12)-(2.13) is inherently biased which leads to aggressive loading which in turn results in the BER constraint violation (see Appendix A).

quencies. Now User 2, by itself has a weak direct channel $H^{22}(f)$ and hence most of its higher frequencies will be unusable. So the only usable subcarriers would be the lower frequencies, but these very frequencies would be swamped with interference arising from the water-filling approach of User 1. Thus User 2 in Figure 2-1(a) or the downtown users in Figure 1-1 will be denied a basic minimum rate. This is known as 'hogging' in DSL. Recognizing the gravity and frequent occurrence of this problem, apart from the total power constraint $\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j \le P_{\max}$, admissible spectral masks $\varepsilon^{\max}(f)$ are also specified for each modem by standardization bodies. Examples of $\varepsilon^{\max}(f)$ in case of current DSL standards are SMClass1 – SMClass9 in [2] and M1FTTCab, M2FTTCab in [3]. This spectral mask acts as the peak power constraint $\varepsilon_j \le \varepsilon_j^{\max};\ 1 \le j \le K$, where $\varepsilon_j^{\max} = \varepsilon^{\max}(j\Delta f)$.

The peak and total power constrained problem is defined as follows:

$$\min_{\substack{\Delta f \sum_{j=1}^{K} \varepsilon_j \le P_{\max} \\ \varepsilon_j \le \varepsilon_j^{\max} \\ \varepsilon_j \ge 0}} \left(-\sum_{j=1}^{K} b(\varepsilon_j \rho_j) \right). \tag{2.15}$$

To gain insight into the solution structure of the above problem, let us temporarily drop the integer bit constraint and assume that the rate function $b(\sigma)$ is a continuously differentiable, strictly increasing and strictly concave function. Note that the SNR-gap rate function satisfies all these properties. Due to the strict concavity assumption the objective function in the above problem $-\sum_{j=1}^{K} b(\varepsilon_j \rho_j)$ is convex and the constraints are affine (i.e., linear and hence convex). Thus (2.15) is a convex optimization problem, and hence once more the optimal energy allocation must satisfy the KKT conditions [41]. This analysis is carried out in [38]. The Lagrangian in addition to (2.5) must include multipliers ν_j to account for the peak power constraint. Thus the Lagrangian is

$$\begin{aligned} &L\left(\{\varepsilon_j\}_{j=1}^{K}; \lambda, \{\mu_j, \nu_j\}_{j=1}^{K}\right) = \\ &-\sum_{j=1}^{K} b(\varepsilon_j \rho_j) + \lambda\left(\sum_{j=1}^{K} \varepsilon_j - \frac{P_{\max}}{\Delta f}\right) - \sum_{j=1}^{K} \mu_j \varepsilon_j + \sum_{j=1}^{K} \nu_j \left(\varepsilon_j - \varepsilon_j^{\max}\right). \end{aligned} \tag{2.16}$$

Applying the KKT conditions, as in Section 2.2.1.1, the following general result can be obtained about the optimal energy allocation [38]:

$$\text{If} \quad \Delta f \cdot \sum_{i=1}^{K} \varepsilon_i^{\max} \le P_{\max}, \quad \text{then } \varepsilon_j^* = \varepsilon_j^{\max}. \tag{2.17}$$

$$\text{If} \quad \Delta f \sum_{i=1}^{K} \varepsilon_i^{\max} > P_{\max}, \quad \text{then}$$

$$\varepsilon_j^* = \varepsilon_j^*(\lambda) = \begin{cases} \varepsilon_j^{\max}, & \text{if } \lambda \le \rho_j b_\sigma\left(\rho_j \varepsilon_j^{\max}\right) \\ \dfrac{1}{\rho_j} b_\sigma^{-1}\left(\dfrac{\lambda}{\rho_j}\right), & \text{if } \rho_j b_\sigma\left(\rho_j \varepsilon_j^{\max}\right) \le \lambda \le \rho_j b_\sigma(0), \\ 0, & \text{if } \lambda \le \rho_j b_\sigma(0) \end{cases} \tag{2.18}$$

where $b_\sigma(\sigma) = \partial b(\sigma)/\partial\sigma$ and $b_\sigma^{-1}(.)$ is the inverse of $b_\sigma(.)$. The parameter λ is the KKT multiplier corresponding to the total power constraint and is the solution to:

$$\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^*(\lambda) = P_{\max}. \tag{2.19}$$

Notice the remarkable result in (2.17) that if $\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^{\max} \le P_{\max}$, the optimal energy allocation ε_j^* is independent of the rate-function $b(\sigma)$ and is equal to the peak power constraint $\varepsilon_j^{\max}$. We refer to this scenario as the Peak Power Only (PPO) scenario henceforth. First we discuss the cause of this result and then its implications. In the problem definition (2.15), we have 3 constraints. The peak power constraint requires that $\varepsilon_j \le \varepsilon_j^{\max}$ and the total power constraint requires that $\Delta f \sum_j \varepsilon_j \le P_{\max}$. Note that if the condition of (2.17) is satisfied i.e., $\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^{\max} \le P_{\max}$; then in (2.15), satisfying the peak power constraint guarantees that the total power constraint is satisfied. So in this case, the total power constraint can be dropped. But this in turn simplifies the problem, because the total power constraint was the only thing that was 'coupling' the optimization. Once this is removed, we have the problem can be solved per subcarrier, e.g., on the jth subcarrier

$$\varepsilon_j^* = \arg \min_{0 \le \varepsilon_j \le \varepsilon_j^{\max}} -b\left(\varepsilon_j \rho_j\right).$$

But since $b\left(\varepsilon_j \rho_j\right)$ is monotonically increasing in ε_j by assumptions of the rate function, we have the final solution as $\varepsilon_j^* = \varepsilon_j^{\max}$. The implication of this result is that in the PPO case, no matter what is the rate function (i.e., what coding and modulation is being used) and what is the method or loading algorithm being used, the solution is the same. Further we shall see in chapter 3 that, in the PPO case, the granularity loss resulting from the integer bit restriction is the highest. When $\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^{\max} > P_{\max}$, we refer to this scenario henceforth as the Total and Peak Power (TPP) case, because here unlike in PPO where peak power constraint completely dominates, both total power constraint and peak power constraint play a role.

For the special case of the SNR-gap rate function $b_{GAP}(\sigma) = \log_2(1+\sigma/\Gamma)$, the result for PPO case, i.e., (2.17) stays exactly the same as discussed above, but for TPP case (2.18),(2.19) simplify to

$$\varepsilon_j^* = \left[\frac{1}{\lambda \ln 2} - \frac{\Gamma}{\rho_j}\right]_0^{\varepsilon_j^{\max}}, \tag{2.20}$$

$$\Delta f . \sum_{j=1}^{N} \left[\frac{1}{\lambda \ln 2} - \frac{\Gamma}{\rho_j}\right]_0^{\varepsilon_j^{\max}} = P_{\max} . \tag{2.21}$$

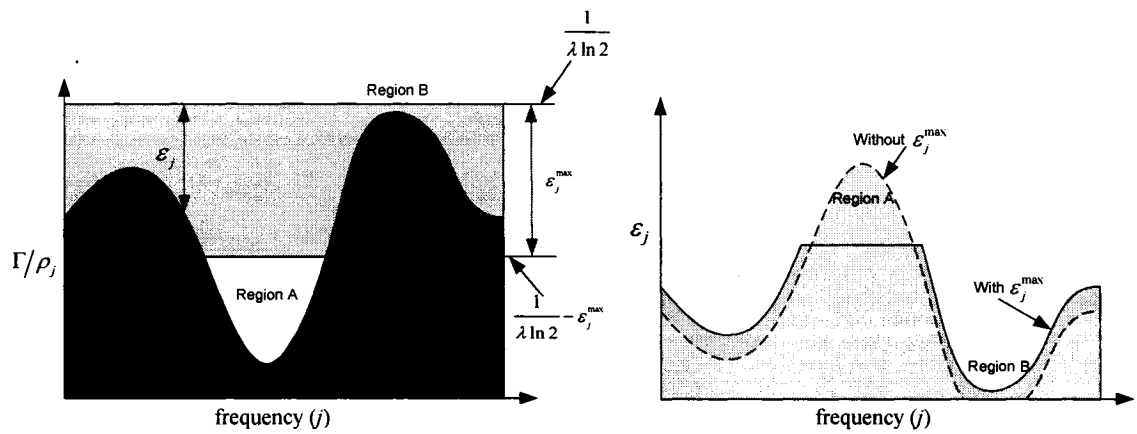


Figure 2-3: Effect of Peak Power Constraint

A graphical interpretation (for the same channel as considered in Figure 2-2) of the TPP case is presented in Figure 2-3. In the figure we have assumed that $\varepsilon_j^{\max}$ is flat (or

constant across j) for simplicity of illustration. Figure 2-3(a) illustrates (2.20) and particularly 'Region A' displays the role of the peak power constraint. In Figure 2-3(b), the energy allocation corresponding to (2.20) and the one corresponding to water-filling (Figure 2-2) is shown for comparison. Notice that the catastrophic interference that would be caused to other users due to high power allocation in 'Region A' is now contained due to $\varepsilon_j^{\max}$. Further the energy saved here is redistributed among the other subcarriers, e.g., in 'Region B'. Previously in the case of water-filling there was not enough power, but now due to the energy saved in Region A, Region B can be loaded. Of course, if we consider only this single user, this redistribution does not mean that we can achieve the same rate as we did in water-filling. By applying peak power constraint the rate for a particular user can only reduce from the water-filling solution. But when viewed from the multiuser perspective, avoiding 'Region A' type of sharp allocations, every user can be guaranteed a basic minimum rate which was the objective of the peak power constraint.

While (2.20) and (2.21) might look deceptively simple due to their closed form, reaching the solution algorithmically is not a straightforward task because of the 2 caps, $[.]_0^{\varepsilon_j^{\max}}$. A very efficient method to reach this solution was proposed in Table I in [38]. This follows the observation that the key equation to be solved is (2.21), because once λ obtained from (2.21), evaluating (2.20) for the optimal ε_j^* is easy. Let B=1/ λln2 and

$$f(B) = \Delta f \cdot \sum_{j=1}^{K} \left[B - \frac{\Gamma}{\rho_j} \right]_0^{\varepsilon_j^{\max}} - P_{\max} . \tag{2.22}$$

It is easy to see that solving (2.21) is equivalent to finding a zero of the function $f(B)$, i.e., B^* such that $f(B^*)$=0. It was observed in [38] that this B^* must lie in the interval $[x_0, x_1]$, where $x_0 = \min_{1 \le j \le K} \{\Gamma/\rho_j\}$ and $x_1 = \max_{1 \le j \le K} \{\varepsilon_j^{\max} + \Gamma/\rho_j\}$ [16]. Further it can be seen that $f(B)$ is a continuous function and in the interval $[x_0, x_1]$ it is monotonically increasing. Thus there exists only one zero of $f(B)$ in $[x_0, x_1]$ which can found by any one-

[16] This can be proven by contradiction, e.g. assume that $B^* < x_0$. This implies $B^* - \Gamma/\rho_j < 0, \forall j$ which implies in turn that $\sum \left[B^* - \Gamma/\rho_j \right]_0^{\varepsilon_j^{\max}} = 0 \Rightarrow f(B^*) = -P_{\max}$ but by definition $f(B^*)$=0. Similarly if we assume $B^* > x_1$, we obtain $\Delta f \sum \varepsilon_j^{\max} = P_{\max}$ but by definition of TPP case $\Delta f \sum \varepsilon_j^{\max} > P_{\max}$

dimensional search technique, e.g., bisection or secant. In [38], a secant method is proposed which converges very fast. A modified version of this algorithm will be discussed in Chapter 3. Also note that the solution for TPO, i.e., water-filling can also be reached by this algorithm as a special case by setting $\varepsilon_j^{\max} = P_{\max}/\Delta f; 1 \le j \le K$.

However the problem just as in the case of water-filling is that even after we find the optimum energy allocation ε_j^*, the corresponding bit-allocation by the SNR-gap function assumption would be $\log_2\left(1+\varepsilon_j^*\rho_j/\Gamma\right)$ which is non-integer and hence cannot be implemented in practice since most known modulation families like M-PSK and M-QAM correspond only to integer bit allocations. Hence we must consider the peak power constraint along with the integer-bit constraint for practical reasons.

Peak Power Constraint with Integer-bit Constraint

Now we are in a position to define the complete rate adaptive problem faced by practical modems at this stage which encounter total power constraint, peak power constraint and integer bit constraint together.

$$\min_{\substack{\Delta f\sum_{j=1}^{K}\varepsilon_j \le P_{\max} \\ \varepsilon_j \le \varepsilon_j^{\max} \\ \varepsilon_j \ge 0}} \left(-\sum_{j=1}^{K} \left\lfloor \log_2\left(1+\frac{\varepsilon_j\rho_j}{\Gamma}\right) \right\rfloor \right) \tag{2.23}$$

Just like in the continuous rate function case considered above, for a PPO scenario, the optimal energy allocation even in the presence of integer-bit constraint is $\varepsilon_j^* = \varepsilon_j^{\max}$. The reasoning behind this is the same as in the continuous case. In recollection, when $\Delta f \cdot \sum_j \varepsilon_j^{\max} \le P_{\max}$ (by definition of PPO), the total power constraint in (2.23) can be dropped altogether. But since this was the only coupling constraint, (2.23) can now decomposed to a per-subcarrier problem of the form $\min_{0\le\varepsilon_j\le\varepsilon_j^{\max}} -\left\lfloor \log_2\left(1+\varepsilon_j\,\rho_j/\Gamma\right)\right\rfloor$ to which a global solution which obeys all constraints is easily seen to be $\varepsilon_j^* = \varepsilon_j^{\max}$ and the corresponding bit allocation is $\left\lfloor \log_2\left(1+\varepsilon_j^{\max}\,\rho_j/\Gamma\right)\right\rfloor$.

The allocation for the TPP case, i.e., when $\Delta f \cdot \sum_j \varepsilon_j^{\max} > P_{\max}$, is more involved. Just as in the TPO case, the integer bit constraint makes the rate-function a step function i.e.,

$b_{GAP}^{\text{int}} \doteq \lfloor \log_2(1+\sigma/\Gamma) \rfloor$ which is clearly not differentiable, and non-linear optimization theory cannot be applied directly and hence the closed form expression for the optimal energy allocation like (2.21) is not possible. Theoretically speaking a greedy algorithm of the form of HH (see Table 2-1) could be designed to reach the optimum, but as discussed in Section 2.2.1.2, greedy algorithms require extensive searching and sorting which makes them impractical in DSL scenarios. Thus bit-rounding methods remain the only practical alternative just as in the TPO case. However the bit-rounding, now has to make sure that the peak power constraint does not get violated.

Note that due to the peak power constraint $\varepsilon_j^{\max}$, we can define the maximum integer number of bits that might be loaded on the j^{th} subcarrier as:

$$b_j^{\max} = \left\lfloor \log_2\left(1+\frac{\varepsilon_j^{\max}\rho_j}{\Gamma}\right)\right\rfloor. \tag{2.24}$$

This is because ε_j will always have to be less than (or equal to) $\varepsilon_j^{\max}$, irrespective of how strict or loose the total power constraint is. Hence the bit allocated (integer) will have to be less than (or equal to) $b_j^{\max}$.

In section 4.3.4 in [30], Chow presented an extension of his flat-power TPO integer bit algorithm, which we discussed in Section 2.2.1.2 by replacing the bit round off by a $b_j^{\max}$ constrained round-off. Thus the equations (2.12)-(2.14) are replaced as follows.

$$\hat{b}_j = \min\left\{round\left(b_j\right), b_j^{\max}\right\};\ \ 1 \le j \le K \tag{2.25}$$

$$\varepsilon_j = \rho_j^{-1} b_{GAP}^{-1}\left(\hat{b}_j\right) = \rho_j^{-1}\left(2^{\hat{b}_j}-1\right)\cdot\Gamma \tag{2.26}$$

Note that as a result of (2.25) we are guaranteed that $\hat{b}_j$ is an integer (due to *round* operation) and further due to the min(.) operation $\hat{b}_j \le b_j^{\max}$. When the energy allocation is done in (2.26), $\hat{b}_j \le b_j^{\max}$ due to (2.24) ensures that $\varepsilon_j \le \varepsilon_j^{\max}$, thus satisfying the peak power constraint.

In Table IV in [38], Baccarelli extended his secant based algorithm (which could handle peak power constraint but not integer bit constraint) to handle the integer bit constraint. The algorithm begins with solving for the zero of the function $f(B)$ in (2.22), i.e., B^* such that $f(B^*)=0$, by the secant method described earlier. From this B^*, ε_j^* is obtained

using (2.20) and $b_j = \log_2\left(1+\varepsilon_j^* \rho_j/\Gamma\right)$ which is a real number. After this, to ensure the integer bit constraint, the following (similar to) round-off procedure is conducted:

$$\hat{b}_j = \min\left\{\left\lfloor b_j + \varphi \right\rfloor, b_j^{\max}\right\} \tag{2.27}$$

where φ is a scalar parameter between 0 and 1 which is adjusted using a bisection method to bring the integer bit allocation throughput i.e., $\sum_j \hat{b}_j$ as close to the continuous-bit optimal throughput $\sum_j b_j$ as possible without violating the constraints. Note that Chow's bit rounding in (2.25) is a special case of (2.27) with φ=0.5. Since Baccarelli uses a bisection method to obtain the best φ instead of fixing it at φ=0.5, the rate achieved by Baccarelli's integer bit algorithm is generally higher than Chow's peak power constrained algorithm for TPP cases which we demonstrate in Chapter 3. The superior rate of Baccarelli is also because of the fact that Chow starts with an ad-hoc assumption of flat power allocation but Baccarelli starts with the optimum continuous–bit energy allocation from(2.20).

The common aspect of both of the approaches is that both are sub-optimal, since after all, bit rounding is a heuristic technique. Also comparing (2.25) and (2.27) to (2.12), we notice that the freedom to do bit rounding and energy rescaling is reduced in the presence of peak power constraint in TPP cases. In the PPO cases this freedom is minimized because $\varepsilon_j^* = \varepsilon_j^{\max}$ and hence the only integer bit allocation is the floor operation i.e., $\left\lfloor \log_2\left(1+\varepsilon_j^{\max}\, \rho_j/\Gamma\right)\right\rfloor$, thus there is no rounding up, there is only rounding down. As we shall see in Chapter 3, this lack of freedom to round up leads to high granularity losses.

2.2.1.4 Granularity Loss

Granularity loss is defined as the loss in rate incurred due to the practical integer-bit constraint in bit loading (when compared with continuous bit loading). For example in the TPO case granularity loss is the difference between the rate achieved by solving (2.4) and (2.11). Similarly in TPP and PPO cases the granularity loss is the difference in rate achieved by solving (2.15) and (2.23).

We discussed bit-rounding strategies for TPO cases in Section 2.2.1.2. It was initially observed, *only* for the TPO case, in [9][30][31] that most (not all) of the integer-bit

granularity losses could be recovered by bit-rounding (2.12) and scaling energies accordingly (2.13). The intuitive explanation (loosely speaking) for this is that when an arbitrary set of continuous b_j is rounded (2.12), some will be rounded up to the higher integer (e.g., when 3.8 is rounded we get 4) and some will be rounded down (e.g., when 5.2 is rounded we get 5) and the gain in rate due to rounding up will some how compensate for most of the loss in rounding down (when summed over all j) and hence the overall granularity loss will be low.

However the freedom for such bit-rounding and energy re-scaling is considerably reduced in the presence of peak-power constraint $\varepsilon_j^{\max}$, i.e., in TPP and PPO cases which we discussed in Section 2.2.1.3. In TPP cases, due to the min(.) operation in (2.25) or (2.27), on many subcarriers, b_j cannot be rounded-up due to the presence of $b_j^{\max}$ (resulting from $\varepsilon_j^{\max}$, see (2.24)). In such a situation only *rounding*-down or doing the *floor* operation is possible. Thus the number of subcarriers where rounding-down happens will be much higher than the rounding up cases. Further in PPO cases, there is absolutely no subcarrier where b_j can be rounded-up without violating the peak power constraint. Thus all subcarriers will have to be rounded-down. This leads us to suspect that granularity losses will be higher in TPP cases and much higher in PPO cases than TPO cases. Certain sporadic results in other works (listed below) when viewed together also strengthen this analysis.

The granularity loss in [38] is reported to be between 6-12% of rate conveyed for ADSL-TPP case. This is significant when compared to the variation of only 0.2-4% in the achievable rates of most existing integer bit algorithms for TPO case (see Fig. 4 in [11]). It is also higher than what would be expected from the 0.2dB *margin* difference due to granularity reported for the ADSL-TPO case in [30][31]. This leads us to believe that granularity losses would grow with increasing strictness in the peak-power constraint. Hence, we examine granularity losses in detail in the presence of peak power constraint in Chapter 3.

2.2.2 Margin Adaptive Loading Algorithms

All predictable impairment sources such as interference from other users or thermal AWGN noise, referred to as capacity limiting noise (discussed in Section 1.3) can be modeled within the framework of $\{\rho_j\}_{j=1}^{K}$. However modems invariably encounter performance limiting noise (discussed in Section 1.4) which consist of geographically variable and temporally and spectrally unpredictable non-stationary noise, e.g., impulse noise, RFI pickup, electric switches etc., the primary defense against which is *margin*. Further in many applications, e.g., tele-conferencing instead of maximizing rate, a fixed target rate is required to be satisfied with the best performance (BER) and best immunity against performance limiting noise.

Margin is physically defined as the amount by which the noise can be uniformly increased across all subcarriers, without violating the target rate requirement or the system BER requirement (which is embedded in the rate function $b(\sigma)$ definition).

Stated mathematically, for an energy allocation $\{\varepsilon_j\}_{j=1}^{K}$, a channel $\{\rho_j\}_{j=1}^{K}$ and a target (demanded) rate T, the *margin* γ_M is defined as the solution to the equation,

$$\sum_{j=1}^{K} b\left(\frac{\varepsilon_j \rho_j}{\gamma_M}\right) = T\,. \tag{2.28}$$

It is expected from the physical definition of *margin* and is apparent from (2.28) that γ_M is common across the subcarriers. The rationale behind this assumption, which is common to all works on *margin* [30]–[36], is that *margin* is the defense against unpredictable and non-stationary noise sources. Hence absolutely no apriori information is available as to which frequency or subcarrier is more likely to be affected by the noise burst. Thus the common approach is to provide the same *margin* to all subcarriers since they are equally likely to be affected. Since *margin* denotes that factor by which noise power can be increased without violating the BER requirement, it naturally appears as a denominator to the SNR $\sigma_j = \varepsilon_j \rho_j$ in (2.28). This is why it is sometimes referred to as SNR *margin* or performance *margin*. Thus maximizing *margin* is equivalent to maximizing the robustness against performance limiting noise. It can also be viewed as maximizing the performance (or minimizing BER) at a fixed target rate (see Lemma 4.1 in [30] for details),

because now at an SNR of $\sigma_j = \varepsilon_j \rho_j$, instead of loading $b(\sigma_j)$ bits we are only loading $b(\sigma_j/\gamma_M)$ bits, thus reducing the BER. Since γ_M is common across the subcarriers, the overall system BER will also be reduced in proportion.

In understanding the *margin* maximization problem, let us first assume that we have only the total power constraint, i.e., TPO case. Then the *margin* maximizing problem can be stated as,

$$\min_{\substack{T-\sum_{j=1}^{K} b(\varepsilon_j \rho_j/\gamma_M) \le 0 \\ \Delta f \cdot \sum_{j=1}^{K} \varepsilon_j - P_{\max} \le 0 \\ -\varepsilon_j \le 0}} (-\gamma_M). \tag{2.29}$$

2.2.2.1 Convex Reformulation of TPO Margin Maximization Problem

The objective in (2.29) is to maximize the *margin* γ_M. The first constraint, $T - \sum_{j=1}^{K} b(\varepsilon_j \rho_j/\gamma_M) \le 0$, denotes that in the process of providing this γ_M, we still satisfy the target rate requirement[17]. Notice that in (2.29), we have a K+1 dimensional problem since our design variables are $\{\varepsilon_j\}_{j=1}^{K}$ and γ_M. The objective function and the second and the third constraint are convex, but the first constraint $T - \sum_{j=1}^{K} b(\varepsilon_j \rho_j/\gamma_M) \le 0$ is concave in γ_M and convex in $\{\varepsilon_j\}_{j=1}^{K}$. Thus (2.29) is not a convex optimization problem, and hence KKT conditions or any of the convex optimization methods cannot be applied. This complication can be circumvented by reformulating the *margin* maximization problem as an energy minimization problem,

$$\min_{\substack{T-\sum_{j=1}^{K} b(\varepsilon_j \rho_j) \le 0 \\ -\varepsilon_j \le 0}} \left(\sum_{j=1}^{K} \varepsilon_j\right). \tag{2.30}$$

Notice that in (2.30), there are only K design variables, namely $\{\varepsilon_j\}_{j=1}^{K}$ and both the objective and the constraints are convex in these variables. Suppose that $\{\hat{\varepsilon}_j\}_{j=1}^{K}$ is the set

---

[17] Note that we could have loaded $b(\varepsilon_j \rho_j)$ bits on the j^{th} subcarrier without violating the BER criterion, but we provide a safety margin by loading less bits i.e., loading only $b(\varepsilon_j \rho_j/\gamma_M)$.

optimal energy allocation which solves the energy minimization problem in (2.30); and that $\{\varepsilon_j^*\}_{j=1}^K$ and γ_M^* solve the K+1 dimensional original *margin* maximization problem in (2.29). Then the aforementioned quantities are related as follows.

$$\gamma_M^* = \frac{P_{\max}}{\Delta f \cdot \sum_{j=1}^K \hat{\varepsilon}_j} \tag{2.31}$$

$$\varepsilon_j^* = \hat{\varepsilon}_j \cdot \gamma_M^* \tag{2.32}$$

The proof of and discussion on the above relationships can be found in Chapter 1 in [33] and [10][36][35] respectively. This equivalence of the *margin* maximization problem and the energy minimization problem arises from the fact that γ_M by its definition is a constant (across the subcarriers) scaling of energies $\{\varepsilon_j\}_{j=1}^K$ in (2.29). Since the final solution must satisfy the total power constraint $P_{\max}$, the highest scaling factor (see denominator in (2.31)) can achieved by minimizing the total power consumed as stated in (2.30).

The advantages of this equivalence are immense. As we noticed, the direct formulation in (2.29) is not convex, thus making both analysis and algorithm formulation difficult. But (2.30) on the other hand is convex and hence easier to solve. Further (2.30) has one less variable and one less constraint than (2.29). Lastly after solving (2.30), converting the solution to the solution of the original problem involves trivial complexity, as shown in (2.31) and (2.32). Due to this reason almost all *margin* maximization algorithms and analysis focus on solving only (2.30), e.g., [33][35][36] and then converting the solution using (2.31) and (2.32).

Let us begin by analyzing the solution for (2.30) when the rate function is our familiar continuous-bit SNR-gap rate function, $b_{GAP}(\sigma) = \log_2(1+\sigma/\Gamma)$.

2.2.2.2 Margin Adaptive Water-Filling

Since now we have a convex problem, we shall use our familiar KKT conditions. For problem (2.30) and the special usage of the continuous SNR-gap rate function, the Lagrangian is

$$L\left(\{\varepsilon_j\}_{j=1}^K;\lambda,\{\mu_j\}_{j=1}^K\right) = \sum_{j=1}^K \varepsilon_j + \lambda\left(T - \sum_{j=1}^K \log_2\left(1+\frac{\varepsilon_j \rho_j}{\Gamma}\right)\right) - \sum_{j=1}^K \mu_j \varepsilon_j . \tag{2.33}$$

Notice in comparison with (2.5), that loosely speaking, the objective function and the constraint have changed roles. λ in (2.5) represented the KKT multiplier for the total power constraint (and we were maximizing rate); here it represents the KKT multiplier of the target rate constraint (and we are minimizing the total power consumed). The KKT conditions corresponding to (2.33) must be satisfied at the optimal solution $\{\hat{\varepsilon}_j\}_{j=1}^K$ for some $\lambda \geq 0$ and $\mu_j \geq 0; 1 \leq j \leq K$, i.e.,

$$\frac{\partial L(.)}{\partial \varepsilon_j} = 1 - \frac{\lambda}{\ln 2} \frac{1}{\hat{\varepsilon}_j + \Gamma/\rho_j} - \mu_j, \tag{2.34}$$

$$\mu_j \varepsilon_j^* = 0; \quad 1 \leq j \leq K, \tag{2.35}$$

$$\lambda \left(T - \sum_{j=1}^{K} \log_2 \left(1 + \frac{\varepsilon_j \rho_j}{\Gamma} \right) \right). \tag{2.36}$$

Simplifying (2.34)-(2.36) in a similar way as we simplified (2.6)-(2.8), we obtain the following *margin* adaptive water-filling conditions.

$$\hat{\varepsilon}_j = \left[\frac{\lambda}{\ln 2} - \frac{\Gamma}{\rho_j} \right]^+ \tag{2.37}$$

where λ is obtained by solving

$$\sum_{j=1}^{K} \left[\log_2 \left(\frac{\lambda \rho_j}{\Gamma \ln 2} \right) \right]^+ = T. \tag{2.38}$$

But this solves only (2.30). To obtain the solution to (2.29), i.e., to obtain $\{\varepsilon_j^*\}_{j=1}^K$ and γ_M^*, the operations in (2.31) and (2.32) must be performed on $\{\hat{\varepsilon}_j\}_{j=1}^K$ which in turn is obtained by solving (2.37) and (2.38).

