System Level Performance of ATM Transmission over a DS-CDMA Satellite Link

By

Tijana Timotijevic

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSPHY

Department of Electronic Engineering

Queen Mary and Westfield College

University of London

1999

To Ivica Antonijevic-Rabusic,

my first maths teacher

ABSTRACT

The thesis provides an insight into the viability of CDMA as an underlying multiple access scheme for a broadband satellite system based on ATM, and is the first step towards the study of the ATM traffic control and management functions in a satellite system employing CDMA.

This thesis introduces a concept of statistical multiplexing in the code domain, by analogy with statistical multiplexing in ATM networks, in order to maximise the CDMA system capacity and therefore provide more efficient use of the available bandwidth. Statistical multiplexing in the code domain is made possible by the use of discontinuous transmission detection (DTX) for all bursty ATM traffic (and not only voice, as is the case in current operational wireless systems and those still being a subject of research). The goal of the thesis is:

- to quantify the system performance by simulation and analytic investigation of the scheme using three scenarios for ATM transmission over CDMA;
- to deduce a methodology for quantitative optimisation of all similar scenarios using the same scheme for CDMA based cell transmission.

To investigate the medium access scheme based on statistical multiplexing in the code domain, the three layers are studied together: physical layer CDMA self-interference is mapped into the network-layer capacity, and the effect that DTX and statistical multiplexing have on the capacity performance of the system is analysed at the ATM layer. The performance measures are defined both on ATM layer (efficiency, buffer state and delay distributions) and network layer (capacity). Physical layer is taken into account through the predefined threshold signal-to-noise ratio (SNR), that is dictated by the quality of service requirement for the maximum allowed cell loss rate. Mathematical analysis of the system performance is validated by cell-level simulations where possible, or by hybrid approach of combined alternative analysis and simulations. For numerical studies three different types of ON-OFF traffic sources are defined.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The finalisation of one's thesis is the point when most people remember not only those who immediately helped carrying it out, but also all those whose incidental or intentional involvement has been essential in this path of learning.

First of all, I want to thank the Department of Electronic Engineering at Queen Mary and Westfield College for sponsoring my research as a postgraduate student. In particular, I am grateful to Prof. Laurie Cuthbert for having confidence in me, thus making my research at QMW College possible. My gratitude also goes to Ferial Chummun and Luigi Bella at the European Space Agency, for their technical support and ESA funding. I will always remember Ferial's humanity, which will remain my strong reference point.

I consider myself fortunate that I have been working with Dr. John A. Schormans as my supervisor, as not only did he give excellent guidance throughout my research, but also continuous encouragement, genuine support and selfless friendship. I know that this friendship will last.

I am indebted to Roger Shaw and Louise Gibbs, whose help, both instrumental and emotional, was crucial to get me through the most difficult times on my path to this degree. Thanks to Tanja Golubovic-Madjarac of Lloyd's Register of Shipping, whose initial help opened the door to a world of opportunities.

Without my family I would not have been able to get this far. My love and gratitude go to my parents, Judita and Boza, from whom I gathered strength, optimism, and faith; to my sister, my guardian-angel Lada; and to always present and loving Nani, Marta, Srecko and Ognjen.

Finally, many thanks to my colleagues and mates at the Telecommunications Research Lab of QMWC: Steven, Arif, Maria, Babul, Antionio, Peter, Paula, Eliane, Husam, Felicia, Sutha, Sammy, and Phil from ISAR Group, for their help, support, and all the laugh that kept us going on.

INTRODUCTION

Within the context of broadband integrated satellite and terrestrial networks, many questions have been raised regarding successful transmission of ATM over satellite links. The main challenges for transmission of ATM over satellite arise from long and varying delays in satellite environment, hostile channel characteristics when compared to optical links, and limitations of satellite spectrum. While a lot of recent research has been focused on mitigating physical layer impairments of satellite channels to develop higher quality links required for ATM transmission, very little attention has been given to the study of higher-layer functions in a telecommunications system carrying ATM. In particular, a problem of medium access scheme that allows efficient use of the limited and expensive satellite spectrum while enabling access to heterogeneous and broadband users has not been adequately addressed. At the same time, the wireless community has been researching the advantages and disadvantages of TDMA and CDMA, the two candidate multiple access schemes for the future broadband wireless communications systems. While CDMA allows cheaper user terminals and the use of diversity techniques to enhance signal reception and thus fight multipath fading, its main shortcoming has been argued to be its inferior capacity when compared to TDMA. Supporters of CDMA argue that if interference-rejection techniques are used in CDMA, its capacity performance can be comparable to that of TDMA. It is crucial for a wireless bandwidth-limited communications system to have a bandwidth-efficient multiple access and corresponding medium access scheme built on it. Given that any medium access scheme in TDMA would inevitably cause long medium access delays in a satellite system due to the necessity to control individual burst transmissions, this thesis takes on to investigate a CDMA-based medium access that can eliminate the need for control of burst transmissions and hence access delay. The reduction of access delay is important for those traffic classes in ATM that are delay sensitive.

The concept of statistical multiplexing in the code domain, presented in this thesis, enables reduction of the access delay while providing the mechanism for maximising the system capacity, a critical parameter in CDMA systems.

The main challenges of the thesis arise from the fact that the system-level performance on the network layer depends on the combined effect of physical and ATM layer

processes: medium access scheme is an ATM-level function, system capacity (maximum number of simultaneous connections) is a measure defined on the network layer, but in a CDMA system, capacity is determined by the physical layer multiple-access interference.

1. SPREAD SPECTRUM AND DS-CDMA

1.1 The principle of spread spectrum

Code division multiple access is a Spread-Spectrum technique used to support simultaneous access of a number of uncoordinated users. Generally, the spread-spectrum techniques are based on the principles described below.

Mathematically, a signal that can be represented as a vector in a signal space of dimension D is masked or "hidden" in the code space of the much larger dimension N, N>>D [Sklar 88]. Physically, a signal is modulated by a noise-like code of much higher bit rate, producing a wide- or spread-spectrum signal that resembles white noise. The receiver can only detect the signal and extract it from the noisy background if it knows the spreading code (also called a signature sequence) that has been used to mask (spread) the signal. Thus despreading is done by correlating the signal with a synchronised replica of the spreading code.

The benefits of using spread-spectrum techniques are:

- Reduced energy density required to achieve low probability of intercept in military communications [Sklar 88]
- Good resistance to multipath fading in wireless communications
- Interference suppression (CDMA)
- Fine time resolution (accurate ranging and tracking) [Sklar 88, Gilhousen *et al.* 91].

The reference (replica of the spreading code) can be transmitted or stored [Sklar 88]. A transmitted reference can utilise a truly random code, while a stored reference must use a pseudorandom or pseudo-noise code (since it has to be stored or generated at the receiver). Most commercial communications systems use a pseudorandom (stored) code

Pseudorandom codes must have certain randomness properties, which are concerned with the number of zeros and ones in a binary code, their distribution and the value of the code auto-correlation (balance property, run property, correlation property) [Sklar 88, Lee 91]. Different code families must all satisfy the randomness properties, but may have different performance properties. These comprise:

- Even and odd auto-correlation function and its mean square value
- Code even and odd cross-correlation function and its mean square value
- Secrecy of code (traceability)
- ease of synchronisation

When designing a system, an appropriate code or code family must be chosen according to their desired performance properties. However, code design is not an issue in this thesis and therefore the details of code design and the mathematical theory behind it will not be covered.

1.2 Types of spread spectrum techniques

There are three types of spread spectrum techniques and their employment is connected to the type of modulation used:

- Direct Sequence (used with phase modulated signals)
- Frequency Hopping (used with M-ary frequency shift keying), and
- Time Hopping (used with pulse position modulation)

1.2.1 Direct Sequence spread spectrum

Direct sequence spread spectrum is used with phase modulated signals. The coded (spread-spectrum) signal modulates the phase of the carrier. In BPSK, it is equivalent to the multiplication of the modulated signal with the (binary) spreading code. Each information bit is coded by a sequence of bits (also called chips) which represent a complete or a part of a PN code. Therefore this PN code is running at a much higher bit rate than the data signal. In the frequency domain, the spectrum of the original signal is spread to the spectrum of the high-rate PN code. Thus the resulting signal that is being transmitted occupies much higher bandwidth than the original signal before the spreading. At the receiver end, the received signal is correlated with the stored replica of the PN code. Provided both ends are synchronised, the correlation of the code with itself results in the constant value 1 for all chip positions, therefore unmasking the original signal. At the same time, all interfering signals that corrupted the transmitted signal become multiplied by the spreading PN code at the receiver for the first time, and therefore they become spread in spectrum and white-noise-like, just like the original

signal prior to transmission. The original signal can now easily be filtered out (see Figure 1.1).

The ratio of the spread spectrum and the original signal spectrum is called processing gain. It is equivalent to the ratio of the data bit period and the chip period, $G_p = \frac{T_b}{T_c} = \frac{W_{SS}}{R_b}$. Processing gain is often expressed in decibels.

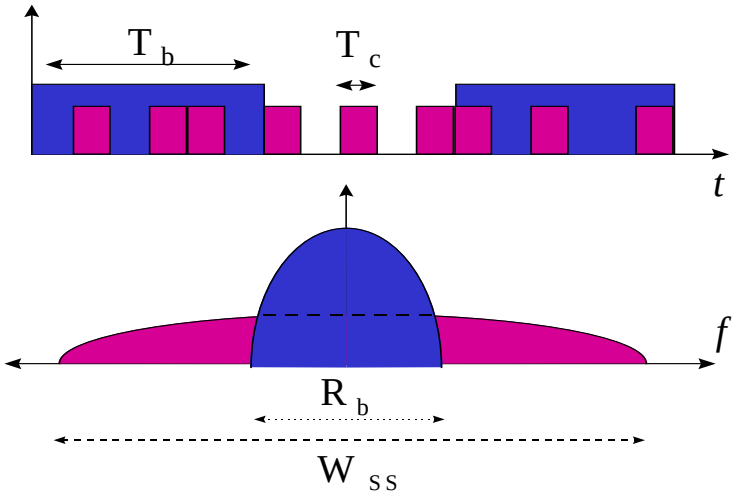


Figure 1.1 Illustration of DS-CDMA spreading and despreading process in time and frequency domain.

1.2.2 Frequency Hopping

In frequency hopping, the modulating frequency follows a random pattern dictated by the spreading code. Thus the frequency of the carrier modulated by the signal is being changed in a random fashion. The number of frequency channels over which the signal hops is the processing gain, since the spectrum occupied by the signal is equal to the original signal spectrum multiplied by the number of channels. Here the frequency channels are called chips, by analogy with the chips in direct-sequence spread spectrum. There are two types of frequency hopping: slow hopping and fast hopping. In slow hopping, a number of symbols is transmitted during one hop, i.e. using one frequency channel (chip). In fast hopping, each symbol is transmitted using a number of chips (hops, or frequency channels) [Lee 91].

1.2.3 Time Hopping

Time hopping is not truly a spread-spectrum technique, but it is similar to spread-spectrum in that the position of a pulse is randomly changed in time [Sklar 88].

1.2.4 Spread spectrum as a multiple access technique

The focus of the thesis is the ATM level system performance when a spread spectrum multiple access technique is used in broadband satellite communications based on ATM.

There are a number of multiple access techniques, depending on the domain in which the signal space is shared among the remote users:

- FDMA - frequency division MA
- TDMA - time division
- CDMA - code division
- SDMA - space division
- PDMA - polarisation division

In FDMA each user gets allocated a frequency channel for its sole use, in TDMA it is a time slot in a time frame. In CDMA all users transmit in the same frequency band for all of the time, but they are separated by the codes or signature sequences allocated to them. The last two multiple access techniques are never used on their own, but in combination with the first three to enhance the system capacity. Space division makes use of the antenna sectorisation and resulting spot beams. The users are separated by the direction from which their signals radiate. Polarisation division relies on orthogonality of different signal polarisations.

As will be explained later, one of the main problems with CDMA is increasing its interference-limited capacity. The advantage of CDMA that can be used for the capacity increase is that it can 'squeeze' extra users into the system without destroying the existing connections but only degrading their quality proportionally. In the other two multiple access schemes this is not possible [Lee 91]. However, frequency hopping can only be used to increase the capacity if fast hopping is employed, because of the possibility of using diversity techniques[Lee 91]. Slow hopping does not allow averaging out across all channels to give on average good performance. If bad channels occur in a slow hopping FH system, all channels get pulled down to unacceptable

quality level [Lee 91]. This is because the symbols going through the bad channels become corrupted, causing the degradation in performance that is in addition to the degradation caused by the CDMA multiple-user interference. Since all signals must use the same set of channels, all signals experience degraded performance if bad channels occur. In fast hopping, one symbol occupies a number of channels and therefore if one of the channels is bad, it can be compensated for by the rest of the channels through diversity combining. However, the technology for fast hopping at 800 MHz (current land mobile band) is very complex and expensive, because of the difficulties in maintaining the frequency and phase coherency during the rapid frequency changes [WEB/MS]. In satellite communications, fast frequency hopping requires technology to actively and quickly compensate for large Doppler shift due to the satellite movement. Because of the technology limitations, fast frequency hopping is mainly used for slow data rates [WEB/MS]. This is why in CDMA it is the direct sequence technique that predominates, and that is the most viable for the use in satellite communications.

CDMA can be asynchronous or synchronous. In asynchronous CDMA a limited number of codes is distributed to the users among whom there is no synchronisation to the common clock, the signals using different PN codes always have some low cross-correlation, which results in low efficiency of the channel capacity utilisation.

In synchronous CDMA perfectly orthogonal PN codes are allocated to users who are fully synchronised (there is a common system clock and time reference). Because of the same starting times of spreading codes and data signals, resulting cross-correlation between them is zero, thus increasing the efficiency of the system compared to the one with asynchronous CDMA. A disadvantage is the requirement for a complicated co-ordination mechanism for PN code allocation and synchronisation of all the users. Propagation delays of the signals need to be constant in order for this technique to be used efficiently.

DS-CDMA systems can also be classified, according to the length of the code period they employ, into [Kärkkäinen 96]:

- short-code DS-CDMA - also termed deterministic (D-CDMA)
- long-code, also termed (pseudo)random DS-CDMA (R-CDMA).

According to [Kärkkäinen 96], the above terminology refers to the "...nature of the spreading code during each data symbol period. In a long-code or R-CDMA system the aim is to randomise the interference from other users and treat them as additive white Gaussian noise, whereas in a short-code or D-CDMA system, an attempt is made to control the amount of MAI [multiple access interference] through suitable choice of a short PN code family."

Long codes are used in military systems and in commercial systems where length of the code allows individual addressing of all users in a large CDMA system. If short codes are used, they need to be re-allocated or re-used since the number of codes in a code family corresponds to the length of the code. However, in short code systems the advantage is that the data-bit to data-bit interference is controlled and lower than in the long-code systems, since the duration of the code period is equal to the duration of the data bit, and the code repeats itself for each data bit. Simplicity of code allocation control and hence network management due to the almost infinite number of codes available in a system with long codes is paid for by the reduction in performance: the data-bit to data-bit interference in a long-code system is higher and uncontrolled [Kärkkäinen 94] resulting in degraded performance, i.e. an SNR on average 1 dB lower than in short-code systems.

From this argument the limitation on the number of users in the system imposed by the number of available codes can be removed if the long codes are used, trading off some of the performance. According to [Kärkkäinen 94], the typical reduction in signal-to-noise ratio when long codes are used is of the order of 1dB.

1.3 CDMA: Issues

CDMA is suitable for fighting interference from other users and for mitigating destructive multipath effects. However, it is an interference-limited scheme, and as such its performance gradually decreases with the increase of the number of users in the system. Research into the capacity of CDMA systems has yielded controversial results. Early work has consistently shown that CDMA system capacity is inferior to the capacity of an equivalent TDMA or FDMA system. However, in early 1990's, when it was shown that the CDMA is only an interference-limited scheme (and not bandwidth-limited, like TDMA and FDMA [Gilhousen *et al.* 91], new research showed that under

certain conditions, capacity of a CDMA system becomes comparable or even superior to that of TDMA and FDMA systems. Such are the results of: Monsen, who found that under conditions of uniformly distributed traffic over many mobile cells [Monsen 95], van Nee *et al.*, who showed that if path diversity techniques are used CDMA is better than the slotted ALOHA [van Nee *et al.* 95], and Elkaheem *et al.*, who showed that at low and medium input traffic loads particular CDMA-based schemes outperform TDMA despite of the fact that voice activity detection and frequency re-use were ignored in this work [Elhakeem *et al.* 94]. In other words, any reduction in interference translates into the increase in capacity [Gilhousen *et al.* 91].

In satellite systems a capacity inefficient scheme is not justified. The main motivation for the choice of CDMA as opposed to TDMA in a satellite system is its resistance to fading and jamming [Geraniotis *et al.* 94], and simplified (cheap) hardware due to less stringent synchronisation requirements [Bella 96]. Since communications became a competitive industry, where the cheap and reliable service will dictate the revenue, CDMA can be expected to become the dominant scheme in future satellite systems if it can achieve a capacity comparable to TDMA.

Thus the main question in CDMA is how can its capacity be increased while maintaining the system performance?

The question of the appropriate performance measure for a CDMA system has been addressed by [Kärkkäinen 96]. Different parameters are proposed as performance measures, but for a commercial communications system (as opposed to a military system) we can assume that the SNR is appropriate since it gives the BER. (In military communications, code secrecy may be more important performance parameter [Kärkkäinen 96].)

CDMA system capacity is limited by two factors: link BER determined by the system self-interference, and the total number of available code words.

1.3.1.1 BER

The average BER depends on the received signal to noise plus interference ratio (SNR). The link BER increases with the number of users in the system, since all users contribute to the system 'self-noise' or self-interference. In the CDMA system the

interference of other users seen by a single (tagged) user results from the following factors:

- noise-like spread spectrum power of other users that falls into the band of the tagged user and thus cannot be filtered out; the amount of spread spectrum power of other users depends on:
  a) accuracy of power control used to mitigate the near-far effect (signals from users that are geographically closer to the resource controller - satellite or base station - arrive at the controller receiver with greater power than the signals from the users which are further away) [Kou & Leib 96, Delli Priscoli & Sestini 96, Vojcic *et al.* 94, Jalali & Mermelstein 94];
  a) chip rate i.e. spreading bandwidth [Noneaker & Pursley 94], which determines a parameter called processing gain;
- cross-correlation of the codes of different users; it depends on the choice of the code family [Pursley 77, Pursley & Roefs 79, Nazari & Ziemer 86, Nazari *et al.* 87a, Nazari *et al.* 87b, Sarwate *et al.* 84, Kärkkäinen 94, Mazzini & Tralli 98]; The code cross correlation and the processing gain are somewhat related, this is explained in Appendix A.

While the accuracy of power control is crucial for the performance of spread spectrum multiple access [Gilhousen et al. 91, Newson & Heath 94], for the study of ATM over CDMA reported in this thesis, power control and fading effects will not be taken into account, since they are effects influenced by the environment and as such are to be solved on the physical layer. However, cross correlation and other-user interference effects are intrinsic to the CDMA concept regardless of the physical medium characteristics, and thus are taken into account in the results for ATM level system performance.

1.3.1.2 Number of codes

The size of the code set (number of codes) for a specified maximum cross-correlation is determined by the code length, L (also known as the code period), and varies between $L^{½}$ and $L^{3/2}$ [Kärkkäinen 94]. Most CDMA applications use the short code sequences with code length equivalent to the number of chips in a data bit; that is also equal to processing gain ($L = T_b/T_c = N$). As mentioned earlier, such code sets have controlled and low cross-correlation, but necessitate code re-use in large systems. On the other

hand, long codes (such that the code period L >> N) result in uncontrolled data-bit-by-data-bit cross correlation, but allow individual user addressing and remove the need for code re-use. Furthermore, in mobile communications the code does not need to be changed in a handover procedure, since the number of codes is large and they can be allocated to users uniquely.

In the following subsections, first multi-user interference will be discussed as obtained from the mathematical analysis of code cross-correlation and the received signal to noise ratio. A problem in determining the theoretical value of the multi-user interference in a multi-rate system will be described. Next, implementation issues concerning the choice of the code family and its influence on the capacity will be briefly discussed. Finally, a problem above labelled as a "second-order" problem for this particular study, but nevertheless important for the successful implementation of CDMA system - the problem of power-control - will be explained, and its coverage in the literature concerning broadband or multi-rate systems will be reviewed.

1.3.2 Capacity and QoS

In ATM, the QoS is measured in terms of the cell loss ratio (CLR - translated from BER on the physical layer), cell transfer delay (CTD) and cell delay variation (CDV). Other performance measures such as cell misinsertion ratio, cell error ratio and severely errored cell block ratio [I.356] are defined. For the purpose of this study we concentrate on cell loss ratio, and on cell delay where the delay is affected by the use of CDMA. Cell delay and cell delay variation resulting from networking functions such as switching and multiplexing, or from physical layer network structure such as long propagation paths in satellite links, will not be addressed since they are results of phenomena unrelated to the use of CDMA.

1.3.2.1 Capacity and processing gain

Spreading of the users' bandwidth gives the individual user effective power advantage over the interfering noise from a jammer (but not white noise). After de-spreading of the wanted signal, the signal power in the resulting signal-to-noise ratio is multiplied G_p times, where G_p, the processing gain, is the ratio of spread to signal bandwidth, W_{SS}/W_s. That is [Pickholtz *et. al*. 87]:

$$SNR = \frac{E^2[U]}{\mathrm{var}(U)} = \frac{E_S}{E_J} \cdot G_P \qquad \text{Eq. 1.1}$$

where:

U is the output of the correlator at the receiver,

E_S is the energy-per-bit of the wanted signal

E_J is the energy-per-bit of the jammer, and

G_P is the processing gain

$$G_P = \frac{W_{SS}}{W_S} = \frac{1/T_c}{1/T_b} = \frac{T_b}{T_c} = N\frac{chips}{bit} \qquad \text{Eq. 1.2}$$

T_c - duration of the chips (bits) in the code sequence;

T_b - bit period of the signal data sequence.

Hence the bigger the processing gain G_p, the greater is the effective power advantage of wanted signal over the jammer and interference power. If we put this into the context of satellite channel dimensioning, this means that the capacity (tolerance to self-interference) is greater if the individual CDMA channels are wider in bandwidth. But that yields fewer channels per satellite beam (since satellite beam bandwidth is limited). Obviously an optimum channel organisation (number of channels and their bandwidth) which gives the highest capacity has to be found for a given satellite band.

1.3.2.2 Capacity and cross-correlation in single-rate systems

PN codes or codes of a particular code family are characterised by their auto-correlation and cross-correlation functions. These functions determine how well the code despreading will extract ("amplify") the spread (hidden) signal from the collection of many noise-like coded signals. The lower the cross-correlation functions between the two codes, the better will be the rejection of the unwanted signals at the receiver during despreading. The cross-correlation function is defined as [Pursley 77, Nazari *et al.* 87a]:

$$R_{k,i}(\tau) = \int_0^{\tau} c_k(t-\tau)\cdot c_i(t)\,dt$$
$$\hat{R}_{k,i}(\tau) = \int_{\tau}^{T} c_k(t-\tau)\cdot c_i(t)\,dt \qquad \text{Eq. 1.3}$$

Here $R_{k,i}(\tau), \hat{R}_{k,i}(\tau)$ are the continuous time even and odd partial cross correlation functions between codes *k* and *i.*

The finite non-zero cross-correlation between the spread signals of multiple users depends on the choice of the family of codes.

Design of a CDMA system requires a trade-off between the number of users allowed in the system, processing gain and user energy-per-bit to noise spectral density (E_b/N_0), in order to achieve the required bit error rate (BER). One of the earliest results in the analysis of CDMA performance gives an approximation for the average signal-to-noise ratio in a system where all the effects apart from cross-correlation are neglected (like power control and near-far effects, fading and multipath), and the result is a function of a total number of users *K*, processing gain G_p and user energy-per-bit to noise spectral density E_b/N_0 [Pursley 77, Turin 84]:

$$SNR = \frac{E^2[U]}{\mathrm{var}(U)} = \left| \frac{K-1}{3G_p} + \frac{N_0}{2E_b} \right|^{-1} \qquad \text{Eq. 1.4}$$

where:

- K is a number of users in the system (limited by the number of codes in the code set);
- N_0 is the single-sided noise spectral density.

The above approximation illustrates how E_b/N_0 of a user and processing gain G_p (code-sequence length in this case) can be chosen to achieve the required signal-to-noise ratio (that is, BER) at the receiver for a given number of users *K*.

Expression 1.4 assumed that all users transmit with equal power, E_b/T_b, and equal data rates, $R_b = 1/T_b$. If all the codes from the code family are used, or if the system self-

interference reached the threshold level, no users will be accepted. If either limiting condition is met, the new connections will be denied access to the system resources.

The total BER of a BPSK CDMA channel is given by the expression [Sklar 88, Pickholtz *et. al*. 87]:

$$BER = Q(\sqrt{SNR}) = Q\left(\sqrt{\left| \frac{K-1}{3G_p} + \frac{N_0}{2E_b} \right|^{-1}}\right) \qquad \text{Eq. 1.5}$$

where Q(x) is a function defined as:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du \qquad \text{Eq. 1.6}$$

The problem of CDMA performance evaluation is complicated in systems where the connections are allowed to have different bit rates, as in ATM. This will be clear when the details of the mathematical analysis of CDMA are described in the next section.

1.3.2.3 Cross-correlation in multi-rate systems

So far only the problem of same-rate systems has been researched and addressed [Pursley 77, Sarwate *et al.* 84, Nazari *et al.* 87a,b]. The signals are assumed to have identical data rates, so that the number of code chips per data bit (i.e. the processing gain) is the same for all users. Due to the condition that the signals have the same data rates, the data bit duration is the same for all users. This has important implications for the value of the code cross-correlation, and eventually BER.

Mean cross correlation value is calculated by observing the pairwise cross-correlation functions of the wanted signal and all other *K*-1 interfering signals. The assumption about identical rates (same duration of data bits) means that during the data period of the desired signal there is at most one data transition of each of the other *K*-1 the interfering signals. That is why the interfering signals could be analysed using only two continuous time partial cross-correlation functions. The assumption leads to certain results of cross-correlation values which then yield an expression for SNR at the receiver of a CDMA system given above by equation 1.4. If a multirate system is to be analysed mathematically, evaluation of cross-correlation gets complicated due to the fact that there may be any number of transitions of the interfering signal during one data period of the desired signal. With multiple rates, pairwise interference of signals will differ

from one interfering user to the next, and will depend on the ratio of the data rates of the wanted and respective interfering users. This raises a question how CDMA can be used in a multirate system.

The question of multi-rate CDMA will be discussed in Chapter 2: ATM over CDMA.

1.3.3 Implementation issues

Implementation of CDMA requires a choice of a suitable code family and efficient power control function to eliminate the near-far effect. A particular problem in broadband i.e. multi-rate systems is how a number of different data rates can be accommodated. A number of solutions have been proposed in the literature.

Fong *et al.* [Fong *et al.* 96] suggested adjustment of actual information rates to some intermediate line rates prior to spreading. When data rates are higher than the line rates, high-bit-rate streams are split into a number of parallel slower streams. Such slower streams are spread using orthogonal codes from a short-length code family. Pseudorandom (PN) codes are then used to scramble orthogonal codes. The system performance is expressed in terms of capacity, but the emphasis is on the analysis of the properties of concatenated orthogonal/PN codes and their performance in different propagation environments. Selection of line rates to achieve capacity increase has been investigated: a single rate system was found to perform better than the system with two or more line rates within a same channel. All traffic shared the same channel, separation between users was achieved by orthogonal codes. Integration of two traffic types was considered, voice and video, where video was assumed to have an activity factor equal to one (transmitting all the time), i.e. only voice traffic is assumed to be bursty in nature. This assumption is limited compared to the ATM multiplexing scenarios. Performance analysis was done on the physical level, in contrast to the performance analysis on ATM level presented in this thesis.

The European CODIT project [CODIT 95, Andermo 95, Andermo & Brismark 94, Baier 93, Baier & Panzer 93] proposed a system where a number of different channel bandwidths are used to accommodate varying user data rates. Long PN codes were used to separate signals within the same channel, and there are three channel bandwidths determined by the three set chip rates of the PN codes. The channel in which a signal will be transmitted is selected according to the signal data rate and the required quality

of service (which depends on the processing gain). That means that within the same channel spread signals with varying processing gains can coexist (since signals of different data rates are spread by the PN sequence of the same chip rate), resulting in their different noise immunity and rejection ability. This is somewhat mitigated by the use of closed- and open-loop power control. In CODIT, variable bit rate voice and bursty data traffic were taken into account. Numerical results for the system capacity were given only for the case of speech and 64 kb/s data, with DTX being applied only to speech [CODIT 95]. System parameters were determined from simulation, no traffic analysis was presented.

McTiffin *et al.* [McTiffin *et al.* 94] suggested implementation of multiple rates similar to both above mentioned approaches. They differentiate between the downlink and uplink: on the downlink, they suggest the use of orthogonal short codes, where the higher rates are supported by parallel base-rate streams (similar to Fong *et al.*) and the processing gain within a channel is constant (i.e. both the chip and data rates are constant), while on the uplink long PN codes are proposed resulting in different processing gains for different data rates (concept similar to the one used in CODIT project).

To summarise, three possible implementations of multiple rates in a DS-CDMA system have been proposed in the literature:

1. Signals of different data rates are all spread to the same spread bandwidth by a long PN code, resulting in varying processing gains. Widely ranging data rates can be accommodated by having a number of channels of different bandwidths (chip rates).
1. Signals of different data rates are adjusted to the same base or line rate, and then spread using orthogonal codes. In case of higher rates, signals are subdivided into parallel lower-rate streams. Processing gains are equal for all signals. All channels have the same bandwidth.
1. Concatenated orthogonal/PN codes are used in a 'combined' configuration: data rates are adjusted to a limited number of line rates that are spread using long PN codes over the entire bandwidth. When data rates are higher than the available line rates, orthogonal codes are used for their splitting into parallel streams prior to spreading by a PN code. Only one chip rate is used (that is, one channel

bandwidth), thus processing gains can vary according to the limited number of intermediate line rates.

2. ATM OVER SATELLITE

This chapter describes the issues that need to be studied and resolved in order to make ATM over satellite work. It focuses on the overview of multiple access schemes used for conveying ATM over satellite, and associated medium access control protocols. The chapter provides a wider background to the problem of ATM over DS-CDMA in a satellite environment, addressed by the thesis.

2.1 Effects of satellite environment on ATM

ATM technology has its own layered structure that is analogous, but not identical, to the ISO-OSI reference model. To find out how ATM can be employed in satellite communications it is necessary to understand the satellite environment and the satellite technology dictated by it, to analyse how this environment affects ATM on different layers in the ATM reference model, and to identify what research has to be done in order to the make the two technologies work together (satellite and ATM).

The effects of the satellite environment on ATM differ depending on the satellite constellation being considered. That is because the environmental phenomena in satellite communications depend on satellite orbit geometry, which is determined by orbit eccentricity, radius (or altitude) and inclination. Satellite orbits used for communications are mostly spherical (eccentricity e=1). Depending on the altitude (orbit radius) three basic types of satellites may be distinguished:

1. low earth orbiting satellites (LEO: altitude up to 5000 km),
1. medium earth orbiting satellites (MEO: 5000 - 20000 km), and
1. geostationary satellites (GEO: approx. 36000 km).

The channel characteristics differ in these three environments.

2.1.1 LEO satellites

Due to the low altitude of these constellations, the global coverage can be achieved only with a large number of satellites. This necessitates frequent handovers between the satellites and consequently complex traffic control functions. LEO satellites allow high frequency re-use and consequently large system capacity. Links have small propagation delays (typically 30 ms) and low free space attenuation, which enables the use of small

terminal antennas. However, there is high Doppler shift which makes synchronisation mechanisms complex.

2.1.2 GEO satellites

GEO constellations eliminate the need for handovers, since the satellites are quasi-stationary. The Doppler effect is significantly reduced, near-global coverage can be achieved by only three satellites, although polar regions are not covered. The trade-off for that coverage is a large propagation delay and free space loss requiring large antennas with high gain and bulky earth stations, and low elevation angles at high latitudes [Vatalaro *et al.* 95, Wu *et al*. 94]. Traffic control functions are simpler than in LEO or MEO constellations thanks to the lack of handovers, but there is high latency in the links.

2.1.3 MEO satellites

MEO constellations trade-off the problems between GEO and LEO constellations. Global coverage can be achieved by 6 satellites, and larger constellations are possible. Handovers are generally less frequent than in LEO constellations, allowing easier connection management.

The effects of a satellite environment on ATM and higher layer protocols can be classified as follows:

1. Physical layer impairments:
    - High channel error (random errors and burst errors) affect ATM cell loss ratio, cell errored ratio, misinserted cell ratio;
    - Varying and long propagation delays affect ATM cell delay and cell delay variation;
1. Data link layer impairments:
    - Retransmission protocols for end-to-end error recovery are affected by the long propagation delays, and will therefore have an impact on ATM QoS.
1. Network layer impairments:
    - Bandwidth is an expensive resource in satellite communications, and this imposes different constraints on the approaches to traffic management functions that allocate network resources.

- Handovers affect the ATM cell stream: the duplication and misinsertion of cells, the required complexity of traffic management functions such as connection admission control, medium access control and routing, and network level performance such as call blocking or dropping.
- Dynamic topology changes and intersatellite links have an impact on routing strategies.
- Multiple access schemes: New systems with regenerative/switching satellites allow the separation of uplink and downlink, and therefore the design choices for the two multiple access schemes may be different. Multiple access schemes determine the performance of connections admission control (CAC) and medium access control (MAC), which should be bandwidth- and time-efficient.

1. Transport layer impairments:
   - transport layer protocols are designed to ensure reliable end-to-end transmission in terrestrial networks where propagation delays are shorter. Different transport protocols that are robust to long propagation delays are required in a satellite environment.

2.2 Multiple Access

Multiple access schemes (MAs) are a concept in which a particular domain i.e. medium of the physical layer representing the network resource (bandwidth, time frame, space, etc.) is divided and shared among a number of uncoordinated users.

Medium Access Control (MAC) is a physical layer dependent protocol which manages the allocation of the resource i.e. the medium using the underlying medium "organisation" defined by the multiple access method. MAC is concerned with the management of traffic on the data layer, and in ATM is analysed on cell or burst level (cell or burst time scale traffic variations).

Connection Admission Control (CAC) is a traffic management function performed on the connection set-up that decides whether or not a new connection can be accepted into the network. It is concerned with the network layer management and is analysed on the connection level (call time scale).

The above classification is intuitive and is derived from the CAC definition and MAC proposals in terrestrial ATM networks [BAF2 92, I.371]. They are clearly functional entities operating on different time scales and ISO-OSI layers. However, in wireless communications (both land mobile and satellite) the clear difference between the two often disappears, and CAC and MAC start to overlap in the description of the 'access' protocols proposed in the literature. The problem of clear differentiation between the two will be addressed later, and the solutions for CAC and MAC operation proposed in the literature under different multiple access conditions will be reviewed.

Satellite communications use FDMA, TDMA, or CDMA, in combination with SDMA and/or PDMA to increase system capacity. More recently combined multiple access schemes have been proposed (like hybrid FDMA/CDMA [Eng & Milstein 94], multi-frequency TDMA, MF/TDMA [Hung *et al.* 96], and some others in [Elkaheem *et al.* 94]). Among FDMA, TDMA and CDMA, TDMA has the highest throughput for a given number of users but requires expensive earth station equipment due to stringent synchronisation requirements [Maral 93], whereas CDMA has the best 'immunity' against interference, allows cheap equipment, but yields low throughput unless additional interference rejection techniques are used. FDMA is generally an inflexible scheme, and in new systems is being used only as part of hybrid techniques (MF-CDMA, MF-TDMA).

Terrestrial networks use mainly TDM and TDMA, apart from some optical networks which are based on wavelength division multiple access. The assignment in terrestrial networks can be fixed (e.g. ATM over SDH), random (pure ATM) or controlled i.e. demand based (like in a permit-based medium access control protocol for passive optical network [BAF2 92]).

The optimal satellite access scheme for ATM, or more generally, broadband communications has not been decided yet. Within the research area of ATM over satellite, and broadband communications via satellite, mostly TDMA was used in tests and demos (except the Vantage project, where FDMA was used). CDMA was not considered at all. The CODIT project [CODIT 95] investigated a broadband application of CDMA in the context of future personal land mobile communications. [Akyldiz & Jeong 97] suggested MF-TDMA since it offers reduced terminal cost and antenna size with TDMA capacity. Generally, an optimal multiple access scheme would need to

improve power efficiency, reduce satellite antenna size, and use satellite bandwidth efficiently. In this study, only the last criteria will be considered, as antenna design and power design are subjects of completely different fields of study.

As CDMA has been described in detail in the first chapter, in the rest of this chapter the focus is only on the remaining two conventional schemes, FDMA and TDMA.

2.2.1 FDMA

In an FDMA scheme, the assignment of a channel to a user is long term or permanent [Sklar 88]. The main disadvantage is its inflexibility in case of reconfiguration: if a frequency plan is to be changed, both transmitting and receiving frequencies as well as filter bandwidths of the earth stations have to be modified [Maral 93]. FDMA is inherently inferior to TDMA schemes from the message delay point of view [Sklar 88]. Furthermore, FDMA experiences loss of capacity due to the intermodulation products when the number of accesses increases [Maral 93]. For these reasons new satellite systems are opting for solutions other than pure FDMA.

2.2.2 TDMA

Here the system resource is the time frame. It is divided into time slots which are allocated to the users according to their needs for bandwidth and the assignment scheme. One problem with TDMA is the need for complex synchronisation among users whose distance from the satellite varies, and is not known a priori when the user first requests access to the network. In order for the user packets to arrive at the satellite at their assigned time slots without collision, their transmission time has to be calculated based on the relative position of their time slot in the time frame, and their distance from the satellite. Furthermore, users have to buffer their data until the transmission time, which in case of high-rate applications requires large buffers. At transmission time, the buffered data has to be transmitted at high rate, which requires high power terminals and high gain antennas, which increases the antenna size. These requirements result in the high cost of TDMA user terminals.

2.3 Medium Access Control

Medium access control is characterised by the multiple access scheme used (physical layer technique), and the method of assignment.

Assignment (of bandwidth or time slots) can be:

- Fixed
- Random
- Controlled
- Combined/Adaptive [Passas *et al.* 97]

Which one of these will be used will depend on the traffic mix envisaged on the link.

2.3.1 Fixed assignment

The system resources are permanently assigned to a number of users. It is most suitable for steady load traffic, which use fully the allocated resources [Maral 87]. Otherwise, if the traffic intensity oscillates, permanently allocated resources may be wasted.

2.3.2 Random access

Users access the resource when they need it. This assignment scheme results in collision between users, but the scheme can be extended by channel sensing in order to provide collision free access. It is most suitable for traffic characterised by short, bursty, randomly generated messages with long interarrival times [Maral 93]. The trade-off in the case of random access is between the channel throughput and transmission delay: as the former increases, so does the latter. ALOHA and Slotted ALOHA are examples of random access in TDMA, with throughputs of 18% and 36% respectively.

2.3.3 Controlled assignment

The access to a resource may be controlled centrally or in a distributed manner, and the resource (a frequency channel, a time slot) is divided into ‘application’ and ‘allocation’ parts. In the ‘application’ part the users apply or request access to the resource. In the ‘allocation’ part they actually use some or all of the resource according to the controller decision. Controlled assignment is best suited to systems where the user activity tends to be long and random. Demand (DAMA) and reservation (RAMA) assignments are examples of controlled access.

The way in which requests are scheduled and bandwidth distributed depends on the scheduler. The design of the scheduler differentiates the various MAC algorithms that belong to the same class of medium access (e.g. DAMA).

When some form of controlled assignment is used, the need arises for a reservation channel. Essentially the problem of low throughput of random or fixed access not solved, but only moved from all of the channels to the reservation channel: the users still have to access the reservation channel in a potentially inefficient way, either randomly or according to fixed assigned reservation slots.

Any MAC protocol that requires initial handshaking between the satellite access controller and the user terminal would encounter an access delay of the order of at least one-hop delay. Due to the long propagation delays in satellite environment, this will have influence on the delay requirement of delay sensitive traffic (cell delay QoS parameter), as well as congestion control mechanisms. Thus a good medium access mechanism for broadband traffic would have to minimise the effects of long propagation delays while still using the bandwidth efficiently.

One of the main performance measures for efficient MAC is access delay. Here ‘access’ refers to the bursts or data packets, and not connections. Connections access delay is the time it takes for a connection to be admitted by a CAC. Medium access delay depends on the location of the control, that is, whether the medium access controller is located on the satellite or on the central ground (hub) station:

- MAC on the satellite: this arrangement calls for distributed control of access to the medium, which implies communication between consecutive satellites in order to maintain connections during handovers.
- MAC on the hub station: this is a centralised control arrangement, which requires the simultaneous monitoring and control of at least two satellites' coverage areas.

2.3.4 Adaptive combined techniques

Adaptive techniques tend to combine all previously mentioned assignment types in order to support many different traffic types. In satellite systems, speed of reaction of the MAC to the changed traffic conditions will be critical. In fact, the problems that arise in adaptive combined techniques are those traditionally found in the design of adaptive algorithms: speed of reaction to the new situation, and the computational load associated

with the desired accuracy of the control. In order to achieve high capacity and utilisation in a satellite system it is required that the accuracy of adaptive control is high. This increases the complexity and processing power of the computations involved, and consequently the required satellite power.

2.4 Connection Admission Control and Medium Access Control in a satellite system

2.4.1 Connection Admission Control

ATM Connection Admission Control (CAC) as defined by [I.371] is:

> "…the set of actions taken by the network at the call set-up phase (or during call negotiation phase) to establish whether a Virtual Channel (VC) connection or a Virtual Path (VP) connection can be accepted or rejected".

In other words, CAC in ATM networks decides to accept a connection request only when sufficient resources are available to establish the connection end-to-end at its required QoS, while maintaining the QoS agreed for the existing connections. To operate effectively CAC should aim at maximising network utilisation. In summary, the main requirements for CAC are [Ramalho 96]:

- The network has to be protected from overload; resources have to allocated such that the QoS requirements are met for all established connections.
- Maximal statistical multiplexing gains should be obtained.
- The required real-time processing should be reasonable.

Once a connection is established, its access to the medium is guaranteed but needs to be synchronised with other users in some way so that the utilisation of the medium is maximised. This is a task of the medium access control.

In a satellite system, depending on the multiple access scheme of the system and associated assignment method, MAC and CAC may retain completely separate functionalities (as defined above), or may overlap partially or fully.

2.4.2 MAC and CAC in a FDMA system

In a conventional FDMA system CAC is performed once, on the system set-up, and no MAC is required. All users have their allotted frequency bands which are reserved for

their exclusive use. Since new systems are based on TDMA or CDMA, more focus will be put on these two schemes.

2.4.3 MAC and CAC in a TDMA system

In some proposals for satellite systems with TDMA where various traffic types are to be supported, the CAC algorithm either does not exist (connection admission is a result of contention and not controller's decision), or overlaps with the MAC partially or fully, depending on whether the medium access is random or controlled.

2.4.3.1 Random access

In the case of random access, the users transmit when they have packets to transmit, without controller arbitration of which connection will be accepted. Thus no CAC takes place, and MAC is reduced to informing users about successful transmission or collision of packets (e.g. ALOHA and slotted ALOHA).

2.4.3.2 Controlled access

In the case of controlled medium access with some form of slot reservation, the CAC overlaps with MAC to some extent: every time the earth station wants to reserve a time slot in the satellite frame, the controller performs a form of CAC algorithm to decide whether or not the station will be assigned a slot in the next frame. The amount of overlap between the MAC and CAC is determined by the frame structure, i.e. the manner in which different traffic classes are catered for. In general, ATM traffic has been divided into the classes which all have different delay, error performance and bandwidth requirements: constant bit rate (CBR), real time and non-real time variable bit rate (rt- and nrt-VBR), available bit rate (ABR) and unspecified bit rate (UBR). Consequently, proposals for MAC/CAC in ATM satellite systems that appeared in the literature are all geared towards accommodating these various traffic types (or at least CBR and VBR) in the most efficient way, according to their QoS requirements [Ors *et al*. 98a,b].

2.4.4 Traffic integration: separation and overlap of CAC and MAC functionality

Research into traffic integration in satellite systems started with attempts to integrate voice and data, the two traffic classes from the opposite ends of the QoS requirements

scale. Voice requires timely delivery and low delay jitter, while it can tolerate high error rates. Data, in contrast, requires high reliability of the link (i.e. very low error rates), but can tolerate long delays. In order to satisfy the stringent delay and cell delay variation requirements for voice (or more generally, CBR traffic), the trend was to reserve part of the frame for this type of traffic, while the rest would be dynamically allocated to delay-insensitive traffic either randomly by contention or in a controlled way by using a form of RAMA or DAMA. The part reserved for delay-sensitive traffic can again be reserved using fixed, random or controlled assignment. But whatever the connection access is, once a connection is admitted to the network, it is granted a certain amount of the system resources (i.e. a predetermined number of slots) in each frame i.e. time-cycle. For delay-sensitive connections the CAC and MAC is performed once (and at the same time), although the MAC scheduler has to take into account the frame space taken up by the CBR connections when distributing the time slots for outstanding, non-CBR connections.

Dynamic allocation of frame slots to delay-insensitive traffic (data, e.g. nrt-VBR, ABR, UBR) means separation of CAC from MAC for this type of traffic. The traffic can be given access to the medium on a per-frame or per-connection basis.

- In per-frame allocation of resources the CAC and MAC essentially overlap: the users need to reserve slots every time they want to transmit, which requires the CAC/MAC entity to simultaneously accept and allocate the slots.
- In per-connection allocation, the CAC decides whether or not the connection is accepted; the MAC controller then keeps track of accepted connections and it ensures the resource allocation (i.e., the distribution of the time slots).

Obviously, the border between the two traffic control functions is not always clear, since different traffic types are treated on different time scales, unlike in terrestrial networks where the two functional entities are clearly defined.

2.4.5 MAC and CAC in a CDMA system

It is possible to have an analogy in CDMA with all the above mentioned TDMA assignment types. However, not all of the assignment types have been implemented in existing CDMA systems. Instead of slots, a user terminal (i.e. ground station) gets a code. In conventional CDMA systems the analogy is only partial, since the bandwidth

taken up by a spreading-code-modulated signal is fixed, and a station gets only one such code.

2.4.5.1 Fixed Assignment

In fixed assignment, a PN code would be allocated to a user (connection) once for ever, regardless of whether the user is active and transmitting or not. This may cause quick exhaustion of the codes in the code set, limiting system capacity, particularly if short codes are used. As the code set size depends on the length of the codes, for very long codes the set size is large and may not be a limitation on the number of users that can be supported by the system.

2.4.5.2 Random Assignment

In random assignment, users compete for PN codes every time they want to set up a connection. Such algorithms have been described in [Abramson 96] and [Makrakis *et al.* 96] for land mobile systems.

- Abramson proposed a form of spread-spectrum ALOHA, where the ALOHA method is combined with the principles of direct sequence spread spectrum CDMA. The users randomly select a phase shift of a PN code. The throughput of the scheme is the same as for ALOHA, 18%, but the advantage is that the transmitting power of the users is reduced due to the spreading, and the scheme can accommodate wide-band traffic. The CAC is non-existent in this scheme, and MAC has a similar function to the MAC in ALOHA: informing the users about the success of their transmissions.
- The MAC concepts described in [Makrakis *et al.* 96] (multiple access scheme + assignment type) is termed Spread Slotted ALOHA and Spread Slotted Random Access with Multipriority (where users randomly choose both a time slot and a PN code available in the network). A collision occurs if:
    1. more than one user transmits in the same slot using the same PN code, and
    1. more than the maximum allowed number of users transmit in the same slot using other PN codes (thus causing too much CDMA self-interference i.e. BER higher than acceptable).

We differentiate between the random assignment in conventional CDMA and random assignment of spread-slotted ALOHA type in hybrid TDMA/CDMA:

- In random CDMA, once a user terminal gets a code, it can use it for as long as the connection lasts (although on bit, i.e. data packet level, the signal may be bursty i.e. discontinuous in nature). CAC algorithm in this case is performed once, on the connection set-up, and MAC function does not exist, since a user, once given a PN spreading code, can use that code to access the medium throughout its duration.
- In spread-slotted ALOHA, whenever a user terminal has some data awaiting transmission, it has to contend for a resource (a time slot *and* a PN code) and hence CAC does not exist as such, while MAC consists of informing the users of successful transmissions (like in ALOHA).