A graphical illustration of the above is presented in Figure 2-4. Figure 2-4(a) represents the relationship between $\{\hat{\varepsilon}_j\}_{j=1}^K$ and $\{\rho_j\}_{j=1}^K$ and Figure 2-4(b) represents the relationship between $\{\hat{\varepsilon}_j\}_{j=1}^K$ and $\{\varepsilon_j^*\}_{j=1}^K$. The $\{\rho_j\}_{j=1}^K$ used here is the same as that considered in Figure 2-2 and Figure 2-3. Note that the water filling level here is λ/ln2 and though the λ here denotes an altogether different quantity, we can say in general that the level in Figure 2-4(a) should be lower than that in Figure 2-2(a) since $\{\hat{\varepsilon}_j\}_{j=1}^K$ here is the

solution to the energy minimization problem. From Figure 2-4(b) it becomes clear that while $\{\hat{\varepsilon}_j\}_{j=1}^{K}$ satisfies the water-filling condition, $\{\varepsilon_j^*\}_{j=1}^{K}$ does not. $\{\hat{\varepsilon}_j\}_{j=1}^{K}$ by virtue of satisfying the water-filling condition in (2.37), naturally loads high power on the better channels e.g., 'Region A' in the figure. As long as only one user used a medium this would not be a problem, but in reality this is never the case. Recall that in Section 2.2.1.3, we discussed how unrestricted water-filling could cause excessive interference to other users sharing the medium, in regions such as 'Region A'. This effect is further aggravated when ε_j^* is used. As indicated by (2.32), in Figure 2-4(b) we see that ε_j^* is a constant scaling of $\hat{\varepsilon}_j$ by a constant scaling factor which is the optimum *margin* γ_M^*. Thus ε_j^* further magnifies the opportunistic loading of $\{\hat{\varepsilon}_j\}_{j=1}^{K}$. Note that (due to the γ_M^* scaling), 'Region B' which was not loaded by $\hat{\varepsilon}_j$, remains unloaded by ε_j^*; but 'Region A' which was allocated the highest energy is further boosted. We shall revisit this problem of protection of other users in Section 2.2.2.4 and 2.3. The immediate concern is to accommodate the integer-bit constraint considering only one user.

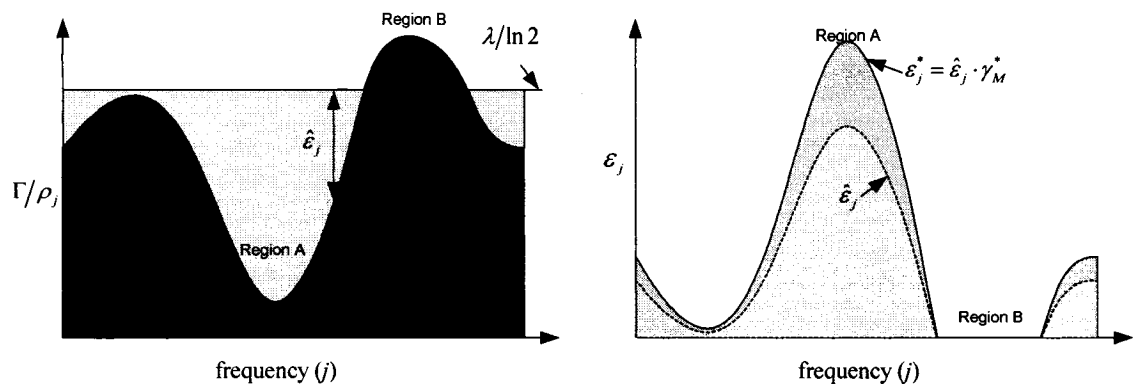


Figure 2-4: Illustration of Margin Adaptive Water-Filling

Thus in a single user optimal solution to (2.29) with continuous-bit SNR-gap rate function, we allocate ε_j^* energy to the j^{th} subcarrier and load

$\log_2\left(1+\frac{\varepsilon_j^* \rho_j}{\gamma_M^* \Gamma}\right) = \log_2\left(1+\hat{\varepsilon}_j \rho_j/\Gamma\right)$ bits[18]. But this term is a real number while most implementable modulation schemes support only integer bits/symbol. Thus we must introduce the integer bit constraint to develop implementable algorithms.

2.2.2.3 Integer-Bit Constraint

The integer bit version of the energy minimization reformulation in (2.30) for the SNR-gap rate function is:

$$\min_{\substack{T-\sum_{j=1}^{K}\lfloor \log_2(1+\varepsilon_j \rho_j/\Gamma)\rfloor \le 0 \\ -\varepsilon_j \le 0}} \left(\sum_{j=1}^{K} \varepsilon_j\right) \tag{2.39}$$

Notice that the floor operation $\lfloor . \rfloor$ in the first constraint ensures that the target (demanded) rate T is met by loading integer bits on each subcarrier. Just like the integer bit rate maximization problem in (2.11), (2.39) is an integer programming problem[19] whose optimal solution is a greedy algorithm. In fact the HH algorithm in Table 2-1 can be directly used to reach the optimal solution for (2.39). To see why, recall that in the HH algorithm in Table 2-1, at each iteration, we added one bit on the subcarrier $\hat{j}$, where it was the cheapest in terms of additional or incremental energy required $\Delta\varepsilon_j$ (see step 6). We continued adding one bit, at each iteration in this fashion, till the target rate T was met. Since while adding each bit we chose the cheapest option in terms of energy, the overall savings in energy is highest, i.e., the total energy $\sum_j \varepsilon_j$ is minimized. There is however one key difference while using the HH algorithm in the Rate-Adaptive (RA) mode (discussed in Section 2.2.1.2) and when using it to solve (2.39). In the RA mode, the target rate T in the stopping condition (see Step 8 in Table 2-1) had no physical meaning since we were maximizing the rate, and hence we had set the T to a very high value so that the HH algorithm keeps loading till it reaches $P_{\max}$. But here we wish to minimize

---

[18] We could have loaded $\log_2\left(1+\varepsilon_j^* \rho_j/\Gamma\right) = \log_2\left(1+\hat{\varepsilon}_j \rho_j \gamma_M^*/\Gamma\right)$ bits. This would have given us a higher rate but no margin. Instead we used a γ_M^* factor of energy on each subcarrier as margin while achieving the target rate T.

[19] To be more precise it is actually a Separable Convex Discrete Resource Allocation (SCDRA) just like the problem in (2.11). For discussions on SCDRA see chapter 1 and 2 in [33].

the energy utilized while satisfying the actual demanded target rate T. Thus for RA mode, the active stopping condition (see Step 8 in Table 2-1) is the total power constraint, where as here the active stopping condition is the target rate constraint. Thus at the end of the HH algorithm we will have the $\{\hat{\varepsilon}_j\}_{j=1}^{K}$, which is the optimal solution to the energy minimization problem in (2.39). But to obtain the solution to our original *margin* maximization problem in (2.29), we must evaluate γ_M^* as in (2.31) and then boost each $\hat{\varepsilon}_j$ by this factor to obtain ε_j^* as in (2.32).

While HH algorithm provides the optimal solution to (2.39), due to adding just one bit at each iteration and extensive searching and sorting, it is slow in converging to the optimal solution. This motivated the works (stated in chronological order) in [31] by Chow et al., in [36] by Krongold et al and in [35] by Campello. In [31], Chow proposed a bit-rounding based sub-optimal (close to optimal HH rate), efficient approach for *margin* maximization. As with all bit-rounding approaches, this started with an infinite granularity assumption to estimate an approximate *margin* and then did bit-rounding and energy rescaling to ensure an integer bit distribution. On the other hand [35][36] guarantee optimality like HH.

The approach in [35], a greedy like method is used but it loads bits on multiple subcarriers in a single iteration, rather than only one, thus making itself faster than HH. The approach in [36] is based on the realization that unlike the *margin* adaptive water-filling solution (for continuous bit loading) discussed in Section 2.2.2.2 where we had a unique λ across all subcarriers, when we have integer-bit constraint, there will be a range of λ for each subcarrier. An algorithmic implementation is presented based on storing the bounds of this range in a look-up table and doing a bisection search which converges much faster than HH. Though both [35] and [36] involve completely different approaches they both reach the same optimal solution $\{\hat{\varepsilon}_j\}_{j=1}^{K}$ for (2.39) as the HH algorithm, but do so much faster. However both approaches just like HH need to boost $\hat{\varepsilon}_j$ by a constant factor γ_M^* obtained from (2.31) to obtain ε_j^* as in (2.32) to obtain the solution to the original *margin* maximization problem. As explained in Section 2.2.2.2, this boosting of energy could be tolerable if there had been only one user, but in most scenarios including DSL, such boost

would cause severe narrowband interferences to other users in subcarriers such as 'Region A' in Figure 2-3 and Figure 2-4.

2.2.2.4 Protection of other users sharing the medium

While single user algorithms described above can provide varying degrees of optimality in terms of *margin* to one user, in most realistic scenarios, the medium is shared by multiple users. As an example recall the DSL scenario discussed in Figure 1-1. Recall that we had encountered this same problem, of one users 'optimal' allocation being too harmful for the other users and vice-versa, while doing rate maximization in Section 2.2.1.3 because of the opportunistic nature of water-filling. As explained in the above paragraph, *margin* adaptive algorithms further aggravate this problem due to a unilateral energy boost (2.32) by a user in the process of *margin* maximization. Now, in the rate adaptive case, we were able to contain the damage caused to other users by employing a peak power constraint $\left\{\varepsilon_j^{\max}\right\}_{j=1}^{K}$ as discussed in Section 2.2.1.3. Theoretically speaking, we could incorporate the peak power constraint into the energy minimization problem (2.30) and analyze the resulting convex problem using KKT conditions similar to the way we did with the peak power constrained rate maximization problem in (2.15).

But unfortunately after introduction of this new constraint the energy minimization problem is no longer equivalent to the original *margin* maximization problem (2.29) with peak power constraint. Recall from section 2.2.2.1 that this equivalence is what led to the development of most analysis and algorithms for *margin* maximization, because (2.30) is convex while (2.29) is not. But if this equivalence does not hold any more, there is very little point in solving the energy minimization problem, because we cannot guarantee any *margin* for protection against performance limiting noise or performance and robustness guarantee for applications that require a fixed (demanded) rate which was the objective we started out with. The reason why this equivalence does not hold is not difficult to see. In the TPO case (when peak power constraint does not exist) the proportion of energy (γ_M^* calculated in (2.31)) saved in the process of solving the energy minimization problem could be directly converted into the maximized *margin* of the system by the boosting the energy allocated to each subcarrier by the factor γ_M^* as in (2.32). But in the presence of peak power constraint this boost will violate the peak power constraint on many if not all

subcarriers. So let alone the question of optimal, we will not even get a feasible solution[20]. A similar problem occurs in multiuser loading algorithms as well (discussed in Section 2.3.4).

2.3 Multiuser Loading Algorithms

We noted in previous sections that most algorithms discussed therein are essentially single user and a single user's channel can be completely characterized by $\{\rho_j\}_{j=1}^{K}$. Assuming that $\{\rho_j\}_{j=1}^{K}$ is available at the transmitter and is fixed, we had developed the entire discussion in Section 2.2 and 2.3 thus making the entire optimization single-user. But in reality most mediums are multiuser and the opportunistic loading resulting from the single-user optimization is not healthy for the system as a whole and as illustrated in Section 2.2.1.3 and 2.2.2.4.

With this realization, in motion towards algorithms which keep multiuser concerns in mind, let us begin by understanding the structure of $\{\rho_j\}_{j=1}^{K}$. For the single-user case we defined ρ_j in Section 2.2 as the normalized received SNR on the j^{th} subcarrier, i.e., when $\varepsilon_j = 1$. In adding the user specification, let us call the received SNR on the j^{th} of the i^{th} user as ρ_j^i and similarly the energy allocation is ε_j^i where $1 \le j \le K$ and $1 \le i \le N$. In the context of the multiuser channel model developed in Section 2.1, the interference power that User i will receive from User l is $\left|H_j^{li}\right|^2 \varepsilon_j^l$. Now, say that every user among the N users sharing the medium except the i^{th} fix their power allocation, i.e., ε_j^l is fixed at some value for all $1 \le j \le K$ and $1 \le l \le N; l \ne i$. Then in retaining our normalized SNR definition,

$$\rho_j^i = \frac{\left|H_j^{ii}\right|^2}{n_j^i + \sum_{l=1, l\ne i}^{N} \varepsilon_j^l \left|H_j^{li}\right|^2}; \quad 1 \le j \le K. \tag{2.40}$$

Note that the numerator denotes the direct channel attenuation of i^{th} user. The denominator denotes the sum of background noise and interference from other users. Thus, clearly

[20] A solution which obeys the constraints is called a feasible solution.

if our chosen User i allocates ε_j^i to its j^{th} subcarrier, and other users do not change their power allocation ε_j^i, then the received SNR for User i on subcarrier j is $\sigma_j^i = \varepsilon_j^i \rho_j^i$ which was intended in the definition of ρ_j^i. But the assumption that other users do not change their power allocation is not a realistic assumption, so though at any instant, (2.40) holds, ρ_j^i is not a fixed quantity and changes with other users' power allocation. Nevertheless ρ_j^i or some estimate of it is required at the transmitter of each User i for any power or bit allocation. This leaves us with 3 options:

- Assume the worst case interference that might result from other users in defining ρ_j^i for any User i and consider this fixed. This is sometimes referred to in DSL specific literature as Static Spectrum Management (SSM) and is discussed in Section 2.3.1.
- Each User i measures through some SNR estimation routine ρ_j^i (repeatedly) each time it expects or is notified that other users have changes their power allocation. This approach leads to what is sometimes referred to in DSL specific literature as distributed Dynamic Spectrum Management (DSM). An example of this is presented in Section 2.3.2.
- One central authority obtains the estimate of all $\left|H_j^{il}\right|$ and all n_j^i. This central authority also controls every single user's and every single subcarrier's power allocation ε_j^i and hence has complete knowledge ρ_j^i at all times. This approach leads to centralized DSM. An example of this is presented in Section 2.3.3.

2.3.1 Worst Case Interference Modeling - SSM

Recollect from Section 2.2.1.3 that presence of peak power constraint, ε_j^{max}; $1 \le j \le K$, controls the extent to which opportunistic loading of one user causes excessive interference to other users. Note that ε_j^{max} changes only with the subcarrier number j. It is the same for any user, since generally ε_j^{max} is specified by a standards body such as ANSI in [2][3] and is common for all users sharing the medium and further ε_j^{max} is always fixed (with respect to time). Thus irrespective of individual user-specific alloca-

tion strategies, $\varepsilon_j^i \le \varepsilon_j^{\max}$; $1 \le i \le N$; $1 \le j \le K$ will always hold. Thus comparing with (2.40), we can say that,

$$\rho_j^i \ge \frac{\left|H_j^{ii}\right|^2}{n_j^i + \sum_{l=1, l\neq i}^{N} \varepsilon_j^{\max} \left|H_j^{li}\right|^2}; \quad 1 \le j \le K \tag{2.41}$$

with equality only in the unlikely scenario when $\varepsilon_j^l = \varepsilon_j^{\max}$; $\forall j$; $\forall l \neq i$. While the right hand side in (2.41) is only a lower bound for the true ρ_j^i, it is very convenient to calculate at any user's transmitter or receiver because $\varepsilon_j^{\max}$ is fixed and already known to each user and other quantities $\left|H_j^{li}\right|^2$ and n_j^i characteristic of the DSL channel are quasi-static. Further the right hand side in (2.41) does not change when other users change their allocation, since each user always obeys $\varepsilon_j^i \le \varepsilon_j^{\max}$. Due to these advantages even though it well known that (2.41) is an inequality, equality is assumed in SSM methods to avoid complication of repeated estimation and coordination required for (2.40). With the assumption that ρ_j^i is equal to the right hand side in (2.41), we essentially have a single-user peak power constrained problem like (2.15) once again, because now our allocation would not change with other users changing their allocation. Thus all the analysis and algorithms discussed in Section 2.2.1.3 directly applies.

But in the process of simplifying the implementation, we assumed more interference than there actually is. This leads each user to load conservatively while they could have loaded more (bits) and hence achieved more rate or *margin* if they had the true ρ_j^i based on (2.40). This realization has motivated significant research in truly multiuser algorithms which base themselves on (2.40) instead of the simplification stated above, and the huge improvements obtained justify [24] the additional complication due to repeated estimation of (2.40). They are collectively referred as DSM techniques (which we discuss next) and have secured recognition in industry and standardization bodies. But at the time of writing of this thesis, still the only standardized and widely deployed method is SSM through the application of peak power constraint [2][3][23][24].

2.3.2 True Interference Modeling - DSM

DSM techniques [23][24] adaptively vary power allocation towards maximizing each user's performance (rate/*margin*) in recognition of the presence and needs of other users. The recognition of presence of other users is facilitated through measurement of the true interference caused, i.e., by estimating ρ_j^i in (2.40) instead of assuming the worst case crosstalk (SSM). As explained before the implementation of any such technique would require some form of coordination among users, assistance from a centralized agent or complete centralized control over transmit spectra. Based on amount of coordination and centralized control necessary, DSM techniques are classified from Level 0 to Level 3 [24][29].

Level 0 coordination essentially implies the SSM case discussed in the previous section. In Level 1 coordination of DSM, only macro parameters such as data rates, total transmit power and *margin* are reported and controlled centrally and other micro parameters such as actual subcarrier specific power and rate allocation are done autonomously (i.e., in a *distributed* manner); an example is [25]. Level 2 denotes the scenario when even micro parameters such as each user's subcarrier specific power allocation are reported and partially or completely controlled by a central agent[21]; an example is [27]. Level 3 is highest level of coordination, where not only power and bit allocation is jointly controlled and coordinates, but also the actual signal sent is jointly transmitted and/or received; an example is [28]. We believe this level should rather be classified as Interference Cancellation techniques rather than DSM which primarily deals with allocation issues. Level 1 schemes are the most popular among DSM techniques because they have the minimum *time to market*. This is because in current DSL deployment scenarios, multiple service providers share the same bundle, a phenomenon referred to as unbundling, see e.g., Figure 1-1. In such a scenario expecting more than Level 1 coordination is less realistic. Among Level 1 DSM techniques, Iterated Water Filling (IWF) [25] is the most popular [23][29], due to its predominantly *distributed* nature and significant rate enhancement from SSM techniques.

[21] The central agent could be the central office, SMC or DSLAM depending on the nature of DSL deployment [23][24].

2.3.2.1 Iterated Water-Filling

In a search for a distributed power allocation scheme for the multiuser DSL scenario, modeling it as a non-cooperative game was natural. This led to the Iterated Water-Filling algorithm [25]. In a pure strategy non-cooperative game, each player's (user's) *strategy* is to respond optimally to the *information* available to it at any stage. When, say User i plays, the *information* that is estimated is ρ_j^i which is dependent on other user's power allocation as stated in (2.40) and the *strategy* is the power allocation ε_j^i . Now if during the play of User i, other users do not change their allocation, then the optimal response to this ρ_j^i from the perspective of this single User i, is obviously the simplest water-filling discussed in Section 2.2.1.1. It was proven in [25] for a 2-user case and later for a N-user case in [26] that for DSL channels a Nash equilibrium exists and further that if the each User i plays in the aforementioned way, the algorithm converges to a Nash equilibrium. In game theory, the Nash equilibrium is defined as a point where no player has an incentive to deviate from the current *strategy* (in our case ε_j^i). This happens of course when each player decides after observing its *information* that its strategy is already the best it can do. In our case this corresponds to the case when all users simultaneously satisfy the water-filling condition, i.e., for N users which use total power (summed across subcarriers) P^i at the Nash equilibrium the following conditions (water-filling conditions - compare with (2.9)(2.10)) are satisfied for each User i:

$$\varepsilon_j^i = \left[\frac{1}{\lambda^i \ln 2} - \frac{\Gamma}{\rho_j^i}\right]^+ \tag{2.42}$$

where the user specific KKT multiplier λ^i is obtained by solving:

$$\Delta f \cdot \sum_{j=1}^{K} \left[\frac{1}{\lambda^i \ln 2} - \frac{\Gamma}{\rho_j^i}\right]^+ = P^i . \tag{2.43}$$

This solution can be reached algorithmically as summarized in Table 2-2.

The sufficient conditions for uniqueness of this Nash equilibrium is found in [25][26] and it is proven that the above algorithm converges to this Nash equilibrium. Note that in any iteration of the game (i.e., steps 2-7 in Table 2-2) each User i first measures ρ_j^i at that stage and then performs a water-filling. This power allocation results in a

SNR profile σ_k^i in step 5 and hence a rate R^i in step 6. This process is repeated till no user's power allocation changes significantly. Note that all of the steps 2-7 could be performed by User i alone as if it were a single user optimization, thus making the IWF purely distributed. When other constraints such as an integer-bit constraint is present then step 4 instead of being the orthodox water-filling would correspond to the energy allocation reach by the algorithms in Section 2.2.1.2. The key idea is that due to the game-theoretic modeling, at each iteration, we have to solve only a single user problem and hence all discussions of Section 2.2 apply and this is what makes the IWF algorithm distributed.

Table 2-2: SubAlgorithm Iterated Water-Filling (IWF)

1.	*repeat*
2.	*for* $i = 1$ to N
3.	Estimate $\rho_j^i = \left\|H_j^{ii}\right\|^2 \Big/ \left(n_j^i + \sum_{l=1,l\neq i}^{N} \varepsilon_j^l \left\|H_j^{li}\right\|^2\right); 1 \le j \le K$
4.	Set $\left\{\varepsilon_j^i\right\}_{j=1}^{K}$ to the water-filling corresponding to noise spectrum $\left\{\Gamma/\rho_j^i\right\}_{j=1}^{K}$ and total power P^i.
5.	Store $\sigma_j^i = \varepsilon_j^i \rho_j^i; 1 \le j \le K$
6.	$R^i = \sum_{j=1}^{K} \log_2\left(1 + \sigma_j^i/\Gamma\right)$
7.	end
8.	*until* the desired accuracy is reached

Two important questions arise here. Even if we reach a Nash equilibrium, at this equilibrium due to step 6, each user will have an achieved rate R^i, referred collectively as the rate tuple **R**. From the uniqueness of the Nash equilibrium for a given total power vector **P**, this **R** is unique. But in reality at different times, different users might demand different rates and hence having just one rate combination is not sufficient. So how do we generate multiple rate tuples **R**? Also how good is this **R**? Just knowing that it is a Nash equilibrium does not tell us anything about whether it is close or far from the optimal or the SSM performance.

Since for a given power combination **P**, the rate combination **R** is unique. So the only way to generate different **R** is to vary the **P**. This must be done by a loop which runs outside the SubAlgorithm IWF and tries to reach the **P** that achieves a certain required **R**

(see Algorithm 1 in [25]). Of course the set of achievable **R** or the rate-region is bounded since the each user's total power P^i is restricted by a total power constraint P^i_{max} just as in the single user case[22].

In terms of comparative performance the rate region generated by IWF is far larger than generated by any SSM technique [25][29][23][24]. This coupled with its predominantly distributed nature makes IWF very popular and a good candidate for future standardization in DSL. But a question of theoretical interest and for future research is that though IWF is spectacularly better than any SSM technique, how much better than it (if at all) is possible.

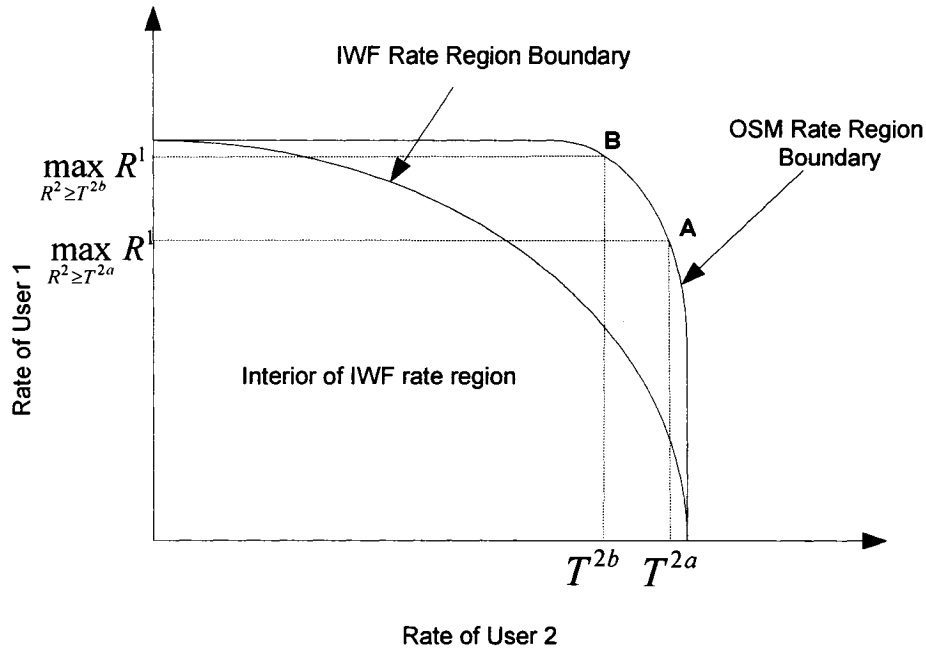


Figure 2-5: Rate Region Boundary

2.3.2.2 Optimum Spectrum Management (OSM)

One way of comparing different multiuser loading algorithms is the rate region boundary. Rate region boundary is the boundary of the achievable rates by an algorithm and hence is a representation of the best capability of an algorithm. This is illustrated in Figure 2-5 for a two user case.

[22] Note that P^i_{max} is specified by standardization bodies and must be respected by all modems under all circumstances. But P^i on the other hand is in a user's control. In a DSM scenario, one user might choose to use a total power P^i less than its constraint out of compassion for other weaker (channel) or more needy (rate/margin) users.

For two methods IWF [25] and OSM [27], we give an illustration of the typical rate region boundary[23]. Clearly from the figure since OSM can achieve an outer rate-region boundary, it can achieve higher rates than IWF. Stated mathematically, the rate region boundary is a solution to the following problem:

$$\max_{R^i \geq T^i; i=2,3,..,N} R^1 \text{ subject to constraints on power.} \tag{2.44}$$

In other words we maximize the rate of one of the users, while constraining the other users' rates to some target rates $T^2, T^3, .., T^N$. By solving (2.44) for a set of $T^2, T^3, .., T^N$, we would get one point on the rate region boundary; but by varying these and solving (2.44) repeatedly the entire rate region boundary can be generated. This is illustrated the 2-user case in Figure 2-5. Notice that by solving (2.44) for the constraint $R^2 \geq T^{2a}$ we obtain the point A on the rate region boundary and by solving (2.44) for the constraint $R^2 \geq T^{2b}$ we obtain point B on the rate region boundary. IWF, which we discussed in the previous section, can be used to generate both its rate region boundary and also the interior of its rate region. But for OSM on the other hand, the only objective is to generate the rate region boundary for comparison purposes. Thus OSM's objective is to find the globally optimal solution, i.e., the optimal $\varepsilon_j^i;\ 1 \leq i \leq N;\ 1 \leq j \leq K$ to the following problem which is a more detailed statement of (2.44).

$$\begin{aligned} &\min\ -\sum_{j=1}^{K} b\left(\varepsilon_j^1 \rho_j^1\right) \\ &\text{subject to}\quad \sum_{j=1}^{K} b\left(\varepsilon_j^i \rho_j^i\right) \geq T^i;\ i = 2,3,...,N \\ &\qquad\text{and}\ \ \Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^l \leq P_{\max}^i;\ l = 1,2,...,N. \end{aligned} \tag{2.45}$$

In the above, ρ_j^i is calculated as in (2.40). It is easy to see that due to ρ_j^1, even though the rate function $b(.)$ is concave, the objective function will be non-convex in all $\varepsilon_j^i;\ i \neq 1$ and a similar argument holds for the constraints $\sum_{j=1}^{K} b\left(\varepsilon_j^i \rho_j^i\right) \geq T^i;\ i = 2,3,...,N$. Thus we have a highly non-convex problem and indeed this problem has many local minima [27]. Hence traditional gradient based methods of optimization which guarantee only local minimum cannot be used to find the global op-

[23] This figure is just an example. Actual simulation results for practical DSL scenarios are provided in Chapter 3.

timum of (2.45). Exhaustive search is simply intractable because we have NK variables ε_j^i, each having many possibilities and N and K are very large in DSL applications (e.g., N is between 20 and 100 and K is close to 1200 for VDSL upstream. In VDSL total number of subcarriers is 4096, about 1200 are for upstream [3]). To make the problem tractable dual decomposition is used in OSM. Using dual decomposition the problem is decoupled to a per-subcarrier problem. But within each such per-subcarrier problem an exhaustive search has to be conducted for each user's power allocation which still keeps OSM far too complex (though tractable) for practical consideration. For more details on the technique see [27] and for complexity comparisons see [42]. However following this dual decomposition and exhaustive search, OSM guarantees the globally optimal solution for (2.45) both in the case when the rate-function is continuous bit and when it is integer-bit [27].

Both the requirement of complete centralized control and coordination and the huge complexity resulting from the exhaustive search make OSM impractical under current DSL scenarios. But the rate region boundary generated by OSM is of important value in 'bench-marking' other rate-maximizing multiuser loading algorithms and hence this answers the question of how much better than IWF we can do.