2.4.5.3 Controlled assignment

Controlled assignment MAC in CDMA has not yet been implemented, and proposals are scarce [Brand & Aghvami 96]. The idea of controlled assignment in CDMA has only appeared since CDMA became the subject of research as a multiple access scheme for integrated traffic [Geraniotis *et al.* 94, Sampath *et al.* 95, Sampath *et al.* 96, Makrakis *et al.* 96]. Unlike in TDMA, controlled assignment in CDMA *does not* imply the controlled allocation of codes, but rather of *bandwidth* i.e. transmission permissions to connections requiring access, so that the overall self-interference is kept low, and system capacity is maximised (in terms of traffic or a number of connections). Controlled assignment often utilises voice activity detection and associated discontinuous transmission, whereby the carrier is switched off or reduced in power in order to decrease the total system self-interference. This increases the total number of acceptable voice connections, while maintaining the average QoS above the required threshold. It would also be possible to allow transmission of delay non-sensitive traffic during the silences between the talk spurts (or generally, during the transmission silences of any delay-sensitive traffic). However, no *data-level* (cell-scale or burst-scale) MAC protocol that employs this idea has been presented in the literature. This is because such a control would not be optimal due to the delays associated with detecting the silence and with commanding the awaiting stations to transmit. The protocol would still be inefficient in resource utilisation and throughput.

This brings into focus the idea of statistical multiplexing: by allowing statistical multiplexing on the wireless link of connections of similar types, and by simply accepting a greater number of such connections based on their activity factors, higher capacity can be achieved with very little computational effort and acceptable QoS. Furthermore, statistical multiplexing in CDMA has an additional attraction: no medium access delay need to be introduced as the nature of CDMA prevents collisions, unless a number of available codes (code set size) is a limitation (there is only 'graceful degradation' if more than the predicted maximum number of users transmit simultaneously). This is the advantage of CDMA-based statistical multiplexing over the TDMA-based one, where transmissions still need to be scheduled in some way in order to avoid collision, and thus long medium access delays inherent in a satellite environment are inevitable.

2.4.6 Traffic integration in CDMA: CAC vs. MAC

A hybrid CDMA/TDMA discrete-time MAC that integrates voice and data for a land-mobile system has been proposed by [Brand & Aghvami 96], called CDMA packet reservation multiple access (CDMA/PRMA). The control arbitration and decision process is done on per-slot basis. The MAC in CDMA/PRMA schedules both data and voice packets according to the voice packet loss probability and the total number of users in the system, in a current time slot. The MAC protocol is distributed and probabilistic, i.e. it is based on broadcasting permission probabilities to transmit. The users perform Bernoulli experiments, the outcome of which is compared to the permission probability in a given time slot. The users can only transmit if the outcome is above the permission probability. Voice and data priorities are regulated by different transmission permission probabilities.

The literature mainly addresses the problem of CAC, and not MAC, although the term 'medium access' is often used. The engineering problem is how to decide if a connection can be admitted, while maximising the capacity (always a problem in CDMA) and optimising performance, given the system constraints. It is implicitly assumed that the code set employed is large, and therefore is not considered a limitation to the system performance. System performance is expressed in terms of delay for data traffic and call blocking probability for voice traffic, and the CAC protocol throughput.

The constraining parameter is the CDMA self-interference, and it is often expressed as a function of a power control mechanism being tested in the investigation. Hence, access control is performed at the connection set-up, as in the CAC in terrestrial networks [Geraniotis *et al.* 94, Yang & Geraniotis 94, Geraniotis *et al.* 95, Sampath *et al.* 95, Sampath *et al.* 96, Sampath & Holtzman 97]. In principle, two types of CAC for integrated voice/data CDMA systems were proposed, both of which attempt to schedule data transmissions when the voice load is low, and curtail them when the voice load is high:

1. Deterministic CAC: the data transmission control can be based on optimisation of a cost function encapsulating the performance objectives of the system (voice call blocking probability and data delay), given the number of active users in the system [Geraniotis *et al.* 94, Yang & Geraniotis 94, Geraniotis *et al.* 95].
1. Probabilistic CAC: the data transmissions are controlled using controller-defined probabilities of permissions to transmit. Transmission permission probabilities are determined from the one or a number of system performance measures (e.g. voice signal to interference ratio [Sampath *et al.* 96], load [Sampath & Holtzman 97]), and are broadcasted on a common channel.

From this survey, it is clear that the problem of medium access in a truly multi-rate, broadband CDMA satellite system has not yet been adequately addressed. While the standardisation of UMTS brought about two proposals for the land mobile air interface, both of which had CDMA (either alone or in combination with FDMA/TDMA) as a multiple access technique, the air interface for the satellite part of UMTS has not been standardised. At the same time CDMA is becoming more widely employed in new satellite systems (e.g. Skybridge [WEB/Alcatel], VoiceSpan [AT&T 95], Globalstar and Arcanet [WEB/Esa]), most of which plan to offer a wide range of data rates (up to 2 Mb/s). The research literature concerning CDMA ‘medium access’ or ‘multiple access’ mainly dealt with either:

- MAC protocols in *land mobile* environments, where propagation delays are low and therefore not a limitation on the protocol performance, or
- within a *satellite environment*, CAC algorithms.

In all cases, the research addressed integrating at most two traffic types from the opposite ends of QoS requirements: voice and data.

The inconsistent use of terminology across the literature covering the problem of ‘medium access’ is a symptom of a multi-layered problem, that requires the study of a number of different time scales and reference layers. In the case of medium access in CDMA, the layers that need to be studied are:

1. physical layer: CDMA self-noise affects the BER perceived by a receiver,
1. data layer: the design of medium access control, i.e. how the packet transmissions from different users are scheduled once the connections have been accepted, and effects on end-to-end QoS;
1. network layer: how many users or connections can be accepted (CDMA capacity) determines the utilisation of the network resources and the network operator’s revenue.

Given all of the above, it is clear that there is a need for medium access schemes that work well with underlying CDMA and can accommodate a wide range of traffic types, in the satellite environment. The satellite environment dictates two important requirements:

1. medium access delay must be reduced in order to improve throughput-delay characteristics of any protocol designed on top of it;
1. expensive satellite bandwidth must be used efficiently, i.e. the number of users that can be accommodated in a given band must be high.

In order to minimise medium access delays in a satellite environment the control has to be on a coarser time scale, i.e. connection level rather than burst or packet level. It was mentioned earlier that statistical multiplexing in CDMA offers such a reduction in delays as the access control is done only on the connection level (by assessing the statistical mixture of existing and newly arriving traffic during the CAC negotiation phase). Statistical multiplexing requires that connections are silent when they are not active (there is nothing to transmit). At the same time, stopping the transmissions while the connections are silent will reduce the waste of bandwidth, and allow an increase in the system capacity. Thus both of the above requirements for a satellite-system CDMA-based medium access mechanism can be met.

This thesis addresses the performance of such a medium access mechanism, based on statistical multiplexing in the code domain. It also proposes a methodology that enables easy and simple satellite system dimensioning for the envisaged traffic mix, for any number and types of traffic.

3. ATM OVER DS-CDMA

This thesis proposes and analyses (at the cell level) three architectural scenarios that exploit the bursty nature of ATM traffic to increase the broadband DS-CDMA satellite system capacity. The bursty nature of ATM traffic is exploited by the use of discontinuous transmission detection (DTX). The novelty of the study is in its focus on the MAC based on *statistical-multiplexing* in the *code domain*. Previous research concerned with multiple access mainly addressed the connection admission control functions (CAC), and medium access strategies on the *call* i.e. *connection time scale* [Yang & Geraniotis 94, Geraniotis *et al.* 94, Sampath *et al.* 95, Sampath *et al.* 96, Sampath & Holtzman 97], and not on the *cell or packet* time scale. Statistical multiplexing, on the other hand, was mainly used and addressed as a concept in the context of fixed, TDM-based terrestrial networks, and not wireless networks. The proposed medium access concept is investigated for the three different system architectures. This medium access concept allows the long delays associated with the satellite environment to be avoided, as there is no need for user-controller handshake prior to the cell burst transmission. Unlike previously published research, this study investigates the effect of the use of discontinuous transmission detection for *all* ATM traffic, and not only voice. Finally, the performance analysis is on the ATM or data packet layer, and not on the physical layer. Numerical results that establish system performance are obtained for a mix of three traffic types, but general conclusions can be drawn for any number of traffic types and any traffic mix.

It is well known that the CDMA system interference depends on the user processing gain, that is the ratio of the spread and user spectra, as this determines the amount of multiple access interference (see Chapter 1: SPREAD-SPECTRUM AND DS-CDMA). The CDMA channel bandwidth is equivalent to the spread spectrum bandwidth. The transmission rate of the user is its data rate as seen by the network. However, the transmission rate does not need to be the actual source bit rate: user data can be buffered and transmitted at a higher or lower rate than they are produced. The choice of the transmission rate should be consistent with the optimum system performance.

We find in this study that instead of ATM just being affected by the physical layer, we have a case of decisions and functionality in the physical layer being affected by the

functioning of ATM. The traffic characteristics of ATM will affect the use and efficiency of the physical layer mechanisms of discontinuous transmission detection. Hence the traffic performance on the ATM layer will determine what organisation of CDMA on the physical layer is optimum for a satellite IBC system.

It has been shown in Chapter 1 that the literature proposed two ways of accommodating multiple data rates in a CDMA system. One proposal was based on keeping the processing gain constant, while splitting the data stream into a number of parallel streams of lower rates and varying the number of bandwidth portions taken up by these streams. A refinement of this concept introduces concatenation of short orthogonal and long PN codes. Short orthogonal codes are used to spread the parallel streams of a connection. Long codes are used for spreading different connections, regardless of whether or not they have been split into streams and modulated by short codes.

The second proposal was based on varying the processing gain, but keeping the spread bandwidth (bandwidth of a CDMA channel) constant. Here connections of different data rates transmit simultaneously in the same channel, but their noise-rejection capability varies due to the different processing gains.

In this thesis an approach is adopted whereby connections of varying data rates are accommodated by adjusting them to an intermediate transmission rate prior to spreading by a pseudo-random code of a much higher rate (chip rate). The focus is on the system performance of the satellite uplink, from the ATM layer viewpoint. Only degradation effects intrinsic to CDMA are taken into account, i.e. CDMA multiple-access interference. Ideal power control is assumed, pure ATM (no framing), and the use of long PN codes, so that the system capacity is not limited by the number of codes available. They also show that the bursty nature of ATM is in this case beneficial rather than detrimental to the system capacity. The performance is assessed with respect to the system capacity and link efficiency (utilisation). The aim is to find an optimum scenario with its system design parameters for a given satellite transponder bandwidth: transmission rate(s), channel bandwidths and number of channels.

The capacity increase in CDMA achievable by DTX has been mainly researched in voice systems [Yang & Geraniotis 94] or multimedia systems with DTX applied only to voice [Fong *et al*. 96]. Work within RACE 2020 CODIT project, as reported in [Baier 93, Andermo 95, Andermo & Brismark 94, Baier & Panzer 93], considered applying

DTX to VBR video traffic as well as voice. However, no detailed cell-level performance analysis has been presented, and the main capacity results (cell capacity, traffic capacity and information capacity) were obtained by ad-hoc simulation and for only two traffic types: 64 kb/s data and 12 kb/s voice, where DTX was only applied to voice [CODIT 95].

In this study, a detailed cell-level performance yields a system of design equations that formalise the decision making process when designing an ATM satellite system. The study investigates the impact of traffic characteristics, as well as the choice of system parameters, on the performance of different scenarios. The thesis is not concerned with source modelling; the main interest lies with understanding how the relative difference in the sources' characteristics and their varying traffic mix affects the performance of statistical multiplexing in the code domain, and how the system parameters that affect this medium access scheme can be chosen to yield optimal performance for the given set of envisaged traffic types. This serves as a starting point for the development of connection-level control schemes in the future satellite IBC networks that use ATM over DS-CDMA.

Where appropriate, mathematical analysis has been validated using a cell level network simulator [MSIM 97]. Alternatively, a hybrid simulation/analytical approach has been used where a full scale cell level network simulation was not an adequate means of validation.

3.1 Introduction

Depending on the transmission rate in the system, two variations of the system architecture are possible:

1. the intermediate transmission rate is fixed across the whole system and is the same for all connections,
1. different connections may have different intermediate transmission rates.

Similarly, depending on the code allocation policy, there are two architectures:

1. code allocation per user and
1. code allocation per virtual channel connection (VCC).

3.1.1 Code allocation in ATM

In a code-per-connection policy each virtual channel connection (VCC) from a source would get a different code. In a code-per-user policy the user would be allocated one code to use for all its connections. Hence this policy would require multiplexing of the connections prior to transmission.

ATM can operate over CDMA in 4 different ways, depending on the system transmission rates and the code allocation policy.

1. All users have the same transmission bit rate (*R(i) = R = const*.) and code per user allocation policy is employed.
1. All users have the same transmission bit rate (*R(i) = R = const*.) and code per VCC is employed.
1. Users are allowed to have different transmission rates (*R(i) =/= R(j), i<>j*), code per user allocation policy is employed.
1. All users have different bit rates (*R(i) =/= R(j), i<>j*), code per VCC is employed.

The third option above (code per user allocation policy with varying transmission rates) does not make much sense unless different user types are specified, in which case the number of traffic types comprising different user types starts to increase, and the system becomes too complex to yield a generally applicable meaningful result. Thus the study focuses on scenarios 1, 2 and 4, renamed as:

- Scenario 1: Code per User, *R = const.*(Figure 3.1)
- Scenario 2: Code per VCC, *R = const.* (Figure 3.2)
- Scenario 3: Code per VCC, *R <> const.* (Figure 3.3)

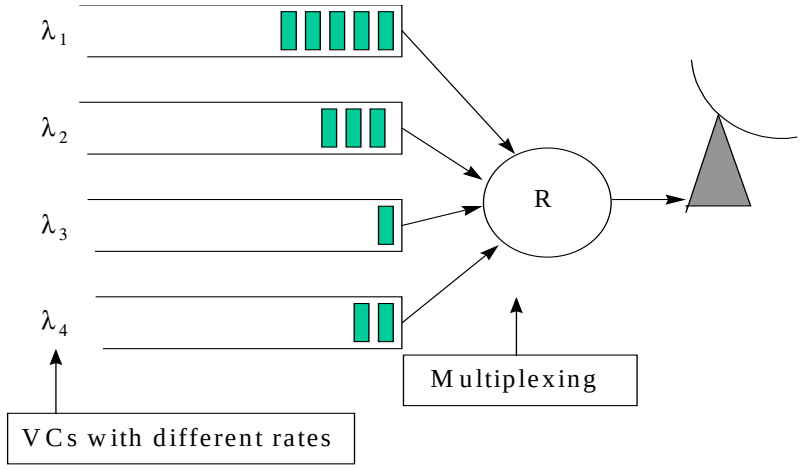


Figure 3.1 Scenario 1: Code per User.

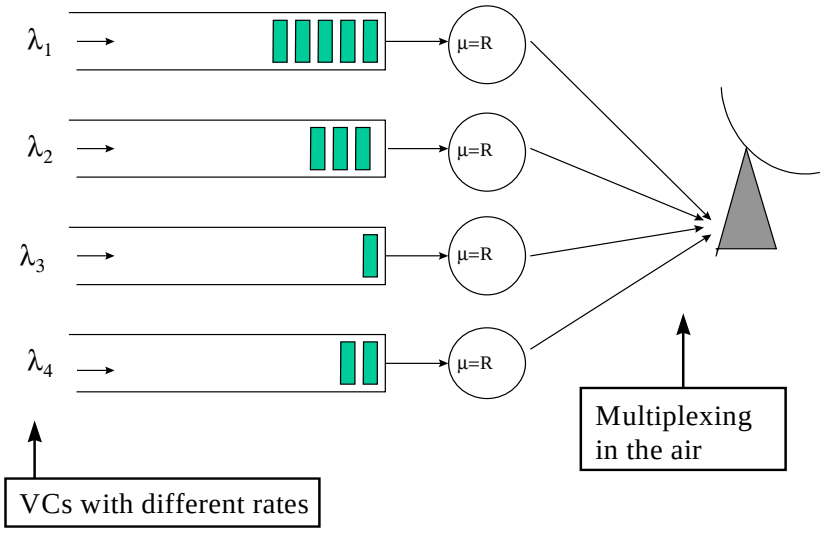


Figure 3.2 Code per VCC, *R*=const.

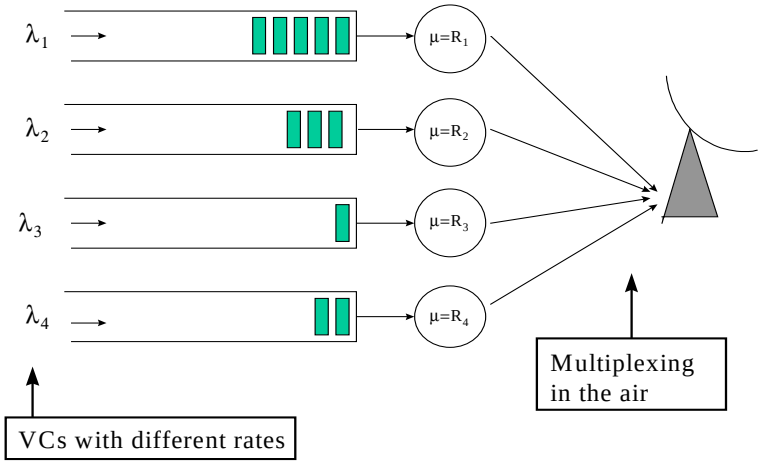


Figure 3.3 Code per VCC, *R*<>*const*.

3.1.2 DTX

DTX relies on the random nature of sources and increases the system capacity by switching off the carrier when there is no data to transmit (the source is idle), thus reducing the overall system self-interference. The drawback of this method is the need for code and carrier re-synchronisation when the source is in the active mode (talk-spurt for voice sources), leading to the requirement for additional overhead and causing increased delay due to the frequent code acquisition. In existing CDMA systems transmission during silent periods is continued at reduced power and bit rate in order to preserve carrier synchronisation. However, the possible increase in capacity in this case may be lower than when the transmission is completely switched off. The trade-off between capacity and efficiency in this case remains to be considered in the future work.

DTX maps to ATM as follows:

When there is no use of DTX, the empty ATM cells are sent even when the source is not active, that is, in its OFF state. With DTX employed, empty cells will still be inserted during active periods of the source (due to the difference in the transmission and source data rates), but during the long periods of silent state no cells are transmitted. The DTX mechanism switches the transmission off when the buffer is empty, and it resumes transmission when the first full cell arrives to the buffer. For pure ATM, the link efficiency is measured as the proportion of useful cells. Link efficiency is increased when the number of transmitted empty cells is decreased. Figure 3.4 illustrates the process of insertion of empty cells. Two links with different source rates (1000 cells/s and 500 cells/s) are shown, and how they can be accommodated over a 2000 cells/s link. Empty cells are represented by white slots.

In order to employ DTX with ATM, it is necessary to investigate the trade-off between the introduced delay and overhead, and the statistical multiplexing gain achieved. Increased overhead leads to decreased link efficiency. Additional acquisition delay affects the user end-to-end delay and therefore the QoS. Illustration of overhead insertion is given in Figure 3.5

3.1.2.1 Code acquisition

Synchronisation overhead is required to re-acquire the code synchronisation when the carrier is switched on. It should be noted that this study didn’t take into account the

overhead required for the carrier synchronisation, as this is done on the physical layer. Code synchronisation, on the other hand, is done on the bit i.e. data level, and its is therefore taken into consideration.

PN code synchronisation depends on the synchronisation method. There are two distinct methods of code synchronisation in CDMA, which represent two ends of a large scale: serial search and parallel search [Sklar 88]. There are many hybrid methods which combine the benefits and trade-off the disadvantages of the serial and parallel search. However, the worst case in terms of search time is the serial search code synchronisation, and this is the one chosen for this study. It is at the same time the cheapest to implement, and therefore is in line with the requirement for the cheap user i.e. VSAT equipment.

The specific result for the code acquisition time in the study was taken from the book *Coherent Spread Spectrum Systems* by J. K. Holmes [Holmes 82]. The mean acquisition time of a code depends on the probability of false alarm and probability of detection.

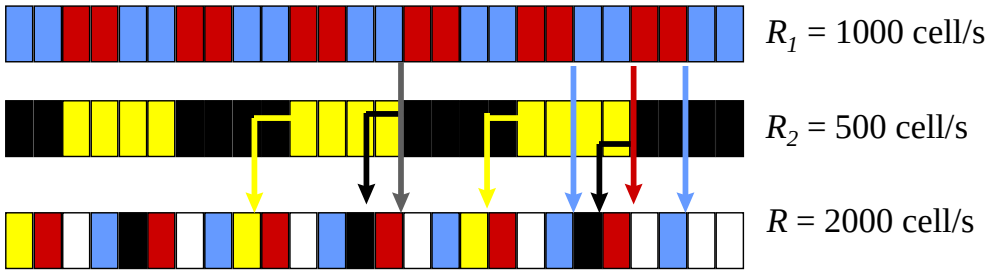


Figure 3.4 Insertion of empty cells between the cells of connection multiplexed streams. R_1 and R_2 are the cell arrival rates of two connections during their active periods, and R is the transmission or cell slot rate at the output of the multiplexer, i.e. on the uplink. The cells are sent on the uplink after they are fully received in the buffer. The arrows show when the arrived cells are sent in cell slots.

Thus it varies according to the QoS requirements of connections (in the Scenarios 1 and 2 it is the QoS of the most stringent traffic type that determines the acquisition time), and a transmission rate. The need for code re-synchronisation causes additional delay of all cells by the mean acquisition time on top of the service time in the multiplexer, expressed as given in equation 3.1:

$$T_{acq} \cong \frac{(2 - P_D)(1 + KP_{FA})}{2P_D} \cdot q\tau_D \qquad \text{Eq. 3.1}$$

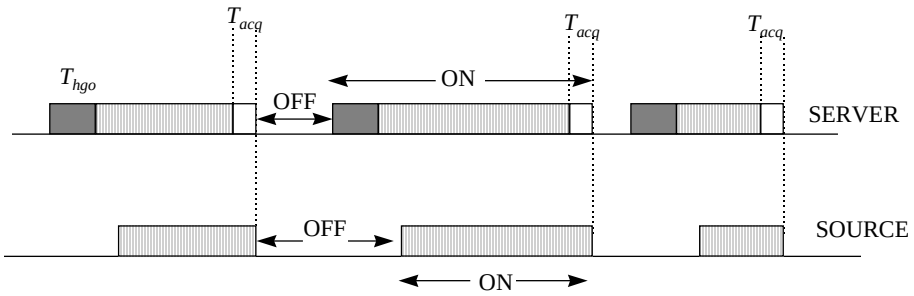


Figure 3.5 Illustration of inserted overhead at the output of the multiplexer, i.e. server.

where the symbols represent:

P_D, P_{FA} probabilities of detection and false alarm, respectively;

τ_D dwell time, i.e. the examination interval over which the locally generated PN sequence is correlated with the incoming PN signal.

q number of positions that are searched through the process; if the search is moved by one-chip updates, then q is equal to the number of chips searched; if the update size is half-chips, then q is twice the number of chips searched.

3.1.3 Capacity, transmission rate and statistical multiplexing

The capacity limit in terms of number of users is set by the QoS i.e. interference limit (BER), and the processing gain. This limit is for the number of users who transmit *continuously*. However, the continuity of transmission depends on a number of factors:

- choice of transmission rate
- connection peak and mean cell rates
- whether or not DTX is employed.

This section discusses the relation between the choice of transmission rates, source rates and use of DTX on one hand, and the system capacity on the other. For clarity, the model of a Markov modulated deterministic ON-OFF source will be used. The source is connected to a buffer of service rate equivalent to the data (intermediate) transmission rate (Figure 3.6).

The lowest possible transmission rate is that equal to the source mean rate, when the signal transmitted is essentially constant bit rate (CBR). In this case the CDMA connection is active all the time. When all connections are CBR, adding new connections above the system capacity would yield deterioration in the QoS of all connections, and the number of cells in the buffer could grow in an unbounded fashion.