2.3.3 Multiuser Margin Adaptation

DSM techniques are able to provide much better rates than SSM techniques because they model the true crosstalk and are able to 'jointly' (relative to SSM) optimize the transmit spectra towards dynamically controlling the crosstalk among users which is the dominant capacity-limiting impairment in DSL. Thus DSM is a promising solution to capacity limiting noise discussed in Section 1.1 and most existing DSM techniques are rate adaptive. But as discussed in Section 1.2 and Section 2.2, the causes of performance limiting noise are not crosstalk and hence still persist. *Margin* remains the primary defense against this form of noise. Further *margin* maximization is desirable for constant bit rate services. Though this was realized early and many single user *margin* adaptive algorithms (discussed in Section 2.2) were proposed, no multiuser *margin* adaptive algorithm has been developed in literature to date. This is because of the following reason.

In the SSM scenario (i.e., in the presence of peak power constraint), we had noted in Section 2.2.2.4 that the energy minimization problem (convex) is no longer equivalent to the *margin* maximization problem and further the unilateral energy boost of (2.32) would violate the peak power constraint, thereby rendering most of the single user MA algorithms unusable. A similar effect occurs in the DSM scenario. Though one might argue that one could solve the multiuser energy minimization problem (in the context of (2.30)), the energy saved as a result of solving this problem cannot be converted to gain in *margin* as it could be in the single user case, see (2.31) and (2.32). This is because the final solution to the multiuser energy minimization problem will be some form of equilibrium (e.g., by a game theoretic approach) or some stationary point (e.g., by a traditional optimization approach). At this point the various users' energy allocations $\hat{\varepsilon}_j^i$ will be in a delicate balance with each other. But in order to convert the saving of energy into a gain in *margin*, if one user unilaterally boosts its energy as was done in (2.32) for the single user case; it will simultaneously boost the crosstalk to all other users (i.e., will break the equilibrium or the stationary that exists between various users). This in turn will result in other users failing to meet their target rate or *margin* or both[24].

Thus we find ourselves at the following paradox. As discussed in Chapter 1, we have two forms of impairment performance limiting noise and capacity limiting noise. To counter performance limiting noise, we need *margin* adaptation. But all available *margin* adaptation methods due to their energy boost feature will sharply increase capacity limiting noise to other users which in turn results in certain users being rendered incapable of meeting their target rate requirement or *margin* or both. On the other extreme, SSM methods to some extent and DSM methods to a large extent can effectively deal with capacity limiting noise by keeping crosstalk in control, but they cannot provide (till now) *margin* adaptation required for constant bit rate services and robustness against performance limiting noise. So the question is: Is it possible to effectively deal with both capacity limiting noise and performance limiting noise at that same time? The answer is affirmative and we shall see in Chapter 4 that with if we stop relying on the convex reformulation of Section 2.2.2.1 and conduct an independent study of *margin* in a multiuser sce-

[24] Due to this reason there is some skepticism about modems working in Margin Adaptive (MA) mode in DSM scenario [29].

nario, we find certain fairness properties which can be exploited to develop strategies which when conjugated with existing multiuser rate maximizing loading algorithms to effectively deal with both forms of noise effectively.

2.3.4 Granularity Loss

In the SSM scenario there is always an explicit peak power constraint $\varepsilon_j^{\max}$. As discussed in Section 2.2.1.4, this peak power constraint restricts the freedom in bit-rounding and energy rescaling and we expect that due to this reason there will be significant granularity loss in the TPP and PPO cases which we investigate in Chapter 3. In DSM scenarios, as discussed above, the peak power constraint is either not present or much less strict when present than SSM schemes. However in (2.44) or (2.45), when $T^i;\ i \neq 1$ increases, the maximum power that User 1 can put on its 'good' subcarriers (i.e., where direct channel attenuation is low) gets increasingly restricted. This is the case because the 'good' subcarriers of various users are very similar frequencies as discussed in Section 2.1, 2.2.1.3 and illustrated in Figure 2-1. Hence the rate demand of one user T^i *implicitly* acts as a strict peak power constraint for the other users, which in turn opens the room for granularity losses. This claim is corroborated by some sporadic results in other works. For example, in [26] it was observed that, for a 24-AWG scenario consisting of 4 loops of 600m and 4 loops of 1200m, when the 1200m loops are constrained to achieve a minimum of 5Mbps, the 600m loops using Iterated Water Filling (IWF) [25] can achieve 3.4Mbps and 7.7Mbps with integer-bit loading and ideal continuous bit-loading, respectively. When Optimum Spectrum Management (OSM) is used in the same scenario, the 600m loops achieve 13Mbps and 15Mbps with integer bit loading and ideal continuous bit loading, respectively. Thus granularity losses are expected to be huge.

Therefore, in Chapter 3, after investigating the granularity losses for TPP and PPO cases arising from the currently standardized SSM scenarios, we shall also investigate granularity losses in DSM scenarios.

Chapter 3

Fine Granularity Loading

3.1 Power, Integer-Bit Constraints and Granularity Loss

The primary indicator of performance and loss will be in terms of rate achieved by a modem. Thus the objective function of the overall rate maximization problem is the total supported rate,

$$R = \sum_{j=1}^{K} b\left(\rho_j \varepsilon_j\right). \tag{3.1}$$

The traditional total power constraint applied across the subcarriers is,

$$\Delta f \cdot \sum_{j=1}^{K} \varepsilon_j \le P_{\max}. \tag{3.2}$$

In addition, many practical systems have limitation on the maximum transmit PSD. This implies the peak-power constraints:

$$\varepsilon_j \le \varepsilon_j^{\max};\ \ 1 \le j \le K \tag{3.3}$$

where $\left\{\varepsilon_j^{\max}\right\}_{j=1}^{K}$ is specified by the admissible transmit PSD mask, e.g., SMClass3 in [2] or M1FTTCab in [3]. The power constraint qualifications TPO, PPO and TPP which were discussed in Chapter 2 are summarized in Table 3-1 because they are fundamental to the discussion on granularity loss in this chapter.

Table 3-1: Power Constraint Qualifications

Condition	Constraint Qualification	Comments
$\varepsilon_j^{\max} = \infty, 1 \le j \le K$	Total Power Only (**TPO**)	Granularity loss observed to be negligible. Refer:- Section 2.2.1.1, 2.2.1.2
$\Delta f \cdot \sum_{i=1}^{K} \varepsilon_i^{\max} \le P_{\max}$	Peak Power Only (**PPO**)	Granularity loss expected to be high. Refer:- Section 2.2.1.3, 2.2.1.4
$\Delta f \cdot \sum_{i=1}^{K} \varepsilon_i^{\max} > P_{\max}$	Total and Peak Power (**TPP**)	Granularity loss expected to be high. Refer:- Section 2.2.1.3, 2.2.1.4

Recall that for the PPO case, the optimal power allocation does not depend on the rate function $b(\sigma)$ and is equal to the peak power constraint. For TPO and TPP cases however, the energy allocation depends on the rate function. Due to the fact that most

practical modulation schemes are integer-bit, this constraint is also to be taken into account. We saw in Chapter 2 that most practical methods of handling this constraint rely on bit-rounding and energy re-adjustment. For the TPO case, it was observed that this bit-rounding and energy-readjustment could recover most (not all) of the granularity losses. But we noted that the freedom to do this energy re-adjustment (corresponding to the bit-rounding) is drastically reduced in the presence of peak power constraint which will lead to high granularity losses in TPP and PPO cases. We quantify this qualitative claim in the next section.

3.2 Quantification of Granularity Loss

Let Ω represent the set of non-trivially loaded subcarriers and Ω_1 be the set of subcarriers in which *ceiling* (or *rounding-up*) the non-integer bit $b(\rho_j\varepsilon_j)$ would cause the corresponding energy allocation to violate the peak power constraint, i.e., $\varepsilon_j > \varepsilon_j^{\max}$. Mathematically, we can say $\Omega \triangleq \{j \in \{1,2,..,K\}: \varepsilon_j > 0\} = \Omega_1 \cup \Omega_2$ where $\Omega_1 \cap \Omega_2 = \varnothing$, $\Omega_1 \triangleq \left\{j \in \Omega: b^{-1}\left(\left\lceil b(\rho_j\varepsilon_j)\right\rceil\right) > \rho_j\varepsilon_j^{\max}\right\}$, $\Omega_2 \triangleq \left\{j \in \Omega: b^{-1}\left(\left\lceil b(\rho_j\varepsilon_j)\right\rceil\right) \le \rho_j\varepsilon_j^{\max}\right\}$, and $\lceil x \rceil$ represents the *ceiling* operation (i.e., $\lceil x \rceil = n$ where n is the smallest integer such that $x \le n$). It follows that $N_\Omega = N_{\Omega 1} + N_{\Omega 2}$, where N_Ω, $N_{\Omega 1}$ and $N_{\Omega 2}$ are the cardinality of the sets Ω, Ω_1 and Ω_2, respectively. The only possibility to satisfy both the integer-bit and peak power constraints for subcarriers in Ω_1 is to use the *floor* (or *rounding-down*) operation $\lfloor b(\sigma_j) \rfloor$. Hence the granularity loss for the j^{th} subcarrier, if it belongs to Ω_1 is

$$\partial b_j^G = b(\sigma_j) - \lfloor b(\sigma_j) \rfloor; \quad \forall j \in \Omega_1. \tag{3.4}$$

For subcarriers, where *rounding* is possible without violation of peak power constraint

$$\partial b_j^G = b(\sigma_j) - round\left(b(\sigma_j)\right); \quad \forall j \in \Omega_2. \tag{3.5}$$

In both cases ∂b_j^G can be treated as a quantization error with a quantization step of 1. Since the variable to be quantized, $b(\sigma_j)$, has a much larger range (up to 15 bits/symbol) than the quantization step, the granularity loss ∂b_j^G can be considered as a uniformly distributed random variable (see p.194 in [12]),

$$\partial b_j^G \sim U[0,1);\ \forall j \in \Omega_1 \text{ and}$$
$$\partial b_j^G \sim U[-1/2,1/2];\ \forall j \in \Omega_2. \tag{3.6}$$

The random variable representing the total granularity loss is $\partial b^G = \sum_{i\in\Omega} \partial b_i^G$ with its average being

$$\overline{\partial b^G} = E\left(\partial b^G\right) = \sum_{i\in\Omega_1} E\left(\partial b_i^G\right) + \sum_{i\in\Omega_2} E\left(\partial b_i^G\right) = N_{\Omega 1}\cdot 1/2 + N_{\Omega 2}\cdot 0 = N_{\Omega 1}/2$$
$$= \eta N_\Omega / 2, \text{ where } 0 \le \eta = N_{\Omega 1}/N_\Omega \le 1 \tag{3.7}$$

where $E(.)$ in (3.7) represents the stochastic expectation operator. The ratio η can be estimated as follows.

- TPO Case: In this case, by definition, there is no peak power constraint or $\varepsilon_j^{\max} = \infty;\ \forall j$, i.e., $\Omega_1 = \varnothing$ and $N_{\Omega 1}$=0, η = 0. Also, due to the denominator being ∞, $P_{\max}/\Delta f \cdot \sum_{i=1}^{K} \varepsilon_i^{\max} = 0$. Thus the average granularity loss is nearly zero, as observed in [9][30][31].

- PPO Case: In this case, $\varepsilon_j = \varepsilon_j^{\max};\ \forall j \in \Omega$, i.e., $\Omega_1 = \Omega$ and $\Omega_2 = \varnothing$. Thus $N_{\Omega 1} = N_\Omega$ and $\overline{\partial b_{PPO}^G} = N_\Omega/2$, η = 1. N_Ω is fairly large in xDSL applications (e.g., more than 1000 in VDSL-DMT). Also from the definition of PPO (see Table 3-1), we have $P_{\max}/\Delta f \cdot \sum_{i=1}^{K} \varepsilon_i^{\max} \ge 1$.

- TPP Case: For the TPP case, the analysis of η is more involved and depends on the specific scenario. However, observing the values of η in TPO and PPO cases, which act as the boundaries of the TPP case and its monotonic nature, we can consider the following approximation

$$\eta \approx \left\lceil \frac{P_{\max}}{\Delta f \cdot \sum_{i\in\Omega} \varepsilon_i^{\max}} \right\rceil^1;\ \text{where } \lceil x \rceil^1 = \min(x,1) \tag{3.8}$$

η represents the relative strictness of the total-to-peak power constraint and we can expect that as η increases due to stricter peak power constraint, granularity losses will be higher. It is worthwhile to note that for a general TPP case, as channel conditions worsen Ω shrinks thereby reducing the denominator of η. Eventually η will increase to 1 and the

TPP case will reduce to a PPO case and all previous inferences shall apply. In VDSL-DMT scenarios, η is seen to be fairly close to 1 in most cases, and N_Ω is large. Thus the granularity loss is expected to be a fairly significant percentage of the supported rate.

But this high granularity loss does not have to be accepted as a necessary evil because of the constraint that we can load only integer bits/ symbols. By using adaptive Reed Solomon coding, most of the granularity losses can be recovered while still using integer bits/symbol.

3.3 Adaptive Reed-Solomon-based Fine Granularity Loading Scheme

In a general setting the sub-carrier specific rate function can be expressed as

$$b(\sigma_j) = r_j \left\lfloor \log_2\left(1+\frac{\sigma_j}{\Gamma_j}\right)\right\rfloor. \tag{3.9}$$

where r_j is the coding rate and Γ_j is the SNR-gap determined by the performance of the modulation and coding schemes in use. The floor operation (i.e., $\lfloor x \rfloor = m$ for the largest integer $m \le x$) arises from the integer bit constraint, since we try to find the largest integer number of bits/symbol that would satisfy the error rate target at SNR of σ_j. When the same FEC coding is applied for all sub-carriers, i.e., $r_j = r$, this floor operation restricts the sub-carrier rate to have steps at nr where n is integer (i.e., integer-bit constraint) and $R = r\sum_{j=1}^{K}\left\lfloor \log_2\left(1+\frac{\sigma_j}{\Gamma_j}\right)\right\rfloor$.

In the current VDSL[25] system [3], as shown in Figure 3-1, there is only one fixed-rate RS (n,k) encoder with n=255 and k=239 in the Physical Media Specific Transmission Convergence (PMS-TC) layer and the bit and energy allocation are carried out only in the Physical Medium Dependent (PMD) layer.

[25] ADSL has a similar structure.

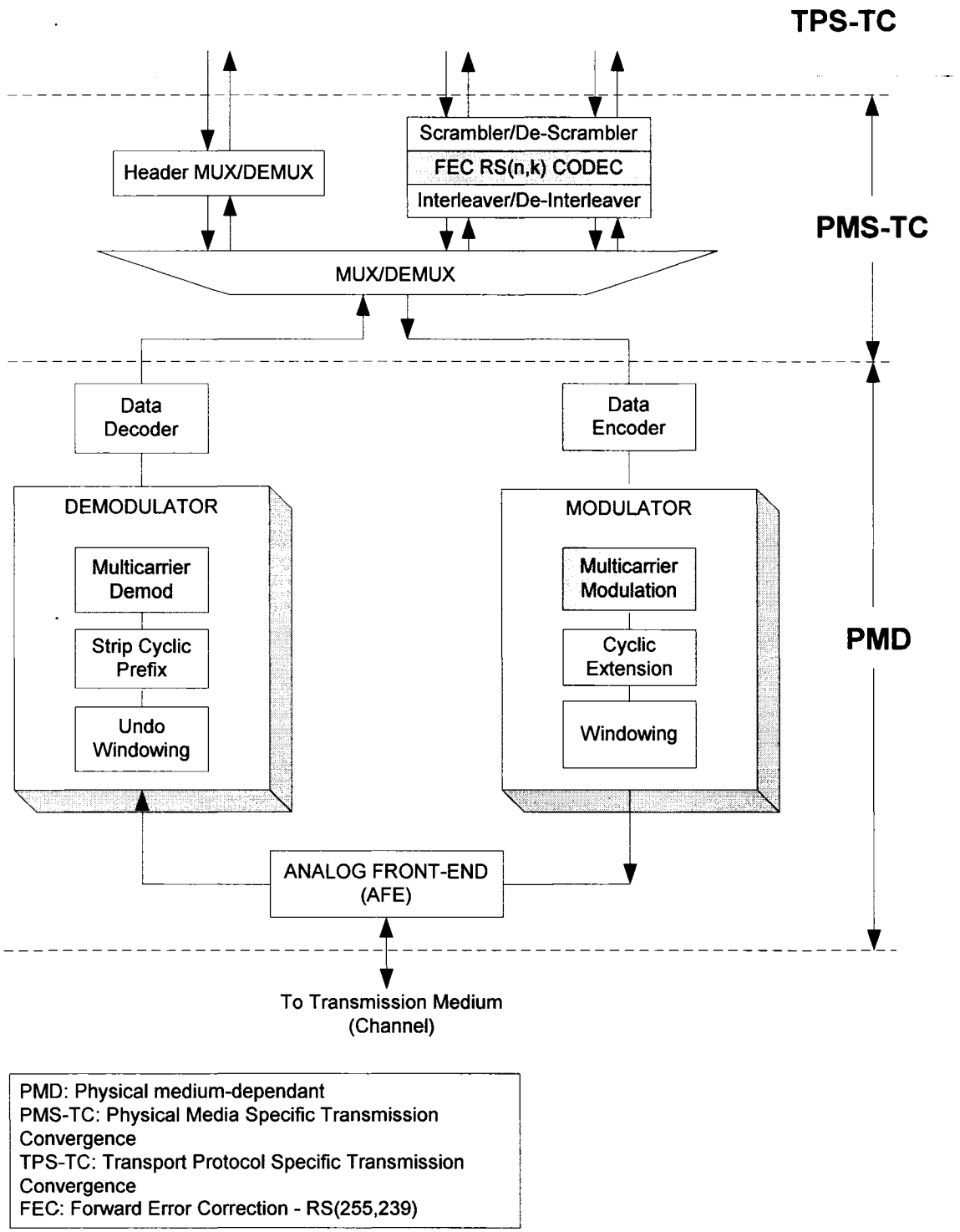


Figure 3-1: Functional Diagram of PMD and PMS-TC layer in current VDSL-DMT System

The RS(255,239) coding is applied to bits that can be transmitted in various sub-carriers. The coding channel is assumed to be a binary symmetric channel with the cross-over bit error probability $\Pr_{e,ch}$, which represents the BER averaged over all K sub-carrier

DMT modem. The final system performance $\Pr_{e,dec}$ is represented by the post-decoding bit-error probability of the RS(n,k) code over GF(2^m) [13]

$$\Pr_{e,dec}\ (\Pr_{e,ch}, n, k) \le \frac{2^{m-1}}{2^m - 1} \sum_{i=t+1}^{n} \frac{i+t}{n} \binom{n}{i} \Pr^i \left(1-\Pr\right)^{n-i} \tag{3.10}$$

where $\Pr = 1-\left(1-\Pr_{e,ch}\right)^m$ and $t = \left\lfloor \frac{n-k}{2} \right\rfloor$. The above upper bound is less than 0.1dB away from the exact BER [18].

For RS(255,239) with m=8, n=255, k=239, t=8, to achieve $\Pr_{e,dec} \le 10^{-7}$, we need $\Pr_{e,ch} < 10^{-3}$(5.65×10^{-4} to be precise). This is ensured indirectly and approximately using the SNR-gap method. Since only M-QAM is used, the uncoded SNR-gap for $P_{e,dec} \le 10^{-7}$ is nearly 9.75dB for a large range of M. The RS(255, 239) code is assumed to provide a uniform coding gain γ_c= 3.75dB. Thus Γ_j = Γ= 9.75–γ_c [6] and the code rate r_j = r = 239/255 in (3.9).

In the proposed Adaptive RS aided Fine-Granularity Loading (ARSFGL) scheme, instead of using a fixed-rate RS (n,k) code for all sub-carriers, we assume a variable rate RS(n,k_i) code for each sub-carrier #i. This can be implemented by replacing the fixed-rate RS codec in Figure 3-1 with a single programmable RS (255,k) codec [20][21], which operates on a per-subcarrier basis. Framing and buffering in MUX/DEMUX (Figure 3-1) will be modified accordingly to support this per-subcarrier RS codec operation and inter-leaving may not be required since independence of error patterns is maintained before decoding unlike in [3]. The loading algorithm provides the allocated *rates* (i.e., k_i, and the number of QAM bits/symbol M_i) and *power* as follows.

3.3.1 Rate Allocation

Figure 3-2 depicts the equivalent model representing the transmission operation for each subcarrier. The complex symbol output of the M-QAM modulator is scaled to an input PSD level of ε_j to achieve the overall *received* SNR, σ_j=$\varepsilon_j\rho_j$. Corresponding to the M_j-QAM demodulator and RS (n,k_j) decoder, the bit error probabilities are $\Pr_{e,ch}(M_j, \sigma_j)$ and $\Pr_{e,dec}[\Pr_{e,ch}(M_j, \sigma_j), n,k_j]$, respectively. Our optimization problem is formulated as follows:

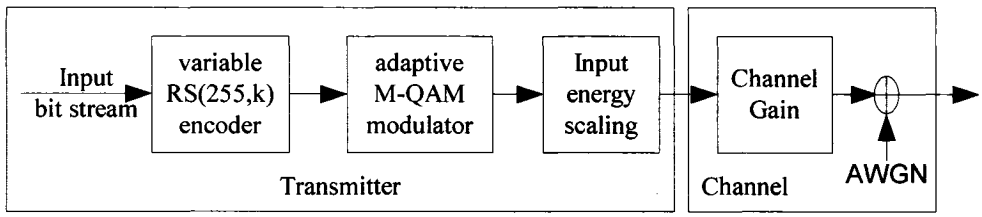


Figure 3-2: Subcarrier Transmission Model

Objective:

$$\underset{k_j,M_j}{\text{Max}}\ b(\sigma_j)=\frac{k_j}{n}\times\log_2 M_j \tag{3.11}$$

Constraints:

$$k=1,3,5....n,\quad \log_2 M_j=1,2,3,...\quad \Pr_{e,dec}[\Pr_{e,ch}(M_j,\ \sigma_j),\ n,k_j]\le 10^{-7} \tag{3.12}$$

$\Pr_{e,dec}[\Pr_{e,ch}(M_j,\ \sigma_j),\ n,k_j]$ is obtained from (3.10) with $k=k_j$ and $\Pr_{e,ch}=\Pr_{e,ch}(M_j,\ \sigma_j)$. $\Pr_{e,ch}(M_j,\ \sigma_j)$ is the BER of M_j-QAM in AWGN channels, i.e., for *odd* $\log_2 M_j$ with cross-QAM using impure Gray encoding [17],

$$\Pr_{e,ch}(M,\ \sigma)\approx\frac{G_{p,M}N_M}{\log_2 M}\cdot Q\left(\sqrt{\frac{2\sigma}{C_{p,M}}}\right) \tag{3.13}$$

where $G_{p,M}$, N_M and $C_{p,M}$ represent the Gray penalty, number of nearest neighbors and packing coefficients, respectively. For validation purposes, we simulated cross-constellations constructed from the above scheme and we observe that (3.13) gives an accurate estimate of BER for all cross-constellations from 2^5, 2^7,... 2^{15} for BERs below 0.07.

For *even* $\log_2 M_j$ with square-QAM using perfect Gray encoding [16], the expression for $\Pr_{e,ch}(M,\ \sigma)$ obtained from [15] as

$$\Pr_{e,ch}(M,\ \sigma)=\frac{2}{\log_2 M}\sum_{s=1}^{\log_2\sqrt{M}}\Pr(s,\sigma) \tag{3.14}$$

where $\Pr(s,\sigma)=\frac{1}{\sqrt{M}}\sum_{i=1}^{(1-2^{-s})\sqrt{M}-1}(-1)^{\left\lfloor\frac{i2^{s-1}}{\sqrt{M}}\right\rfloor}\left[2^{s-1}-\left\lfloor\frac{2^{s-1}i}{\sqrt{M}}+\frac{1}{2}\right\rfloor\right]erfc\left((2i+1)\sqrt{\frac{3\sigma}{2(M-1)}}\right)$.

Note that $b(\sigma_j)$ is a monotonically increasing with k_j and M_j. $\Pr_{e,ch}(M_j,\ \sigma_j)$ and $\Pr_{e,dec}[\Pr_{e,ch}(M_j,\ \sigma_j),\ n,k_j]$, on the other hand, are monotonically increasing with M_j and k_j respectively. Thus we can search for M_j and k_j in a sequential manner. At first, M_j is found to be within the limits specified by the *uncoded* case and the ideal Shannon limit,

i.e., $\lceil \log_2(1+\sigma_j/\Gamma) \rceil \le \log_2 M_j \le \lfloor \log_2(1+\sigma_j) \rfloor$. We then search for k_j in descending order, i.e., from n to $(n-2)$, $(n-5)$,... until $Pr_{e,dec}[Pr_{e,ch}(M_j, \sigma_j), n, k_j] \le 10^{-7}$. The optimum values for k_j and M_j for given σ_j can also be pre-calculated and stored in a table such as Table 3-2, so that the search for k_j and M_j can be done by table look-up technique during actual modem operation.

Table 3-2: Example of Rate Look-Up Table

σ (dB)	Optimum k_j (1-255)	Optimum $\log_2(M_j)$
30.0	245	8
30.5	247	8
31.0	249	8
31.5	251	8
32.0	229	9
32.5	235	9
33.0	239	9
33.5	223	10
34.0	229	10
34.5	235	10
35.0	239	10

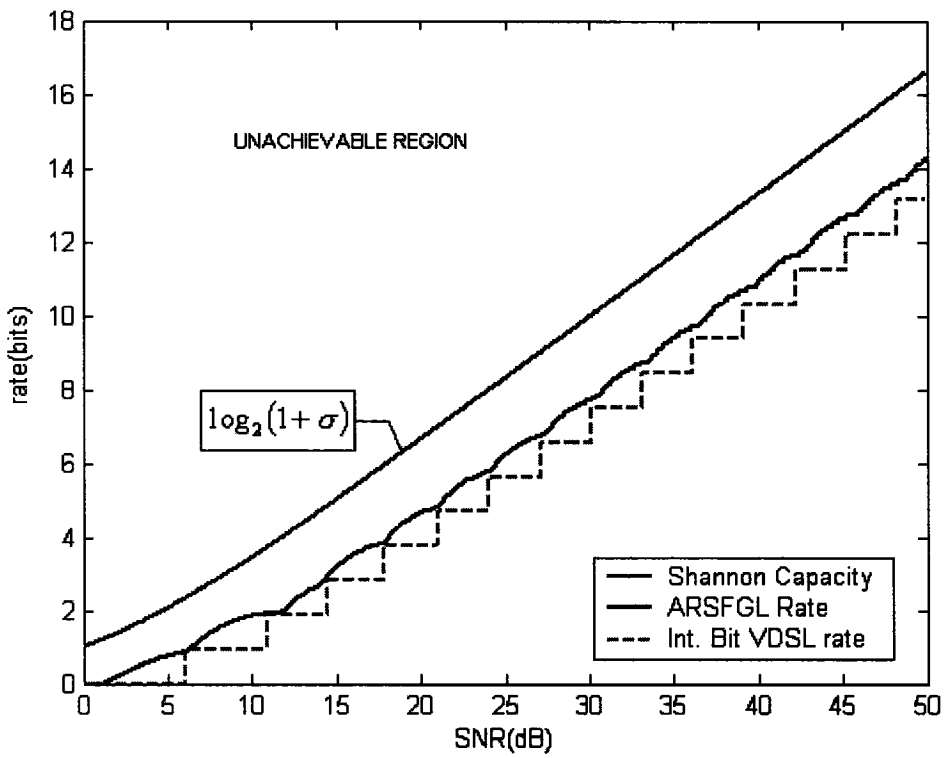


Figure 3-3: ARSFGL rate function: Rate vs. SNR

The optimized rate-function (3.11) of the proposed ARSFGL is plotted along with that of the integer-bit loading for VDSL in Figure 3-3. The finer granularity and inherent gains[26] in rate can be clearly seen. The gains stem from the fact that while k and hence, $\mathrm{Pr}_{e,ch}$ are fixed in the existing VDSL schemes, the proposed ARSFGL scheme varies $\mathrm{Pr}_{e,ch}(M_j, \sigma_j)$, jointly optimizing the adaptive coding and modulation schemes to achieve the maximum information rate. The gain in rate offered by the proposed ARSFGL is larger at higher SNR due to the fact that the proposed ARSFGL uses the bit-error probability (BER) criterion while the existing VDSL loading scheme is based on symbol-error probability [5]. As SNR increases, higher M can be used and the difference between BER and symbol error probability becomes significant. Hence the BER-based ARSFGL is closer to the constraint $\mathrm{Pr}_{e,dec} \leq 10^{-7}$. Another reason for choosing the BER based scheme is that for the choice of RS(255,k_j) on each subcarrier, the input BER $\mathrm{Pr}_{e,ch}(M_j, \sigma_j)$ is a more meaningful quantity than the M_j-ary symbol error probability (see (3.10)).

3.3.2 Energy Allocation

As can be seen from Figure 3-3, the ARSFGL rate-function is non-decreasing and can provide near-continuous rate adaptation. These conditions are sufficient for (2.17) to be satisfied[27]. Thus, for the PPO case, the optimal power allocation will be the PSD constraint. For the TPP case, however, the energy allocation depends on the rate function. Note that the solution for a continuously differentiable, strictly increasing and strictly concave rate-function is already available in (2.18) and (2.19). Furthermore, the ARSFGL rate function is close to obeying the above properties. Therefore, we consider the rate function approximated by

$$b_e(\sigma) = \alpha \log_2(\beta\sigma + \gamma). \qquad (3.15)$$

The approximation[28] is achieved by curve-fitting and the values of α=0.9597, β=0.2736 and γ=0.8232 yield a mean squared error of less than 0.0076 bits.

---

[26] Based on [3], no other code than RS is assumed in Figure 3-3. When additional or higher-performance coding is used, the gap between the Shannon limit and both curves in Figure 3-3 would be reduced by the same amount due to the additional coding gain. However, the granularity loss would remain the same.

[27] It is straightforward to verify (2.17) to hold for a continuous and increasing case. For the continuous and non-decreasing case, the only change is that (2.17) is no longer the unique optimum and solutions with smaller total energy might exist.

[28] The approximation is done only for the purpose of energy allocation so that (2.17)-(2.18) can be directly used. However, the rate allocation following this energy allocation is done using the look-up table.

From (2.18)-(2.19) with $b_e(\sigma)$ as the rate-function, the final solution to the TPP energy allocation problem is:

$$\varepsilon_j = \left[B - \gamma/\beta\rho_j\right]_0^{\varepsilon_j^{\max}} \tag{3.16}$$

where B is the solution to:

$$\Delta f \cdot \sum_{j=1}^{K} \left[B - \gamma/\beta\rho_j\right]_0^{\varepsilon_j^{\max}} = P_{\max} . \tag{3.17}$$

Here B relates to λ in (2.18) and (2.19) as $B=\alpha/\lambda\ln 2$. Thus the energy allocation problem reduces to the evaluation of B. This is done by using the low cost secant-based search method proposed in [38], with minor changes to suit our usage of (3.15), instead of the traditional SNR-gap function. These changes, when incorporated into the secant algorithm presented in Table I in [38], results in the algorithm described in Table 3-3.

Table 3-3: ARSFGL Energy Allocation

1.	$x_0 = \min_{1\le l\le K}\{\gamma/\beta\rho_l\};\ x_1 = \max_{1\le l\le K}\{\varepsilon_l^{\max} + \gamma/\beta\rho_l\}$
2.	$f_0 = -P^i;\ f_1 = \Delta f \cdot \sum_{l=1}^{K} \varepsilon_l^{\max} - P^i;\ f = f_1$
3.	while $\lvert f\rvert > eps$
4.	$B = (f_1 \cdot x_0 - f_0 \cdot x_1)/(f_1 - f_0)$
5.	$f = \Delta f \cdot \sum_{l=1}^{K} \left[B - \gamma/\beta\rho_l\right]_0^{\varepsilon_l^{\max}} - P^i$
6.	if $(f > 0)$ $\quad x_1 = B;\ f_1 = f$
7.	else $\quad x_0 = B;\ f_0 = f$
8.	End
9.	$\varepsilon_l = \left[B - \gamma/\beta\rho_l\right]_0^{\varepsilon_l^{\max}};\ \sigma_l = \varepsilon_l\rho_l;\ l = 1,2,....,K$

It is worthwhile to note that by virtue of providing near-continuous rate adaptation, in the ARSFGL energy allocation, a secondary iterative procedure characteristic to integer-bit algorithms (e.g., bit-*rounding* and energy-adjustment in [30] [9] or bisection search in [38]) is not necessary. Thus the energy allocation for ARSFGL is simpler.

3.4 Illustrative Examples for Application to VDSL-DMT Systems

We consider the four transmit PSDs specified for VDSL-DMT (see Section 7.1.2 in [3]) in both upstream (US) and downstream (DS) and total power budget (see Table 7-1 in [3]) over the same band to classify the case as PPO or TPP based on Table 3-1. As shown in Table 3-4, all five shaded sections which include all upstream cases and the

M1FTTCab downstream case represent PPO cases and the remaining three cases represent TPP cases. The TPO case does not occur in practice because most applications have peak power constraint [2][3], but has been presented here for the sake of completeness.

Table 3-4: Occurrence of PPO and TPP cases in VSDL-DMT

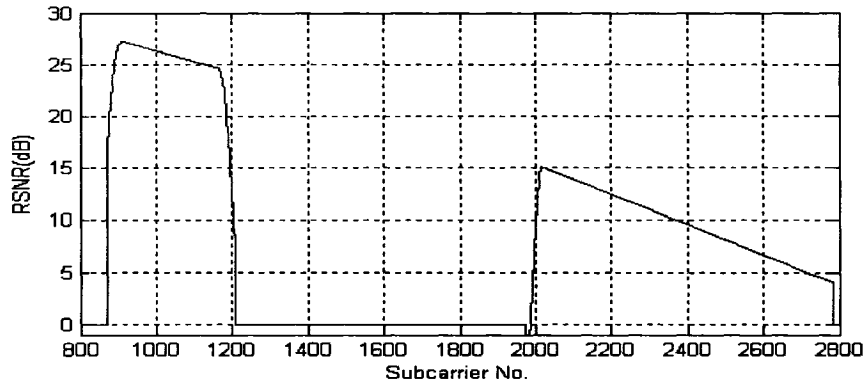
PSD	UPSTREAM		DOWNSTREAM	
	$\Delta f \cdot \sum_i \varepsilon_i^{max}$ (dBm)	P_{max} (dBm)	$\Delta f \cdot \sum_i \varepsilon_i^{max}$ (dBm)	P_{max} (dBm)
M1 FTT Cab	6.94	14.5	8.39	11.5
M2 FTT Cab	13.26	14.5	14.47	11.5
M1 FTT Ex	6.94	14.5	20.54	14.5
M2 FTT Ex	13.26	14.5	21.52	14.5

3.4.1 Evaluation of PPO case

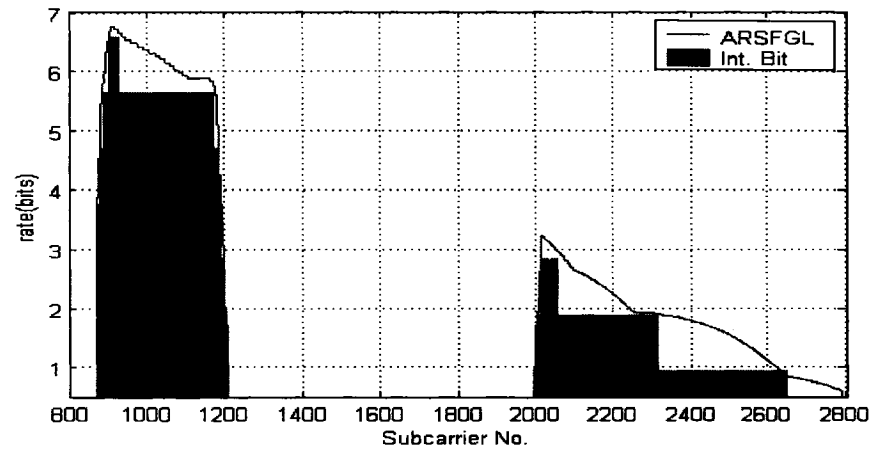
For PPO cases, (2.17) indicates that the energy allocation is independent of the rate-allocation function. Thus all existing algorithms would result in the same solution because they strive for optimization in the energy domain and in this case the energy distribution is completely decided by the peak-power constraint. The received SNR profile as a result of any loading algorithm would be $\sigma_j = \varepsilon_j^{max} \rho_j$.

Table 3-5: Simulation Parameters

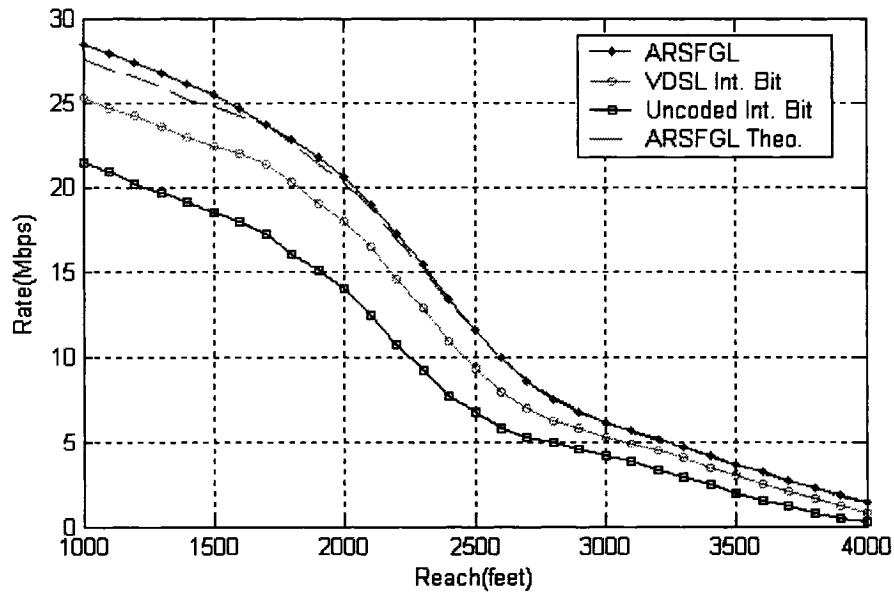
Number of Subcarriers:	4096	
Cyclic Prefix Length:	640 samples	
Upstream carriers:	Band1: 870-1205, Band2: 1972-2782	
Downstream carriers:	Band1: 33-869, Band2: 1206-1972	
Loop & Basic Noise:	Loop 1 with AWGN(-140dBm/Hz) + 20 VDSL xTalkers	
	PPO	**TPP**
Direction:	Upstream	Downstream
Total Power Constraint:	14.5dBm	11.5dBm
Transmit PSD Constraint:	M1FTTCab	M2FTTCab
Additional Noise :	+ Alien Noise A [3]	+ Alien Noise F [3]



(a) *Channel* signal-to-noise ratio, ρ, at 2400ft



(b) Sample rate allocation $b(\sigma_j)$ at 2400ft



(c) Rate versus Reach

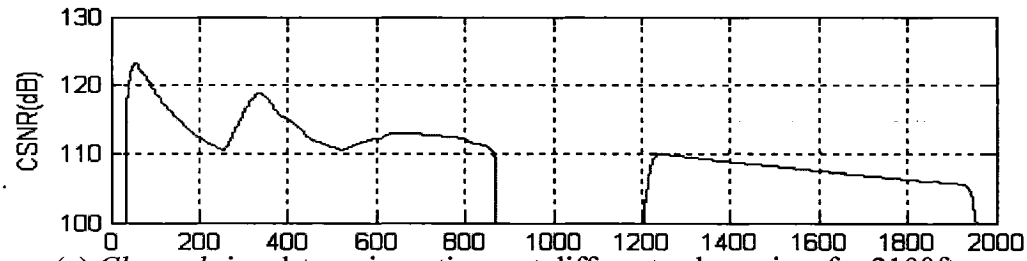
Figure 3-4: ARSFGL Performance for PPO Case

The general simulation parameters and those specific to the PPO case are presented in Table 3-5. This configuration resembles Test Case -1 in [19] except that we do not fix the data rate at 10Mbps, and study its variation over a wide range of loop lengths. The received SNR profile $\{\sigma_i\}_{i=1}^{K}$ and rate allocation over the subcarriers for this configuration at 2400ft are presented in Figure 3-4(a) and (b), respectively. The resulting data rates offered by the integer-bit loading and proposed ARSFGL schemes are 10.94 Mbps and 13.41 Mbps[29], respectively. In other words, the proposed ARSFGL scheme provides an increase in rate of 22.6% (=13.41/10.94-1). The rate-reach curves for different schemes are presented in Figure 3-4(c). Any integer-bit loading algorithm would result in this same distribution as shown for the coded and uncoded cases. The proposed ARSFGL offers a much better rate-reach curve than the integer-bit loading algorithm. The "theoretical expectation" curve is generated by adding $\overline{\partial b_{PPO}^{G}}$, i.e., (3.7) with $\eta = 1$, to the rate-reach curve of the integer-bit loading algorithm for coded case at each reach value. The ARSFGL curve closely follows the "theoretical expectation" for distances larger than 1800ft. However, for distances shorter than 1800ft, it is noticeable that the ARSFGL curve is better due to the improvements arising from a BER-based loading. Shorter distances allow higher SNR and hence higher M_j. Therefore, the BER-based improvement is more pronounced as previously discussed (Figure 3-3). The improvements offered by the proposed ARSFGL are 23.6% and 27.5% and 70% at loop lengths of 2500ft., 3600ft. and 4000 ft. respectively.

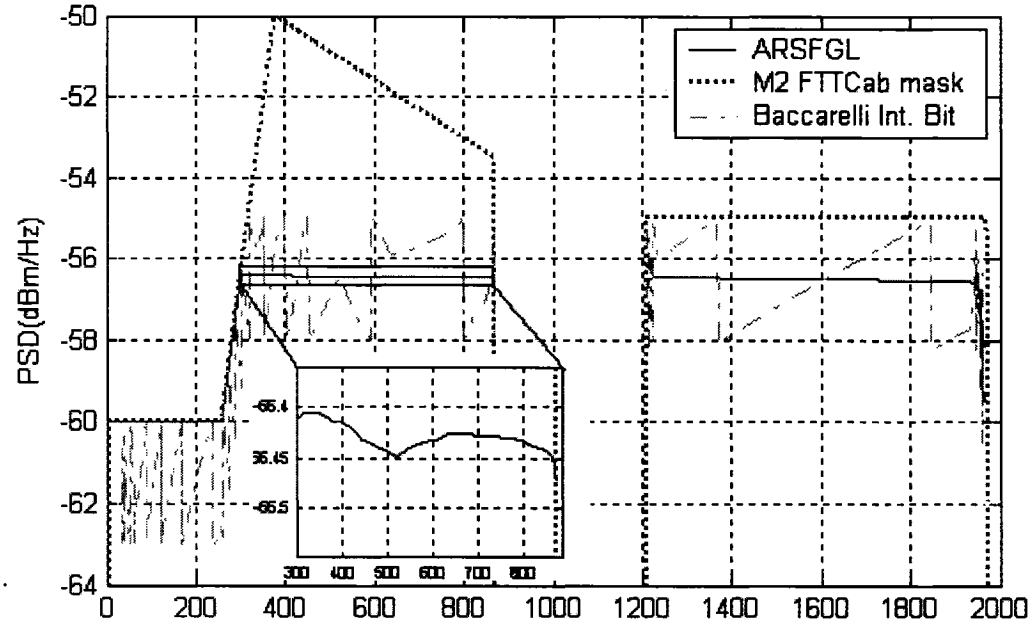
3.4.2 Evaluation of TPP cases

In TPP cases, the peak-power constraint is less stringent than the PPO case, and hence there is some room for maneuverability in the energy domain to recover some of the granularity losses.

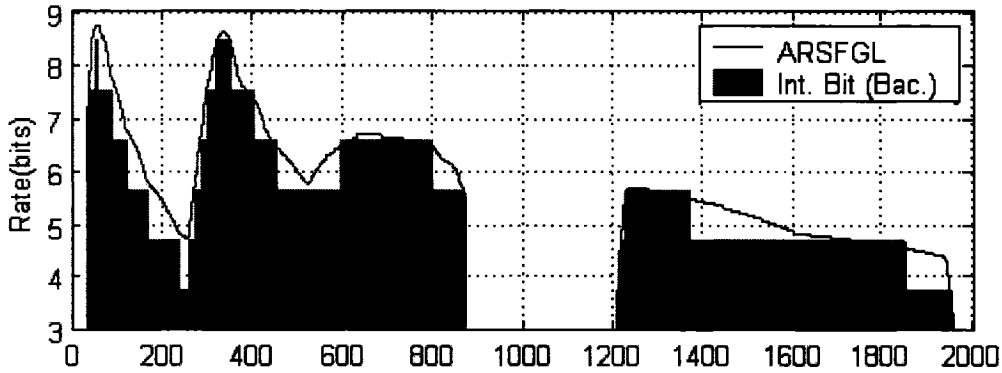
[29] It is worth noting that to achieve this increased rate with the integer-bit loading algorithm, a coding gain of 8.6dB would be required, assuming 1 bit redundancy per subcarrier characteristic of TTCM schemes [14].



(a) *Channel* signal-to-noise ratio, ρ, at different sub-carriers for 2100ft.



(b) Transmit PSD ε_j at different sub-carriers for 2100ft.

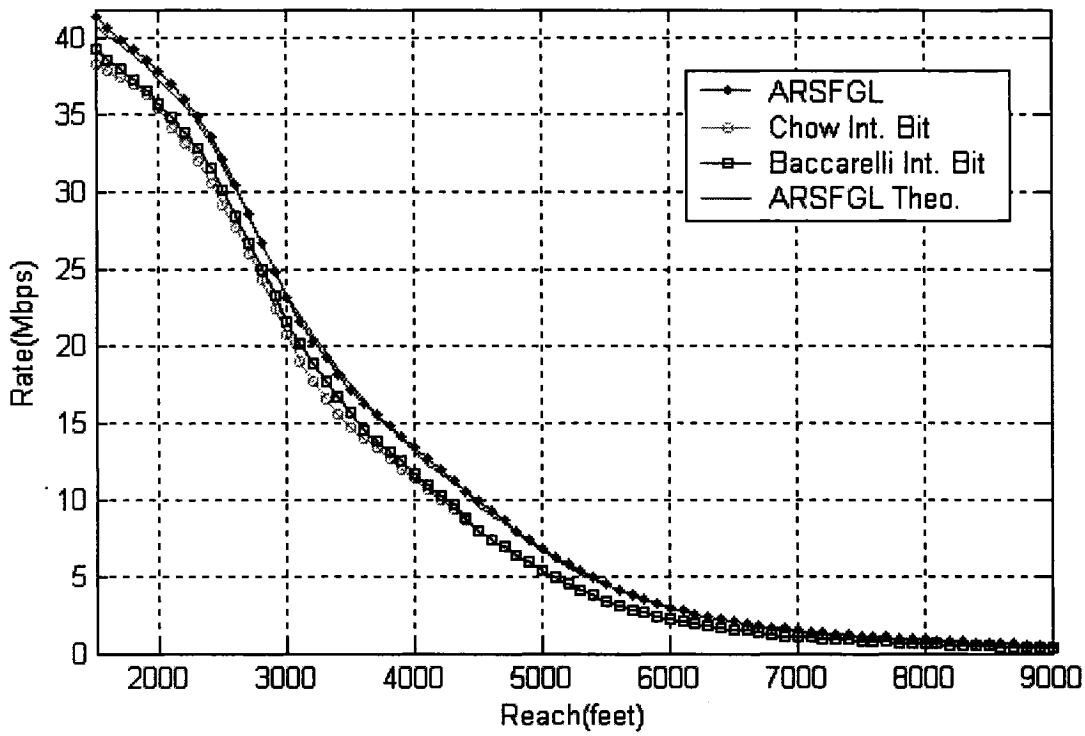


(c) Bit distribution $b(\rho_j\varepsilon_j)$at different sub-carriers for 2100 ft.

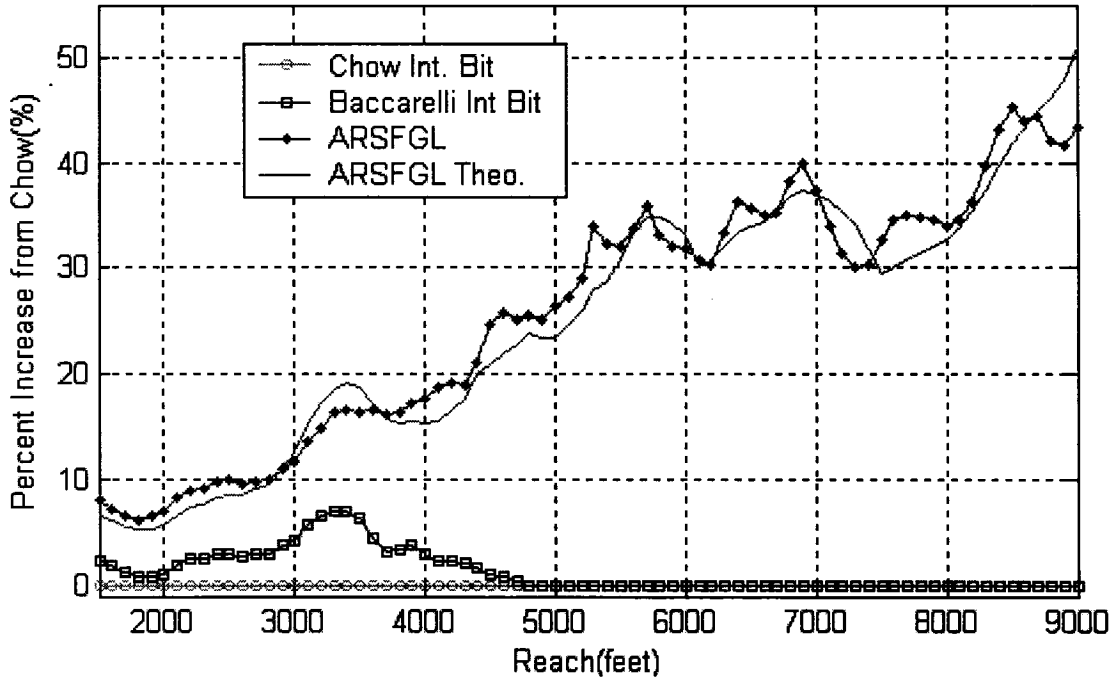
Figure 3-5: ARSFGL Performance for TPP case

The simulation parameters specific to the TPP case are presented in Table 3-5. This configuration resembles Test Case -25 in [19] except that we do not fix the data rate at 22Mbps, and study its variation over a wide range of loop lengths. The channel SNR ρ_j for the above configuration and a loop length of 2100ft. is shown in Figure 3-5(a). In

Figure 3-5(b), the PSD constraint in the form of M2FTTCab mask is presented along with the transmit PSD allocated by the ARSFGL scheme and integer-bit scheme by Baccarelli [38]. The integer-bit scheme leads to the characteristic saw-tooth distribution, which deviates on both sides of the smooth distribution of the ARSFGL scheme. In Figure 3-5(c), the resulting bit-distributions are presented. Unlike in the PPO case (where $\Omega_2 = \varnothing$), here we observe sets of subcarriers (belonging to Ω_2), where the integer-bit scheme is able to allocate more bits than the ARSFGL scheme due to the saw-tooth nature of the energy distribution. This is what we have referred to as recovery of granularity loss through energy re-adjustment in earlier parts of the chapter. It can be seen that, in the subcarriers 33-300 where the M2 mask is particularly stringent at -60dBm/Hz, the ARSFGL scheme always is able to allocate more bits just like in PPO cases. These subcarriers form a part of set Ω_1.



(a) Rate-Reach Curves



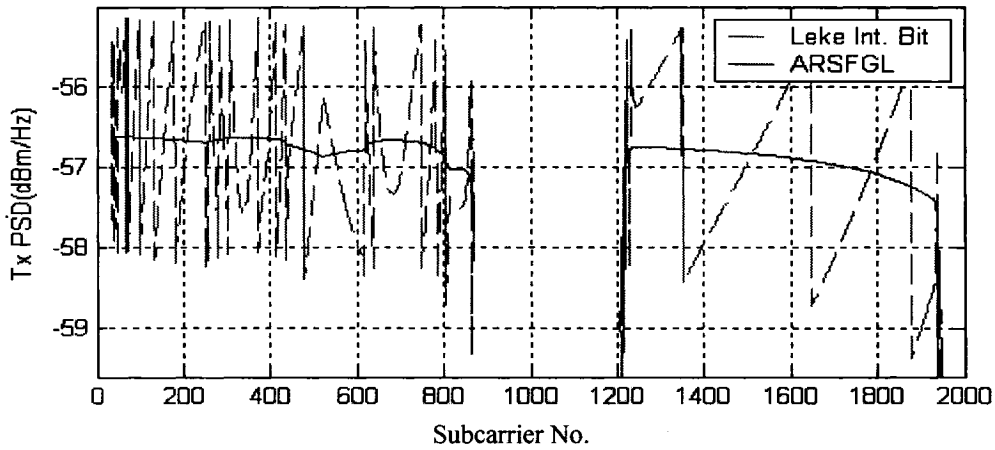
(b) % Increase in data rate as compared to Chow's algorithm

Figure 3-6: Performance of various schemes for TPP

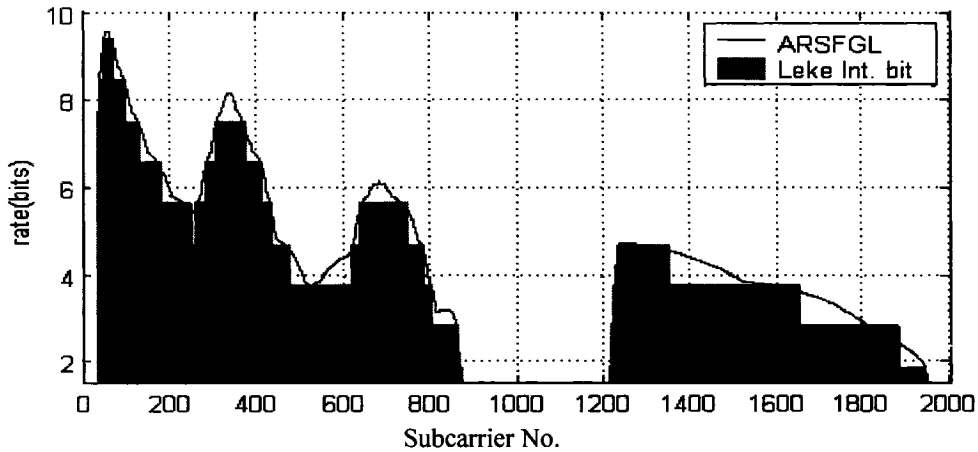
The rate-reach curves are presented in Figure 3-6(a). For easier comparison of schemes, the percentage improvements over Chow's algorithm have been presented in Figure 3-6(b). From Figure 3-6(a), we can see that on an average, the ARSFGL scheme provides about 2Mbps improvement over the integer bit schemes for loops shorter than 5500ft. As expected from (3.8), for loops longer than 4700ft., η becomes 1 and this case reduces to a PPO case as shown in Figure 3-6(b), with both Chow's and Baccarelli's schemes giving exactly the same performance. As reach increases, both granularity loss (that depends on N_Ω), and rate are reduced. However, the reduction in rate is much faster than that in N_Ω (and hence granularity loss). Since the proposed ARSFGL draws most of its improvement from the granularity loss, its percentage of improvement increases with reach as shown in Figure 3-6(b). The theoretical curves are generated by adding $\overline{\partial b^G}$ from (3.7), (3.8) to the rate provided by the integer-bit Baccarelli's algorithm at different reach values. It is observed that the rate-reach curve of the ARSFGL follows the theoretical expectations closely and thereby the assumption on η in Section 3.2 is validated.

3.4.3 Evaluation of TPO case

Though the TPO case does not occur in practice, it has been presented here for the sake of completeness. The hypothetical TPO scenario is constructed by removing the PSD constraint from the TPP configuration shown in Table 3-5.



(a) Transmit PSD ε_j at different sub-carriers



(b) Bit distribution $b(\rho_j\varepsilon_j)$at different sub-carriers

Figure 3-7: Power and Rate allocation for TPO Case – 2400 ft. loop

The *power* and *rate* allocation results of the Leke's algorithm [9] and proposed ARSFGL scheme are shown in Figure 3-7(a) and (b), respectively, for a 2400ft. loop.

Figure 3-8 shows the percentage increase in rate as compared to Chow's algorithm [30] versus loop lengths offered by the Leke [9], Baccarelli [38], optimal (greedy) integer-bit Hughes-Hartogs (HH) [8] algorithms and proposed ARSFGL scheme. It indicates that the rate increase offered by the Leke, Baccarelli, and Hughes-Hartogs algorithms is less than 1% while the proposed ARSFGL scheme can provide 4%-6% rate improvement for distances up to 7000ft. This improvement is explained by the fact that though in Sec-

tion 3.2, we have assumed bit-*rounding* to be an unbiased operation for simplifying the analysis, *rounding* up a bit always costs more in terms of energy than *rounding* down for the same difference due to the logarithmic (concave) nature of the rate function. This bias leads to the granularity loss being positive even for the TPO case due to Ω_2 set of sub-carriers. However in PPO and TPP case, as we observed, this effect is strongly dominated by loss due to Ω_1.

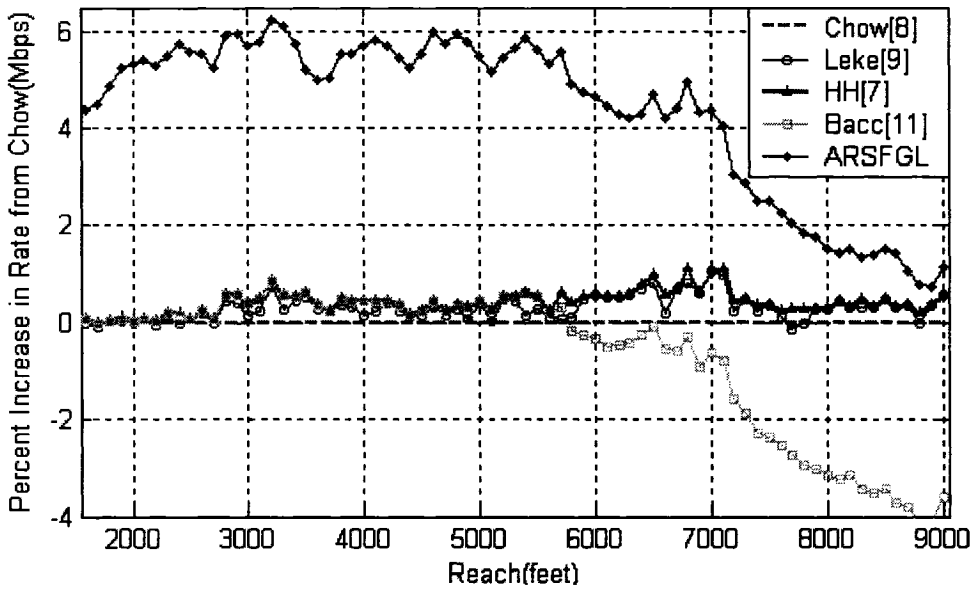


Figure 3-8: Performance of various schemes for TPO case

3.5 Application to Dynamic Spectrum Management (DSM)

The above results and analysis have been presented for the case when spectrum management is performed through specification of spectral masks for all users, which is the currently standardized form of spectrum management in ADSL [1][2] and VDSL [3], known as static spectrum management. In Dynamic Spectrum Management (DSM) which was discussed in Section 2.3.2, the interference modeling is not worst case and the peak power constraint ε^{max} is either not present or much less strict when present than SSM schemes. We had also noted in Section 2.3.2 that one of the ways to compare multi-user schemes is by plotting the rate-region boundary (Recall the discussion on Figure 2-5). The rate region boundary was noted to be the solution to the following problem:

$$\begin{aligned} &\min\left(-\sum_{j=1}^{K} b\left(\varepsilon_j^1 \rho_j^1\right)\right) \\ &\text{subject to } \sum_{j=1}^{K} b\left(\varepsilon_j^i \rho_j^i\right) \geq T^i; \quad i = 2,3,...,N \\ &\qquad \text{and } \Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^l \leq P_{\max}^l; \quad l = 1,2,...,N. \end{aligned} \tag{3.18}$$

Though the peak power constraint is absent, in (3.18), when T^i increases, the maximum power that User 1 can put on its 'good' subcarriers is restricted. This is the case because in DSL all users have low attenuation at low frequencies. So the 'good' subcarriers of various users are very similar frequencies (recall discussion on Figure 2-1) and hence the rate demand of one user T^j implicitly acts as a strict peak power constraint for the other users which opens the room for granularity losses as discussed in this chapter. That's why we would once more expect the ARSFGL scheme to provide significant improvements if molded into the DSM cast.