As the transmission rate increases from the source mean rate to the source peak cell rate,

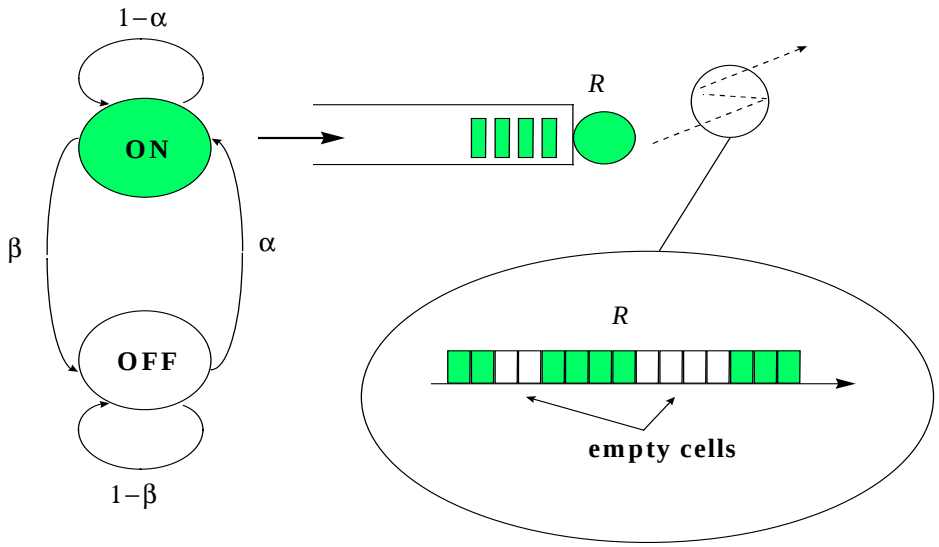


Figure 3.6 Modelling of a source, multiplexer, and cell stream on the uplink.

the activity on the transmission link of the connection decreases. Statistical multiplexing becomes possible by the use of DTX, and the capacity limit as defined by the theoretical result of Eq. 1.4 can be overcome. However, the capacity limits differ in the cases when the transmission rate is higher than the mean rate, and when it is equal to the mean rate, provided the spread or channel bandwidth is kept constant. This is due to the fact that the capacity limit is also determined by the processing gain, i.e. the ratio of the channel (spread) bandwidth and the transmission rate. As the transmission rate increases, the processing gain decreases, provided the channel bandwidth is kept constant. Consequently, as the transmission rate increases, the theoretical capacity limit decreases; this limit can, however, be somewhat increased by the use of statistical multiplexing and DTX.

The achievable channel capacities will be bounded by the two extreme cases:

- Case 1: transmission rate equal to the source mean rate, continuous transmission on the link

- Case 2: transmission rate equal to the peak cell rate, discontinuous transmission on the link, and statistical multiplexing employed.

Intuitively it is clear that the achievable channel capacity will be inevitably higher in the first case, as in the second case additional synchronisation overhead is required at the beginning of each new transmission. However, the queueing effects could be much worse in the first case than in the second. In the thesis a comparison has been made among a number of system instances ranging between the two extreme cases, and the results confirmed the expected trends in the capacity and queueing performance (see Chapter 5: RESULTS).

Benefits of statistical multiplexing in the code domain

Statistical multiplexing in the code domain allows the reduced packet end-to-end delay and buffer requirements. When a bursty ON-OFF source is forced to transmit at its mean rate, the buffer size required for a given cell loss rate in the buffer can be very large, depending on the peak rate and burst size of the source (in fact, it tends to infinity for low cell loss rates). This would yield long mean waiting time in the buffer, and consequently increased end-to-end cell delay. However, if a source can transmit at its peak cell rate, the burst of cells is transmitted as it is formed (with a very small delay associated with inserting the synchronisation overhead). Thus the buffer size required is much smaller, and the cell loss is less likely to occur, as the cells are shifted out of the buffer at the same speed at which they arrive to it.

In multi-rate systems, a mix of a large number of different sources is envisaged. In such a system a discrepancy is inevitable between the large number of source types and the transmission rates that could be implemented. Statistical multiplexing in the code domain, enabled by the use of DTX, is an effective measure of increasing the multi-rate system capacity.

The relation between the sources' peak cell rate, mean cell rate, system transmission rate(s) and the theoretical value of CDMA channel capacity will have implications on the values of loads achievable at the satellite in each of the scenarios.

3.2 System description

In this thesis, users are assumed to be very small aperture terminals (VSATs - portable and fixed terminals). User requirements for bandwidth vary, and so do the source data rates. The QoS of the most stringent traffic type must be respected. Uniform user population is assumed. Each user (VSAT) comprises K=3 different traffic sources, namely voice, video and data. Thus there are K virtual channel connections (VCCs) originating in each of the user terminals (see Figures 3.1-3.3, where K=4). In the rest of the text and the analysis, a general number K of traffic types i.e. connections will be used, for the sake of generality. Numerical results are obtained for K=3, and presented in Chapter 5: RESULTS.

Multiplexing of connections, and their rate adjustment to the intermediate transmission rate, is via a buffer. In case of VCC multiplexing (in Scenario 1), the buffer is the multiplexer.

3.2.1 Scenario 1

One code is allocated to all connections of a terminal, and the terminal uses this code to transmit all of its generated traffic. Different VCCs are first multiplexed at the user terminal, prior to the spectrum spreading by a PN code (Figure 3.1). We assume therefore that the user terminals in this scenario are identical (have the same number of voice, video and data connections), although in reality not all of the connections may be active at any one time, and therefore the users would produce slightly different traffic mixes. Transmission continues for as long as there are data to transmit, i.e. for as long as the buffer (multiplexer) contains cells. When the buffer is empty, the DTX mechanism waits for a hangover period, which is equivalent to the cell interarrival time of the slowest connection. If all connections became silent, no cell will be generated during this hangover period. Thus after the hangover period has expired, the DTX mechanism, that is a transmission controller, switches off the carrier, and transmission ceases until the next full cell arrives to the buffer. When any of the connections become active again, they start to generate cells. When a new cell arrives to the buffer, the transmitter (or DTX) generates the overhead required for code and carrier re-synchronisation, and inserts it before the first cell of the new burst. The overhead is followed by the cells that have arrived from the source into the buffer. As the transmission cell slot rate is higher

than the cell arrival rate, empty cells are being inserted into the transmitted cell stream between the useful cell arrivals.

The performance of the scenario depends on the traffic mix at the multiplexer, and the choice of intermediate transmission rate *R*. For instance, if the users in the system have predominantly only one (say, voice) type of connection, and the difference between the transmission and source rates is large, little gain can be achieved since the bandwidth of the link will be used inefficiently. If many different types of connections are present, the problem of proper dimensioning of the multiplexing buffer arises, with a possible increase in buffering delay when server load approaches 1. However, high loads are achieved when the intermediate transmission rate is close to the mean aggregate cell arrival rate, and therefore the link is used 100% of the time, resulting in efficient bandwidth i.e. capacity usage. Hence, there is a trade-off between the gain that can be achieved by multiplexing many connections, and the decreased performance in terms of the delay.

The BER perceived by a receiver will fluctuate, depending on the background noise produced by other users in the same CDMA channel, but on average it will be above the required threshold, if the system has been properly dimensioned. Link BER will dictate a requirement for CLR at the buffer, which may lead to increase of the queueing delay. The link has to maintain the QoS requirements for the most stringent traffic type. This yields $G_p = const.$ across the system. As transmission rates R are the same for all users, all channels have equal bandwidth $W_k = RG_p$. If C is the total number of users, M - number of channels, C_{max} - maximum number of signals in a channel allowed by a CDMA system, and W the satellite transponder bandwidth, then the total spot-beam capacity is:

$$C_{\max} = M \cdot K_{\max} = \frac{W}{B} \cdot K_{\max} = \frac{W \cdot K_{\max}}{R \cdot G_p} \qquad \text{Eq. 3.2}$$

3.2.2 Scenario 2

Each connection gets its own spreading code, hence multiplexing at the terminal is avoided (Figure 3.2). The buffering delay will be reduced in comparison to the first scenario for the same choice of the transmission rate, as each connection has its own

buffer. Uniform transmission rate only makes sense if all the channels in the system are of the same bandwidth, that is, if the processing gain is constant. Equal channels allow a simple system architecture. Since the transmission rate is the same for all connections (R=*const.*), and so is the processing gain, different connection types can be transmitted in the same channel. QoS cannot vary since all processing gains within a channel must be equal, and thus the QoS for the most stringent traffic class must be respected. The aggregate number of ON-OFF state changes during some fixed period will be greater than in Scenario 1, and this will result in a larger total overhead, since the codes will have to be re-synchronised more often. However, this is not directly proportional to the efficiency (ratio of the number of useful and total number of transmitted cells). Depending on the source type (its mean burst size) and transmission rate, the total number of inserted empty cells may be smaller than in the first scenario, yielding lower efficiency than in the second scenario.

The spot-beam capacity is the same as in Scenario 1, but the scenario exhibits different performance, see Chapter 5.

3.2.3 Scenario 3

Each connection has its own spreading code, and different traffic types have different transmission rates (Figure 3.3). Therefore channel bandwidths correspond to the traffic types they carry. This scenario allows for the different channels (i.e. traffic types) to have different processing gains, resulting in greater system flexibility. Intuition suggests that Scenario 3 should be the most efficient since a more bandwidth efficient transmission rate can be chosen for each traffic type. Intuitively it is clear that the optimum transmission rate R_k (with respect to efficiency and capacity) for the channel carrying traffic type k in this scenario would be the peak cell rate of that traffic type, as the number of inserted empty cells for that rate is minimised. Once the optimum R is determined, system channels need to be dimensioned, i.e. the appropriate values of processing gain and channel bandwidths need to be determined for different traffic types, and the number of such channels for envisaged traffic loads. In other words, depending on how the load is distributed among different traffic types, the required number of channels for each type may change accordingly. The number of channels with

their respective theoretical and achievable capacities must correspond to the traffic offered, and must satisfy the resource demand of different traffic types.

In order to assess the performance of the third scenario, the number of channels carrying each traffic type must be first optimised for the maximum system capacity. This brings about a question of integer combinatorial optimisation, which will be addressed in Chapter 4.

In Scenario 3, variation of QoS can be achieved by choosing appropriate values of the processing gains, yielding different channel bandwidths: $W_k = R_k G_{p,k}$, where k is a traffic type, $k:=1, ..., K$. However, in order to compare the three scenarios, the processing gain in the third scenario will be kept constant and equal to that used in the second and first scenarios. It should be noted that keeping the processing gain identical for all traffic types does not allow full exploitation of the richness and flexibility that Scenario 3 offers. By varying the processing gains, it is possible to optimise the system for different traffic load distributions, and this is explained in Chapter 4.

3.3 Performance measures

In total, the available performance measures are:

- buffering requirements
- delay due to the waiting time in the buffers
- end-to-end delay
- throughput
- complexity of control required.

The main performance measures by which the scenarios will be evaluated are the system capacity, link and system efficiency, and the increase of these measures when the DTX is used relative to when it is not used (i.e. transmission is continuous, as in conventional CDMA). This follows from the main motivation of the study: the need to devise a medium access mechanism that uses the scarce and expensive satellite bandwidth efficiently, and the need to increase the capacity of the otherwise attractive CDMA access scheme in order to make it viable for transmission of broadband traffic over satellite. As it was mentioned in previous chapters, the critical performance parameter in

CDMA is its capacity. Efficiency is a measure of the additional overheads required on top of the pure ATM cells to make the scheme work. The relative change of these values with and without DTX gives a measure of the scheme viability.

The end-to-end delay and throughput should be studied within a context of link layer end-to-end error control mechanisms such as automatic repeat request (ARQ) algorithms, while the complexity of control requires the study of possible hardware implementations of the statistical multiplexing in the code domain for all three system configurations. Therefore the study of end-to-end delay, throughput and complexity of control is outside the scope of the research reported herein and is not covered by the thesis.

The system performance parameters are defined as follows:

Efficiency (*Eff*) (of the link or a system) is a total data rate (useful information) expressed as a fraction of the total traffic (data plus control and overheads rate) through the system. This is different to spectral efficiency used in performance assessment of different modulation schemes.

$$Eff = \frac{useful}{empty + overhead + useful} \qquad \text{Eq. 3.3}$$

Relative efficiency increase (*REI*) is the ratio of the difference of the system efficiencies when DTX is used and when it is not used, to the efficiency of the system when it is not used (transmission is continuous).

$$REI = \frac{Eff_{DTX} - Eff_{CONT}}{Eff_{CONT}} \cdot 100\% \qquad \text{Eq. 3.4}$$

Capacity (*C*) is the maximum number of users or connections that can be accommodated in a specified system bandwidth. The theoretical limit for CDMA capacity is set by equation 1.4 in Chapter 1: SPREAD SPECTRUM AND DS-CDMA. The expressions for achievable limit in each scenario are derived in the next section.

Relative capacity increase (*RCI*) is the ratio of the difference of capacities achievable with and without the use of DTX, and the capacity achievable without the DTX.

$$RCI = \frac{C_{DTX} - C_{CONT}}{C_{CONT}} \cdot 100\% \qquad \text{Eq. 3.5}$$

3.4 Analysis

The objective of the analysis is to establish how the performance of the three configurations changes with the traffic mix, transmission rates and channel bandwidths. The analysis allows a numerical comparison to identify the best of the three configurations.

The analysis is performed on the cell level and in discrete time. Sources are modelled as ON-OFF streams with deterministic interarrival times and geometrically distributed state sojourn times. The multiplexer is modelled as an infinite queue with fixed deterministic service time $d = (R \text{ [in cells]})^{-1}$, which is determined by the slot rate at the output of a multiplexer. The slot rate is the CDMA intermediate transmission rate expressed in cells per second, i.e. multiplexer service time is determined by the CDMA intermediate transmission rate. The state sojourn times are geometrically distributed and are governed by state transition probabilities, $\alpha(k)$, $\beta(k)$ of connection k. Each of the source connections has different probabilities of ON and OFF periods, with different cell rates in ON states (Figure 3.6).

$$p_{on,k} = \frac{\alpha_k}{\alpha_k + \beta_k} \qquad \text{Eq. 3.6}$$

$$p_{off,k} = \frac{\beta_k}{\alpha_k + \beta_k} \qquad \text{Eq. 3.7}$$

$$\lambda_k = p_{on,k} \cdot PCR_k \qquad \text{Eq. 3.8}$$

The process at the output of a multiplexer is also modelled as an ON-OFF process with geometrically distributed ON and OFF (i.e. BUSY and IDLE) periods.

The analysis starts from the known relations between the source parameters (such as mean duration of ON and OFF periods expressed in terms of time units and in terms of

cell slots) and the transitional probabilities of a two-state Markov-modulated process. Queueing theory is applied to establish the mean duration of ON and OFF periods of the equivalent two-state process at the output of a multiplexer. The probabilities of ON and OFF states (active and silent periods) of the uplink lead to the derivation of the unused link capacity and the exact formulae for the system performance measures defined in section *3.3 Performance measures*. The sources are defined by the duration of their mean ON and OFF periods (expressed in seconds), and the peak cell rate (expressed either in bits per second or cells per second).

The analysis aims at finding the mean busy and idle periods of the process at the output of the multiplexer. The mean busy and idle periods of an ON-OFF process can be easily found [Schormans and Pitts 96], and therefore this part of the analysis is presented in Appendix A.

Once the mean busy and idle periods at the output of the multiplexer are found, the analysis proceeds by establishing the expressions of the performance measures (efficiency, capacity, REI, RCI), given the required overheads.

A note should be made here that the only physical layer parameter that is indirectly taken into account in the analysis are the *SNR*, E_b/N_0 and processing gain G_p, which determine the capacity limit (Eq. 1.4, Chapter 1). The *SNR* defines the link BER (and consequently the CLR i.e. QoS) before any decoding and error correction. Since satellite channels are very prone to errors, some form of forward error correction (FEC) (possibly combined with interleaving), and a data link automatic repeat request protocol, will probably be employed. It was shown in [Cain & McGregor 97] that significant BER improvements can be achieved for ATM over wireless disturbed channel with very simple FEC mechanisms, such as pure convolutional or concatenated convolutional and Reed Solomon codes. The improved error performance at a low *SNR* has a cost in bandwidth, which in terms of ATM cells means that the FEC either adds bits to the existing ATM cells, thus extending the cell size, or alternatively “packs” its bits into additional cells, thus increasing the number of cells transmitted. As the analysis presented here assumes pure ATM with 53 octets, an FEC can be taken into account through the duration of the mean ON periods of the sources. Thus the analysis will not lose generality if FEC is not explicitly modelled. The study of ARQ schemes within the given scenarios is outside the scope of this study and will not be covered.

Variables used in the analysis are (subscript k used in Scenarios 2 and 3 analysis):

K	number of traffic types (for numerical analysis K=3)
λ	mean aggregate arrival rate at the multiplexer input
λ_k	mean arrival rate of the kth connection
ρ	load at the multiplexer
SR, SR_k	slot rate expressed in [cells/s]; it is equal to the transmission rate R divided by cell size: $SR=R/424$
PCR, PCR_k	peak cell rate of a source during busy (ON) period
B, B'	random variable representing duration of the busy period with and without overhead
I, I'	random variable representing duration of the idle period with and without overhead
T_B, T_{Bk}	mean duration of a busy period of a multiplexer including overhead transmission time
T_B, , $T_{Bk'}$	mean duration of a busy period of a multiplexer without overhead
T_I, T_{ik}	mean duration of an idle period of a multiplexer including overhead transmission time
T_I, , $T_{ik'}$	mean duration of an idle period of a multiplexer without overhead
T_{hgo}, $T_{hgo,k}$,	mean hangover period
T_{acq}	mean acquisition period
$p_{ON,k}$, $p_{OFF,k}$	probability that the connection k is ON/OFF,
P_{ON}, P_{OFF}	probability that the multiplexer (process at its output) is ON/OFF
P_{ACTIVE}	probability that the multiplexer is active
$T_{ON,k}$, $T_{OFF,k}$	mean ON and OFF periods of the k^{th} connection,
N, N_k, $N_k(i)$	number of connections (of traffic type k in Scenario 2; connections of traffic type k in i^{th} channel in Scenario 3)
C_{max} ,$C_{max,k}$	maximum theoretical number of signals (connections) in a DS-CDMA channel
$\Delta C(\rho)$	unused capacity for load ρ i.e. difference between maximum possible CDMA capacity and the capacity used when DTX is employed
m_k	number of channels for carrying traffic type k
M	total number of channels

W, W_k	total bandwidth, bandwidth of channel carrying type k
$E[.]$	expected (mean) value operator
Eff_{DTX}	efficiency when DTX is employed
Eff_{CONT}	efficiency for continuous transmission (no DTX)

In the following sections it is assumed that $E[I']=T_{I'}$ and $E[B']=T_{B'}$ are found as described in Appendix A.

3.4.1 Scenario 1

It is important to note that the analysis presented in Appendix A for Scenario 1 does not take into account the effects of queueing in the case of significant loads in the multiplexer. Queueing in the multiplexer may lead to the extended busy periods of the ON-OFF process at its output, i.e. the busy periods may become longer than predicted by the analysis. As Queueing theory does not provide the tools to solve the queueing problem in the case when a buffer is fed by a small number of heterogeneous sources, the effect of queueing at higher loads on the busy periods cannot be found analytically. Thus the numerical results obtained by analysis for high loads in the multiplexer for Scenario 1 must be viewed with caution. This will be elaborated in Chapter 5: RESULTS.

The DTX mechanism turns the transmission off after a hangover period T_{hgo} when the buffer becomes empty. The hangover period is equal to the cell inter-arrival time of the slowest connection. This is because the slowest connection will have the longest cell inter-arrival time, and if there is no cell arrival into the multiplexer during that period after the last cell arrival, it follows that all connections have gone silent. Code synchronisation overhead is inserted at the beginning of every active period. This overhead contains the number of cells that ensures that the code is acquired within the desired mean acquisition time T_{acq}. Due to the hangover period and code acquisition overhead, the busy and idle periods of a multiplexer, $T_{B'}$ and $T_{I'}$, as determined by equations A.11 and A.12 (see Appendix A) are extended by T_{hgo} and T_{acq}:

$$T_B = T_{B'} + T_{acq} + T_{hgo}$$
$$T_I = T_{I'} - T_{acq} - T_{hgo} \qquad \text{Eq. 3.9}$$

We are interested in efficiency of one user (link) and some equivalent 'aggregate' efficiency of the whole channel. Since in Scenario 1 all users and channels are identical, efficiency of a channel will be the same for all users, and equal to the efficiency of the system. Efficiency is measured through the number of useful, empty and overhead cells in one activity cycle. The number of useful cells during one activity cycle are found as the total number of cells that arrive on average during the busy and idle periods. The aggregate mean cell arrival rate is equivalent to the sum of the mean arrival rates of connections. Total number of cells transmitted on the uplink is determined by the slot rate and the multiplexer ON time. Efficiency of the system when the DTX is employed is given by:

$$Eff_{DTX} = \frac{useful\ cells\ in\ one\ duty\ cycle}{total\ cells\ in\ one\ duty\ cycle}$$

$$= \frac{\bar{\lambda} \cdot \left(T_B + T_I\right)}{SR \cdot \left(T_B + T_{hgo} + T_{acq}\right)} \qquad \text{Eq. 3.10}$$

$$= \frac{\left(T_B + T_I\right) \cdot \sum_{k=1}^{K} \lambda_k}{SR \cdot \left(T_B + T_{hgo} + T_{acq}\right)}$$

Efficiency of the system when transmission is continuous is:

$$Eff_{CONT} = \frac{\sum_{k=1}^{K} \lambda_k}{SR} \qquad \text{Eq. 3.11}$$

3.4.1.1 Available capacity and capacity increase

It is necessary to assess the available unused capacity in order to find the capacity increase achievable by statistical multiplexing in the code domain. If $U(\rho)$ is a maximum number of users required to achieve the total load on the satellite of ρ, then the proportion of the unused available capacity can be expressed as:

$$\frac{\Delta C(\rho)}{C_{\max}} = 1 - \frac{U(\rho)}{C_{\max}} \cdot P_{ACTIVE} = 1 - \frac{U(\rho)}{C_{\max}} \cdot \frac{T_B + T_{acq} + T_{hgo}}{T_B + T_I} \qquad \text{Eq. 3.12}$$

P_{ACTIVE} is a probability that the multiplexer is active. This result follows from the fact that capacity is not used during the silent periods on a transmission link.

It can be seen that the relative available capacity will depend on the load on the system, code acquisition time and the mean busy and idle periods of the multiplexer at a user terminal. Capacity increase achievable by statistical multiplexing will depend greatly on the activity factors of different traffic types. In Scenario 1, the "equivalent" activity factor that results from multiplexed connections is equal to the probability P_{ACTIVE} that the multiplexer is active. In the system model all users have an equal number of all connection types, since they are first multiplexed prior to CDMA spreading (i.e. all VSATs have equal number of voice, video and data connections, equal to 1 in the numerical results), and therefore the number of users $U(\rho)$ that produces the load ρ is $U(\rho)=U_k(\rho)$. The load of the system, from which we can find $U(\rho)$, is:

$$\rho = \frac{\lambda}{SR} = \frac{\sum_k \lambda_k \cdot U_k(\rho)}{SR} = \frac{U(\rho) \cdot \sum_k \lambda_k}{SR} \qquad \text{Eq. 3.13}$$

The number of extra users that can be admitted into the network above C_{max} is found as a ratio of the available capacity and the equivalent activity factor P_{ACTIVE}:

$$\Delta U(\rho) = \frac{\Delta C(\rho)}{P_{ACTIVE}} \qquad \text{Eq. 3.14}$$

$$RCI(\rho) = \frac{\Delta C(\rho)}{P_{ACTIVE} \cdot C_{\max}} \cdot 100\% \qquad \text{Eq. 3.15}$$

Expression Eq. 3.14 for capacity increase forms the basis for a connection admission control (CAC) scheme in a CDMA satellite multirate system. The controller can assess the available capacity using the known activity factors. These factors can be measured on-line, or provided as traffic descriptors. When a new user requests access, the CAC algorithm finds out if $\Delta U(\rho) > 1$. If it is, the connection is accepted.

3.4.2 Scenario 2

In this scenario the server mean idle and busy periods differ from source ON and OFF periods by the hangover period and the mean code acquisition time. The hangover of the

server after the source has gone into silent state is different for different connections, and is equal to the cell inter-arrival time of the connection being served by the server. The cells are delayed by the mean acquisition time, which is the same across the system since all connections transmit at the same transmission rate *R*. The QoS requirements for all sources are assumed identical for the sake of clarity and tractability of results.

Mean busy and idle periods of the server are :

$$T_{Bk} = T_{ON,k} + T_{hgo,k} + T_{acq} \qquad \text{Eq. 3.16}$$

$$T_{Ik} = \max(0, T_{OFF,k} - T_{hgo,k} - T_{acq}) \qquad \text{Eq. 3.17}$$

The mean number of overhead and hangover cells of a connection type *k* in a single duty cycle is:

$$\left(T_{acq} + T_{hgo,k}\right) \cdot SR$$

Over some arbitrary long period Δt the mean number of duty cycles of a connection type *k* will be:

$$\frac{\Delta t}{T_{Bk} + T_{Ik}}$$

Thus for a traffic type *k* with total number of connections equal to N_k the mean number of overhead and hangover cells over a period Δt is:

$$N_k \left(T_{acq} + T_{hgo,k}\right) \cdot SR \cdot \frac{\Delta t}{T_{Bk} + T_{Ik}}$$

The rate of empty cells at the output of the multiplexer during a connection ON state is equal to the difference between the multiplexer service i.e. slot rate and the connections peak cell rate during the ON period. Thus over a period Δt the total number of empty cells for N_k connections of traffic type *k* with probability $p_{ON,k}$ of being active will be:

$$N_k \cdot \left(SR - PCR_k\right) \cdot p_{on,k} \cdot \Delta t$$

The total number of useful cells during Δt are those transmitted by a source during its ON periods, i.e.:

$$PCR_k \cdot p_{on,k} \cdot \Delta t = \lambda_k \cdot \Delta t$$

Efficiency of the system when the DTX is employed is therefore:

$$\begin{aligned} Eff\,(DTX) &= \frac{useful\ cells}{useful\ cells + empty\ cells + overhead\ cells + hangover\ cells} \\ &= \frac{1}{1 + \dfrac{(overhead + hangover + empty)\ cells\ in\ some\ \Delta t}{useful\ cells\ in\ \Delta t}} \\ &= \frac{1}{1 + \dfrac{\sum_{k=1}^{K} N_k \cdot \left(T_{acq} + T_{hgo,k}\right) \cdot SR \cdot \dfrac{\Delta t}{T_{Bk} + T_{Ik}} + \sum_{k=1}^{K} N_k \left(SR - PCR_k\right) \cdot p_{on,k} \cdot \Delta t}{\sum_{k=1}^{K} N_k PCR_k \cdot p_{on,k} \cdot \Delta t}} \\ &= \frac{1}{1 + \dfrac{\sum_{k=1}^{K} N_k \cdot p_{on,k} \left[SR \cdot \left| 1 + \dfrac{T_{acq} + T_{hgo,k}}{T_{Bk}} \right| - PCR_k \right|}{\sum_{k=1}^{K} N_k \cdot \lambda_k}} \end{aligned} \qquad \text{Eq. 3.18}$$

Efficiency of the system without DTX, (continuous transmission case):

$$Eff_{CONT} = \frac{\sum_{k=1}^{K} N_k \cdot \lambda_k}{C_{\max} \cdot SR} \qquad \text{Eq. 3.19}$$

3.4.2.1 Available capacity and capacity increase

The activity factor of a traffic type k is $p_{on,k}$, and $U_k(\rho)$ the number of connections (producing load ρ) of that type. The capacity increase in terms of any of the present traffic types can be found. Let $\Delta U_k(\rho)$ be a number of additional connections of type k that can be accepted above the theoretically predicted value C_{max} by equation 1.4 to fully utilise the CDMA system capacity. Then we have:

$$\sum_k U_k(\rho) = C_{\max} \qquad \text{Eq. 3.20}$$

$$\sum_k p_{on,k} \cdot \left(U_k(\rho) + \Delta U_k(\rho)\right) = C_{\max} \qquad \text{Eq. 3.21}$$

$$\sum_k p_{on,k} \cdot \Delta U_k(\rho) = C_{\max} - \sum_k a_k \cdot U_k(\rho) \qquad \text{Eq. 3.22}$$

Let $\Delta U_{j,max}(\rho)$ be the number of users of traffic type j that can be accepted above C_{max}:

$$\Delta U_{j,\max}(\rho) = \frac{C_{\max} - \sum_{k} p_{on,k} \cdot U_k(\rho)}{p_{on,j}} \qquad \text{Eq. 3.23}$$

Expression Eq. 3.23 treats a general case, and can be used as a concept for a CAC in the CDMA system based on Scenario 2.

Relative capacity increase for each traffic type k, given the traffic mix that produces the load ρ, is:

$$RCI_k(\rho) = \frac{\Delta U_{k,\max}(\rho)}{C_{\max}} \qquad \text{Eq. 3.24}$$

3.4.3 Scenario 3

In this scenario, for each connection of type k there are m_k channels of corresponding bandwidth W_k.. Channel bandwidths are determined by the QoS requirements of different connection types and their transmission rates. The total spot-beam bandwidth W can be expressed as:

$$W = \sum_{k=1}^{K} m_k \cdot W_k \qquad \text{Eq. 3.25}$$

Each channel can have a theoretical maximum of $C_{max,k}$ connections, and a number of active users in the channel is $N_k(i)$, $i=1,\ldots, m_K$. Theoretical maximum number of users is:

$$C_{\max} = \sum_{k=1}^{K} m_k \cdot C_{\max,k} \qquad \text{Eq. 3.26}$$

Each connection type has a different code mean acquisition time $T_{acq,k}$ and hangover period $T_{hgo,k}$. The derivation of the expressions for efficiency with and without DTX for Scenario 3 follow the same reasoning as the one presented for Scenario 2. The difference arises from the fact that in Scenario 3 different traffic types have separate channels, and each traffic type k has m_k channels allocated to it. Thus total number of connections of type k is:

$\sum_{i=1}^{m_k} N_k(i)$.

Expression for efficiency for a system with DTX is similar to the one in the second scenario:

$$Eff(DTX) = \frac{useful\ cells}{useful\ cells + empty\ cells + overhead\ cells + hangover\ cells}$$

$$= \frac{1}{1 + \frac{(empty + overhead + hangover)\ cells\ in\ some\ \Delta t}{useful\ cells\ in\ \Delta t}}$$

$$= \frac{1}{1 + \frac{\sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot\left(T_{acq,k} + T_{hgo,k}\right)\cdot SR_k \cdot \frac{\Delta t}{T_{Bk} + T_{Ik}} + \sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot\left(SR_k - PCR_k\right)\cdot p_{on,k}\cdot\Delta t}{\sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot\lambda_k\cdot\Delta t}}$$

$$= \frac{1}{1 + \frac{\sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot p_{on,k}\left[SR_k \cdot \left| 1 + \frac{T_{acq,k} + T_{hgo,k}}{T_{Bk}} \right| - PCR_k\right|}{\sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot\lambda_k}}$$

Eq. 3.27

Efficiency for continuous transmission:

$$Eff_{CONT} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{m_k} N_k(i)\cdot\lambda_k}{\sum_{k=1}^{K} m_k \cdot C_{\max,k}\cdot SR_k}$$

Eq. 3.28

3.4.3.1 Available capacity and capacity increase

Within each channel, the increase of capacity is measured in terms of the connections of that particular type. Thus the increase in capacity depends on the maximum number of users $C_{max,k}$ in each of m_k channels of type k, QoS for traffic type k, and the activity factor of the traffic type $p_{on,k}$, and is measured in the same way as in Scenario 2, only separately for each channel type k:

$$\Delta C_k = \frac{C_{\max,k} - p_{on,k}\cdot C_{\max,k}}{p_{on,k}} = \frac{p_{off,k}}{p_{on,k}}\cdot C_{\max,k}$$

Eq. 3.29

$$RCI_k = \frac{\Delta C_k}{C_{\max,k}} = \frac{p_{off,k}}{p_{on,k}} \qquad \text{Eq. 3.30.}$$

3.5 Delay and buffer size

It was mentioned in section *Capacity, transmission rate and statistical multiplexing* that a trade-off in terms of the delay incurred and buffer space required ought to be assessed for the two extreme cases:

- the case of transmission rate being equal or close to the source mean rate, $R=\lambda$, and
- the transmission rate being equal to the source peak cell rate, $R=PCR$.

It is well known from Queueing theory (for cell and burst scale queueing) that as the service rate of a queue decreases from peak cell arrival rate to the mean cell arrival rate of a source the queue is serving, and the utilisation of the server approaches 1, the buffer size required for the desired cell loss probability increases to infinity, and so does the mean waiting time in the queue. In this section the results of the assessment of the three scenarios will be presented, obtained by simulations and discrete fluid flow analysis [Schormans *et al*. 94, Pitts & Schormans 96]. The comparative results show how the mean number in the queue, mean delay and required buffer size change as the service rate increases from the mean cell arrival rate to the peak cell rate.

The insertion of overhead affects simulation and analysis in the following way:

As the hangover is only the extended transmission of empty cells at the end of a busy period of a multiplexer, it will not affect the delay and state distributions of the actual cells within the queue. The code acquisition overhead is minimal: the required number of bits is small (less than a size of one ATM cell) and the overhead can be assumed to take only a single cell. As the overhead takes only a single cell during a whole duty-cycle of the multiplexer output process, its effect on the buffer state and delay distribution is insignificant and can be disregarded in the simulation without the loss of accuracy.

3.5.1 Analysis of delay and buffer requirements for Scenario 1

Scenario 1 is very difficult to analyse mathematically, as there is no acceptable approximation of its input traffic. Queueing theory gives a variety of results either for a single isolated source (e.g. for a single ON-OFF or Poisson source), or for a large number of sources feeding a single queue. Scenario 1, however, comprises only three heterogeneous sources feeding a queue. Thus we resort to simulation to find the behaviour of the queue (the multiplexer) when fed by three different sources. The simulator used is MICROSIM [MSIM 97], described in Chapter 4: SIMULATION TOOLS, and in Appendix B. The service rate of the queue was varied between

$SR = \lambda + 0.2 \cdot (PCR - \lambda)$ [cells/s] and

$SR = \lambda + 0.8 \cdot (PCR - \lambda)$ [cells/s] ,

where λ represents the sum of the mean arrival rates of the three sources feeding the queue, and *PCR* is the worst case peak cell rate, equal to the sum of peak cell rates of all the sources.

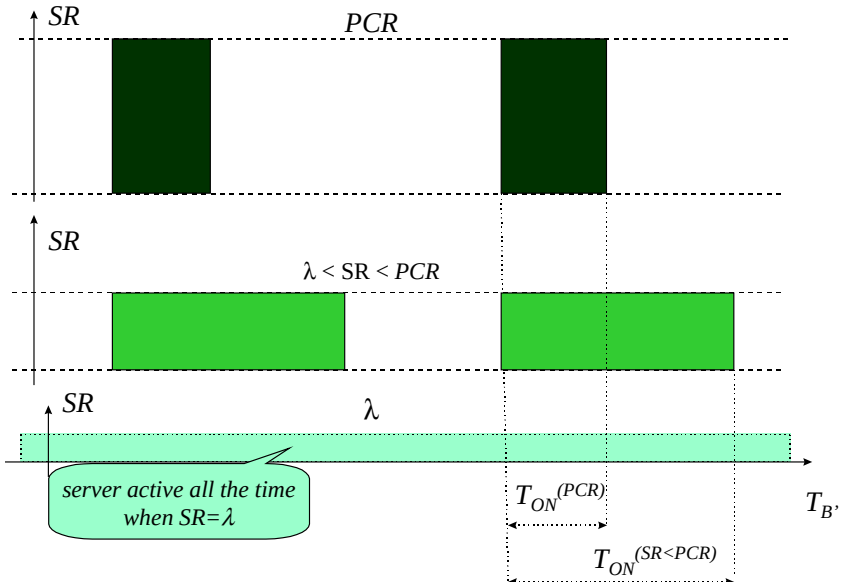


Figure 3.7 Illustration of simulated values of multiplexer slot (i.e. service) rate, and how the slot rate influences the duration of the busy period of the multiplexer.

The simulation produces the mean and the maximum wait in the queue (including the service time), number of cells blocked due to the queue overflow, and state distribution.

From the mean delay and the service time it is possible to determine the mean number in the queue. The delay experienced by the cells is:

$$D = T_q + T_{acq} \approx T_q$$

where D is the total mean delay, T_q is the mean delay in the queue including the service time.

3.5.2 Analysis of delay and buffer requirements for Scenario 2

Discrete fluid-flow analysis as presented in Chapter 7: *Burst Scale Queueing* of [Pitts & Schormans 96] was used to analyse a queue fed by a single ON-OFF source. The analysis finds state probabilities for excess rate cells. Schormans and Pitts found that this analysis gives an accurate, but always slightly overestimated prediction for the cell loss rate [Schormans *et al*. 96]. The queue state probabilities are expressed as:

$$p(X) = \frac{1}{1 + \sum_{i=1}^{X} \left| \frac{s}{a} \right|^{i} \frac{1-a}{s}} \qquad \text{Eq. 3.31}$$

where

$$a = 1 - \frac{1}{T_{ON} \cdot (PCR - SR)} \qquad \text{and}$$

$$s = 1 - \frac{1}{T_{OFF} \cdot SR}.$$

Cell loss probability is found as:

$$CLP = \frac{PCR - SR}{PCR} \cdot p(X)$$

Mean number in the queue is given as the mean of the distribution *p(X)*:

$$E[X] = L = \sum x \cdot p(x)$$

and the mean delay is given by Little's formula:

$$D = \frac{E[X]}{\lambda} = \frac{L}{\lambda}.$$

To find how the CLP varied with the transmission rate, the service rate of a queue was again varied for each individual source *k* (*k*=1,2,3, Figure 3.2 andFigure 3.6), from

$$SR_k = \lambda_k + 0.2 \cdot (PCR_k - \lambda_k)$$

to

$$SR_k = \lambda_k + 0.8 \cdot (PCR_k - \lambda_k)$$

In case of $SR_k=PCR_k$ almost no buffering is required (only a couple of cells). In scenario 2, the buffers for the three sources will have different utilisation values, as only one transmission rate has to be chosen. The chosen rate will have to be higher than any of the mean cell arrival rates. The results of efficiency give the same indication, as will be seen later. The resulting plots of CLP vs. Buffer size for all three sources are given in Chapter 5: RESULTS.

3.5.3 Analysis of delay and buffer requirements for Scenario 3

This is the most efficient scenario with respect to the buffering requirements and delay incurred, as each source can transmit at its own peak cell rate. The same buffer state and delay distribution analysis applies as for Scenario 2, only no compromise is necessary in the choice of transmission rate. The delay can be negligible, and it is only caused by the acquisition overhead, and not by the queueing in the buffers.

4. SOFTWARE TOOLS DEVELOPED DURING THIS RESEARCH

As appropriate, mathematical analysis was verified using simulation. Similarly, system optimisation in the third scenario was carried out using optimisation tools. This chapter describes the two tools used in the study: their structure, functionality, suitability and limitations in performing the tasks they were intended for.

4.1 Method

The analysis of the three scenarios was presented in Chapter 3: ATM OVER CDMA. In general, the problem in simulating the scenarios arises from the different time scales and the reference levels that affect the performance of the medium access technique described: CDMA interference that affects the capacity takes place on the physical layer; the analysis of the transmission is from the point of view of the ATM cell layer (equivalent to the data layer in ISO-OSI reference model) and on the cell time scale; the capacity increase is measured in terms of the number of connections, the effect being 'perceived' on the network layer, i.e. call time scale. The intertwined reference levels and time scales of the mechanism studied imposes difficulties on the choice of an appropriate simulator and on the system simulation itself, and calls for a hybrid approach that involves both simulation and analysis.

As the analysis was performed at the cell level, the available cell-level network simulator developed by QMW College - MICROSIM - was chosen. With some modifications, MICROSIM was used to obtain experimental results for the verification and validation of the mathematical analysis. If a particular performance parameter could not be measured directly, it was evaluated by substitution of the measured parameters obtained by simulation into analytical formulae.

Furthermore, in order to have a fair comparison of the scenarios it was necessary to optimise the system parameters in the third scenario. An optimisation program based on simulated annealing was developed for this purpose. The simulated annealing optimisation program was validated against a genetic algorithm now publicly available on the Web.

4.2 Microsim

MICROSIM [MSIM 97] is an event-driven cell-level simulator that models networks as a collection of interconnected sources and queues, which represent network elements. It was developed at QMW College for simulation and modelling of terrestrial packet-switched networks, and later modified in a collaborative project with the European Space Agency (ESA) to enable modelling of the effects of satellite-specific phenomena on the ATM cell stream. For the purpose of this research, the version with the minimum number of modifications has been used.

4.2.1 How MICROSIM works

In MICROSIM, queues model links (with or without propagation delay), switch delay elements and node buffers. Sources initiate calls and generate an associated cell stream by producing one call at a time. They send their cells only to one queue or delay element (multicasting is not supported). Network elements are grouped into nodes, that define the topology of the network (Figure 4.1). Time units are not specified, but they have to be consistent across all network elements and sources, i.e. throughout the configuration specification.

Queues are specified by their buffer size, service time probability distribution function and mean. If a queue represents a link transmission delay, or a node switching delay, the buffer size does not matter (it is set to a high value), and only the service delay is defined.

The user specifies the sources in MICROSIM by the probability distribution function (PDF) of the state sojourn times in ON and OFF states, the mean duration of ON and OFF states, and the cell interarrival time (its mean and PDF) of the cell stream generated in the ON state. This allows various traffic models to be supported: Poisson, CBR, VBR, bursty. Each source generates a cell during its ON period at time instants which are generated as random variables from the PDF of the cell inter-arrival time. Cells are routed to their destinations via the queues defined by the routing information supplied to the simulator as an input file. When a cell arrives to a queue, the time of its departure is determined as a function of the current queue size and the service time. (Service time is also a random variable generated from its PDF.) Upon determining the cell departure

time, an event for cell departure is scheduled for in the event list. By this mechanism cells get moved through the network.

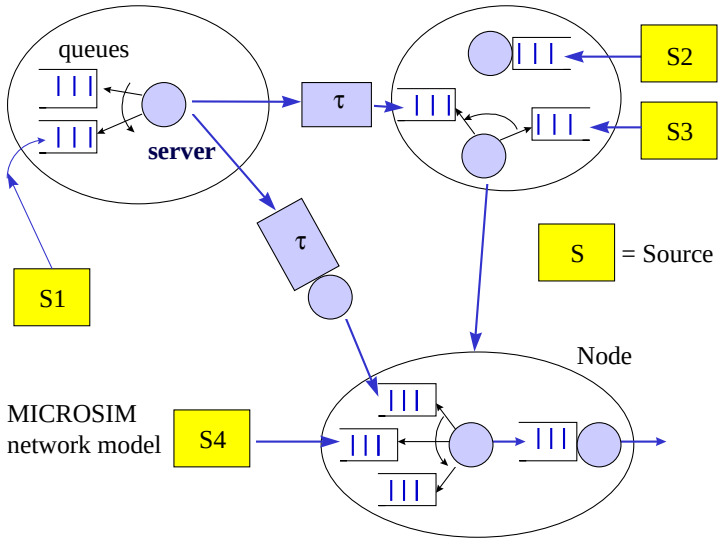


Figure 4.1 MICROSIM network model: nodes comprise input queues, zero-delay switch, node delay element that models switching delay in the node, and output queues. A node must have at least one input queue. Delay on transmission links is modelled using a delay element, but links may have zero delay.

It is clear that this simulator did not allow modelling of empty cells, insertion of overhead, and a DTX mechanism, so MICROSIM was modified to allow measurements of the mean busy and idle periods of the selected queues in order to simulate DTX mechanism. However, it was not possible to model transmission of overhead and empty cells, as it would require writing a new simulator based on a completely different concept - time-driven simulation instead of event-driven - and this would also result in prohibitively long simulation duration. Time-driven simulators simulate every time slot (or clock tick). In scenarios with high transmission (clock) rates long streams of empty time slots without any events taking place would be generated, and consequently the execution times would be very long for the simulation of the required number of ATM cells (around 1 million). Furthermore, for the measurement of the capacity, simulation of empty cells is not required, while the effect of overhead can be taken into account by simply extending the measured mean busy periods (and reducing the mean idle periods) by an analytically derived value for the mean code acquisition time and the hangover period. As it will be seen later, the worst case mean code acquisition time for the

scenarios studied is short in comparison to the duration of a single ATM cell. Thus its effect on both the measured and analytical results is negligible for the traffic types studied. However, it is included in all analytical expressions as it may have a significant influence on performance of the scenarios with traffic types of very short active and silent periods, where the mean acquisition time may be comparable to the duration of the server mean busy period.

The number of generated empty cells can be estimated from the duration of the mean busy periods of a queue, the number of cells passed through the queue, and the queue service time. The number of empty cells transmitted is required for the evaluation of the link and system efficiency only.

In order to validate the theoretical results, a number of simulations were performed.

4.2.2 What was measured

The following performance measures are available in MICROSIM: end-to-end cell delay distribution of tagged sources, queue state and delay distributions, queue mean and maximum delay (waiting time including the service time), cell loss ratio, mean busy and idle periods of the queues modelled to have DTX. Of these, the following are relevant for validating the system performance and analysis validation:

- mean duration if busy and idle periods
- mean delay in the queue
- state distribution of the server queues, whether multiplexing or not,
- cells lost and cells that passed through the queue.

The achievable capacity and effect of DTX on efficiency in the proposed CDMA system scenarios were found using a hybrid simulation/analytical approach, explained in CHAPTER 5: RESULTS.

4.2.2.1 Capacity

Achievable capacity depends on the mean busy and idle periods of a multiplexer server. Thus the first measurement was a validation of analytical results for the mean busy and idle periods. This was easily achieved with MICROSIM. By putting the simulated busy and idle periods into the formula for the system capacity increase, the results close to the theoretical ones were obtained (see Chapter 5: RESULTS). The next point was to show

that the actual formula (equations 3.13, 3.24 and 3.30 in Chapter 3) for the capacity increase based on the busy and idle periods was valid. In other words, the number of ON-OFF sources that could be multiplexed together above the capacity of the satellite server had to be found and compared to the theoretically obtained value. The principle of statistical multiplexing was validated using server configurations simpler than those imposed by the scenarios studied without loss of generality. This was a necessity imposed by the nature of the problem, described next.

A satellite server receives both empty and full cells. It discards the empty ones. The server must work at least at the rate of $C_{max}SR$, that is C_{max} faster than the slot rate of individual users (VSATs). The utilisation of the satellite server is only equal to 1 when all C_{max} sources transmit cells continuously at the slot rate SR corresponding to the CDMA intermediate transmission rate R. Then there are no empty cells and no idle periods on the user transmission uplinks. However, when the source rates are lower than the transmission i.e. slot rate, empty cells are inserted. Even with continuous transmission the satellite server will have a utilisation lower than 1 due to the presence of empty cells (i.e. the mean cell arrival rate at the satellite server will be lower than the satellite server rate). When DTX is introduced, the number of sources that could be statistically multiplexed together is determined though the duration of the busy periods (that contain empty cells) and idle periods or the process at the output of the multiplexers (i.e. on VSAT uplinks). The utilisation of the server will be pushed up, but will not reach 1. As MICROSIM cannot simulate empty cells, scenarios studied could not be directly measured for the capacity by just simply increasing the number of sources until the satellite buffer starts to overflow. The satellite, in the case of the three scenarios, is always under-utilised due to the source types chosen and the system parameters: theoretical capacity and transmission rate. Thus a different approach was taken.

The principle of statistical multiplexing in the code domain can be validated by increasing the number of sources feeding a single queue until the CLP of the queue reaches a threshold level. Ideally there should be no cell loss in the queue. With a reasonably large queue, there will be no cell loss for as long as the queue server can cope with the input load and its fluctuation. This is in fact a classical example of statistical multiplexing in the time domain, i.e. in ATM networks. If the connection

between the sources' activity factors and relative capacity increase can be shown on a simple model of a number of sources feeding a single queue, then it is valid for a more complex configuration where the satellite queue is fed by streams from individual users or connections (where each connection stream can be seen as some "equivalent" source).