To illustrate this, we conduct the comparison between 3 schemes the integer-bit version of IWF, integer bit version of OSM and the adaptive Reed-Solomon aided version of IWF by plotting their rate region boundaries.

3.5.1 Adaptive RS aided IWF (ARS-IWF)

This paragraph recalls the IWF discussion in Section 2.3.2.1. IWF considers the DSL channel as an *interference* channel and models the power allocation problem as a non-cooperative game theoretic problem with each user being a player. At each turn each player implements the optimal response to the sum of noise and interference profile, it observes at that stage, which is the traditional water-filling. A sequence of such moves eventually converges to a unique Nash Equilibrium point for each **P**. The rate region is generated by varying **P** within the constraints.

IWF can be combined with the ARSFGL scheme by making each user implements its response power allocation to the effective sum of noise and interference profile ρ_l^i using Table 3-3. At the end of the IWF procedure (i.e., after convergence) each user has its SNR profile σ_l^i stored from Step 9 of Table 3-3. This is used to do the bit and code allocation from the look-up table similar to Table 3-2. This is summarized in Table 3-6. In comparing ARS-IWF in Table 3-6 with the original IWF Table 2-2, it should be noted

that the above changes preserve the purely distributed nature of IWF, since both ARSFGL power allocation and the look-up table reading can be carried out autonomously by the users.

Table 3-6: Adaptive RS aided IWF (ARS-IWF)

1. *repeat*
2. *for* $i = 1$ to N
3. Find $\{\varepsilon_l^i\}_{l=1}^K$, $\{\sigma_l^i\}_{l=1}^K$ from $\{\rho_l^i\}_{l=1}^K$, P^i using Table 3-3.
4. *end*
5. *until* the desired accuracy is reached
6. User i loads M_l-QAM and RS(n, k_l) corresponding to σ_l^i from the ARSFGL lookup-table.

3.5.2 Integer-Bit IWF (IB-IWF)

IB-IWF is currently implemented by at least four companies [29]. The key step is 4 in Table 2-2. When this step is replaced by an integer bit approximation of WF e.g., Leke's algorithm [9] we have IB-IWF. We choose the integer bit loading algorithm in Table IV in [38] for this purpose. Notice that in such a scenario, no final step, like step 6 in Table 3-6, would be required since integer bit allocation would already have been done during the energy allocation. But the step 3 itself, for integer bit case, takes more iterations, due to the additional bisection search [38] or bit-rounding [9], than the ARSFGL scheme.

3.5.3 Integer-Bit OSM (IB-OSM)

OSM [27] uses dual decomposition to transform the total power constraint into a subcarrier-specific problem. But within each subcarrier still an exhaustive search has to be carried out for all integer bit allocations. The inversion of the SNR-gap function is used in OSM; see Section IIC in [27]. While the IB-OSM needs centralized control it ensures that it is the true optimum that can achieved by any integer bit DSM scheme.

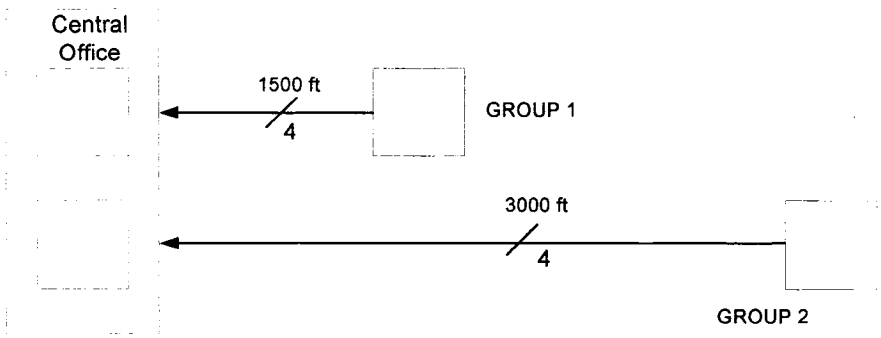


Figure 3-9: 8 user VDSL upstream scenario

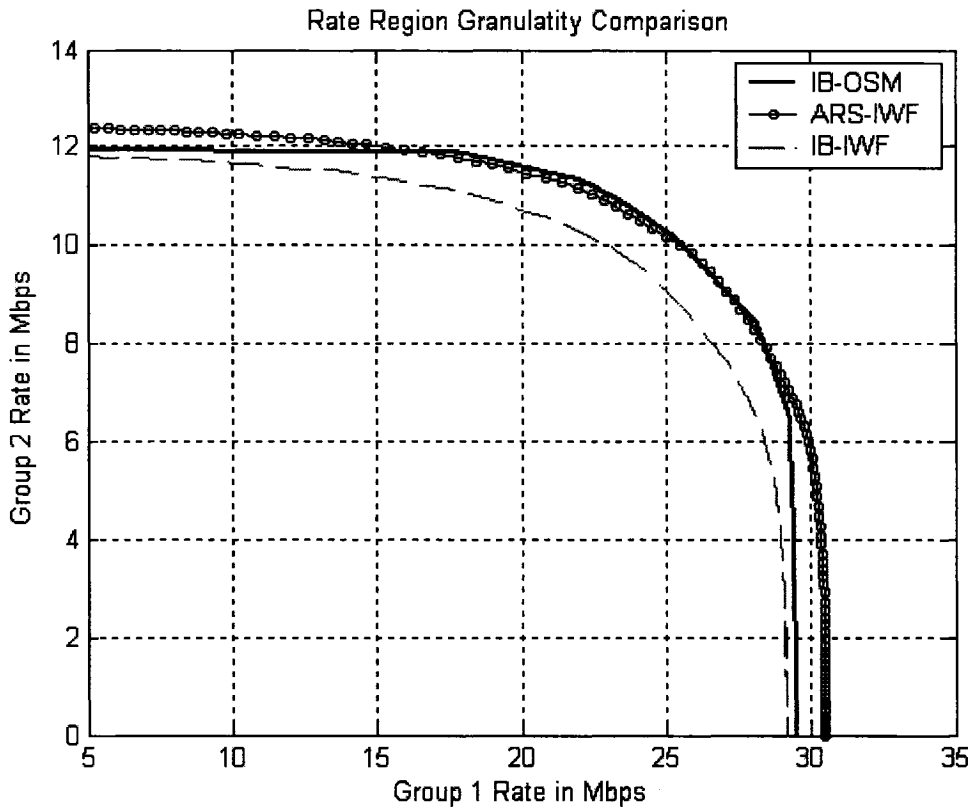


Figure 3-10: Rate Region Boundary of ARS-IWF

3.5.4 DSM ARSFGL Simulation Results

We consider the 8 user upstream VDSL-DMT scenario similar to [25] as shown in Figure 3-9. Each of the lines is 26 AWG. For each user P^{i}_{max} =11.5dBm [3] and K=4096. Crosstalk Noise Model A is assumed and the FEXT transfer functions and upstream frequency band specifications are chosen as specified in [3] and the optional band is not

used. For each of these scenarios the 3 schemes are ARS-IWF, IB-IWF and IB-OSM are tested. IB-OSM generates the rate region boundary by varying the weights in the weighted rate sum. The ARS-IWF and IB-IWF rate region boundaries are generated by keeping $P^i = P^i_{\max}$ for at least one of the 2 groups and executing Table 3-6 and Table 2-2(with integer bit loading). The rate region boundaries are plotted in Figure 3-10. The ARS-IWF curve is significantly above the IB-IWF rate region. More importantly, on most occasions ARS-IWF provides greater rate (on other occasions it is nearly equal) than even IB-OSM which represents the best any integer bit DSM scheme can do. Further the number of IWF loops for the Nash equilibrium to be reached in case of ARS-IWF took 3.018 iterations on an average and the maximum was 4 iterations (measured over 167 realizations). But the IB-IWF took 5.359 iterations on an average and a maximum of 15 iterations. In each such iteration of IWF, each user will have to measure the sum noise and interference profile at its receiver front-end, which is a costly and lengthy procedure.

3.6 Chapter Summary

We examined the granularity loss due to the integer-bit restriction that can contribute in a significant percentage in reducing the achievable data rates, especially in peak power constrained cases, and developed a fine-granularity BER-based loading scheme to recover these losses. This is done by jointly optimizing the coding rate of a programmable RS (n,k) code and the bit and energy allocation on each sub-carrier. Illustrative examples of applications to VDSL-DMT systems indicate that the proposed scheme outperforms various existing integer-bit loading algorithms with an increase in rate of about 20% in most cases. This is a large rate increase as compared to the variation in achievable rates of less than 4% between various existing integer-bit loading algorithms. This improvement is in a good agreement with the theoretical estimates developed to quantify the granularity loss. The theoretical estimates also present an insight into how the granularity losses increase with rising strictness in peak power constraint, in comparison to the total power constraint and with the number of subcarriers in use.

We also observed that in the DSM scenario when a peak power constraint is not explicitly present, the demanded rates of other users implicitly act as peak power constraint which leads to granularity loss. ARSFGL applied to the DSM scenario also gives very

encouraging results. When combined with IWF it results in ARS-IWF - a **purely distributed** scheme which can provide data rates as high as the **centralized** and **optimal** integer bit scheme. Further the convergence of IWF is faster when used with this technique than with integer bit loading.

Throughout this chapter, rate achieved was used as the indicator of performance and loss. Rate functions summarize the capability of one's modulation and coding. Clearly having a better rate-function (like we do in the ARSFGL scheme) implies that both RA algorithms and *margin* adaptive algorithms can or have the potential to produce better results. We have demonstrated this for RA algorithms, in this chapter and were able to combine and adapt various rate-adaptive loading scenarios with the ARSFGL scheme which led to good results. But from the perspective of *margin* adaptation, there is a more fundamental problem to be addressed.

As explained in Section 2.2.3, a multiuser *margin* adaptation algorithm currently does not exist. Thus the possession of ARSFGL scheme from the perspective of *margin* adaptation is analogous to a having a pond full of fish but not having a fishing net (i.e., the means to harness the resource)[30]. Even viewed independently of ARSFGL scheme, as explained in Section 1.4 and Section 2.3.4, a multiuser *margin* adaptation is required to answer the fundamental question of whether or not it is possible to deal with capacity limiting noise and performance limiting noise effectively at the same time. Thus filling this void is the primary objective of the next chapter.

[30] Please note that this comment is only from the perspective of *margin* adaptation. ARSFGL even without *margin* adaptation has been proven to be valuable enough for rate adaptive applications in this chapter.

Chapter 4
Enlightened Margin Maximization

Margin maximization is desirable for constant bit rate applications and provides protection against non-stationary and bursty noise sources (impulse noise), radio frequency interference pickup, mode transition of other modems, switching of electric devices etc. The definition of *margin*, assumptions involved therein, the rationale behind those assumptions, motivation behind *margin* maximization and existing *margin* maximization algorithms were discussed in Section 2.2.2. In this chapter we develop algorithms for multiuser *margin* optimization for usage in multicarrier Digital Subscriber Loop (DSL) systems employing Dynamic Spectrum Management (DSM).

As we saw in Sections 2.2.2 and 2.3.3, most single-user *margin* maximization algorithms rely on a fixed crosstalk assumption. But in DSM, each user's power allocation dynamically determines the other users' crosstalk. Thus, with direct extension of single-user algorithms in DSM scenarios, one user's *margin* maximization can lead to the failure of other users in meeting their target rates.

Hence in this chapter[31], we conduct an independent study of multiuser *margin* which we observe to possess certain favorable monotonicity and fairness properties. We use these properties to develop a box-constrained Nonlinear Least Squares (NLSQ) formulation which is efficiently solved using a scaled gradient trust region approach with Broyden Jacobian update. This algorithm converges to a solution providing the *best common equal margin* to all users while explicitly guaranteeing that each user's target rate requirement is satisfied. The algorithm is also practically viable because it can be implemented in current DSL-DSM scenarios with only Level 1 coordination. Levels of coordination were discussed in Section 2.3.2.

[31] Throughout the chapter a continuous SNR-gap based rate-function is assumed. This assumption is generally made in many DSM algorithms [25], for simplicity. No doubt this makes our analysis simpler as well. But in our case, we actually have a method to implement a near-continous rate function. As stated in Chapter 3, the ARSFGL scheme can provide near-continuous rates (see Figure 3-3). Thus the results presented here have more accurate physical meaning and algorithms are closer to reality, since the developed algorithms can actually be implemented in conjunction with the ARSFGL scheme.

4.1 Problem Formulation and Margin Basics

Let us consider the DSL *interference channel* [25] of N users (i.e., N transmitters and N receivers). Each of the N users is a multicarrier system with K subcarriers. Throughout this thesis, unless otherwise specified, superscripts refer to user number and subscripts to subcarrier number. The channel from User i to User l on the j^{th} subcarrier is denoted by H_j^{il}. For User i on Subcarrier j the controlled transmit power spectral density (PSD) used is denoted by ε_j^i and the background noise PSD encountered is n_j^i. The set of all ε_j^i is denoted by $\boldsymbol{\varepsilon} = \left\{\varepsilon_j^i\right\}_{j=1,i=1}^{j=K,i=N}$. The inter-carrier spacing Δf is assumed to be small enough for the aforementioned PSDs to be nearly flat. The total *utilized* power by User i is P^i,

$$P^i = \Delta f \cdot \sum_{j=1}^{K} \varepsilon_j^i . \tag{4.1}$$

The total *utilized* power vector for the N users is denoted by $\mathbf{P} = \left\{P^i\right\}_{i=1}^{N}$. The total power *constraint* on User i is $P_{\max}^i$ and together for the N users $\mathbf{P}_{\max} = \left\{P_{\max}^i\right\}_{i=1}^{N}$. It is important to note that $P_{\max}^i$ is specified by standards e.g., in [3] $P_{\max}^i$ =11.5Bm for upstream. But P^i in a certain scenario depends on the particular DSM algorithm in use, e.g., a user might choose $P^i < P_{\max}^i$ out of compassion for more needy users. However, independent of the allocation method in use the following must hold:

$$\mathbf{0} \le \mathbf{P} \le \mathbf{P}_{\max} \tag{4.2}$$

where the above vector inequality is component-wise.

The available received signal-to-noise ratio (SNR) if we use other users' interference as noise is σ_j^i and is evaluated as:

$$\sigma_j^i = \frac{\varepsilon_j^i \left|H_j^{ii}\right|^2}{n_j^i + \sum_{l=1,l\neq i}^{N} \varepsilon_j^l \left|H_j^{li}\right|^2} . \tag{4.3}$$

The maximum rate achievable R^i given a SNR profile $\left\{\sigma_j^i\right\}_{j=1}^{K}$ is:

$$R^i = \sum_{j=1}^{K} \log_2\left(1+\sigma_j^i/\Gamma\right) \tag{4.4}$$

where Γ is the SNR-gap [4][31] and $\mathbf{R} = \{R^i\}_{i=1}^{N}$. The rate region is defined as the set of achievable rate combinations by a particular algorithm *ALG*; $\Re_{ALG} = \{\mathbf{R} : \exists \boldsymbol{\varepsilon}$ reachable by *ALG* satisfying (2), (3) and (4)$\}$ and the boundary of the rate region is defined as $\overline{\Re_{ALG}} = \{\mathbf{R} : \mathbf{R} \in \Re_{ALG} \wedge \mathbf{R} + \boldsymbol{\delta} \notin \Re_{ALG} \forall \boldsymbol{\delta} > \mathbf{0}\}$, the inequality being component-wise.

At any stage, each of the N users has a target (demanded) rate T^i referred together as $\mathbf{T} = \{T^i\}_{i=1}^{N}$. The performance or SNR *margin* of the i^{th} user denoted by γ_M^i is defined[32] as the factor by which the noise can be increased before the system error-rate rises above acceptable threshold (specified implicitly in Γ). Mathematically γ_M^i is the solution to:

$$T^i - \sum_{j=1}^{K} \log_2\left(1 + \sigma_j^i / \gamma_M^i \Gamma\right) = 0 . \tag{4.5}$$

For a given T^i and $\{\sigma_j^i\}_{j=1}^{K}$; the left hand side of (4.5) is monotonically increasing in γ_M^i (and obviously continuous); hence (4.5) uniquely (though implicitly) specifies γ_M^i. Noticing the similar equations (4.4) and (4.5), one might be tempted to think that σ_l^i can be eliminated and an expression for γ_M^i can be obtained only in terms of R^i and T^i. Such an expression would make the *margin* independent of the particular algorithm in use and would remain the same for single-user or multiple users. While no such exact expression is possible, an approximation is available [4],

$$\gamma_M^i \approx \frac{2^{R^i/K} - 1}{2^{T^i/K} - 1}. \tag{4.6}$$

Reliability of (4.6) is discussed in Section 4.2. Further using the monotonicity of (4.5), the continuity of (4.4) and (4.5), and a few operations, the following 3 conclusions can be proven about γ_M^i, R^i and T^i irrespective of the $\{\sigma_j^i\}_{j=1}^{K}$ that generated them (i.e., independent of algorithm and scenario).

(C1) $\gamma_M^i = 1$ if and only if $R^i = T^i$.

(C2) $\gamma_M^i > 1$ if and only if $R^i > T^i$.

[32] The subscript *M* here indicates *margin* and is not a subcarrier number. Margin is a single positive scalar for a user, i.e., it is a user-specific quantity and not subcarrier-specific.

(C3) $\gamma_M^i < 1$ if and only if $R^i < T^i$.

Margin by its nature is a safety *margin* [37] and simultaneously a performance (in error rate terms) enhancer [32]. Thus going by the philosophy of distributed DSM whereby our concern is each user's well-being and not some single overall system metric, it is natural to model the problem as a min-max problem. Our objective will be to maximize the minimum *margin* among $\gamma_M^i; 1 \le i \le N$ by varying $\boldsymbol{\varepsilon}$ while respecting the total power constraint (4.2) and the target rate constraints (4.5) in as distributed a manner as possible. Stated formally we have:

$$\begin{aligned} &\text{Objective Function: } \min_{\boldsymbol{\varepsilon}} \left(\max_i \left\{ -\gamma_M^i \right\}_{i=1}^N \right) \\ &\text{Constraints: } \Delta f \cdot \sum_{l=1}^K \varepsilon_l^i \le P_{\max}^i; \quad 1 \le i \le N \\ &\sum_{j=1}^K \log_2 \left(1 + \frac{\varepsilon_j^i \left| H_j^{ii} \right|^2}{\gamma_m^i \Gamma \left(n_j^i + \sum_{l=1, l \ne i}^N \varepsilon_j^l \left| H_j^{li} \right|^2 \right)} \right) = T^i; \quad 1 \le i \le N \end{aligned} \tag{P1}$$

By explicitly accounting for the target rate constraint in (P1) we guarantee that the luxury of *margin* maximization does not come to any user at the unacceptable cost of failure of meeting target rate for any other user. One might suspect from the min-max formulation that the users closer to the central office will always dominate and get a better $\hat{\gamma}_M^i$ at the solution to (P1). But we observe in Section 4.2 that for a convex rate region $\Re_{ALG}$; the solutions $\hat{\gamma}_M^i$ to (P1) have the beautiful property of $\hat{\gamma}_M^1 = \hat{\gamma}_M^2 = \ldots = \hat{\gamma}_M^N$ thus inherently guaranteeing fairness among users.

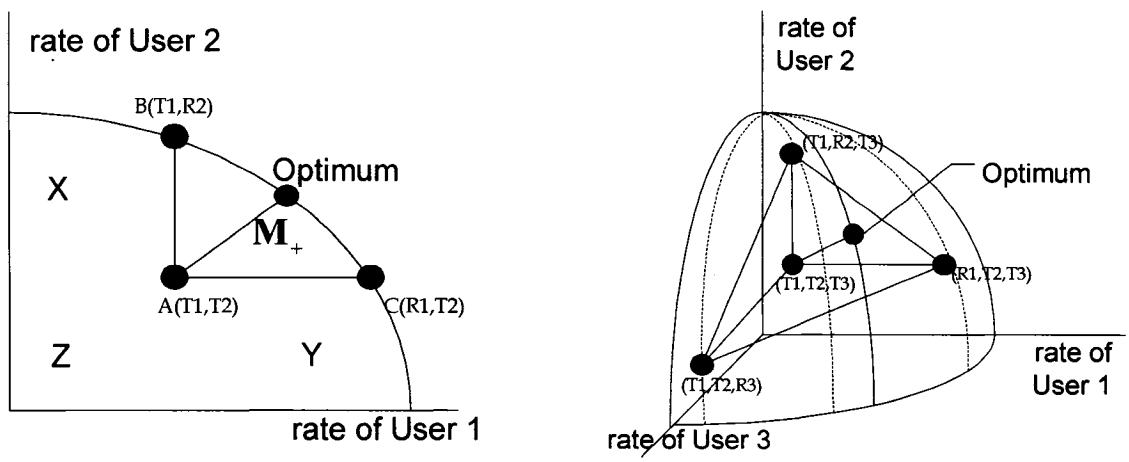
4.2 Behavior of Margin over a Rate Region

In all cases where the approximation in (4.6) holds accurately the following intuitively obvious conclusion can be made for DSL channels:-

(C4) γ_M^i is monotonically increasing in R^i for any $\mathbf{R} \in \Re_{ALG}$ for fixed $\mathbf{T}$.

(C4) follows directly from the functional form in (4.6). Consider the 2-user rate region in Figure 4-1(a). Point A represents $\mathbf{T}$. For any $\mathbf{T} \in \Re_{ALG}$, it is possible to find 2 points B, $C \in \overline{\Re_{ALG}}$. B represents the maximum rate achievable for user 2 when user 1 has a mini-

mum rate of T1. C represents the analogous situation for user 1. First of all it is important to understand that once $\mathbf{T}$ is specified the entire rate region $\Re_{ALG}$ is transformed into a *margin* region $\mathbf{M}$. This is because by definition of $\Re_{ALG}$ each point in $\Re_{ALG}$ specifies a corresponding ε which in turn translates into N user *margin* γ_M^i by (4.3) and (4.5).



(a) 2-user Case (b) 3-user case

Figure 4-1: Relationship between Rate Region and Margin

By (C1), for any point taken on the straight line AB; User 1's margin is zero (dB). For any other point which is not on AB, User 1's margin is not zero (dB). An analogous statement for User 2 and straight line AC can be made. By (C2), the region demarked by A, B, C and $\overline{\Re_{ALG}}$ (excluding points on AB and BC) represents the subset of $\mathbf{M}$ where all the users are guaranteed to have positive (dB) *margin*. Let us call this region and its closure as $\mathbf{M}_+$. Finally by (C3), all other points represent negative (dB) *margin* for at least one of the users and hence cannot contain the solution to (P1); e.g., X has negative *margin* for user 1, but positive *margin* for User 2, Y has the reverse situation and Z has negative *margin* for both users. Thus the solution must exist in $\mathbf{M}_+$. For any point $\left(\tilde{R}^1, \tilde{R}^2\right)$ in $\mathbf{M}_+$ and not on $\overline{\Re_{ALG}}$, it is possible to find a point $\left(\hat{R}^1, \hat{R}^2\right)$ on $\mathbf{M}_+ \cap \overline{\Re_{ALG}}$ such that $\hat{R}_1 > \tilde{R}_1; \hat{R}_2 > \tilde{R}_2$. Hence due to (C4), $\left(\hat{R}^1, \hat{R}^2\right)$ will have a greater minimum *margin* than $\left(\tilde{R}^1, \tilde{R}^2\right)$. Thus the solution must exist in $\mathbf{M}_+ \cap \overline{\Re_{ALG}}$ represented by BC in

Figure 4-1(a). At point B, the *margin* achieved by User 1 and User 2 is $\left(0,\gamma_M^{2,\max}\right)$, where $\gamma_M^{2,\max}$ is the maximum *margin* User 2 can have while User 1 having a non-negative *margin*. Similarly at point C, the *margin* achieved is $\left(\gamma_M^{1,\max},0\right)$. If the $\Re_{ALG}$ is concave; as we move from B to C along $\overline{\Re_{ALG}}$, γ_M^2 monotonically decreases from $\gamma_M^{2,\max}$ to 0 and γ_M^1 monotonically increases from 0 to $\gamma_M^{2,\max}$. Thus by continuity arguments, there must exist a point $\mathbf{R}^*$ between B and C on $\overline{\Re_{ALG}}$ (labeled as 'Optimum' in Figure 4-1(a)) with the *margin* $\left(\gamma_M^*,\gamma_M^*\right)$; i.e., equal *margin* for both users and $0<\gamma_M^*<\min\left(\gamma_M^{1,\max},\gamma_M^{2,\max}\right)$. For any point between 'Optimum' $\mathbf{R}^*$ and B, $\gamma_M^1<\gamma_M^*$ and hence $\mathbf{R}^*$ is a better solution to the min-max in (P1). Similarly for any point between $\mathbf{R}^*$ and C, $\gamma_M^2<\gamma_M^*$. Thus $\mathbf{R}^*$ is the unique global solution to (P1) over $\Re_{ALG}$.

The argument is easily extended step by step for N>2. As an example see Figure 4-1(b) for a 3-user case. Here the argument would begin by the identification of 3 points that demarcate $\mathbf{M}_+$ instead of the 2 points B and C in the 2 user case and in the general N user case there would be N such points. Thus we have the following conclusion:

(C5) For every $\mathbf{T}\in\Re_{ALG}$, $\exists$ a unique point $\mathbf{R}^*\in\mathbf{M}_+\cap\overline{\Re_{ALG}}$ which is the solution to (P1) and further at this point $\gamma_M^1=\gamma_M^2=\ldots=\gamma_M^N=\gamma_M^*$.

In summary we have the following mapping $\mathbf{R}\xrightarrow{ALG}\boldsymbol{\varepsilon}\xrightarrow{(4.3)}\sigma_k^i\xrightarrow{T^i}\gamma_M^i$ from a rate point on the rate region to the *margin* for a given algorithm *ALG* and target rate T^i. The mapping from $\boldsymbol{\varepsilon}$ to γ_M^i is inherently distributed (as we shall shortly see). For *ALG* the best available distributed choice is IWF (see Table 2-2). As discussed in Section 2.3.2.1 from the uniqueness of the Nash equilibrium, IWF further has the convenient property that the total power tuple has a one to one correspondence with the rate tuple, i.e., $\mathbf{P}\xleftrightarrow{IWF}\mathbf{R}$. Thus by adjusting $\mathbf{P}$ alone we can span entire $\Re_{ALG}$ and control γ_M^i. The Step 4 in Table 2-2 is the one which depends directly on the choice of $\mathbf{P}$ and the water filling can be implemented using Table I in [38]. Step 5 ensures that at the end of the

IWF, each User i has its own SNR profile available, i.e., $\{\sigma_k^i\}_{k=1}^K$ which it would anyway need to calculate implicitly or explicitly for Step 6.