4.2.2.2 Efficiency

Server utilisation and efficiency without the DTX is found as the ratio of the arrival rate to the service rate. Arrival rate is simply the number of cells arriving divided by the period of simulation. Service rate is an input parameter to the simulation.

Efficiency when DTX is used requires knowledge of the total number of empty cells generated at the output of the queue. As the simulator measures total duration of active (busy) periods of the server as well as the number of state changes of the server, the number of empty cells, and consequently the efficiency when DTX is employed, can be estimated as follows:

Total number of cells transmitted by the multiplexer server is found as:

$$n_{total} = N_B \cdot SR \cdot \left(T_{acq} + T_{hgo}\right) + SR \cdot \sum_{i=1}^{N_B} B_i \qquad \text{Eq. 4.1}$$

Total number of empty cells is a difference between the last term in the above equation Eq. 4.1 and the total number of cells through the queue n_q measured in the simulation:

$$n_{empty} = SR \cdot \sum_{i=1}^{N_B} B_i - n_q \qquad \text{Eq. 4.2}$$

Efficiency with DTX employed is found as:

$$Eff_{DTX} = \frac{n_q}{n_{total}} \qquad \text{Eq. 4.3}$$

B_i a random variable representing busy period during one server activity cycle (simulated)

N_B a measured number of server state changes (activity cycles)

n_{total} total number of cells transmitted on a link by a server

n_q a measured number of useful cells that passed through the queue

4.2.3 What was simulated

Scenario 1 was simulated in order to find the multiplexer mean busy and idle periods, mean delay in the queue, and the state distributions as the slot rate varies from the mean cell arrival rate to the worst case arrival rate (sum of all peak cell arrival rates). This gave measured values for the system capacity, required buffer sizes and incurred delay for different slot rates and desired CLP.

For scenarios 2 and 3 a hybrid simulation-analysis approach was used to validate the analysis presented in Chapter 3. Scenarios 2 and 3 were simulated for the mean busy and idle periods of the servers, where each server queue is fed by a single source of one of the three traffic types, while their mean delay, mean number in the queue and queue state distributions were obtained by well established discrete fluid flow analysis.

Due to the scale of the system to be modelled, only smaller configurations could be tested and validated using MICROSIM. For instance, scenarios with up to 16 users were simulated, and those with higher number of users were only mathematically analysed using the analysis presented in Chapter 3 and Appendix A. However, the analytical and simulated results agreed for all of the simulated configurations, which confirmed the validity of the mathematical analysis and numerical results obtained by it.

4.3 Optimisation

Optimisation of the system parameters was required for Scenario 3. This was an integer linear combinatorial problem with a very large state space; furthermore, the generalisation of the problem that would arise from a real system would be a large-scale combinatorial problem.

An integer programming problem is one that can be formulated as [Nemhauser & Wolsey 88, pp.3-4]:

$$\max\{cx: \quad \boldsymbol{A}x \le b, \quad x \in Z_n^+\}$$

where Z_n^+ is the set of non-negative n-dimensional vectors and $x=(x_1, \ldots, x_n)$ is the unknown variable. c is an n-vector, $\boldsymbol{A}$ is an $(n \times m)$ matrix, and b is an m-vector.

A generic combinatorial optimisation problem as defined by [Nemhauser & Wolsey 88, pp.3-4] is given by the following:

" Let N=(1, ... , n) be a finite set and c=(c_1, ... , c_n) be an n-vector. For $F \subseteq N$ define $c(F) = \sum_{j \in F} c_j$. Suppose we are given a collection of subsets Φ of N. The combinatorial optimisation problem is: $\max\{c(F) \therefore F \in \Phi\}$ "

A good overview of the classification of optimisation problems can be found in the Introduction chapter of [Papadimitriou & Steiglitz 82].

Many theoretical and practical combinatorial problems belong to the class of NP-complete problems [Aart and Korst 89]. NP-complete problems are those which are:

> "...of such inherent complexity that any algorithm, solving each instance of such a problem to optimality, requires a computational effort that grows superpolynomially with the size of the problem."

Such problems are therefore difficult to solve to optimality due to the prohibitive computational time required [Laarhoven & Aart 87]. According to [Papadimitriou & Steiglitz 82], the general integer linear programming problem is NP-complete.

An alternative approach to solving an NP-complete or large scale combinatorial optimisation problems is to use approximation algorithms for which near-optimal solution can be found within limited amount of computation time. The price to pay is the limited knowledge of the quality of the final solution.

The two previously mentioned approaches classify the optimisation techniques into two classes [Laarhoven & Aart 87, Aart and Korst 89]:

1. optimisation algorithms, which attempt to find an optimal solution at a risk of very long computation time, and
1. approximation or heuristic algorithms, where the accuracy of the optimal solution is traded-off for the reduction in computation time. Within this class, the largest groups are:
    - the local search (or neighbourhood search, or iterative improvement [Papadimitriou & Steiglitz 82]) algorithms, and
    - randomisation algorithms.

In both classes of techniques, two types of algorithms can be distinguished:

1. tailored algorithms, that are tailored to solving a specific problem

1. general algorithms, that can be easily adjusted to a broad range of optimisation problems and therefore are generally applicable i.e. problem independent.

Given the class and size of the system parameter optimisation problem of Scenario 3, a general approximation algorithm was sought which could yield a near-global minimum within a reasonable computation time. The choice was clearly between the local search algorithms, and randomisation algorithms.

Algorithms based on a local search are widely used general approximation algorithms, but they are often of low quality in that the solution obtained is not close enough to the global optimum [Aarts & Korst 89]. These algorithms have a tendency to converge towards a local rather than a global minimum (or maximum), and the solution of the problem depends heavily on the initial starting point.

Among randomisation techniques, there are two classes of algorithms that can find near-global optima of objective functions and are robust in complex state-spaces [Laarhoven & Aart 87, Aart and Korst 89, Drakos 97, Hart 97, Coddington 96]: simulated annealing and evolutionary algorithms. These also overcome the above mentioned disadvantages of local search algorithms, with which they otherwise share some common features. However, the analysis of randomisation algorithms in most cases is very complex, and therefore the values of the parameters required for the quick convergence and accurate results are difficult to determine analytically [Coddington 97, Press 89]. Parameter adjustment is therefore done empirically, through experimentation (trial and error).

Evolutionary algorithms and simulated annealing draw their concepts from natural phenomena: evolution processes in the case of evolutionary algorithms, and the physical process of cooling and crystal formation in metals, in case of the simulated annealing. Their main difference from the traditional optimisation techniques is that they allow an occasional change of the search direction towards the less favourable region of the state space, which allows the algorithm to "jump out" from the local minima (or maxima) and move to a global optimum. Simulated annealing achieves this by accepting the less favourable states to influence the progression of the "cooling" process. Evolutionary algorithms achieve this by searching from a population of solutions, rather than a single point, and by allowing such manipulation of the individual members of the population that could yield less optimal members or solutions. The manipulation operations involved are: mutation, recombination and competitive selection. Mutation makes small

variations in a single element of a solution, thus taking wider state space into the search path. Recombination swaps pairwise the parts of the "fit" members of the population. Competitive selection eliminates less fit members of the population (i.e. less optimal solutions) and keeps the fittest members from the joint set of both old and new solutions generated by mutation and recombination. In this way, mutation and recombination are used to generate solutions that are biased towards the regions of state space in which good solutions have already been found [Hart 97]. According to [Hart 97], evolutionary algorithms are classified into:

1. evolutionary programming, which optimises continuous functions without using recombination,
1. evolutionary strategies, which optimise continuous functions with recombination, and
1. genetic algorithms, which optimise general combinatorial problems, and
1. genetic programming, which applies genetic algorithms to a state space of functions to determine a solution expressed as a function.

For the purpose of this study a simulated annealing algorithm was developed. To validate it, a genetic algorithm was downloaded from the web and the results compared.

The details of simulated annealing and its comparison to the chosen genetic algorithm are described in Appendix B: SCENARIO 3 PARAMETER OPTIMISATION BY SIMULATED ANNEALING.

5. RESULTS

5.1 Introduction

Numerical results were obtained for three traffic types. The number of traffic types was kept small for the sake of clarity of results. However, the formulae of Chapter 3 allow results generalised for any number and combination of traffic types. The parameters of the traffic types used in all the experiments are given in Table 5.1.

	Sources			
	T_{ON} [s]	T_{OFF} [s]	P_{ON}	PCR [cells/s]
Voice	1.54	2.75	0.36	167
Video	20	6.5	0.75	905.66
Data	0.1	1	0.09	1207
Mean arrival rate	$\sum_{k=1}^{3} \lambda_k = 851.5$ cells / s			
Maximum PCR	$\sum_{k=1}^{3} PCR_k = 2279.66$ cells / s			

Table 5.1. Parameters of the sources

A number of different transmission rates were considered: 64 kb/s, 384 kb/s, 512 Kb/s, 1 Mb/s, 1,5 Mb/s, 1,8Mb/s, 2Mb/s, 3Mb/s and 4 Mb/s. The satellite transponder bandwidth is constant and equal to 500 MHz. For the sake of results' tractability and clear conclusions, the quality of service requirements were assumed to be the same for all traffic types. This is expressed in the requirement that the *SNR* at the receiver is the same for all traffic types, and equal to 4.5 dB [ESTEC-QMW]. An *SNR* of 4.5 dB gives, with appropriate FEC coding, a BER between 10^{-5} and 10^{-10} (e.g. convolutional code with rate ½ and length 7 or concatenated convolutional Reed Solomon code [Cain & McGregor 97, Bella 96]). The E_b/N_0 of the transmitted signals is assumed to be 12 dB, and together with the value for *SNR* is based on discussions with ESTEC personnel [ESTEC-QMW].

In the first and second scenarios, the bandwidths of the channels were changed from 25 MHz to 50 MHz in increments of 5 MHz for each system instance. Each combination of the transmission rate R and the channel bandwidth W yields the processing gain $G=W/R$ of the system that determines what theoretical CDMA channel capacity C_{max} is allowed for the given QoS requirements (Eq. 1.4 in Chapter 3), in terms of the maximum number of simultaneous continuous connections. In the third scenario, the transmission rates vary according to the traffic type. In order to have a fair comparison, a constant processing gain that corresponds to the values encountered in the Scenarios 1 and 2 was used to determine the required channel bandwidths for a particular set of parameter values. The optimisation of the number of channels of each type to maximise system capacity was performed for each value of the processing gain obtained by the pairs (R,W). This yielded different channel structures and maximum capacity values, allowing fair comparison of scenarios with one another. Table 5.2 shows the values of different system parameters used in obtaining numerical results.

Transmission rates *R* [kb/s]	64, 384, 512, 1024, 1536, 1802, 2048, 3072, 4096		
SNR	4.5 dB		
E_b/N_0	12 dB		
Channels bandwidths *W*	25 MHz - 50 MHz, in increments of 5 MHz		
	Scenario 1	*Scenario 2*	*Scenario 3*
Processing gain *G* within a system for one instance of parameter values	*W/R=const.*	*W/R = const.*	constant, all values found in Scenarios 1 & 2 used: $G=W_i/R_i$

Table 5.2. Numerical values of system parameters.

It is important to recall how the total load on the system was modelled. As interest was focused on statistical multiplexing in the code domain, which "packs" more connections into the channel than is theoretically allowed, the satellite end-point (information sink in the model) was modelled as a single server with server rate C_{max} times higher than the transmission (slot) rate in the system (Scenarios 1 and 2). That is, the Scenario 1 system load (*m* represents the total number of CDMA channels) is:

$$\rho^{(1)} = \frac{m \cdot \bar{\lambda} \cdot C_{\max}}{m \cdot SR \cdot C_{\max}} = \frac{\lambda_k}{SR} \qquad \text{Eq. 5.1}$$

Scenario 2 system load (U_k represents average number of type *k* in each channel) is:

$$\rho^{(2)} = \frac{m \cdot \sum_k U_k \cdot \lambda_k}{m \cdot SR \cdot C_{\max}} = \frac{\sum_k U_k \cdot \lambda_k}{C_{\max} \cdot SR} \qquad \text{Eq. 5.2}$$

In the case of Scenario 3, the server rate is determined by the sum of slot rates of all connections. The load in this case is the ratio of the aggregate mean rate to the aggregate peak cell rate (k is a traffic type index, m_k number of channels of type k):

$$\rho^{(3)} = \frac{\sum_k \lambda_k \cdot m_k \cdot C_{\max,k}}{\sum_k SR_k \cdot m_k \cdot C_{\max,k}} \qquad \text{Eq. 5.3}$$

Since the loads on different connections depend on the mean arrival and slot rates, and the mean arrival rates in the model are fixed, then in Scenarios 1 and 2 for higher transmission (slot) rates only low loads can be achieved. This is an implication of the fact that the maximum number of simultaneous continuous users or connections C_{max} is fixed.

Once the achievable capacity increase is determined, new achievable loads on the satellite server can be found. Scenario 1:

$$\rho_{new}^{(1)} = \frac{m \cdot \bar{\lambda} \cdot \left(C_{\max} + \Delta C\right)}{m \cdot SR \cdot C_{\max}} = \rho^{(1)} \left| 1 + \frac{\Delta C}{C_{\max}} \right| \qquad \text{Eq. 5.4}$$

Scenario 2 (i represents the traffic type whose extra load is being estimated):

$$\rho_{new}^{(2)}(i) = \frac{\sum_k U_k \cdot \lambda_k + \Delta U_i \cdot \lambda_i}{C_{\max} \cdot SR}, \quad i \in [1, K] \qquad \text{Eq. 5.5}$$

Scenario 3:

$$\rho_{new}^{(3)} = \frac{\sum_k \lambda_k \cdot m_k \cdot \left(C_{\max,k} + \Delta C_k\right)}{\sum_k SR_k \cdot m_k \cdot C_{\max,k}} \qquad \text{Eq. 5.6}$$

The higher the channel bandwidth for the same transmission rate, the higher the processing gain, and hence more users are allowed to transmit simultaneously (i.e. the capacity per channel is increased). However, as the channel bandwidth increases, the number of such channels decreases. Similarly, as the transmission rate increases for a given channel bandwidth, the processing gain decreases, and hence capacity per channel

is reduced. So the question for all scenarios is: what is the ideal (optimal) combination of the transmission rate and channel bandwidth, that yields maximum capacity?

When comparing the scenarios, it is necessary to compare "like with like", that is, only the system instances with equal or very similar system parameter values should be compared: the loads ought to be the same (or very close) and the processing gains (and therefore the immunity against noise of the transmitted signals) should be similar in order to compare fairly the achievable capacities of the three system scenarios.

In the first two scenarios the lowest transmission rate of 64 kb/s is not sufficient for video and data traffic. Cell loss due to buffer overflow would occur, and the queue would be unstable. Those values of the transmission rates that would lead to an unstable queue were not considered.

In all graphs featuring load, efficiency without DTX is essentially the load of the system i.e. satellite server (ratio of cell arrival and cell service rates).

5.2 Comparison with TDMA

To compare fairly the capacity of a statistical multiplexing based CDMA scheme to the capacity of a TDMA scheme, it would be necessary to allow similar mechanisms of efficient bandwidth use in the TDMA scheme (e.g. statistical multiplexing in the time domain). That means that the TDMA MAC candidate for capacity comparison would have to satisfy the following criteria:

- The scheme would have to accommodate heterogeneous traffic.
- Some form of bandwidth (i.e. time slot) allocation as and when required should be supported.
- The sources' silences should be utilised to release the bandwidth and make it available to other connections, thus making efficient use of bandwidth.

Different assignment schemes that make the TDMA MAC protocols were presented in Chapter 2: ATM OVER SATELLITE (sections 2.1.3: Medium Access Control and 2.1.4.3: MAC and CAC in a TDMA system). The assignment schemes were classified into random and controlled, the latter being either demand or reservation assignment, or any combination.

Because those proposed TDMA medium access protocols that use bandwidth efficiently and increase the throughput are based on some form of demand or reservation assignment, some handshaking will take place, introducing the MAC delay, which the statistical multiplexing in the code domain is trying to avoid (this is not the access delay that is inevitable on the connection set-up during the CAC negotiation phase). Furthermore, TDMA MAC protocols for broadband or heterogeneous traffic have to be carefully designed for high efficiency and bandwidth utilisation (which is a subject for a separate study).

On the other hand, random-assignment schemes such as ALOHA or slotted ALOHA, which could be used to support a form of statistical multiplexing (all connections transmit data packets when and as required), have been analysed only for uniform i.e. homogeneous traffic. Besides, random access protocols have low throughputs due to collisions. The medium access scheme presented in this thesis does not allow collisions as the number of codes is very high (long PN codes are used). A high number of PN codes means that no two connections or users would get the same code causing a collision in the code domain (code allocation is performed during CAC). The form of "collision" that may occur is an occasional increase in interference due to statistical traffic fluctuations, which temporarily increases the BER and might therefore require re-transmissions. (The effect of such collisions on message throughput and delay is again a subject for a separate study).

Given the existing TDMA MAC schemes, it is clear that none of them satisfies the requirements outlined above for fair comparison, and hence direct comparison of statistical multiplexing in CDMA with any of the mentioned TDMA schemes is not possible. To enable an accurate but indirect comparison, a capacity analysis of idealised TDMA statistical multiplexing where the resource allocation is based on effective bandwidth was used. Effective bandwidth is the amount of bandwidth higher than the mean rate and lower than the peak rate that ensures the required probability of cell loss due to buffer overflow, given the size of the buffer which the source feeds. The parameters required to determine the effective bandwidth of a connection are mean cell rate, peak cell rate, desired probability of cell loss, and buffer size. The following reasoning was applied:

A TDMA time frame is of fixed length T_F, and is divided into a fixed number of time or burst slots of duration T_{burst}. To allow effective bandwidth allocation to individual connections, each connection of type *i* requiring effective bandwidth $W_{eff}(i)$ must get N_i burst slots in the frame. The transmission rate during a burst is fixed i.e. $R_{burst} = const.$, and so is the number of cells n_{cells} sent in a single burst:

$$n_{cells} = R_{burst} \cdot T_{burst} \qquad \text{Eq. 5.7}$$

In Eq. 5.7 R_{burst} is in [*cells/s*]. The amount of information (number of cells) sent during one time frame must be equivalent to the effective bandwidth allocated to a single connection of type *i*, which can be expressed as:

$$W_{eff}(i) = \frac{N_i}{T_F} \cdot R_{burst} \cdot T_{burst} \quad [cells/s] \qquad \text{Eq. 5.8}$$

The sum of all effective bandwidths is equal to the transmission rate R_{burst}:

$$\sum_i W_{eff}(i) = R_{burst}. \qquad \text{Eq. 5.9}$$

As each frame serves all *K* connections, then the total capacity is:

$$C_{stat.mux.} = K \cdot (\text{Number of channels}) = K \cdot \frac{W_{sat}}{R_{burst}} \qquad \text{Eq. 5.10}$$

Thus if we know the total required effective bandwidth we can determine the transmission rate during a time burst, and consequently the capacity. The following formula developed in [Schormans *et al*. 94] was used to determine the effective bandwidth:

$$W_{eff} = \frac{-\left(T_{ON}T_{OFF} \cdot PCR \cdot \ln P_L + B \cdot \left(T_{ON} + T_{OFF}\right)\right)}{-2 \cdot T_{ON} \cdot T_{OFF} \cdot \ln P_L} \pm \pm \frac{\sqrt{\left[T_{ON}T_{OFF}PCR \cdot \ln P_L + B\left(T_{ON} + T_{OFF}\right)\right]^2 - 4BT_{ON}^2 T_{OFF} PCR \cdot \ln P_L}}{-2 \cdot T_{ON} \cdot T_{OFF} \cdot \ln P_L} \qquad \text{Eq. 5.11}$$

B = Buffer size in cells. Value B=1000 used.

P_L = Cell loss probability required (due to buffer overflow).

For comparison, the effective bandwidth required for the cell loss probabilities of 10^{-4}, 10^{-5}, and 10^{-6} was calculated. As will be shown in the following sections, the TDMA

capacity plots for the three values of BER are very close and yield a capacity much greater than that achievable with statistical multiplexing in CDMA. However, any realistic TDMA scheme would require guard times between the burst transmissions, as well as synchronisation of burst transmissions from different users, which would reduce the achievable TDMA capacity found from Eq. 5.10, i.e. the maximum number of connections that can be served for the required CLR.

For a CDMA system in this study, the cell loss considered is interference-related i.e. caused by the channel. Any cell loss due to buffer overflow would be an additional degradation of the end-to-end cell loss performance on top of the cell loss that was considered in this study. As mentioned earlier, a BER lower than 10^{-5} can be achieved with the assumed value of SNR=4.5 dB and an adequate FEC coding (e.g. convolutional code or concatenated convolutional Reed Solomon code [Cain & McGregor 97]), yielding a CLR of the order lower than 10^{-7}, based only on ATM cell header error correction, and disregarding effect of error in the payload on the total CLR. The figure of 10^{-7} follows from the result of COST226 project [COST226 94], which gives the approximation formula that relates random bit error rate and cell loss rate:

$$CLR = 2380 \cdot (BER)^2$$ [COST226 94].

5.3 Scenario 1

As mentioned in Chapter 3, multiplexer delay and state distributions for Scenario 1 are difficult to analyse using Queueing theory, as there are no results for the case where the sources feeding a single queue are heterogeneous and their number is small. Thus in order to obtain accurate results for the performance of Scenario 1 a hybrid analytical-simulation approach was used.

Finding the capacity increase from using DTX requires the mean busy and idle periods of a multiplexer. However, the analysis represented in Chapter 3 did not take into account the effects of queueing on duration of the busy and idle periods of the multiplexer. Queueing effects in a multiplexer are prominent in cases where the slot rate is very close to the aggregate mean cell arrival rate, i.e. the traffic load in the multiplexer is very high (close to 1). For the traffic input considered (Table 5.1), high loads correspond to the lower values of transmission rate R (384 kb/s). For loads close

to 1 (e.g. ρ=0.94) it can be expected that the multiplexer is busy virtually *all the time*. That is, the probability of a busy period is close to 1 (contrary to the results of Chapter 3 analysis, which is based entirely on the aggregate process of the three sources and does not include the queueing effects, and therefore yields the probability of ON period of 0.86 even for very high loads). For the case of high load, given that the multiplexer is busy nearly all the time, no capacity increase can be achieved since the transmitted signals become continuous rather than ON-OFF. The system becomes the conventional continuous-transmission CDMA system with capacity limit given by C_{max}. Analytical results for R=384 kb/s showed irregularities, such as efficiency higher than 1, which was caused by the inaccuracy of the Scenario 1 analysis for loads close to 1 (these inaccuracies in the analysis resulted from omitting the effect of queueing in the multiplexer at high loads on the busy periods at multiplexer output). Thus to verify the analysis and find the confidence limits within which it can be safely applied simulation was required. It was shown by simulation that already for R=512 kb/s (1207 cells/s), although the slot rate is still below the aggregate peak cell rate (2279 cells/s), the queueing is low and therefore has little effect on analytical results. Hence the analytical results for R=512 kb/s and higher are close to those obtained by simulation.

5.3.1 What was simulated for Scenario 1

The system configuration in Scenario 1 is simple and uniform. Thus it is sufficient to simulate one user terminal to find out how the system behaves as a whole. As all user terminals are identical, simulation of one VSAT terminal gives the mean busy and idle periods at the output of the terminal multiplexer, which in turn gives the relative capacity increase. Since the simulator models the system on the cell level, there is no scope to simulate the effects of different channel bandwidths. However, as the system is uniform, any variations versus the channel bandwidth are in fact due to the bandwidth quantisation error, i.e. the bandwidth waste caused by dividing the total bandwidth into channels. For example, when the total bandwidth of 500 MHz is divided into the 35 MHz channels, 20 MHz remains unavailable for capacity in the analysis.

In the simulation experiments a MICROSIM input configuration file consisted of three sources feeding a single queue. The queue had a model of a DTX mechanism that dictated the busy and idle periods of the multiplexer i.e. its output stream. Insertion of

empty cells during ON periods between the cell arrivals as well as the hangover periods was modelled. The code re-synchronisation overhead was not modelled as it comprised at most a single cell, and thus did not contribute significantly to the queueing in the multiplexer and the numerical results. Hangover overhead has a much greater impact on the performance than the code re-synchronisation overhead.

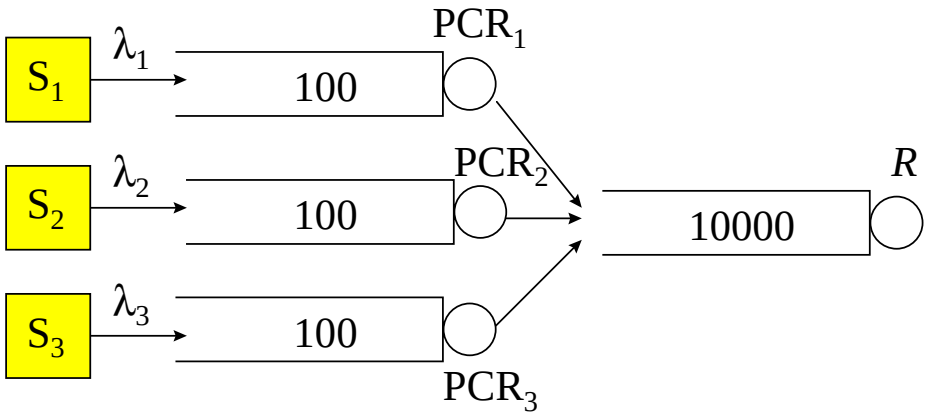


Figure 5.1 Simulated terminal configuration.

The illustration of the configuration is given in Figure 5.1.

In the simulation experiments the buffer size was set to 10000 cells, a figure consistent with dimensioning of modern ATM switches [Winstanley 96]. No cell loss at all was detected at $\rho<0.96$. When $\rho=0.96$ the CLP was found to be 10^{-2}. This sudden increase in the CLP value is consistent with what is known about the asymptotic behaviour of queueing systems as $\rho->1$ [Bruneel & Kim 93, Pitts & Schormans 96].

5.3.2 Capacity for Scenario 1

Figure 5.2 shows the comparison between the values for the busy period of a multiplexer when obtained by analysis (Appendix A, Eq. A.12) and by simulation. It is these busy periods of the aggregate cell stream coming out of a multiplexer that determines how much of the CDMA capacity will be used, and therefore how many extra cell streams (users) the system can accept for the given BER. The consequent relative capacity increase is shown on Figure 5.3.

Figure 5.2 demonstrates that the analysis is very accurate for most but the highest loads: for loads lower than ρ=0.75 the simulation and analysis give very similar values. For loads higher than ρ=0.75, the queueing effects become significant and the analysis, which does not take into account queueing in the multiplexer, loses accuracy. The simulated results very quickly converge to the analytical line following an exponential decrease at higher loads. As the server load should not exceed 0.7 in practical applications [Winstanley 96, Schormans 97], the analysis give excellent prediction of

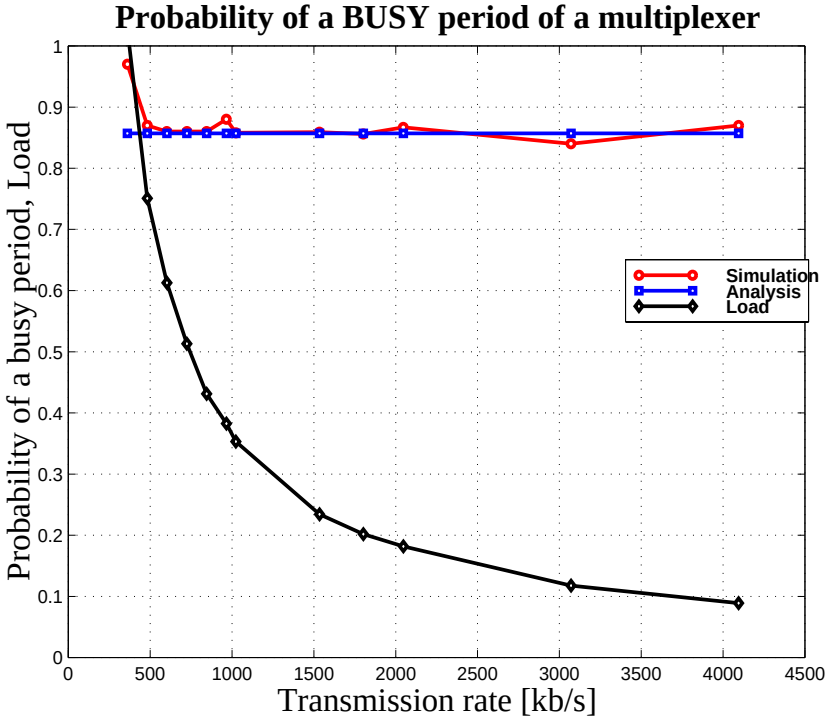


Figure 5.2 Probability of a multiplexer being busy, *P(B)*.

the probabilities of busy and idle periods of the multiplexer, and hence be used within connection admission control mechanisms.

Figure 5.3 shows the relative capacity increase that corresponds to the simulated and analytical busy periods shown on Figure 5.2. The simulated (thick red) line on Figure 5.2 is obtained by substituting the values for the probability of the multiplexer busy period P(B) from Figure 5.2 into the formula:

$$\frac{\Delta C}{C_{\max}} = \frac{P(I)}{P(B)} \qquad \text{Eq. 5.12}$$

The blue thin lines represent the analytical results for various channel bandwidths. They were obtained by calculating the actual number of users for a given processing gain (i.e. ratio of the channel size and transmission rate, channel size being the parameter of the

plots), and by finding how many more users could be added in each channel on the basis of the equation 5.11. For very small values of processing gain (G<20), the integer rounding effect can be observed: the capacity of a single channel is so small (4 or 5 users) that the 15% of it is less than 1, yielding no capacity increase. This is why some of the curves drop to 0 at high transmission rates (small values of G). This is what was termed the bandwidth quantisation error. As explained in the previous section What was simulated for Scenario 1, due to system uniformity, the variations in channel size should not alter the system relative capacity increase.

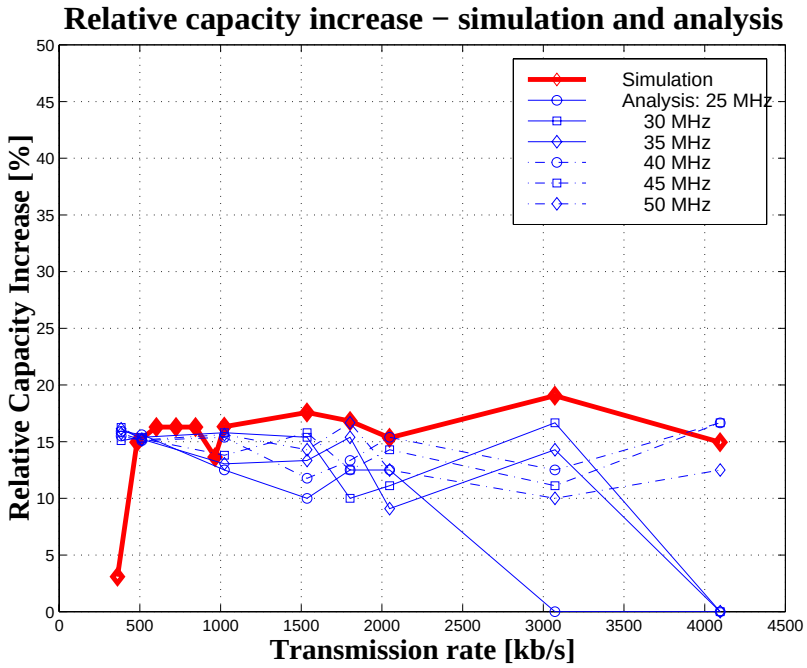
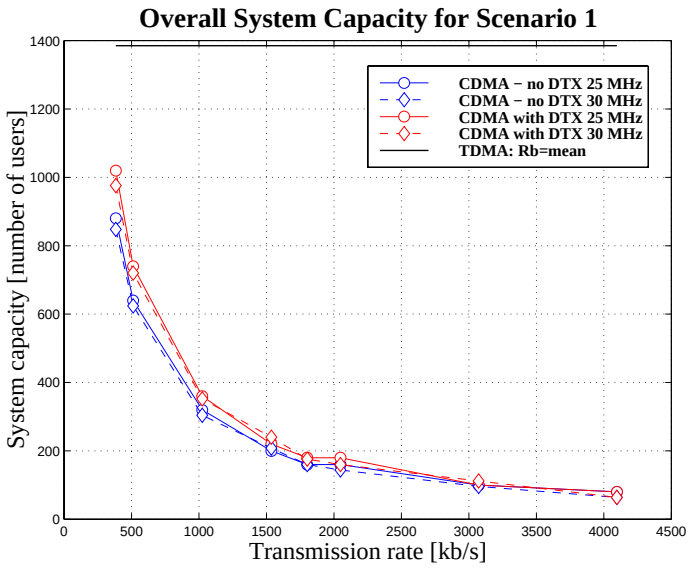
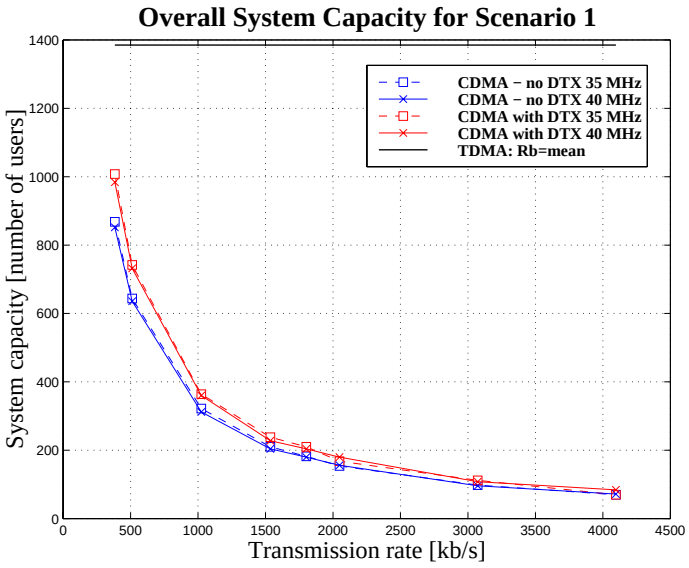


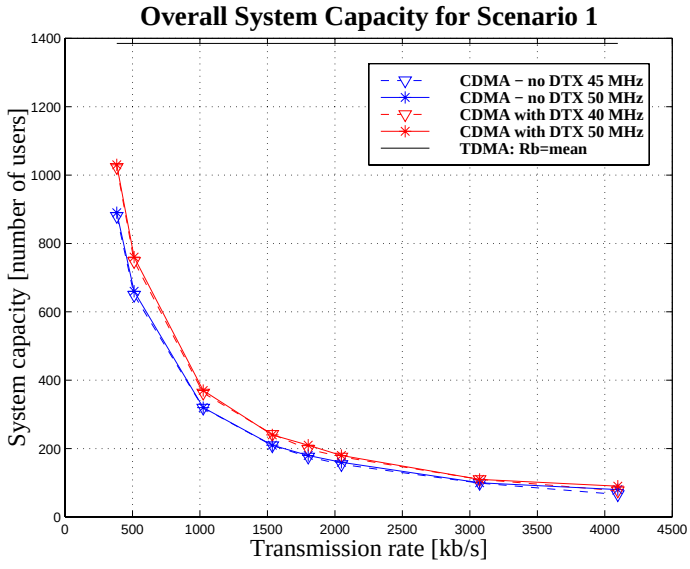
Figure 5.3 Relative capacity increase vs. transmission rate. The thin blue lines represent analytic results for various channel bandwidths.



a)



b)



c)

Figure 5.4 Overall system capacity vs. transmission rate R for various channel bandwidths W_{ss}, and comparison to TDMA when data rate is equal to the mean rate (R_d=λ [kb/s]): a) W_{ss}=25 MHz, W_{ss}=30 MHz, b) W_{ss}=35 MHz, W_{ss}=40 MHz, c) W_{ss}=45 MHz, W_{ss}=50 MHz

Total system capacity

Figure 5.4 gives a comparison of the total system capacity achievable with statistical multiplexing in the code domain, versus the idealised statistical multiplexing-based TDMA scheme described in section *Comparison with TDMA*. For Scenario 1, due to system uniformity and simplicity, the equivalent statistical multiplexing TDMA capacity can be directly derived by using the mean aggregate cell arrival rate from the three traffic types and substituting this rate into R_{burst} of expression 5.10. The resulting equivalent TDMA capacity is represented by the black line on Figure 5.4.

The total CDMA system capacity decreases with transmission rate, as expected (Figure 5.4). This is due to the decrease of processing gain $G = W_{ss}/R$, which determines the capacity per channel. For a given channel bandwidth as the transmission rate increases, the processing gain decreases, and so does the capacity per channel, and the overall capacity drops. On the other hand, for a fixed transmission rate, the channel bandwidth varies only slightly (in increments of 20% or less of the starting value of 25MHz), and so does the processing gain for a fixed R. At the same time, increase in channel bandwidth reduces the number of channels in the system. That is why the total system capacity does not vary much with the change of the spreading bandwidth for the same transmission rate (976 users for 30 MHz to 1030 users for 50 MHz channels at 384 kb/s, or 720 users for 30 MHz to760 users for 50 MHz at 512 kb/s; i.e. up to 5% difference from the lowest capacity value). The extent to which the system capacity is influenced by the channel bandwidth is dictated by the bandwidth "waste" that is not taken into account by the analysis, and which is due to the ratios of the total bandwidth and the studied channel bandwidths. This was termed “bandwidth quantisation error” in the previous section. E.g. for channels of 30 MHz, there is an unused portion of the satellite band of 20 MHz (500 MHz = 16 channels of 30 MHz + 20 MHz). Had this portion of bandwidth been used, the equivalent system capacity would have come close to the maximum observed values. In other words, when the whole available bandwidth is used, the partitioning of this band into the channels of identical bandwidth does not increase the capacity of a single-channel system. However, spreading to 500 MHz requires approx. 500 Mb/s transmission i.e. chip rate, which is too high a rate for cheap implementation (the requirement for a very high-power user terminals will make the terminals expensive, similar to the TDMA system, which is exactly the problem CDMA

attempts to eradicate). Furthermore, the channel partitioning increases both the system efficiency and the capacity when non-uniform channels are used to support different traffic types, as it will be shown later in Scenario 3.

Disregarding the transmission rate of 384 kb/s, which causes the multiplexer to be busy almost 100% of the time (load ρ=0.92), the total system capacity is highest for the lowest transmission rate of 512 kb/s in all cases of channel bandwidths.

Relative capacity increase

The relative capacity increase (Figure 5.3) from the old to the new capacity obtained by analysis is between 10% - 15%, and is lower for higher transmission rates. Due to the integer rounding effect the trend cannot be clearly observed. This is again a consequence of reduced processing gain and associated capacity per channel. For high transmission rates and low processing gains, the capacity per channel can be as low as 5 or 4 connections. For such a low number of connections a potential 15-17% increase becomes a number of extra connections less than 1. (The increase can be estimated using the idle and busy periods: for probabilities of approximately P_{IDLE}=0.14 and P_{BUSY} = 0.86, as is the case in Scenario 1, the relative capacity increase is approximately P_{IDLE}/P_{BUSY} = 0.14/0.86 = 0.16). Thus the rounding causes the analysis to give 0% increase of capacity for cases where transmission rate is high and channel bandwidth is low.

Results based on simulation show that the relative capacity increase is approximately constant for loads equal and lower than ρ=0.7 (transmission rate R=512 kb/s and higher), which is in agreement with intuition. Constant relative capacity increase is to be expected, since the probability of the busy and idle periods is constant when the queueing effects are not significant. The relative capacity increase was found by substituting the measured busy and idle periods into the formulae 3.10-3.13. Simulation results for 384 kb/s show that when the system is operating close to the 100% load no significant capacity increase can be achieved (i.e. increase is around 3%), which, again, is in agreement with intuition.

5.3.3 Efficiency for Scenario 1

The nature of Scenario 1 is such that only the system capacity changes with the change of the channel bandwidth. As the P_{ON} and P_{OFF} are found from the aggregate process of

the three sources, and this process does not change by altering the channel (spreading) bandwidth, the efficiency values do not change with varying the CDMA channel bandwidth W_{ss}.

In Scenario 1 efficiency of the system with DTX decreases with the increase of the

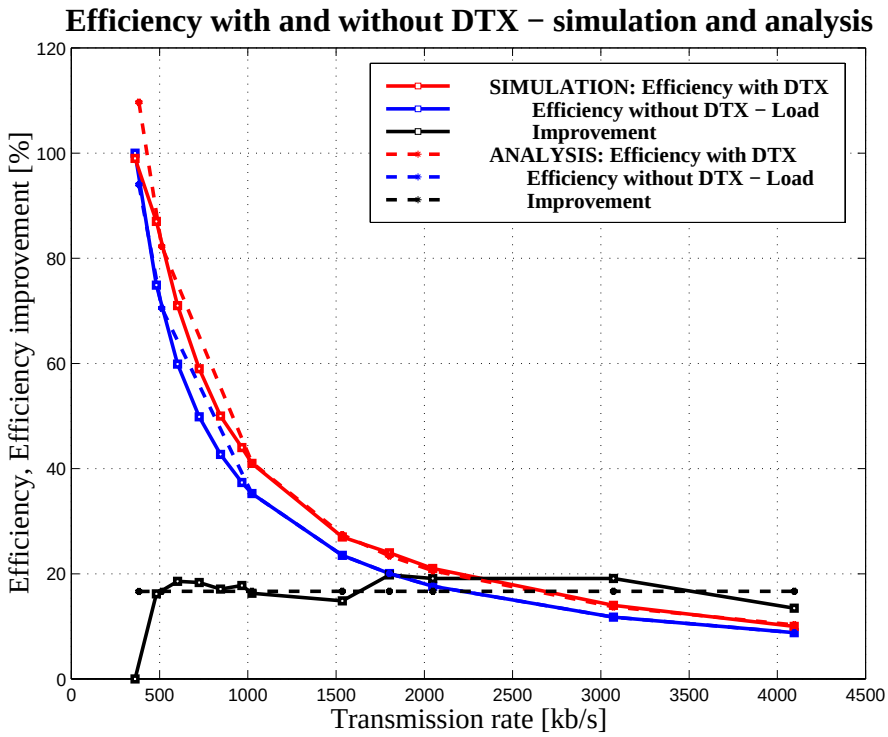


Figure 5.5 Comparison of analytical and simulation results for efficiency and efficiency improvement.

transmission rate (from 0.82 at 512 Kb/s to 0.10 at 4Mb/s, Figure 5.5). This is intuitive as more empty cells are inserted during the active periods of the multiplexer at higher service (slot) rates. The relative efficiency improvement is constant (17%) for all transmission rates; simulation showed the same trend. Hence for a more efficient system it is necessary that the transmission rate is as low as possible (so that fewer empty cells are transmitted). However, very low transmission rates cause high loads in the multiplexer, heavy queueing, extended busy periods of the multiplexer and consequently reduced relative capacity increase. It was shown in the previous section that for loads of 0.7 and lower the queueing effects are not significant, so for the best efficiency the transmission rate should be chosen around an operating point of ρ=0.7. This also ties up with the capacity results in the previous section.

From the formulae derived in Chapter 3 it is clear that the statistics of the multiplexed sources will have significant effect on the simulated results of efficiency and efficiency increase. The efficiency improvement produced by DTX increases as the probability of an ON period of a multiplexed stream of traffic decreases, regardless of the transmission rate. Efficiency is also affected by the *time scales* of the sources' ON and OFF periods, i.e. the mean length of their duty cycle. The sources that have shorter periods of ON and OFF states will require insertion of greater aggregate overhead, as they will change states more often over a fixed period. The insertion of overhead generates additional cell rate, and increases the 'aggregate' load on a link i.e. satellite server. This aggregate load must not exceed 1 for a system to be stable. In the cases studied, the aggregate load was always very close to the offered traffic load, as the additional load due to the overheads was insignificant. The hangover periods equivalent to the cell inter-arrival time of the slowest source and the code acquisition periods were not comparable to the periods of the ON and OFF states of the sources. The hangover represented a greater part of the overhead than code re-synchronisation. However, in sources with short duty cycles and only a small number of cells in a burst, the duration of the overhead will become significant and comparable to the duration of the actual active states of the sources.

In general, the traffic mix is important in determining the optimum transmission rate for Scenario 1. The transmission rate has to be low enough to keep the system efficient (small number of inserted empty cells), while maintaining the aggregate load below 1.

5.3.4 Buffer sizes and delay for Scenario 1

Simulation was used to determine the queue and delay state distributions for various transmission rates (Figure 5.6 and Figure 5.7). The results obtained are typical examples of queueing behaviour described extensively in literature on Queueing theory [Pitts & Schormans 96, pp.134, Bruneel & Kim 93, pp.177]. Buffer states distributions (Figure 5.6) exhibit a typical shape for cell-scale and burst-scale queueing.

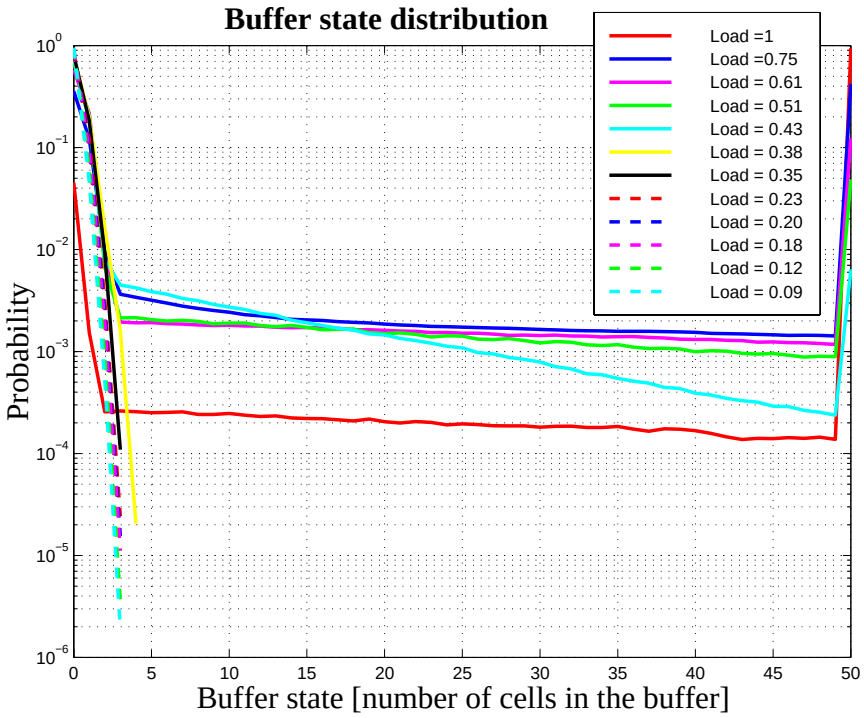


Figure 5.6 Buffer state distribution - simulation.

The buffer state probability distributions show that for very high loads the probability of any of the states between 1 and 50 is very low, and the buffer is more likely to be in a very high state (the probability distribution function was truncated at state 50 and all states above 50 were accounted for in the probability for state 50; the buffer size was 10000 cells). As the load decreases, the probability of any of the lower states increases (Figure 5.6), as expected. For load ρ=1 (full red line) the graph shows that the state probabilities are an order of magnitude lower than for the loads between 0.75 and 0.51, and constant. This means that a high proportion of cells find the buffer full or in states higher than 50. For loads between 0.75 and 0.51 the intermediate states are more probable (the lines are shifted up by an order of magnitude), although the distribution function is still pretty flat, meaning that the buffer is more likely to be in high than in low states. A load of 0.43 produces an inclined curve whereby high states are less probable than the low ones. For loads lower than 0.40 the queue never exceeds 5 cells.

Simulation did not produce any cell loss for loads equal to and lower than 0.75.

Delay distributions are similar to buffer distributions. For the load of ρ=1 (full red line) the probability of delay between 1 and 50 cells is very low. Higher loads show longer delays on average.