Table 4-1: SubAlgorithm NRME (Newton Raphson Margin Estimation)

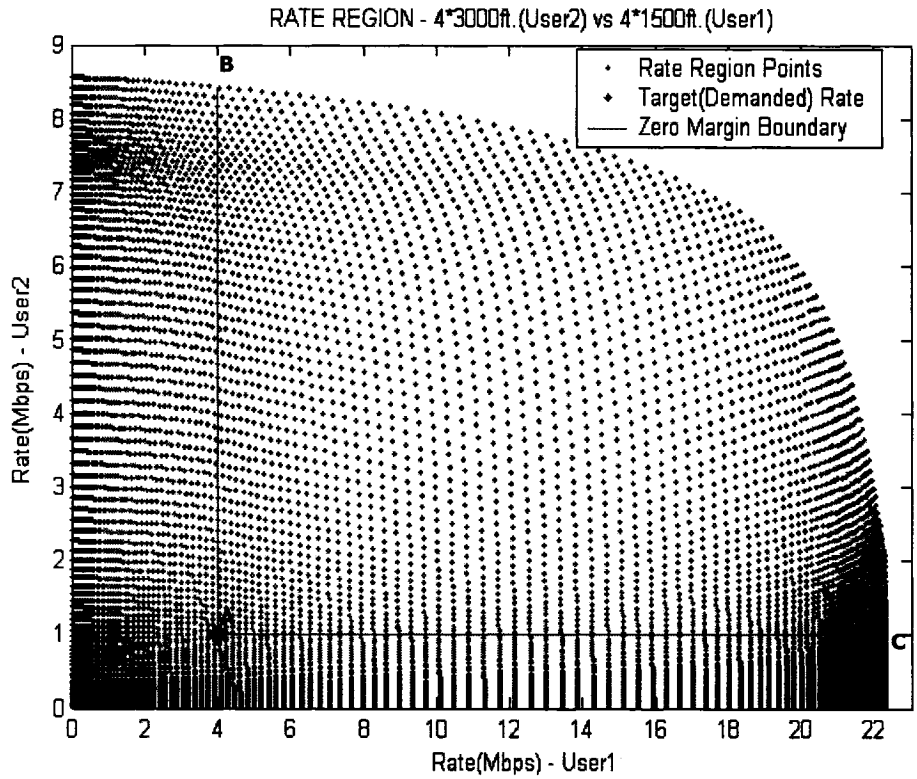
function $\gamma_M = NRME\left(\{\sigma_j\}_{j=1}^K; T\right)$

1. $R = \sum_{j=1}^K \log_2\left(1+\sigma_j/\Gamma\right)$
2. $\gamma_M = \left(2^{R/K}-1\right)/\left(2^{T/K}-1\right)$
3. *Repeat*
4. $\quad T_{new} = \sum_{j=1}^K \log_2\left(1+\sigma_j/\gamma_M\Gamma\right)$
5. $\quad \nabla = \dfrac{-1}{\gamma_M \ln 2}\sum_{j=1}^K \dfrac{1}{1+\gamma_M\Gamma/\sigma_j}$
6. $\quad M0 = \gamma_M$
7. $\quad \gamma_M = \gamma_M - \left(T_{new}-T\right)/\nabla$
8. $\quad$ *if* $\gamma_M < 0$
9. $\quad\quad \gamma_M = M0/10$
10. $\quad$ *End*
11. *until* $\left|T_{new}-T\right| > eps_{NRME}$

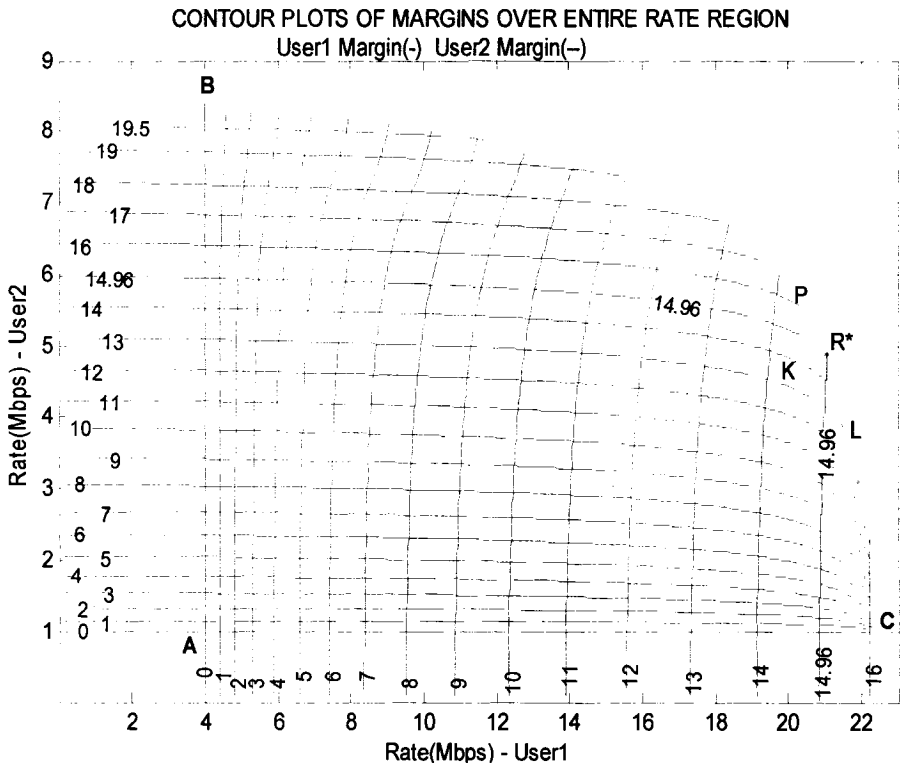
The approximation in (4.6) follows from the representation of the entire SNR profile $\{\sigma_j^i\}_{j=1}^K$ by its geometric mean. This single parameter representation is seen to be fairly accurate when we have a nearly continuous SNR profile, e.g., in ADSL. But in VDSL, since upstream bands sandwich a downstream band in between and vice-versa, there is a sharp discontinuity in the SNR profile. The geometric mean being a single parameter cannot reflect properties of both bands. In such a situation, the only alternative is an iterative approach to solve (4.5). Since (4.5) has continuous first derivative and a unique solution; and a reasonable good starting point is available in (4.6), we propose the usage of Newton Raphson method which is one of the fastest methods to solve nonlinear equations. This is presented in SubAlgorithm NRME in Table 4-1. Note that in the algorithm all user-related superscripts are dropped because, once the SNR profile is available, the *margin* calculation does not need to distinguish between users. The ∇ in step 5 denotes the differentiation of (4.5) with respect to γ_M. Step 7 is the Newton Raphson update step.

Steps 8-10 are required while measuring very small γ_M (close to zero), to prevent an ambitious Newton Raphson step from pushing the *margin* to negative (not in dB), which is unacceptable. The algorithm was tested in many different scenarios and reached the solution in less than 4 iterations in most cases for $eps_{NRME} = 0.1$ bits.

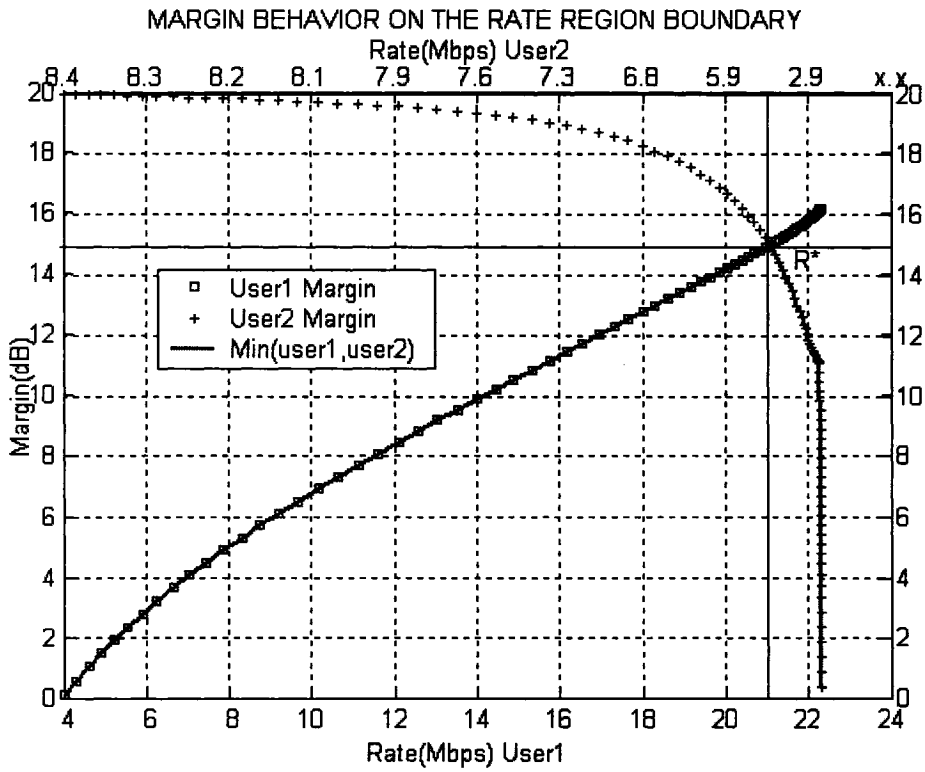
It is now our objective to verify (C4) and (C5). Reconsider (from Section 3.5) the scenario of 8 VDSL 26 AWG lines (similar to [25]), 4 of which are at a distance of 1500ft. (collectively referred to as User 1) from the Central Office and the other 4 at a distance 3000ft. (collectively referred to as User 2) as shown in Figure 3-9. Now we conduct an extensive experiment by analyzing every power combination of User 1 and User 2 from -60dBm to 11.5dBm at a step size of 0.5dBm; where P'_{max} =11.5dBm [3] in the upstream direction. Crosstalk noise model A is assumed and the FEXT transfer functions and upstream frequency band specifications are chosen as specified in [3].



(a) Rate Region



(b) Margin Contours



(c) Min-Max Margin Behavior

Figure 4-2: Behavior of Margin on Rate Region

With each power combination we execute SubAlgorithm IWF in Table 2-2 and each resulting unique rate pair is plotted as dots in Figure 4-2(a) and thus span entire $\Re_{ALG}$. The target (demanded) rate vector is 4 Mbps for each of the 1500ft. lines and 1Mbps for each of the 3000ft. lines. This point is shown in Figure 4-2(a) and corresponds to point A in Figure 4-1(a) and the straight lines correspond to lines AB and AC in Figure 4-1(a). Now for each of these points in $\Re_{ALG}$, we evaluate the γ_M^i for each user by using SubAlgorithm NRME of Table 4-1. The constant *margin* contours over $\Re_{ALG}$ for both users are plotted in Figure 4-2(b). We plot only the positive *margin* (in dB) contours. The nearly horizontal contours represent the contours for User 2 and the nearly vertical ones are for User 1. The contours are labeled with the *margin* value in dB that they represent. We can see that the 0 dB *margin* contours are perfect straight lines following from (C1) and the zero *margin* boundaries predicted in Figure 4-2(a). The region where the horizontal and vertical contours intersect represent $\mathbf{M}_+$.We see that the contour values of User 2 keep increasing with R^2 and the same effect is observed for User 1. Thus (C4) holds. For verifying (C5), we highlight the contour representing 14.96dB for User 1 and User 2. As we can see, both these contours intersect on $\overline{\Re_{ALG}}$ at R*. For any point like P above this point it can be seen that (vertical contour) γ_M^1 <14.96dB, any point below like L γ_M^2 <14.96dB, and any point to the interior like K both γ_M^1 and γ_M^2 <14.96dB. The behavior of γ_M^1 and γ_M^2 is very important particularly on $\mathbf{M}_+ \cap \overline{\Re_{ALG}}$. This is plotted in Figure 4-2(c). The lower x-axis denotes the R^1 set for which γ_M^1 and γ_M^2 are positive (in dB) and the upper x-axis denotes the corresponding R^2 on the rate-region boundary, i.e., $\overline{\Re_{ALG}}$. γ_M^1, γ_M^2 and $\min(\gamma_M^1, \gamma_M^2)$ are plotted. Recall that $\min(\gamma_M^1, \gamma_M^2)$ is the function we want to maximize in (P1). We can clearly see the display of (C4), in that γ_M^1 increases with R^1 (lower x-axis) and γ_M^2 increases with R^2 increases with R^2 (upper x-axis). Finally, as noticed in Figure 4-2(b), $\min(\gamma_M^1, \gamma_M^2)$ has a unique maximum at R* at which $\gamma_M^1 = \gamma_M^2$ =14.96 dB. Thus (C5) is verified. Also in other examples presented in Section

4.4 and many others that we tested (C4) and (C5) are always seen to hold for DSL channels.

4.3 Enlightened Margin Maximization

Given SubAlgorithm IWF and SubAlgorithm NRME, we now have the following mapping from the design variable **P** to the objective function:

$$\mathbf{P}(\mathbf{R}) \xrightarrow{IWF} \boldsymbol{\varepsilon} \xrightarrow{(4.3)} \sigma_k^i \xrightarrow{NRME} \gamma_M^i \tag{4.7}$$

P and **R** have one to one mapping in IWF; however **P** is naturally easier to control. In the last step of the mapping the usage of SubAlgorithm NRME causes a simplification of (P1). The termination condition in Step 11 of Table 4-1 SubAlgorithm NRME ensures that the target rate constraints are met; thus we can eliminate this constraint from (P1) as long as we use *NRME* to evaluate γ_M^i. This is similar in optimization to the substitution of one of the constraints within the objective function, e.g., consider the 2 equivalent problems:

$$\min_{\substack{x+y=1 \\ x\geq 0}} \left(x^2+y^2\right) \equiv \min_{x\geq 0} \left(x^2+(1-x)^2\right)$$

With the removal of the target rate constraint we are left with the constraint $\mathbf{0} \leq \mathbf{P} \leq \mathbf{P}_{\max}$. These forms of constraints are much simpler than general nonlinear constraints and are referred to as 'bound constraints' or 'box constraints'. Significant research has been done for nonlinear minimization with box constraints [39][40]. The min-max problem (P1), while being the most natural modeling of the situation, cannot directly apply these methods. Further, the trust region methods in [39][40] require the Jacobian and the Hessian. In our case the *margin* is a very complicated multi-stage mapping from **P** for which even the Jacobian, let alone the Hessian, is impossible to obtain in closed form. Nonlinear Least Squares (NLSQ) problems present a convenient class of problems whereby dependence on Hessian of component functions is minimized under certain conditions. Further, the techniques in [39][40] have been implemented with good results for NLSQ with bound constraint; and (C4) and (C5) help us reformulate our problem into such a problem.

4.3.1 Bound Constrained NLSQ Formulation

We know from (C5) that at the optimal point $\gamma_M^1 = \gamma_M^2 = ... = \gamma_M^N = \gamma_M^*$. In other words the variance of the vector of *margin*s will be zero. Further, from (C5), this optimal point lies on $\mathbf{M}_+ \cap \overline{\Re_{ALG}}$. So we would want the iterates to be always on $\overline{\Re_{ALG}}$. This can be ensured by maximizing the mean of the vector from (C4) because the mean has continuous growth radially towards $\overline{\Re_{ALG}}$. Thus we have the following NLSQ formulation:

$$\min_{\mathbf{0} \le \mathbf{P} \le \mathbf{P}_{\max}} F(\mathbf{P}) \qquad \text{(P2.a)}$$

$$F(\mathbf{P}) = \tfrac{1}{2}\sum_{i=1}^{N}\left(f^i(P)\right)^2 = \tfrac{1}{2}\mathbf{f}(\mathbf{P})^T\mathbf{f}(\mathbf{P}) \qquad \text{(P2.b)}$$

$$\text{where } f^i(\mathbf{P}) = \left(\gamma_M^i \Big/ \tfrac{1}{N}\sum_{j=1}^{N}\gamma_M^j - 1\right) \qquad \text{(P2.c)}$$

In the above formulation $F(\mathbf{P})$ is the variance of the *margin* vector divided by the square of its mean. γ_M^i in (P2.c) is a function of $\mathbf{P}$ through (4.7). Clearly at the optimal point $\mathbf{R}^*$, by (C5), $F(\mathbf{P})$=0. This is sometimes called the residual of the NLSQ problem, which is zero in our case. The Jacobian $\mathbf{J}(\mathbf{P})$ is an $N \times N$ matrix with the (i,j) component $\left(\mathbf{J}(\mathbf{P})\right)_{ij} = \partial f^i / \partial P^j$. From this the gradient and the Hessian of $F(.)$, i.e., $\mathbf{F}'(\mathbf{P})$ and $\mathbf{F}''(\mathbf{P})$ are calculated as:

$$\mathbf{F}'(\mathbf{P}) = \mathbf{J}(\mathbf{P})^T\mathbf{f}(\mathbf{P}) \qquad (4.8)$$

$$\mathbf{F}''(\mathbf{P}) = \mathbf{J}(\mathbf{P})^T\mathbf{J}(\mathbf{P}) + \sum_{i=1}^{N} f^i(\mathbf{P})\mathbf{f}^{i''}(\mathbf{P}) \qquad (4.9)$$

where $\mathbf{f}^{i''}(\mathbf{P})$ is the Hessian of the component function $f^i(\mathbf{P})$. Due to our problem being a zero-residual problem, the $f^i(\mathbf{P})$ in the summation in (4.9) will be very small near $\mathbf{R}^*$ and hence $\mathbf{F}''(\mathbf{P})$ can be approximated by $\mathbf{J}(\mathbf{P})^T\mathbf{J}(\mathbf{P})$. The functional dependence on $\mathbf{P}$ is dropped henceforth in notation whenever obvious from context. Thus, we need to only form an estimate of $\mathbf{J}(\mathbf{P})$ or $\mathbf{J}$. This could be done using finite-differencing. But finite-differencing would increase the number of objective function evaluations by N at each iteration. Since in our case each objective function evaluation in (P2.a) will require an

IWF operation due to (4.7), doing this at each iteration is unaffordable. Instead we use a secant method known as Broyden's rank one update, which is stated below:

$$\mathbf{J} = \mathbf{J}_{prev} + \mathbf{u}\mathbf{h}^T \tag{4.10}$$

where $\mathbf{h} = \mathbf{x} - \mathbf{x}_{prev}, \mathbf{u} = \frac{1}{\mathbf{h}^T\mathbf{h}}\left(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_{prev}) - \mathbf{J}\mathbf{h}\right)$.

This leaves the problem of getting an initial estimate of the Jacobian. For the first estimate we could use finite-differencing once, but from (C4) and (C5) a good initial *guess* is also possible, which we discuss little later. The above NLSQ problem with bounds is a special case of algorithms in [39][40] and is also solved in many optimization packages including lsqnonlin in MATLAB Optimization Toolbox, which uses [40] discussed below.

4.3.2 Trust Region Reflective Newton Methods

In any unconstrained trust region method we try to define a simple quadratic *model* and define the *neighborhood* where this model can be *trusted*; i.e., portray the more complicated nonlinear objective function reasonably well. Thus the 4 key requirements are the (1) gradient and (2) the Hessian, which are used to define the *model* and (3) the scaling matrix and (4) the trust region size, which define the *neighborhood.* At each iteration solving the above problem yields a step update and then both the *model* and the *neighborhood* are updated. For a constrained trust region method, at each iteration, a constrained version of the above problem would have to be solved. One of the primary contributions of [39] was being able to define a special scaling matrix for the bound constrained case for which, at each iteration, the quadratic problem to be solved was unconstrained. We define this for our problem as follows:

$$\mathbf{D}(\mathbf{P}) = diag\left(\left|\mathbf{v}(\mathbf{P})\right|^{-\frac{1}{2}}\right) \tag{4.11}$$

where $\mathbf{v}(\mathbf{P}) = \left(v^1(\mathbf{P}), v^2(\mathbf{P}), .., v^N(\mathbf{P})\right)$ and $v^i(\mathbf{P}) = \begin{cases} P^i - P^i_{\max}, & \left(\mathbf{J}^{\mathrm{T}}\mathbf{f}\right)_i < 0 \\ P^i, & \left(\mathbf{J}^{\mathrm{T}}\mathbf{f}\right)_i \geq 0 \end{cases}$ using (4.8)

and the bounds in (P2.a).

At any iteration let the power vector used be $\mathbf{P}$ and let $\mathbf{D} = \mathbf{D}(\mathbf{P}), v^i = v^i(\mathbf{P})$ and so on. Thus at any iteration the quadratic subproblem to be solved is:

$$\min\left\{\psi(s):\|\mathbf{D}s\|_2 \le \Delta\right\} \tag{4.12}$$

where $\psi(s)=s^T\mathbf{J}^T\mathbf{f}+\frac{1}{2}s^T\mathbf{\Pi}s$, $\mathbf{C}=\mathbf{D}diag\left(\mathbf{J}^T\mathbf{f}\right)\mathbf{J}^v\mathbf{D}$, $\mathbf{\Pi}=\mathbf{J}^T\mathbf{J}+\mathbf{C}$ and $\mathbf{J}^v$ is the Jacobian of $\mathbf{v}(.)$. At each iteration all the above quantities are updated. In [40], the authors primarily focus on solving (4.12). It is proposed to choose among the 'best of 3' possible solutions: - (i) original solution to (4.12); (ii) scaled gradient solution (both (i) and (ii) is truncated when necessary to remain strictly feasible) (iii) reflected solution, where the solution component, which violates the bound, is reversed in bound. Both (i) and (iii), being complicated, the authors proposed subspace based preconditioned conjugate gradient techniques. The 'best of 3' approach is currently implemented in MATLAB Optimization Toolbox. While in the option (ii), the scaled gradient i.e., $-\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}$ converges to zero and thus first order optimality; it is observed in [40] that predominantly (iii) and (i) lead to much faster convergence for large scale problems. However, we choose to use only the Scaled Gradient (SG) method rather than the 'best of 3' approach. Firstly we must notice that the SG method is far simpler (in terms of complexity and computation) and guarantees convergence to first order optimality (see p. 8 in [40]). The simplicity of the SG method arises from the fact that the *step direction* is already determined to be $-\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}$. The only operation to be performed is determining the optimum *step size* (along the above direction), which is a one dimensional problem. The operations involved are summarized in SubAlgorithm SGTRNLSQ in Table 4-2.

Table 4-2: SubAlgorithm SGTRNLSQ: SG Trust Region method for NLSQ

$\mathbf{h}=SGTRNLSQ(\mathbf{P},\mathbf{f},\mathbf{J})$	
1.	$\tau^* = \arg\min_\tau \left\{\psi\left(-\tau\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}\right):\left\|\tau\mathbf{D}^{-1}\mathbf{J}^T\mathbf{f}\right\| \le \Delta\right\}$
2.	$\alpha = \min\left\{\max\left\{\frac{P^i}{\left(\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}\right)_i}, \frac{P^i - P^i_{\max}}{\left(\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}\right)_i}\right\}:1\le i\le N\right\}$
3.	$\mathbf{h}=-\min\left(\tau^*,\alpha\right)\cdot\mathbf{D}^{-2}\mathbf{J}^T\mathbf{f}$

The problem in step 1, with a few operations, is easily seen to be a one-dimensional problem of the form $\arg\min_x\left\{a_1x+\frac{1}{2}a_2x^2 : -\kappa \le x \le \kappa\right\}$, where each of the quantities is a

scalar. As is well known, the solution is $-a_1/a_2$ if $|a_1/a_2| \leq \kappa$ else it is either $-\kappa$ or κ. Thus, τ^* in step 1 is obtained in less than 3 evaluations of a 1-D quadratic. The subsequent steps 2 and 3 ensure that the step results in a strictly feasible **P+h**, i.e., $\mathbf{0} < \mathbf{P} + \mathbf{h} < \mathbf{P}_{max}$. As we shall see in section 4.4, the choice of SG method will cost us between 5-10 iterations only. In most cases the 'best of 3 approach' gives the same number of iterations and only rarely better by 1-2 iterations. This can be because of the cumulative of 3 reasons.

- The number of variables for which [40] showed the great improvements of (i) and (iii) was in hundreds and thousands, in DSL N is much smaller.
- The 'best of 3' approach can have faster convergence with exact Jacobian availability, while in our problem, we can only have approximate Jacobian.
- The accuracy or tolerance for the stopping condition in [39][40] is 10^{-12}. For DSL application the maximum accuracy in *margin* is 10^{-2} [31] or at best 10^{-4} [36].

Thus while the 'best of 3' approach is very good for many problems, for our problem the usage of SG method more than satisfies our needs and the minor improvement on rare occasions does not justify the much higher computational requirement of the 'best of 3' approach at each iteration. Thus the primary utility of the trust region method for our problem is the scaling matrix **D**, which ensures appropriate behavior of SG around "correct" and "incorrect" bounds [40]. Also at each step of solving for τ^* we deal with a simple 1-D quadratic rather than having to evaluate γ_M^i after using SubAlgorithm IWF and SubAlgorithm NRME.

4.3.3 Implementation in Current DSL Scenario

Now we have all the tools required and the complete algorithm is presented in algorithm EMM in Table 4-3. The shaded regions denote the operations taking place at the *center* which could be the central office in VDSL upstream, or a spectrum managernent center (SMC) or the DSLAM depending on the nature of deployment, see [23]. The unshaded regions denote the autonomous operation of the N modems. As is easily seen, the only exchange between the center and the modems is of macro-parameters such as γ_M^i

and P^i. The assignment of micro-parameters such as PSD and bit allocation over subcarriers is done autonomously and hence the level of coordination [24] is only Level 1.

Table 4-3: Algorithm EMM: Enlightened Margin Maximization

0. $\mathbf{P} = \mathbf{P}_{\max}; \mathbf{J} = \mathbf{J}_0$.*Iter* = 0.
1. repeat
1. Read P^i from center
2. Run SubAlgorithm IWF
3. for $i = 1$ to N
4. Report $\gamma_M^i = NRME\left(\left\{\sigma_j^i\right\}_{j=1}^K; T^i\right)$ to center.
5. end
7. $\mu = \frac{1}{N}\sum_{i=1}^N \gamma_M^i$
8. Evaluate **f**, where $f^i = \gamma_M^i / \mu - 1$
9. if *Iter*>0
10. $\mathbf{u} = \frac{1}{\mathbf{h}^T\mathbf{h}}\left(\mathbf{f} - \mathbf{f}_{prev} - \mathbf{J}\mathbf{h}\right); \mathbf{J} = \mathbf{J} + \mathbf{u}\mathbf{h}^T$
11. end
12. $\mathbf{h} = SGTRNLSQ(\mathbf{P}, \mathbf{f}, \mathbf{J}); \mathbf{P} = \mathbf{P} + \mathbf{h}$
13. $\mathbf{f}_{prev} = \mathbf{f}; Iter = Iter + 1$
14. until $\{\mathbf{f}^T\mathbf{f} > eps_{EMM}$ and $\|\mathbf{h}\|_2 > eps_{EMM}\}$

All modems initially start at their maximum power constraint $\mathbf{P}_{\max}$ as in [25]. This is the most natural starting point. Further, in our case we know that the optimal solution is on $\overline{\Re_{IWF}}$, so it is suitable to start with $\mathbf{P}_{\max}$, which is guaranteed to be a point on $\overline{\Re_{IWF}}$. The estimation of the initial Jacobian $\mathbf{J}_0$ is a non-trivial task. Finite-differencing could be used as a first step, which would involve N iterations of EMM to estimate $\mathbf{J}_0$. But with experience, a reasonable 'guess' is also possible. For all our results in Section 4.4, we chose $\mathbf{J}_0 = 11 \times \mathbf{I}_N - \mathbf{1}_N$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{1}_N$ is the $N \times N$ matrix

of all ones. This results in a $\mathbf{J}_0$ where all diagonal elements are 10 and all off-diagonal terms are -1. This choice is (loosely) explained by the fact that each user's *margin* is increasing in its own power and decreasing in other users' power (hence -1). Nothing is special about the number 10; it is chosen for diagonal dominance because the DSL channel matrix by itself is diagonal dominant (recall from Figure 2-1 that $H^{11}(f) > H^{12}(f)$ and $H^{22}(f) > H^{21}(f)$), whereby P^i has more effect on γ_M^i than P^j. Though heuristic, this choice (motivated primarily by numerical experiments) gives good results in all tested cases (Section 4.4) and saves the N iterations that would be required for finite-differencing.

At each *Iter* (see Table 4-3) each user reads P^i to be used from the center and then performs IWF after which ε_j^i and σ_j^i are known to each User i for all its subcarriers. Obviously each user knows its T^i. From this, using *NRME*, γ_M^i for each user is determined autonomously and reported to the center. Once the center receives the γ_M^i from all N users, it computes the NLSQ component functions (P2.c) and updates the Jacobian using the Broyden update (4.10). Following this, an appropriate scaled gradient step for $\mathbf{P}$ is calculated. This new set of P^i is read by the users and IWF is conducted and so on. The stopping condition is determined by the value $\mathbf{f}^T\mathbf{f}$ (the residual), which we know must go to zero at $\mathbf{R}^*$ by (C5) and section 4.3.1 and also the Euclidean norm of the step $\mathbf{h}$. At the end of the EMM, each users' *margin* is expected to reach an equal γ_M^*. At this point the PSD allocation ε_j^i is decided by the last *Iter*'s IWF process and the bit loaded b_j^i is determined during the NRME process to be $\log_2\left(1+\sigma_j^i/\gamma_M^*\Gamma\right)$. At each iteration (and hence at the last) the usage of NRME guarantees that T^i of each user is met as explained in Section 4.3. Thus the *margin* maximization does not come at a cost of any of the users not meeting their T^i and the common equal *margin* γ_M^* ensures no user looses *margin* due to another.

4.4 Performance and Convergence

We consider the 4×1500ft. (User1) and 4×3000ft. (User2) upstream scenario described in Section 4.2 with $P^i_{\max}$=11.5dBm. We consider 6 target tuples fairly well dispersed over $\Re_{IWF}$ as shown in Table 4-4 (shaded columns) and Figure 4-3. For each of these target rates we run the EMM algorithm and the results are reported in Table 4-4. The maximized common equal *margin* γ^*_M is reported in dB. The number of iterations taken corresponding to *Iter* in EMM algorithm is reported next. $\left(\mathbf{f}^T\mathbf{f}\right)^*$ denotes the value of the residual of the NLSQ function in (P2.a) at the end of EMM. This is the stopping condition in EMM and used to test if γ^*_M is truly the *common equal margin* among all users. Next the vector of total power that achieves the optimal point $\mathbf{P}^*$ is reported in dBm. Lastly $\mathbf{R}^*$ denotes the actual location of the optimal point on the rate region. As denoted by (4.7), the energy allocation used $\boldsymbol{\varepsilon}$, corresponds to this point. Also $\mathbf{R}^*$ verifies that the optimal point lies on $\overline{\Re_{IWF}}$.

The (4,1) Mbps target rate example was discussed in Section 4.2 and we had noticed that a unique optimum existed in Figure 4-2(b), where the 14.96 dB *margin* contours of both users intersected on $\overline{\Re_{IWF}}$. Table 4-4 confirms that EMM actually reached this optimum point with $\gamma^*_M = 14.96$ in 7 iterations. The $\mathbf{R}^*$ of (21.07,4.87) corresponds to the actual location of the optimal point on $\overline{\Re_{IWF}}$ as seen in Figure 4-2(b) as well.

Table 4-4: Performance and Convergence of EMM

T (Mbps)		γ^*_M (dB)	*Iter*	$\left(\mathbf{f}^T\mathbf{f}\right)^*$	$\mathbf{P}^*$ (dBm)		$\mathbf{R}^*$ (Mbps)	
User1	User2				User1	User2	User1	User2
4	1	14.96	7	1.0486e-6	-0.077	11.500	21.07	4.87
8	6	3.79	10	2.8807e-7	-14.28	11.500	12.30	7.85
10	4	6.38	8	8.2651e-8	-6.918	11.499	18.37	6.71
14	3	5.04	6	2.6253e-7	0.026	11.500	21.09	4.84
20	5	0.54	6	6.3448e-6	-1.232	11.500	20.78	5.23
15	1	5.15	5	5.6057e-6	11.500	2.0403	22.30	1.21

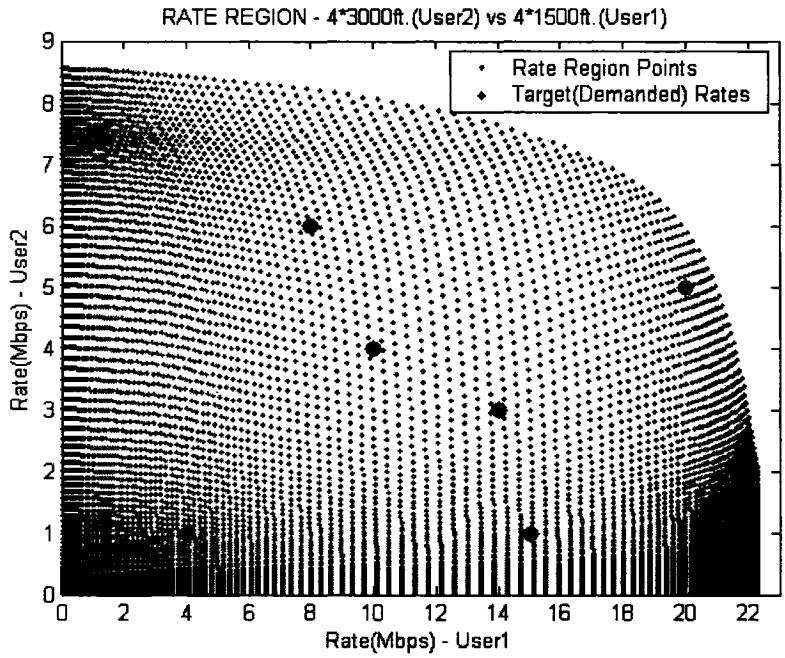
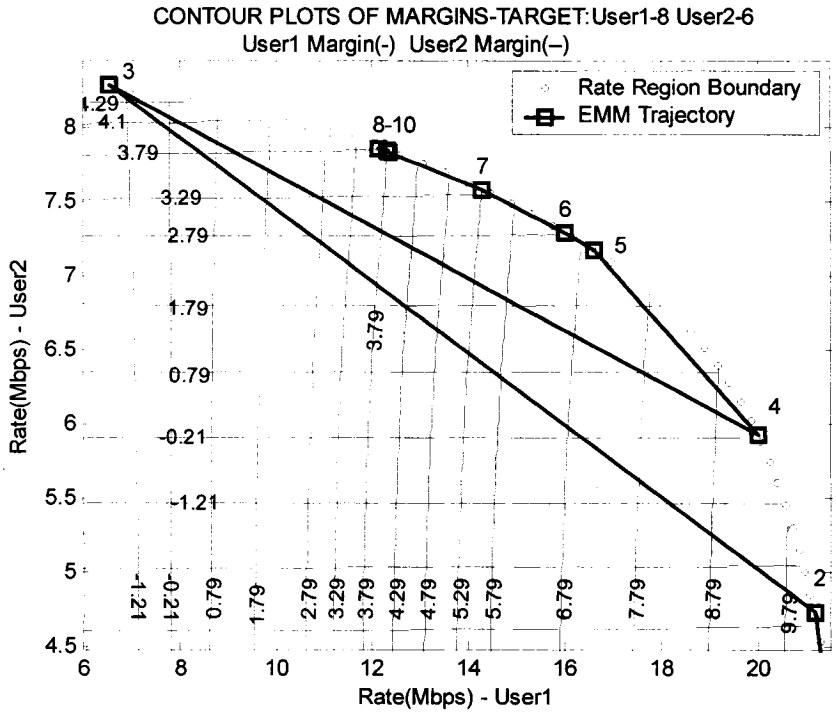
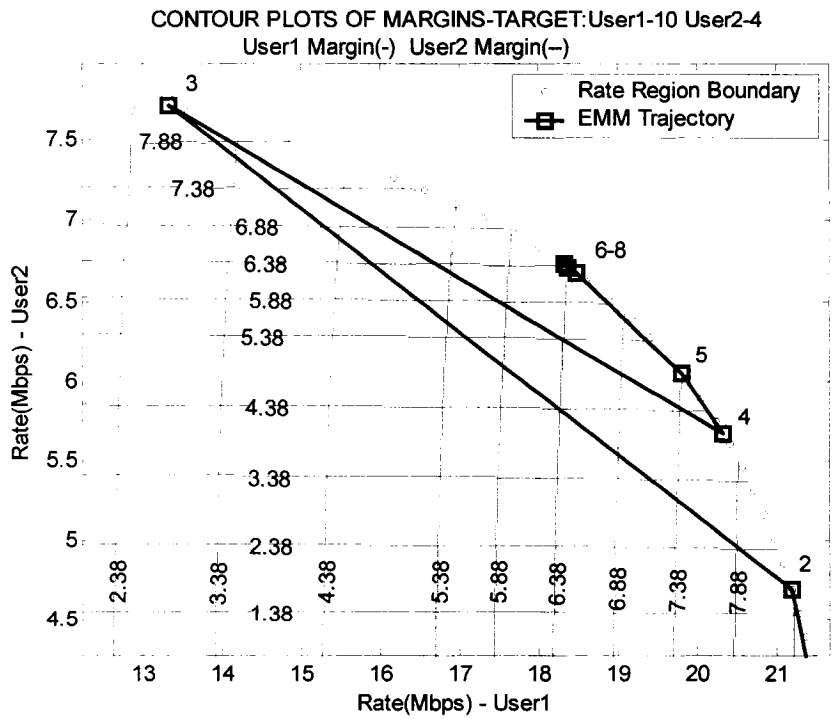


Figure 4-3: Spread of Target Rate points on the Rate Region

For the next two target rate pairs i.e., (8,6) and (10,4) we display the contour plots of both users' *margin*s similar to Figure 4-2(b) in Figure 4-4(a) and Fig. Figure 4-4(b) respectively. A survey of the contour values shows that there is an intersection of 3.79dB *margin* contour lines for both users in Figure 4-4(a) on $\overline{\Re_{IWF}}$ and the same is observed for 6.38dB *margin* contours in Figure 4-4(b). A quick survey of the contours in the neighborhood of these intersection points in both figures shows that, for each feasible point around this intersection at least one of the *margin*s drops and hence cannot be the solution to (P1), thus establishing the unique optimality of the intersection of equal *margin* on $\overline{\Re_{IWF}}$, as expected from (C5). The trajectory of EMM in both figures shows how EMM approaches and eventually converges to this optimal solution, which is also confirmed by Table 4-4 with γ_M^* =3.79 and 6.38 dB respectively. One more important observation from Figure 4-4 is that not only is the final solution on $\overline{\Re_{IWF}}$, but also each iteration in the trajectory of EMM corresponds to a point on $\overline{\Re_{IWF}}$. Recollect that this is what we had aimed for in the NLSQ reformulation in (P2), because from (C5) we knew that the solution $\mathbf{R}^*$ lies on the $\overline{\Re_{IWF}}$ and hence restricted the search to only that set.



(a) for target pair (8Mbps, 6Mbps)



(b) for target pair (10Mbps, 4Mbps)

Figure 4-4: Margin Contours and EMM Trajectory

For each of the target rates in Table 4-4 we can see that the EMM algorithm converges in 5-10 iterations. A study of the $\mathbf{P}^*$ column reveals that for each target rate pair the EMM algorithm converges to $\mathbf{P}^*$ with at least one of the two users using their maximum power constraint of 11.5dBm, which itself guarantees that the solution point is on $\overline{\Re_{IWF}}$ [25]. The positive value of γ_M^* verifies that the optimal point lies in $\mathbf{M}_+ \cap \overline{\Re_{ALG}}$ and lastly, the low value of $\left(\mathbf{f}^T\mathbf{f}\right)^*$ in each case below eps_{EMM} =10^{-5} ensures that this point indeed provides a common equal *margin* to all users. Thus all features of (C5) are satisfied. We further observe that, loosely speaking, the value of optimum γ_M^* depends on how far the target rate point **T** is from the $\overline{\Re_{IWF}}$ or equivalently, the area of $\mathbf{M}_+$. We can see in the Figure 4-3 that the (4,1) target point is the inner most among **T** considered and has a γ_M^* =14.96dB. On the other hand, for the (20,5) target point, which is very close to the $\overline{\Re_{IWF}}$, the corresponding area of $\mathbf{M}_+$ is minimum and γ_M^* =0.54. A similar ordering of γ_M^* for intermediate points can be explained based on corresponding area of $\mathbf{M}_+$.

However, irrespective of how close or how far **T** is from $\overline{\Re_{IWF}}$, the EMM algorithm is seen to behave well and reach the unique optimum, which ensures the best equal *margin* for all users, while guaranteeing each users' target rate requirement due to NRME.

4.5 Practical viability of EMM

In the multiuser *margin* maximization problem considered in this chapter, the multiuser scenario directly leads to a constrained min-max optimization problem. However, in general, algorithms which are available for constrained min-max optimization are more complicated than those for constrained nonlinear minimization problems[33]. Among general constrained nonlinear minimization problems, algorithms for constrained NLSQ minimization problems converge faster and require less computation because the Hessian

[33] This is because in min-max, the objective we are minimizing is already a max of more than one component function. This makes the objective non-differentiable at some points (even if the individual component functions are differentiable). Hence the objective function is more difficult to minimize than in the case of simple minimization where the objective function is differentiable.

does not need explicit calculation (see Section 4.3.1). Finally, among general constrained NLSQ problems, for the ones with bound or box constraints, efficient (in terms of number of iterations to converge to solution) algorithms are available in literature [39][40]. In this chapter we discovered properties (C4) and (C5) of *margin* which helped us simplify the original min-max problem to a NLSQ minimization with bound constraints. Further, we used a simplified version of the algorithm in [40] by reducing the trust region problem to a 1-D step size problem (as explained in Section 4.3.2) and thereby avoiding the computationally expensive sections of the algorithm in [40]. Hence we believe the EMM algorithm is an economical solution to the multiuser *margin* maximization problem.

A rigorous characterization of the computational complexity of the EMM algorithm will be an interesting task for future work. At this point, loosely speaking, we can say that among the various components of the EMM algorithm, SubAlgorithm IWF is the most computationally expensive. This is because, in a single execution of SubAlgorithm IWF, on an average $3N$ to $4N$ water-filling operations are performed [25]. For each of these water-filling operations the complexity will be similar to that of the algorithm presented in Table 3-3. This is clearly much more taxing computationally than the SubAlgorithm NRME and Broyden Jacobian update. Thus the number of executions of SubAlgorithm IWF is a suitable indicator of overall algorithm complexity. The EMM algorithm only required 5-10 executions of SubAlgorithm IWF. But some implementations of DSM power control, e.g., with the outer power loop in [25] can use up to 20-30 executions of SubAlgorithm IWF. Since at least 4 companies have produced implementations [29] of the power minimization algorithm in [25], we believe that the EMM algorithm with only 5-10 executions of SubAlgorithm IWF will definitely be practically acceptable.

In the discussed multiuser DSL scenario, what is more critical than low computational complexity for the practical viability of a multiuser algorithm is whether or not it renders itself in a predominantly *distributed* implementation (Level 1 coordination in DSM terminology). This is because, multiple competing service providers share the same bundle currently (see Figure 1-1). As we have demonstrated throughout this chapter, the EMM algorithm can be implemented with only Level 1 coordination. Further, since at the end of EMM algorithm, each user will have equal *margin*, it will be acceptable to all ser-

vice providers, who share the bundle, as a fair scheme. Hence the EMM algorithm is a practically viable method for multiuser *margin* maximization.

Chapter 5

Conclusions

In this work, we have highlighted that the granularity loss arising out of the integer-bit restriction is non-negligible. Particularly with the peak power constraint, these losses turn out to be a significant percentage of the supported data rate. We developed a fine-granularity BER-based loading scheme called ARSFGL to recover these losses. This is done by jointly optimizing the coding rate of a programmable RS (n,k) code and the bit and energy allocation on each sub-carrier. Illustrative examples of applications to VDSL-DMT systems in (currently standardized) SSM scenarios indicate that the proposed scheme outperforms various existing integer-bit loading algorithms with an increase in rate of about 20% in most cases. This is a large rate increase as compared to the variation in achievable rates of less than 4% between various existing integer-bit loading algorithms. This improvement is in a good agreement with the theoretical estimates developed to quantify the granularity loss. The theoretical estimates also present an insight into how the granularity losses increase with rising strictness in peak-power constraint, in comparison to the total-power constraint and with the number of subcarriers in use.

We then developed extensions of the ARSFGL scheme for the DSM case (not yet standardized, but very popular in research and industry). This technique, combined with IWF results in a purely distributed scheme which can provide data rates as high as the centralized and optimal integer bit scheme. Further the convergence of IWF is faster when used with this technique than with integer bit loading.

Through our review in Chapter 2, we noticed that although many single user *margin* maximization algorithms exist, they all rely on a convex reformulation of the problem. This convex reformulation does not hold in the multiuser case and hence any extensions of these single user algorithms to multiuser scenarios will lead to one user's *margin* maximization leading to other users' failure in meeting their rate or *margin* requirement. This motivated us to devote an independent study into the multiuser *margin* maximization theory which has not been developed in the literature yet.

We studied the problem of multiuser *margin* maximization in the min-max sense. We discovered and demonstrated certain monotonicity and fairness properties that *margin* exhibits over the rate region. Based on these properties we remodeled the problem as a NLSQ problem. This problem is efficiently solved using a scaled gradient trust region method and Broyden Jacobian update. For estimation of *margin* and subcarrier bit allocation, at each iteration, a Newton Raphson method is used. For subcarrier power allocation, the game-theoretic IWF technique is used. The combination of these leads to the Enlightened Margin Maximization (EMM) algorithm. Due to the near-continuous performance of the ARSFGL scheme developed in Chapter 3, we are able to assume continuous bit loading in development of the EMM algorithm, which however can be extended to purely integer bit loading as well. This algorithm reaches the best *common equal margin* for all users. This *common equality* does not follow from a compromise by any user; rather is an inherent fairness property that *margin* possesses. Further by using the Newton Raphson *margin* estimation as a SubAlgorithm, EMM explicitly guarantees that all users meet their target rate requirements. The method is verified to converge to unique min-max optimal point in 5-10 iterations for a variety of cases. Further the method can be implemented in current DSL-DSM scenario with only Level 1 coordination. Level 1 coordination, which allows for a distributed implementation, is desirable because in current DSL scenarios multiple service providers share the same cable and coordination higher Level 1 is unrealistic to expect among competing service providers.

Suggested Future Research Topics

In this thesis, one layer of adaptive Reed-Solomon coding was used to recover the granularity losses. But still, there is a gap between the smoother rate-SNR curves we achieved from the Shannon capacity curve (see Figure 3-3). In recent years, the turbo code concept, initially proposed for recursive convolutional codes, has been extended to block codes and these are known as turbo product codes. In our case as well an additional RS code across the frequency can be employed in conjunction with the proposed ARSFGL scheme. Turbo product codes have been devised specifically for RS codes as well, but in existing designs the 2 RS codes that are used in product along the 2 dimensions are fixed rate. The problem here however is that in our case one of the dimensions

of RS code will be variable rate. Designing a good turbo product code for such a scenario while keeping complexity practical will be challenging. But it will undoubtedly have the potential of reaching closer to the capacity of a frequency selective channel (not just an AWGN channel like existing turbo codes). This will be an interesting area to look into from the perspective of transmission.

From the perspective of resource allocation, we saw that for the multiuser case, a competitive game theoretic approach for power allocation (and interference balancing) has provided encouraging results. However, the rate region is still inferior to the optimal multiuser rate region. We believe this sub-optimality can be combated using two approaches, while still retaining the merits of a game-theoretic approach (distributed and simpler implementation). The cause of suboptimality is that, given that the Nash equilibrium point is unique for a given set of total power constraints, the only way the rate combinations are varied is by varying the total power constraints of users. We have observed that this form of generating multiple rate combinations (i.e., multiple Nash equilibria) often results in significant portions of unused resources (power and bandwidth). This can be prevented by grouping users, and conducting games among these groups in sequenced stages to reuse these resources. Initial results of this approach have been very encouraging and are reported in [42]. The second approach is, instead of being purely competitive in the games, users can be slightly compassionate to other users. Competitive games imply that each user makes the best possible move (i.e., power allocation) in response to other users' moves (i.e., their power allocations which is seen as interference). We can add compassion by each user knowingly performing a slightly less than its best possible move in response to other users. When this is done in a controlled and intelligent manner, it prevents deadlocks in the game, just like being open-minded prevents deadlocks among groups of people. This approach has shown the potential of better utilization of resources and hence better rates.

Appendix A
Aggressive Loading in Bit-Rounding Algorithms

Bit-rounding algorithms are practically attractive because they are relatively *computationally* efficient and *implementationally* realizable since they ensure integer-bit distribution. However, due to the concave nature of the rate-energy relationship, these algorithms often tend to load bits *aggressively*, which leads to the system BER constraint violation. To counter this violation we propose an efficient *moderate* algorithm based on *biased* bit round-off, which ensures BER fidelity and total power constraint equality, essential for rate optimality.

A.1 Introduction

Rate adaptive loading algorithms [4] for multicarrier systems [5] have the objective of maximizing the rate conveyed, while satisfying two main constraints – the total power constraint (TPC) and the performance or bit error rate (BER) constraint. The optimal power allocation for parallel independent Gaussian channels has long been known to be the 'water-filling' solution [7]. The Lagrange multiplier that leads to the water-filling solution explicitly ensures that the TPC is satisfied with *equality*. Since the rate is a monotonically increasing function of power, this *equality* is a necessary condition for rate optimality. However, while the water-filling method assumes infinite granularity, most known modulation schemes support only integer number of bits/symbol. With the integer-bit constraint, we now have an integer-programming problem, the optimal solution for which is a greedy-search approach [8]. But the extensive searching and sorting characteristic of greedy methods renders them *computationally* impractical for usage as loading algorithms in current scenarios such as Digital Subscriber Loops (DSL) [31]. Hence popular methods involve heuristic approaches, which start with an 'infinite granularity' assumption for computational efficiency as their first step [4][9][30][31]. As a second step, they round these fractional bits off to the nearest integers, and adjust the energy accordingly to ensure BER constraint satisfaction. Though after the first step, the TPC is satisfied with equality, the equality is almost certainly violated in either direction after the

second step. Most algorithms deal with this problem by an ad-hoc rescaling of all subcarrier energies, by a common factor (which is treated as *margin* of the system) to re-ensure the TPC equality. This final rescaling factor or *margin* generally turns out to be negative (in dB). We point out that the negative *margin* leads to the unacceptable BER constraint violation. We identify the source of the problem to be the biased nature of the bit round-off operation (resulting from the concavity of the rate-energy relationship). To rectify this problem, we propose an efficient algorithm to estimate the optimum biasing parameter for round-off operation.

The remainder of this appendix is organized as follows. Section A.2 presents the overall optimization problem formulation and introduces the notation. Section A.3 provides a unifying review of bit-rounding algorithms and identifies the causes of the *aggression*, which lead to BER constraint violation. Based on these causes, a computationally efficient algorithm is proposed in Section A.4. Section A.5 presents the results obtained in diverse VDSL-DMT scenarios and concluding remarks are made in Section A.6.

A.2 Problem Formulation

Consider a multicarrier system of K sub-carriers. Let ε_j be the controllable transmit power spectral density (PSD) and ρ_j be the normalized received SNR when ε_j=1 over the j^{th} sub-carrier. The inter-carrier spacing Δf is assumed to be small enough for all the aforementioned PSDs to be nearly flat over each sub-carrier. The rate-function $b(\sigma)$ is defined as the maximum information rate in bits that can be supported at SNR of σ keeping the BER below a certain target. $b(\sigma)$ is generally assumed to be the SNR-gap function.

$$b(\sigma) = b_{GAP}(\sigma) = \log_2\left(1+\frac{\sigma}{\Gamma}\right) \tag{A.1}$$

where Γ is the SNR-gap [4][5]. For QAM schemes Γ=9.8dB for a target BER of 10^{-7} [31]. The object function is the total supported rate,

$$R = \sum_{j=1}^{K} b\left(\rho_j \varepsilon_j\right). \tag{A.2}$$

The traditional constraints of total power, P_{max}, and integer bit constraint can be expressed as follows:

$$\text{power distribution: } \Delta f \sum_{j=1}^{N} \varepsilon_j \le P_{\max}, \varepsilon_j \ge 0, 1 \le j \le K \tag{A.3}$$

$$\text{integer-bit restriction: } b\left(\rho_j \varepsilon_j\right) \in W, \ 1 \le j \le K \tag{A.4}$$

where W is the set of non-negative integers.

A.3 Aggressive Loading in Bit-Rounding Algorithms

The Rate Adaptive (RA) loading algorithms (by Chow et al.) proposed in [4][30] and (by Leke et al.) in [9] start with an Optimum Transmit Bandwidth Identification (OTBI) procedure. This identification is done through a costly iterative procedure in [4][30] (presented in Table A.1). In [9] a non-negative energy assignment criterion by virtue of which the identification is achieved in a single iteration over the subcarriers is introduced (presented in Table A.2). Furthermore, while initial energy distribution is flat in [4][30], algorithm in [9] starts with a water-filling distribution.

Table A-1: OTBI Procedure for Chow's Algorithms

1.	Sort $\{\rho_j\}_{j=1}^{K}$ in descending order and save the index mapping in ρ_{index}
2.	$b_{old} = 0$, $i = K$
3.	$\{\varepsilon_j\}_{j=1}^{i} = P_{\max}/i\Delta f$ and $\{\varepsilon_j\}_{i+1}^{K} = 0$
4.	$\{b_j\}_{j=1}^{K} = \log_2\left(1+\varepsilon_j\rho_j/\Gamma\right)$
5.	$b_{curr} = \sum_{j=1}^{K} b_j$
6.	if $b_{curr} < b_{old}$
7.	Unsort $\{\rho_j\}_{j=1}^{K}, \{\varepsilon_j\}_{j=1}^{K}$ and $\{b_j\}_{j=1}^{K}$ using ρ_{index}
8.	else
9.	$i = i-1$, $b_{old} = b_{curr}$
10.	GOTO Step 3.
11.	end

Table A-2: OTBI Procedure for Leke's Algorithm

1.	Sort $\{\rho_j\}_{j=1}^{K}$ in descending order and save the index mapping in ρ_{index}
2.	$i = K$
3.	$NSR = \frac{1}{i} \cdot \left(\frac{P_{\max}}{\Delta f \cdot \Gamma} + \sum_{j=1}^{i} \frac{1}{\rho_j} \right)$
4.	if $\rho_i < 1/NSR$
5.	$NSR = \frac{1}{i-1} \cdot \left(i \cdot NSR - \frac{1}{\rho_i} \right)$
6.	$i = i-1$, GOTO Step 4
7.	end
8.	$\{\varepsilon_j\}_{j=1}^{i} = \Gamma \cdot \left(NSR - \frac{1}{\rho_j} \right)$; $\{\varepsilon_j\}_{i+1}^{K} = 0$
9.	Unsort $\{\rho_j\}_{j=1}^{K}$ and $\{\varepsilon_j\}_{j=1}^{K}$ using ρ_{index}
10.	$\{b_j\}_{j=1}^{K} = \log_2 \left(1 + \varepsilon_j \rho_j / \Gamma \right)$

The common aspect of both algorithms is that following the OTBI, the bits allocated to each of the K subcarriers, $\{b_j\}_{j=1}^{K}$ obtained from the usage of SNR-gap function in (A.1) are rounded to integers, i.e.,

$$\hat{b}_j = round\{b_j\}\ ;\ j = 1,2..K \tag{A.5}$$

and energy allocated to each subcarrier $\{\varepsilon_j\}_{j=1}^{K}$ is adjusted to make the BER nearly equal in the sub-carriers, leading to the saw-tooth energy distribution, i.e.,

$$\varepsilon_j = \rho_j^{-1} b_{GAP}^{-1}\left(\hat{b}_j\right) = \rho_j^{-1}\left(2^{\hat{b}_j} - 1\right)\Gamma . \tag{A.6}$$

However, following this adjustment of energies in (A.6), the total utilized energy is no longer equal to the total energy constraint of the system $P_{\max}$. To re-achieve this equality, in the final step of both the algorithms, $\{\varepsilon_j\}_{j=1}^{K}$ are multiplied by an energy re-scaling factor,

$$\varepsilon_{RF} = P_{\max}\left[\Delta f \sum_{j=1}^{K} \varepsilon_j\right]^{-1} . \tag{A.7}$$

The energy re-scaling factor, ε_{RF}, can be treated as the additional performance *margin* of the system. When $\varepsilon_{RF}=1$, it is truly the optimum point for the rate-maximization problem. However, given the integer bit constraint, this would be a rare occurrence. When $\varepsilon_{RF}>1$, it is sub-optimal in terms of rate-maximization, because the remaining $\{P_{\max}\text{-}\Delta f.\sum\varepsilon_j\}$ could have been used to load additional bits rather than enhancing the performance *margin* using (A.7).

A.3.1 Energy Rescaling Factor Problem

The problem occurs when $\varepsilon_{RF}<1$. Due to (A.6), following the energy adjustment, the SNR on each loaded subcarrier is already at a bit transition boundary, i.e., at the minimum SNR required to satisfy the error rate constraint for the specific $\hat{b}_j$ loaded on that subcarrier. Thus, when $\varepsilon_{RF}<1$, each ε_j being reduced by (A.7), the BER constraint on each subcarrier will be violated. As a result, though the loading algorithm predicts that $\sum\hat{b}_j$ bits can be supported, this bit distribution will violate the error-rate criterion. This effect is illustrated in the following example. For the sake of illustration, we consider the following simplified parameters:

$$N=4,\Gamma=1,\Delta f=1,P_{\max}=1,\rho_1=40,\rho_2=30,\rho_3=20,\rho_4=20.$$

The values of the involved parameters for 4 sub-carriers are computed and recorded as follows:

j	After OTBI Procedure		After Bit Rounding		After Rescaling	
	ε_j	b_j	$\hat{b}_j$	ε_j(A.6)	ε_j	$\varepsilon_j\rho_j$
1	0.25	3.4594	3	0.1750	0.1579	6.3158
2	0.25	3.0875	3	0.2333	0.2105	6.3158
3	0.25	2.5850	3	0.3500	0.3158	6.3158
4	0.25	2.5850	3	0.3500	0.3158	6.3158
Σ	1	11.72	12	1.11	1	

After OTBI procedure, the TPC is satisfied with equality. But after rounding, the bit loading is aggressive and $\sum\varepsilon_j$= 1.11 > $P_{\max}$=1. Thus, the energy is finally rescaled by ε_{RF}=1/1.11=0.9. This reduces each ε_j and causes the final SNR on each subcarrier to be

6.3158 < $b_{GAP}^{-1}(3)$= 2^3-1=7. Thus on each subcarrier, the SNR being lower than the required threshold, the BER constraint will be violated and thereby so will the overall system BER.

A.3.2 Aggression due to SNR Gap Rate Function

The function $b_{GAP}^{-1}(x), \forall x \in W$ is supposed to provide the minimum SNR required to support 2^x-QAM at a certain error rate requirement. In Figure A-1 we plot the minimum SNRs for various QAM constellations obtained from exact or near-exact formulae $b_{BER}^{-1}(.)$ and $b_{GAP}^{-1}(.)$ for 10^{-7} BER. For the exact minimum SNR calculation, we assume perfect gray encoded square QAM constellations [15] for even number of bits per symbol and impure gray encoded QAM cross-constellations for odd number of bits per symbol [17][34]. For most QAM constellations the SNR predicted by SNR-gap function is slightly more than the exact value. This is safe since in many practical situations perfect gray encoding is not used, e.g., in VDSL [3].

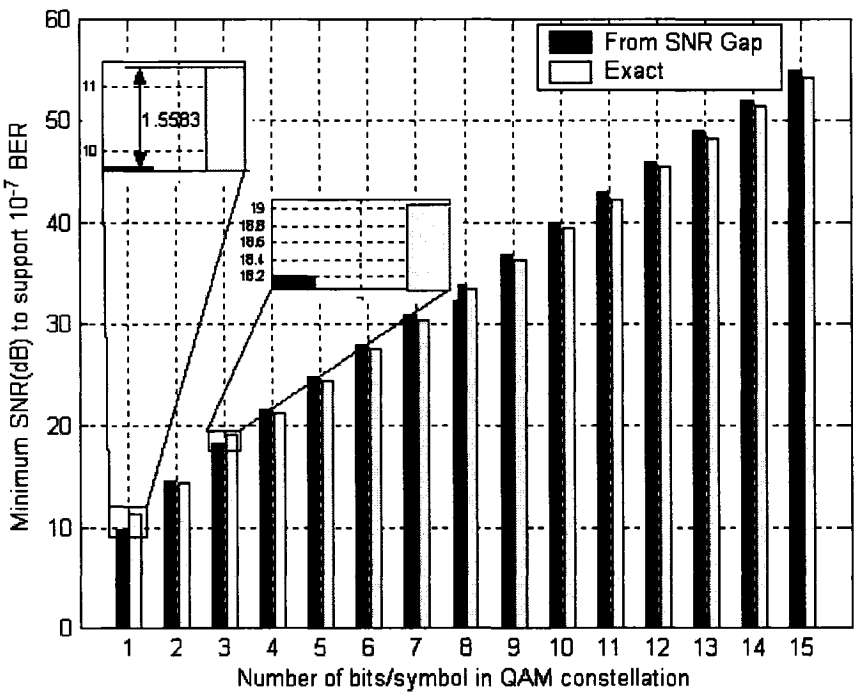


Figure A-1: Accuracy of SNR-GAP Approximation

[34] For 8-QAM neither a square nor a cross-constellation is possible. However, a perfectly gray encoded rectangular QAM constellation [15] gives lower BER than the constellation in [3] for a given SNR.

The major problem occurs for 8-QAM and BPSK, where the SNR-gap function predicts 0.8328 and 1.5583dB less than what is exactly obtained. It is important to note that this negative *margin* of -1.5583dB sharply moves the BER from 10^{-7} to 0.7×10^{-5}. This problem would affect any loading algorithm that relies on the SNR-gap rate function and the BER constraint would be violated whenever 8-QAM and/or BPSK are predominantly loaded.

A.4 A Moderate Algorithm with Biased Round-Off

The case of $\varepsilon_{RF} < 1$ occurs because the round-off process is inherently biased in terms of energy. The additional energy *required* to send 4 bits instead of 3.51 bits is proportional to $2^4 - 2^{3.51} = 4.6076$; however, the energy *saved* by sending 3 bits instead of 3.49 bits is proportional to $2^{3.49} - 2^3 = 3.2356$. In other words, due to the logarithmic nature of the SNR-gap function, rounding up by the same fractional amount of bits always costs more energy than rounding down. Thus, just after the OTBI, the equality $\Delta f\sum\varepsilon_j = P_{\max}$ is satisfied, but, due to its inherent bias, the rounding process generally makes $\Delta f\sum\varepsilon_j > P_{\max}$, and hence $\varepsilon_{RF} < 1$. To overcome this problem, we propose to negatively bias the round-off procedure by a parameter $\alpha \in [0, 0.5]$, which will make the $\varepsilon_{RF} \geq 1$, i.e.,

$$\hat{b}_j = round(b_j - \alpha). \tag{A.8}$$

While $\alpha > 0$ is essential to avoid the cases of $\varepsilon_{RF} < 1$, choosing an unreasonably large α (e.g., 0.3) will cause significant loss in rate. Hence, we select α that would make ε_{RF} approach 1. The pseudo-code to do this iteratively is presented in Table A-3. The OTBI in the first step can be performed using Table A-1 or Table A-2. Bisection search is employed to find the optimum α. From (A.8) and (A.6), we can see that $\hat{b}_j$ and ε_j are monotonically decreasing functions of α, $\forall j : 1 \leq j \leq N; \varepsilon_j > 0$. Thus, $(\Delta f.\sum\varepsilon_j)$ is also a monotonically decreasing function of α and due to (A.7) and $P_{\max}$ being a constant, ε_{RF} is a monotonically increasing function of α. For $\alpha = 0.5$, (A.8) is equivalent to $\hat{b}_j = \lfloor b_j \rfloor \Rightarrow \hat{b}_j \leq b_j$. Therefore, if $\hat{\varepsilon}_j$ and ε_j are the corresponding energy allocations,

the TPC is satisfied with equality after OTBI $\Delta f \sum_{j=1}^{K} \varepsilon_j = P_{\max}$ and $\hat{\varepsilon}_j \le \varepsilon_j \forall j \in \{1,2..K\}$; $\Delta f . \sum \varepsilon_j \le P_{\max} \Rightarrow \varepsilon_{RF} \ge 1$. At $\alpha = 0, \varepsilon_{RF} < 1$ (otherwise, the loop breaks in Table A-3). For $\alpha = 0.5$, $\varepsilon_{RF} \ge 1$ is guaranteed and in the interval $\alpha \in [0, 0.5]$, ε_{RF} is a monotonically increasing function of α. Thus, there exists a unique $\alpha \in [0, 0.5]$ where $\varepsilon_{RF} = 1$ is established and hence the convergence of the bisection search in Table A-3 is proven. This solves the energy rescaling factor problem in Section A.3.1.

Table A-3: Pseudocode for Moderate Algorithm

```
1.  Perform OTBI which provides {b_j}_{j=1}^K
2.  ε_RF = 0.9, α_min = -0.5 , α_max = 0.5
3.  while |ε_RF - 1| > 0.01
4.     α = (α_min + α_max)/2
5.     for j = 1 to K
6.        b̂_j = round{b_j - α}
7.        Allocate ε_j using (A.9)
8.     end
9.     ε_RF = P_max / (Δf Σ_{j=1}^K ε_j)
10.    if ε_RF < 1
          α_max = α
       else
          α_min = α
       end
11. end
```

To counter the aggression due to the SNR gap function anomaly for BPSK and 8-QAM cases explained in Section III.B, we specifically identify the sets $\tilde{J}_1 = \{j \in \{1,2,..,K\} : \hat{b}_j = 1\}$ and $\tilde{J}_3 = \{j \in \{1,2,..,K\} : \hat{b}_j = 3\}$ and perform the following energy allocation and thereby guarantee BER fidelity:

$$\varepsilon_j = b_{BER}^{-1}(i)/\rho_j, j \in \tilde{J}_i, i = 1,3 \text{ or } b_{GAP}^{-1}\left(\hat{b}_j\right)/\rho_j, j \notin \tilde{J}_1 \cup \tilde{J}_3 \qquad \text{(A.9)}$$

A.5 Simulation Results

We consider a VDSL-DMT system [3] with the parameters stated in Table A-4. In Figure A-2(a), the rate-reach curves for the algorithms in [30]-Sec. 4.3.4 and [4]-Sec. 7.2.3.1 are presented. The bit rounding is conducted within the framework of the OTBI in [30], and after OTBI in [4]. For easier comparison in Figure A-2(b), the percent decrease of rate of various schemes with reference to algorithm in [30] is plotted. In the worst case, the rate reduction is only 4%. In Figure A-2(c) the BER achieved by the loading resulting from various schemes is plotted. In most cases, the aggressive bit loading of [30] and [4] results in a significant system BER violation. In contrast, the BER of the proposed moderate algorithm (Table A-3) stays nearly flat and consistently below the target BER of 10^{-7}.

Table A-4: Simulation Parameters [3][19]

Number of Subcarriers:	4096	
Length of Cyclic Prefix:	640 samples	
Upstream carriers:	U1: 870-1205	U2: 1972-2782
Downstream carriers:	D1: 33-869	D2: 1206-1972
Transmit PSD constraint:	None	
Direction	Downstream	
Loop Type/ Length	Loop 2/ 1500-9000ft. at steps of 100ft.	
Noise Type	AWGN(-140dBm/Hz) + 20 VDSL xTalkers + Alien Noise F	
Total Power Constraint	11.5dBm	

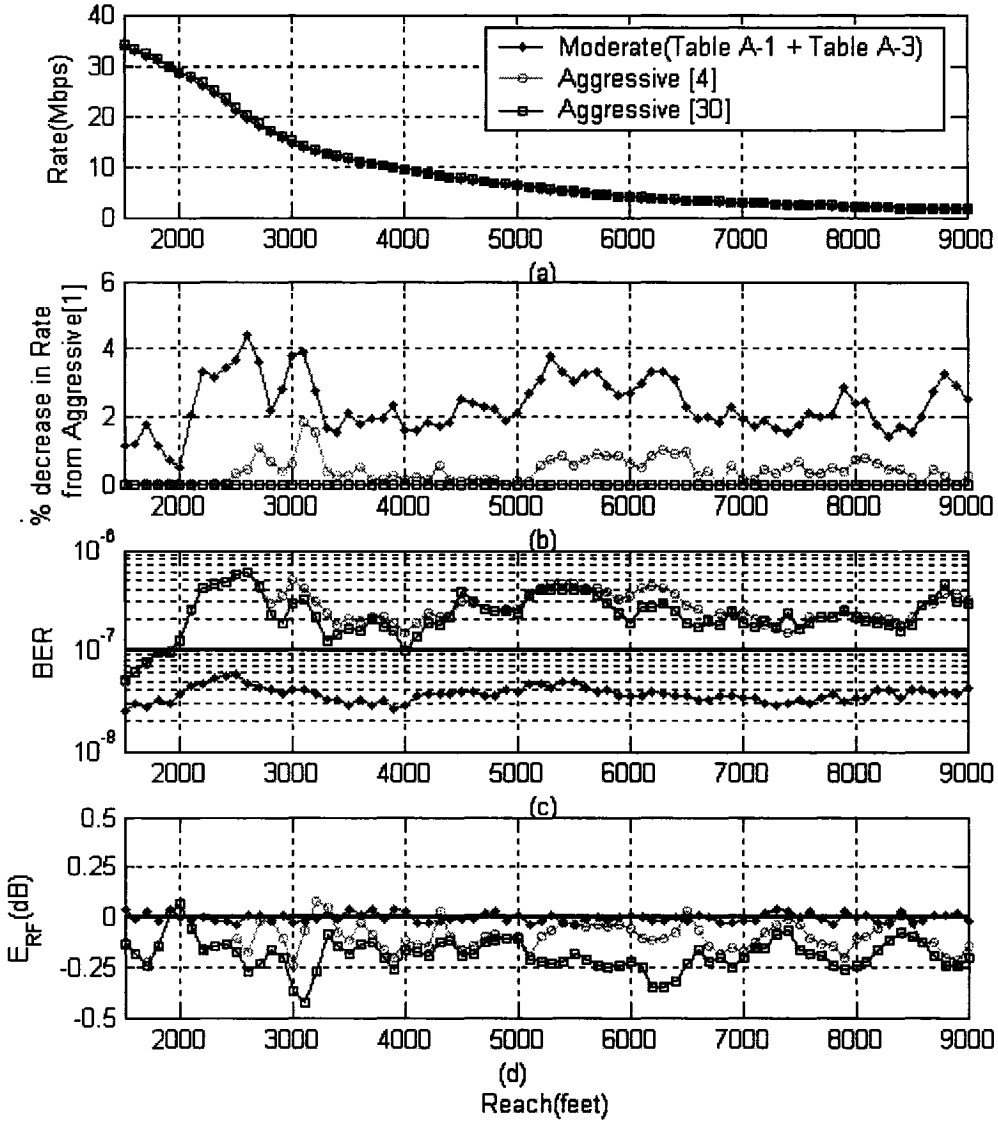


Figure A-2: Performance of Algorithms in Tables A-1 and A-3

Similar results for algorithm in [9] are presented in Figure A-3. Again, the moderate algorithm (Table A-3) can guarantee a nearly flat BER profile consistently below 10^{-7} with a rate loss of less than 4% from the aggressive algorithm. Also we notice in Figure A-3(c) that the BER violations of algorithm in [9] have higher magnitude than algorithms in [30] and [4], e.g., at 5400ft., algorithm in [9] results in a BER of 10^{-6}. This occurs because algorithm in [9] is based on water-filling. The energy allocation slowly approaches zero for higher frequency, thus a higher percentage of subcarriers load BPSK, which

causes the violations due to the anomaly explained in Section A.3.2. Due to the gray encoding, the BER results are slightly optimistic.

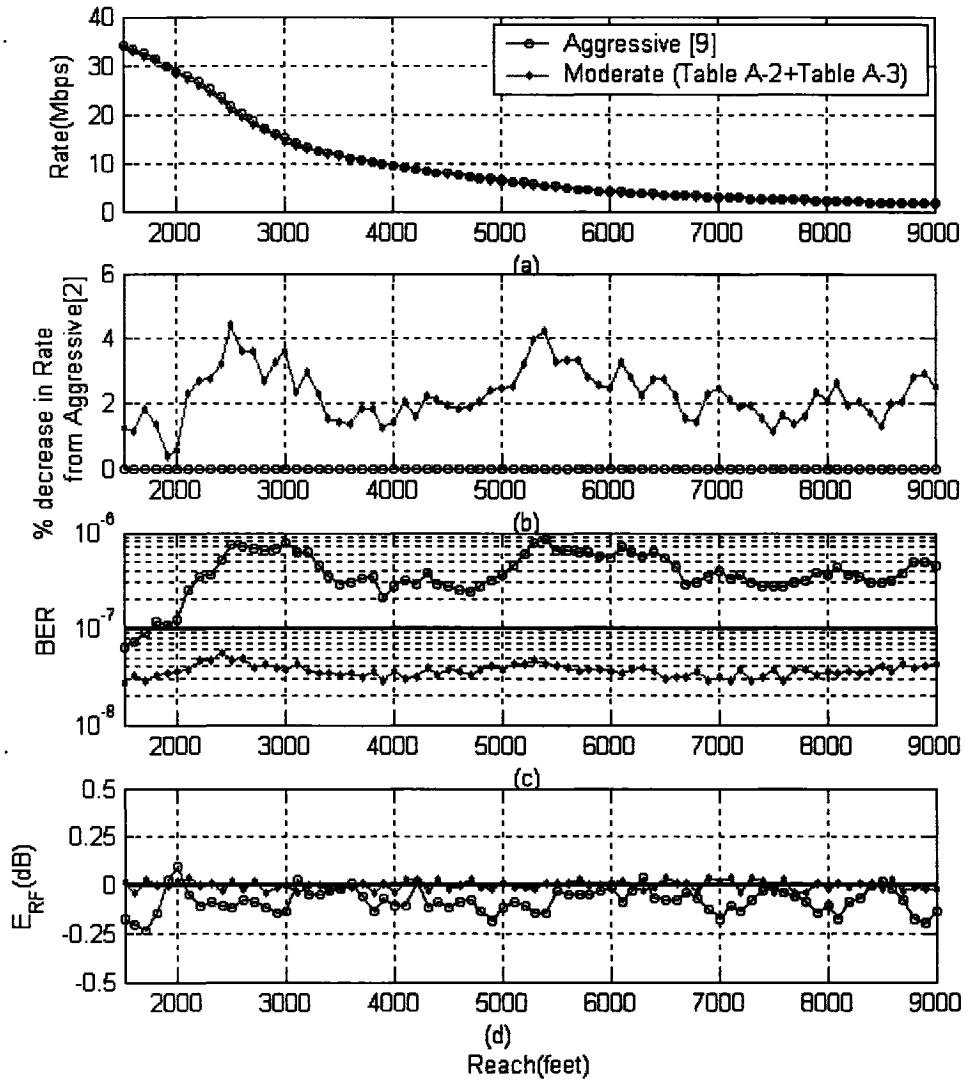


Figure A-3: Performance of Algorithms in Tables A-2 and A-3

In applications such as VDSL-DMT, where the constellation labeling does not follow gray coding [3], all the BER curves would shift slightly higher. Thus BER violations of the aggressive algorithms would be higher and at particular points will be drastic, e.g., at 2600ft. in Figure A-2 or 5400ft. in Figure A-3. However for the moderate algorithms, Figure A-2(c) and Figure A-3(c) show that there is some margin to absorb the effects of non-gray coding while still keeping the BER below 10^{-7}. In addition, the nearly flat na-

ture would ensure that no drastic violations occur at any point. If we consider the testing procedure in [19], when the measured BER is within the range of the 0.85×10^{-7} to 1.15×10^{-7}, the test case is declared as 'marginal'. But any BER above 1.15×10^{-7} would be declared certain failure. Thus most of the cases resulting from the aggressive algorithms shown in Figure A-2 and Figure A-3 would fail.

In Figure A-2(d) and Figure A-3(d), the ε_{RF} in dB, is presented. As expected from Section A.3.1, for the aggressive algorithms $\varepsilon_{RF}<1$ (and not $\varepsilon_{RF}\geq1$) is the predominantly occurring case. As aimed in Table A-3, ε_{RF} for the moderate algorithms stay close to 1 (or 0dB).

The average number of additional iterations taken by Table A-3 is 4.8 for algorithm in Table A-1, and 4.6 for algorithm in Table A-2 averaged over 75 channel realizations.

A.5 Summary

In this appendix, we have highlighted the fact that bit-rounding algorithms tend to be *aggressive* in loading bits due to the concave nature of the rate-energy relationship. This problem and the SNR gap function anomaly for BPSK and 8-QAM cause significant BER violation for the 75 different VDSL-DMT channel realizations that were tested. We proposed a moderate algorithm with biased bit rounding coupled with the explicit handling of BPSK and 8-QAM. The proposed algorithm was seen to require only 4-5 additional iterations on an average and with only 4% reduction in rate in the worst-case was able to guarantee BER below the target for all 75 tested cases when compared to its *aggressive* counterparts.

Appendix B
Karush-Kuhn-Tucker (KKT) Conditions

Consider the following general optimization problem:

$$\min_{f^i(\mathbf{x})\le 0; i\in\{1,2,...,Q\}} f(\mathbf{x}) \qquad \text{(P)}$$

where $\mathbf{x}\in\mathbb{R}^V$ and the Q+1 functions $f(.), f^1(.), f^2(.),..., f^Q(.)$ are differentiable mappings from $\mathbb{R}^V$ to $\mathbb{R}$. $f(\mathbf{x})$ is the objective function to be minimized and $f^1(.), f^2(.),..., f^Q(.)$ are the Q inequality constraints of the problem. The problem in (P) is called a convex program if all the functions i.e., $f(.), f^1(.), f^2(.),..., f^Q(.)$ are convex.

The Lagrangian of the above problem is: $L(\mathbf{x}, u_1, u_2, .., u_Q) = f(\mathbf{x}) + \sum_{i=1}^{Q} u_i f^i(\mathbf{x})$.

Then we have the following theorem which characterizes the optimality of a point $\mathbf{x}^*$.

THEOREM: Consider the program (P) where $f(.), f^1(.), f^2(.),.., f^Q(.)$ are assumed to be convex and differentiable. Also assume that the constraints satisfy Slater's condition[35]. Then a point $\mathbf{x}^*$ is globally optimal if and only if:

1. $\mathbf{x}^*$ is feasible, i.e., $f^1(\mathbf{x}^*)\le 0, f^2(\mathbf{x}^*)\le 0,..., f^Q(\mathbf{x}^*)\le 0$.
2. There exists a set of multipliers $u_1, u_2,..., u_Q$; such that the following hold:

$$\begin{aligned} &\nabla_{\mathbf{x}} L(\mathbf{x}^*, u_1, u_2, .., u_Q) = 0 \\ &u_i f^i(\mathbf{x}^*) = 0; \forall i \in \{1,2,..,Q\} \\ &u_i \ge 0; \forall i \in \{1,2,..,Q\} \end{aligned} \qquad \text{(KKT)}$$

Proof: The proof is available in almost every book on optimization, e.g., [41].

---

[35] A set of constraints $f^i(\mathbf{x}) \le 0; i \in \{1,2,...,Q\}$ is said to satisfy Slater's condition, if there exists a Slater point $\mathbf{x}'$ such that $f^i(\mathbf{x}') < 0; \forall i \in \{1,2,...,Q\}$.

The equations in (KKT) system are known as KKT conditions. We provide an example below of how these can be used, to find an optimal point:

Example: $\min_{x_1+x_2+5\leq 0} x_1^2 + x_2^2$

Consider the above convex problem. Given the notation above, here in this example, we have $V = 2; Q = 1; \mathbf{x} = [x_1, x_2]^T; f(\mathbf{x}) = x_1^2 + x_2^2; f^1(\mathbf{x}) = x_1 + x_2 + 5$. The constraint satisfies Slater's condition because $\mathbf{x}' = [-5, -5]^T$ is a Slater point, i.e., $f^1(\mathbf{x}') < 0$. The corresponding Lagrangian is $L(x_1, x_2, u_1) = x_1^2 + x_2^2 + u_1(x_1 + x_2 + 5)$.

For an optimal $\mathbf{x}^*$, the KKT conditions have to be satisfied.

$$\frac{\partial L(\mathbf{x}^*, u_1)}{\partial x_1} = 2x_1^* + u_1 = 0$$

$$\frac{\partial L(\mathbf{x}^*, u_1)}{\partial x_2} = 2x_2^* + u_1 = 0$$

$$u_1(x_1^* + x_2^* + 5) = 0$$

$$u_1 \geq 0$$

The only feasible solution to the above system is $\mathbf{x}^* = [-2.5, 2.5]^T$ and the corresponding $u_1 = 5$.

Remarks:

- If all the conditions stated in the above theorem are satisfied and further the objective function $f(.)$ is STRICTLY convex, then $\mathbf{x}^*$ is not only the global optimum it is also the unique global optimum. Notice that all the continuous rate functions we have considered in this thesis are STRICTLY convex[36].
- For all the problems considered in this thesis it is always easy to find a Slater point and hence repeated references are avoided in discussions. We state here that all the constraints in the optimization problems considered in this thesis satisfy Slater's condition.

[36] A twice continuously differentiable function is strictly convex if and only if the Hessian of the function is positive definite.

References

[1] "Asymmetric Digital Subscriber Line (ADSL) Metallic Interface", *ANSI Std. T1.413 –1998.*

[2] "Spectrum Management for loop transmission systems", *Committee T1 Std. T1.417-20011*, Jan. 2001.

[3] "Very-high Speed Digital Subscriber Lines (VDSL) Metallic Interface", *ANSI Std. T1E1.4/2003-210R5, 2003.*

[4] T. Starr, J.M. Cioffi, and P.J. Silverman, *Understanding Digital Subscriber Line Technology*, Prentice Hall, 1999.

[5] J.M. Cioffi, "A Multicarrier Primer", *ANSI T1E1.4 Committee Contribution*, pp. 91-157, Nov. 1991

[6] Telcordia Technologies, "Proposed Bit Rates for Spectral Compatibility with VDSL", *ANSI T1E1.4 Committee Contribution*, T1E1.4/2002-159R2, Aug. 2002.

[7] R.G. Gallager, *Information Theory and Reliable Communication.* Wiley, New York 1968

[8] D. Hughes-Hartogs, "Ensemble Modem Structure for Imperfect Transmission Media". *U.S. Patents Nos. 4,679,227* (July 1987), *4,731,816* (March 1988) and *4,833,706* (May 1989).

[9] A. Leke and J.M. Cioffi, "A maximum rate loading algorithm for discrete multitone modulation systems", *GLOBECOM'97*, pp.1514-1518, Nov. 1997.

[10] B.S. Krongold, K. Ramachandran, and D.L. Jones, "Computationally Efficient Power Allocation Algorithms for Multicarrier Communication Systems", *IEEE Transactions on Comm.*, Vol. 48, pp. 23-27, Jan. 2002.

[11] J. Jang, K.B. Lee, Y.H. Lee, "Transmit Power and Bit Allocations for OFDM Systems in a Fading Channel", *GLOBECOM'03*, pp. 858-862, Dec. 2003.

[12] Sec. 4.8.3 "Analysis of Quantization Errors" - Alan V. Oppenheim, Ronald W. Schafer and John R. Buck, *Discrete-Time Signal Processing, 2nd edition*, Prentice Hall, New Jersey 1998.

[13] Robert H. Moreles-Zaragoza, *The Art of Error Control Coding*, John Wiley and Sons. 2000.

[14] J.P. Lauer and J.M. Cioffi, "A Turbo Trellis Coded Discrete Multitone Transmission System", *GLOBECOM'99*, pp. 2581-2585, Nov. 1999.

[15] K. Cho and D. Yoon, "On the General BER Expression of One- and Two-Dimensional Amplitude Modulations", *IEEE Transactions On Comm.*, Vol. 50, pp.1074-1080, Jul. 2002.

[16] P.K. Vitthaladevuni and M. –S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations", *IEEE Transactions on Inform. The.*, vol. 49, pp.297-307, January 2003.

[17] J.G. Smith, "Odd-bit Quadrature amplitude shift keying", *IEEE Transactions On Comm.*, vol. 23, pp. 385-389, March 1975.

[18] L.Zhang, C. Gao, and Z. Cao, "Exact Analysis of Bit Error Rate of Maximum-Distance-Separable Codes", *GLOBECOM'00*, pp. 816-819, Nov.-Dec. 2000.

[19] "VDSL Test Specification for VDSL Olympics", *ANSI T1E1.4 Contribution, T1E1.4/2003-036R4*, Feb.2003.

[20] M.A. Hasan and V.K. Bhargava, "Architecture for a Low Complexity Rate-Adaptive Reed-Solomon Encoder", *IEEE Transactions on Computers*, Vol. 44, pp. 938-942, Jul. 1995.

[21] Y.R. Shayan and T. Le-Ngoc, "A Cellular Structure for a Versatile Reed-Solomon Decoder", *IEEE Transactions on Comp.*, Vol. 46, pp. 80-85, Jan.1997.

[22] J.W. Cook, R.H. Kirkby, M.G. Booth, K.T. Foster, D.E.A. Clarke, and G. Young, "The Noise and Crosstalk Environment for ADSL and VDSL Systems", *IEEE Comm. Mag,* Vol. 37, pp. 73-78, May 1999.

[23] K.B. Song, S.T. Cheung, G.Ginis and J.M. Cioffi, "Dynamic Spectrum Management for Next-Generation DSL Systems," *IEEE Comm. Mag.*, Vol. 40, pp.101-109, Oct. 2002.

[24] K.J. Kerpez, D.L. Waring, S. Galli, J. Dixon and P. Madon, "Advanced DSL Management," *IEEE Comm. Mag.*, Vol. 41, pp. 116-123, Sept. 2003.

[25] W. Yu, G.Ginis and J.M. Cioffi, "Distributed Multiuser Power Control for Digital Subscriber Lines," *IEEE JSAC*, Vol. 20, pp. 1105-1115, June 2002.

[26] S. Chung, *Transmission schemes for frequency selective Gaussian interference channels*, Ph. D. dissertation, Stanford University, Stanford, CA, 2003.

[27] R. Cendrillon, W. Yu, M. Moonen, J. Verliden, and T. Bostoen, "Optimal Multiuser Spectrum Management for Digital Subscriber Lines", to appear in *IEEE Transactions on Comm.*, 2005.

[28] G. Ginis and J.M. Cioffi, "Vectored Transmission for Digital Subscriber Line Systems", IEEE JSAC, Vol. 20, pp. 1085-1104, June 2002.

[29] T. Starr, M. Sorbara, J.M. Cioffi and P.J. Silverman, *DSL Advances*, Prentice Hall, 2003.

[30] P.S. Chow, *Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise*, Ph. D. dissertation, Stanford University, Stanford, CA, May 1993.

[31] P.S. Chow, J.M. Cioffi, and J.A.C Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels", *IEEE Transactions On Comm.*, Vol. 43, pp. 773-775, Feb./Mar./Apr. 1995.

[32] R.F.H Fischer and J.B. Huber, "A new loading algorithm for discrete multitone transmission", in *Proceedings IEEE GLOBECOM'96*, pp.724-728, Nov. 1996.

[33] J. Campello de Souza, *Discrete Bit Loading for Multicarrier Modulation Systems*, Ph. D. dissertation, Stanford University, Stanford, CA, June 1999.

[34] J. Campello, "Optimal Discrete Bit Loading for Multicarrier Modulation Systems", *Proceedings IEEE International Symposium on Info. Theory*, 1998. pg. 193, Aug. 1998.

[35] J. Campello, "Practical Bit-Loading for DMT", *Proceedings IEEE ICC'99*, pp. 801-805, Jun. 1999.

[36] B.S. Krongold, K. Ramachandran, and D.L. Jones, "An efficient algorithm for Optimal Margin Maximization in Multicarrier Communication Systems", *Proceedings IEEE GLOBECOM'99*, pp. 899-903, Dec. 1999.

[37] D. Toumpakaris, J.M. Cioffi, and D. Gardan, "Reduced-Delay Protection of DSL Systems Against Nonstationary Disturbances", *IEEE Transactions On Comm.*, Vol. 52, pp.1927-1938, Nov. 2004.

[38] E. Baccarelli, A. Fasano, and M. Biagi, "Novel Efficient Bit-Loading Algorithms for Peak-Energy-Limited ADSL-Type Multicarrier Systems," *IEEE Transactions on Signal Processing*, Vol. 50, pp.1237-1247, May 2002.

[39] T.F. Coleman and Y. Li, "An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIAM Journal on Optimization*, Vol. 6, pp. 418-445, May 1996.

[40] M.A. Branch, T.F. Coleman and T. Li, "A Subspace, Interior and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems," *SIAM Journal on Scientific Computing*, Vol. 21, pp. 1-23, 1999.

[41] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, New York. 1997.

[42] Y. Xu, S. Panigrahi and T. Le-Ngoc, "Selective Iterated Water-Filling for Digital Subscriber Loops," submitted to *IEEE GLOBECOM' 2005*.

[43] N. Damji and T. Le-Ngoc, "Adaptive downlink multi-carrier resource allocation for real-time multimedia traffic in cellular systems", Proceeding IEEE International Conference on Communications (ICC) 2004, pp. 4258-4262, June 2004.