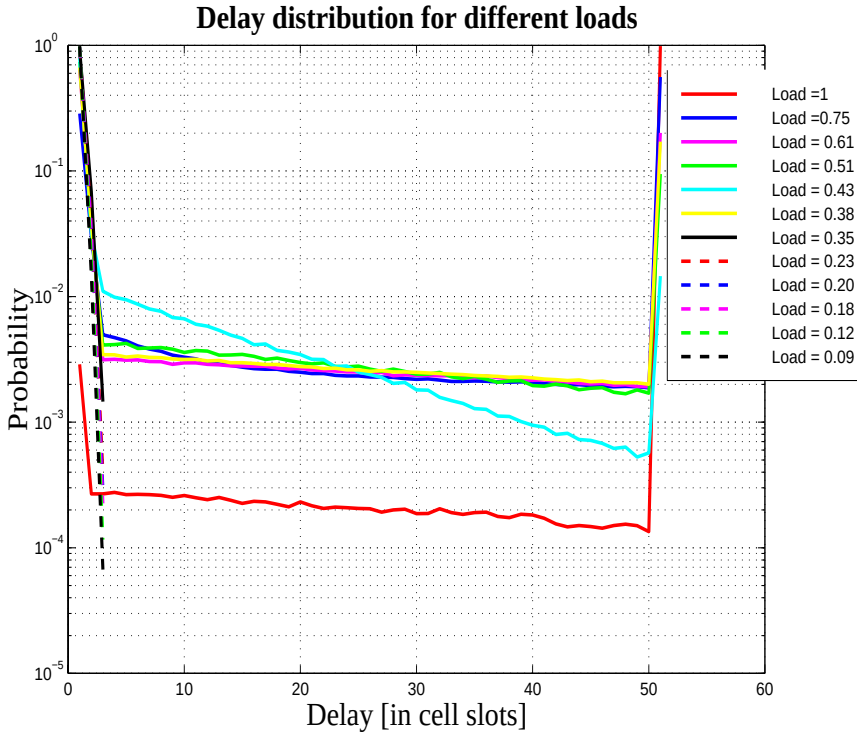


Figure 5.7. Delay distribution - simulation.

5.4 Scenario 2

In Scenario 2 different connection types coexist within the same channel in different numbers. The load within a channel and within a system is determined by the relative distribution of connection types, and by the predominant traffic type (if there is one). For instance, if the majority of connections are voice traffic, the system load will be in the vicinity of 0.35 (voice activity factor), and with a majority of video traffic, given the model of Table 5.1, the load will move towards 0.75. As there are many combinations of connections that could give the same or similar load, the plots shown (Figure 5.12-Figure 5.19) represent the results arising from an "average" combination required to produce a particular load, given the transmission rate and channel bandwidth. That means that for each pair (R, W_{ss}) all combinations of connection types (U_1, U_2, U_3) were analysed and then ordered by the load. The representative combination of the load $\left(\overline{U_1},\overline{U_2},\overline{U_3}\right)$ was then used for finding the average values of the unused capacity that would allow capacity increase by statistical multiplexing, the efficiency when DTX is employed, the exact load - that is, efficiency - when DTX is not employed, and the percentage of the total maximum number of extra voice, video or data connections that could be admitted to fully utilise the unused capacity of the system. It should be recalled

that as the channel capacity is shared among different connection types, one cannot give a definite value of capacity increase, but the upper limit values for each connection type, i.e. the increase achievable when the unused capacity is completely allocated either to voice, video, or to data connections.

5.4.1 What was simulated for Scenario 2

Figure 5.8 shows simulated configurations. The simulation experiment was focused on measuring busy and idle periods of each source as the transmission rate varied (Figure 5.12-a), and efficiency, that is the total number of useful and empty cells (Figure 5.18-a). For verification of the capacity results the most adequate simulation would be the one that would measure the interference levels on the physical layer that would determine the (interference limited) capacity available. As simulating two layers at the same time would be difficult due to the different time scales of events (i.e. it would take far too long to complete for a single instance of the system), alternative ways of verifying the analytical results for capacity and relative capacity increase were developed. It should be noted that once the busy and idle periods of individual simulated sources are shown to comply with the analytical results when simulated on the cell level, it is pointless to further measure the aggregate busy and idle periods of a bunch of connections that make a single channel to verify the amounts of unused capacity in it, as this would logically agree with the analysis. In Scenario 2, queueing effects do not alter analytically obtained results as the transmission rates considered produce reasonable loads.

It was already mentioned that the number of combinations (U_1, U_2, U_3) is high, particularly for high values of capacity per channel $C_{max}=U_1+U_2+U_3$. The number of combinations becomes computationally unmanageable when all channels are allowed to have different traffic distributions i.e. combinations (U_1, U_2, U_3) of connection types. Therefore, for the sake of computational manageability, both in simulations and numerical results, it was assumed that the traffic distribution is identical in all channels for a single system instance. That is, all W_{sat}/W_{ss} channels are assumed to have the same traffic distribution (U_1, U_2, U_3). Thus the simulations were performed only for a single channel. The configurations comprised of $\overline{U_1}, \overline{U_2}$ and $\overline{U_3}$ sources feeding their respective queues all having a DTX mechanism, and that all fed a single queue

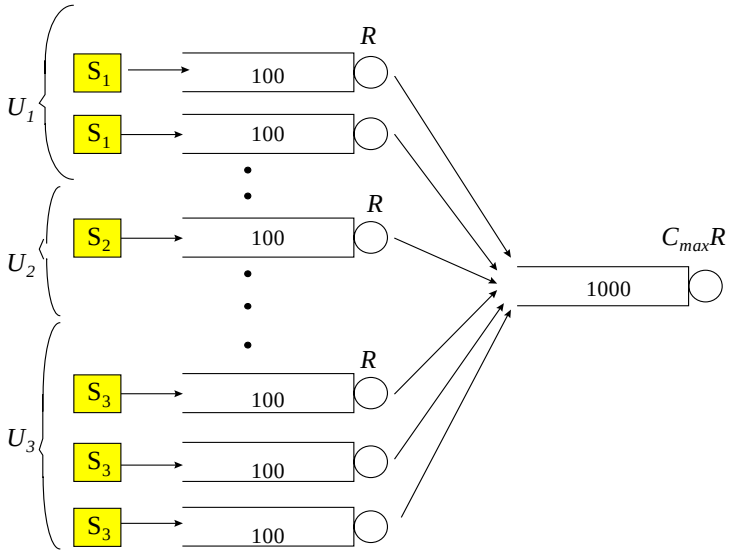


Figure 5.8 Simulated configurations. The notion of "terminal" is not relevant in Scenario 2, as the codes are allocated to VCCs, and thus a terminal can have any number and combination of connections.

representing a satellite buffer (a sink or destination node). Only the configurations for a channel bandwidth W_{ss} =25 MHz were simulated (C_{max} = 44, 32, 16, 10, 8 for R = 384, 512, 1024, 1536, 1802 kb/s). Useful and empty cells were counted to find efficiency when DTX was employed.

5.4.2 Verification of the results for Scenario 2 by alternative analysis

The analysis for Scenario 2 developed in Chapter 3 works with time averages of the total generated CDMA self-interference. The capacity is defined as an average number of connections that could be statistically multiplexed together, given the time average of the produced self-interference. The process on which the analysis is based is illustrated on Figure 5.9. Here the *y*-axis represents the amount of interference generated by the random activity of the existing connections, shown in the top part of the picture. C_{max} is determined by the limit on SNR that gives the maximum acceptable BER. For clarity C_{max}=4 in Figure 5.9. The $K\Delta$ is a simplified representation of the amount of interference produced by K simultaneously active connections ($K \le C_{\max}$). The actual interference is not a linear function of the number of simultaneously active connections. The plot of SNR in decibels versus the number of active connections K is shown on Figure 5.10. However the principle of time averaging the activity of all connections to find the unused capacity remains the same.

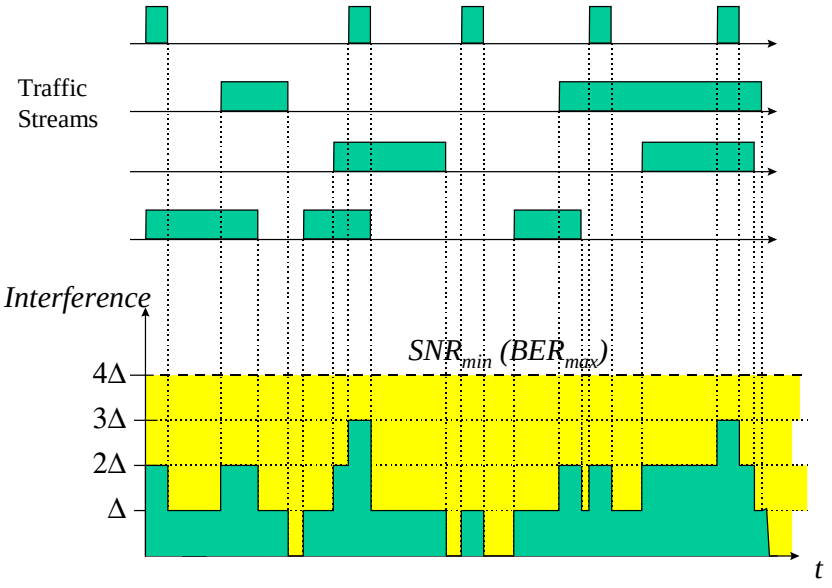


Figure 5.9 Illustration of the interference-limited randomly changing CDMA capacity.

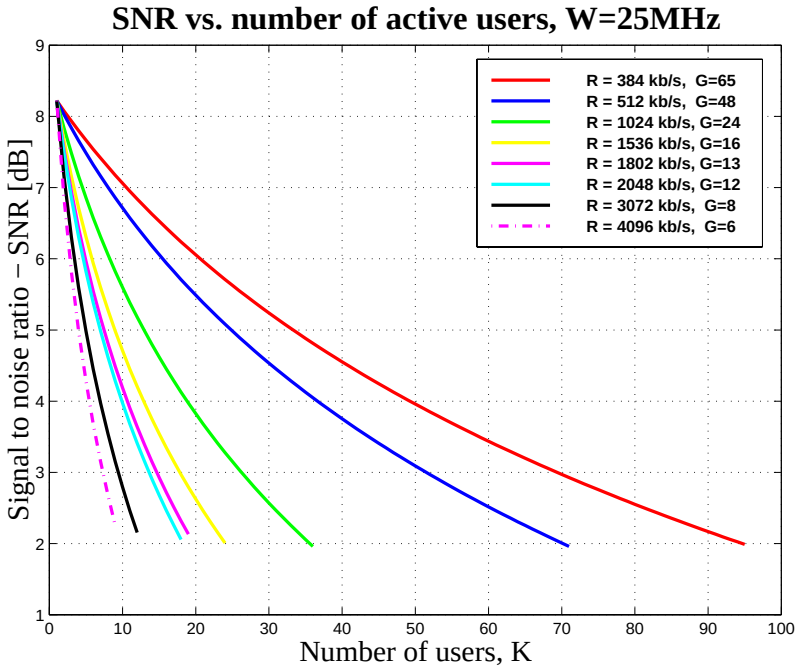


Figure 5.10 SNR versus number of simultaneously active users.

If the average number of active connections for individual system instances can be found using a different approach, then the validity of the analysis presented in Chapter 3 can be proven. Using the knowledge of the activity factors of the connection types, it is possible to find the probability distribution function of the number of active users for a given combination (U_1, U_2, U_3). The mean of the probability distribution function would give the mean used CDMA capacity, and the difference C_{max} - $\overline{K}$ would represent the mean utilised capacity. The probability distribution function of the number of

connections active at any one time can also be mapped to the plot of SNR vs. K to give the probability distribution function of the interference levels i.e. SNR.

The probability distribution function for a system instance with C_{max} limit for the maximum number of simultaneous transmissions and a combination of connection types (U_1, U_2, U_3) can be expressed as:

$$P[K] = \sum_{n_1+n_2+n_3=K} \left| \binom{U_1}{n_1} \right| \left(\binom{U_2}{n_2} \right| \left(\binom{U_3}{n_3} \right| \cdot \left(p_{on,1}\right)^{n_1} \cdot \left(p_{off,1}\right)^{U_1-n_1} \cdot \cdot \left(p_{on,2}\right)^{n_2} \cdot \left(p_{off,2}\right)^{U_2-n_2} \cdot \left(p_{on,3}\right)^{n_3} \cdot \left(p_{off,3}\right)^{U_3-n_3} \quad \text{Eq. 5.13}$$

where n_1, n_2 and n_3 represent the number of connections of type voice, video and data simultaneously active.

The plots for the probability distribution function of K for various (U_1, U_2, U_3) and C_{max}=44 (corresponding to W_{ss}=25 MHz and R = 384 kb/s) are given on Figure 5.11. It can be seen that as the number of more active connections increases (i.e. the number of video connections U_2 with highest activity factor of 0.75 increases), the peak of the PDF moves towards the right, i.e. higher K. The plots show the mean of the PDF i.e.:

$$\overline{K} = \left\lfloor \sum_{i=1}^{C_{\max}} i \cdot P[i] \right\rfloor \quad \text{Eq. 5.14}$$

and the corresponding percentage of the unused capacity: $\frac{C_{\max} - \overline{K}}{C_{\max}}$.

The comparison of the analysis presented in Chapter 3 and Appendix A, and this approach, is given in Figure 5.12-b. The two curves are very close to each other, small differences arise from the averaging and rounding errors. Figure 5.12-a shows simulated busy periods of the three traffic sources as transmission rate is varied.

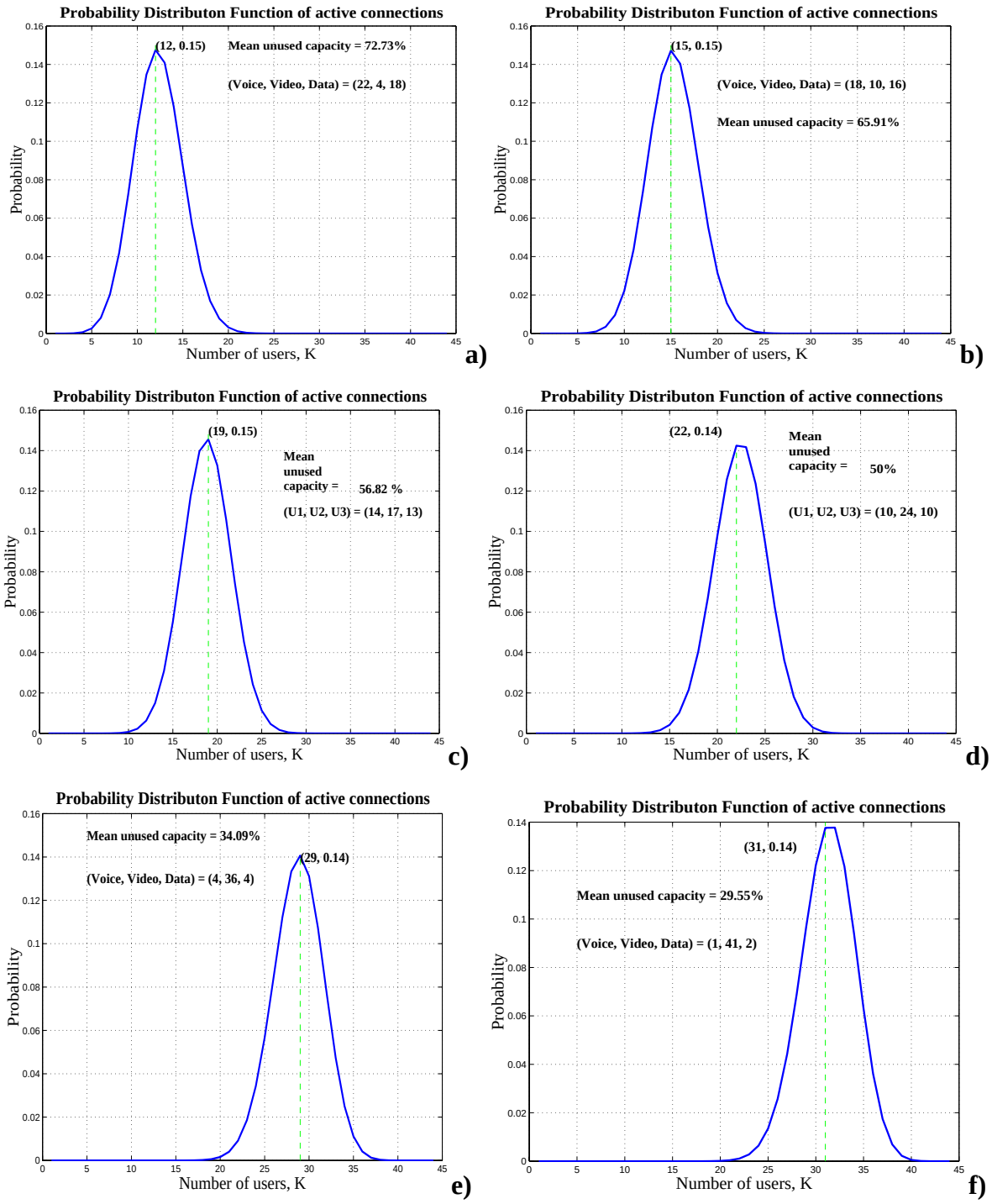
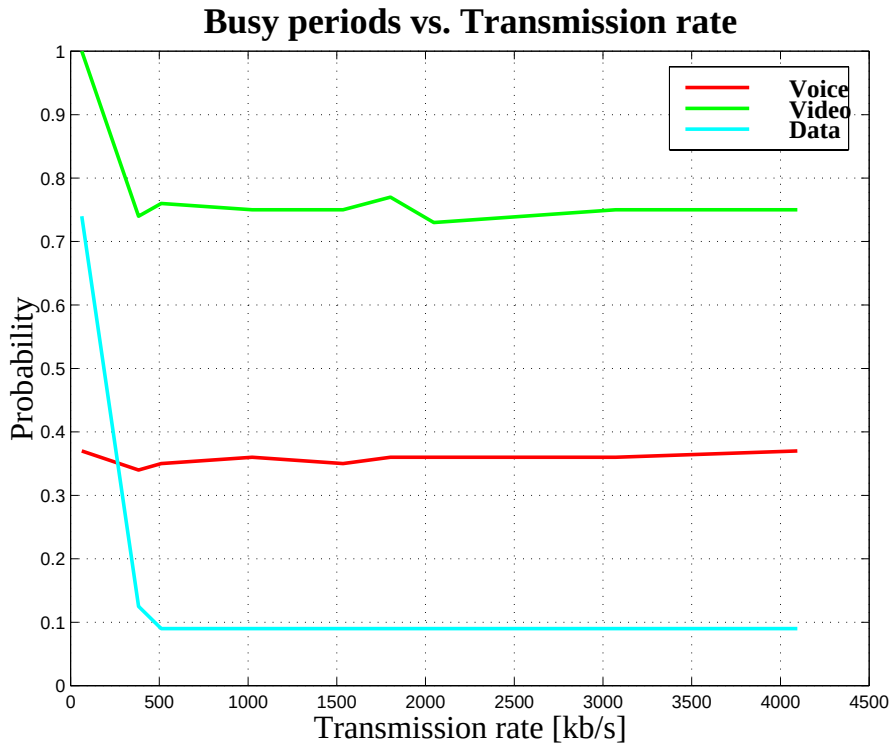
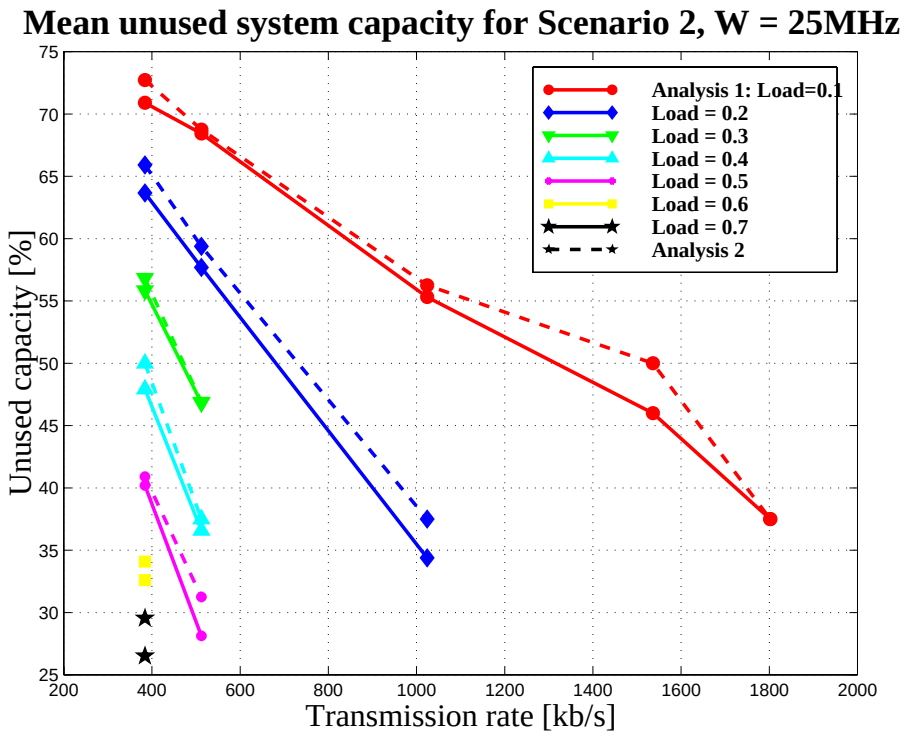


Figure 5.11 Probability distribution function of the number of active users *K*, for various loads i.e. combinations of users *(U_1, U_2, U_3)* from: a) voice predominant (satellite load ρ=0.1 for *R*=384 kb/s) to f) video predominant (satellite load ρ=0.7 for *R*=384 kb/s)

As for Scenario 1, it shows that for very high loads (transmission rate < 384 kb/s) the buffer server is busy most of the time. For loads less than 0.75 the analysis gives an accurate prediction of the duration of the busy periods of the server. Figure 5.13 shows analytic results for channel bandwidth of 30 MHz and 35 MHz.



a)



b)

Figure 5.12 a) Busy periods of sources when transmission rate is varied, and b) how these busy periods translate into the mean unused capacity: dotted lines represent the results from the analysis of the p.d.f. of the mean number of active users at any one time for channel bandwidth of 25 MHz.

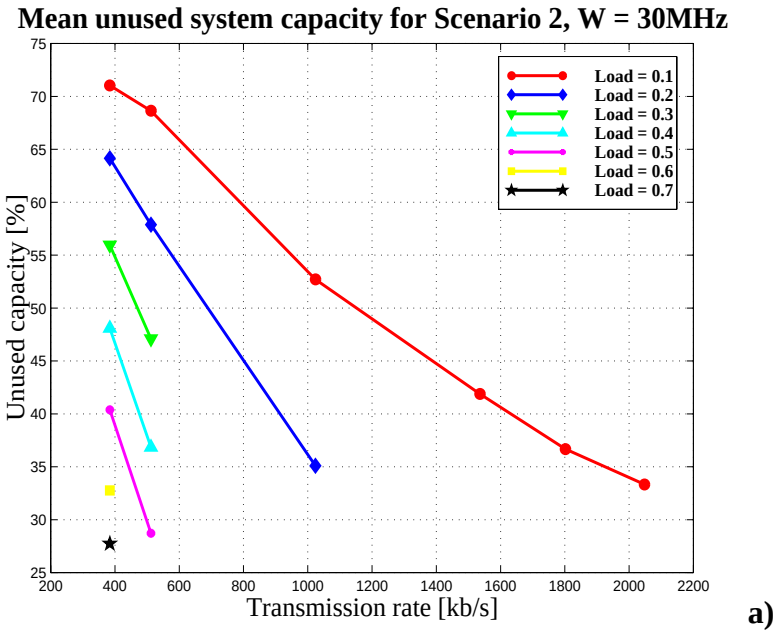


Figure 5.13 Unused available capacity for channel bandwidths of a) 30 MHz and b) 35 MHz.

5.4.3 Capacity for Scenario 2

Total system capacity

Figure 5.14 and Figure 5.15-a show the maximum capacities that can be achieved when all unused capacity is allocated to one particular connection type, and Figure 5.15-b shows what happens when the unused capacity is equally shared among all traffic types. The maximum capacity that can be achieved is around 2000 connections, which is lower than for Scenario 1, as Scenario 1 has a maximum capacity of around 750 users at 512

kb/s, which is equivalent to 2250 connections (for each user terminal 3 connections of the three traffic types are first multiplexed together). The effect of the traffic activity factor is obvious: as the activity factor decreases, more connections can be statistically multiplexed together to increase the overall capacity.

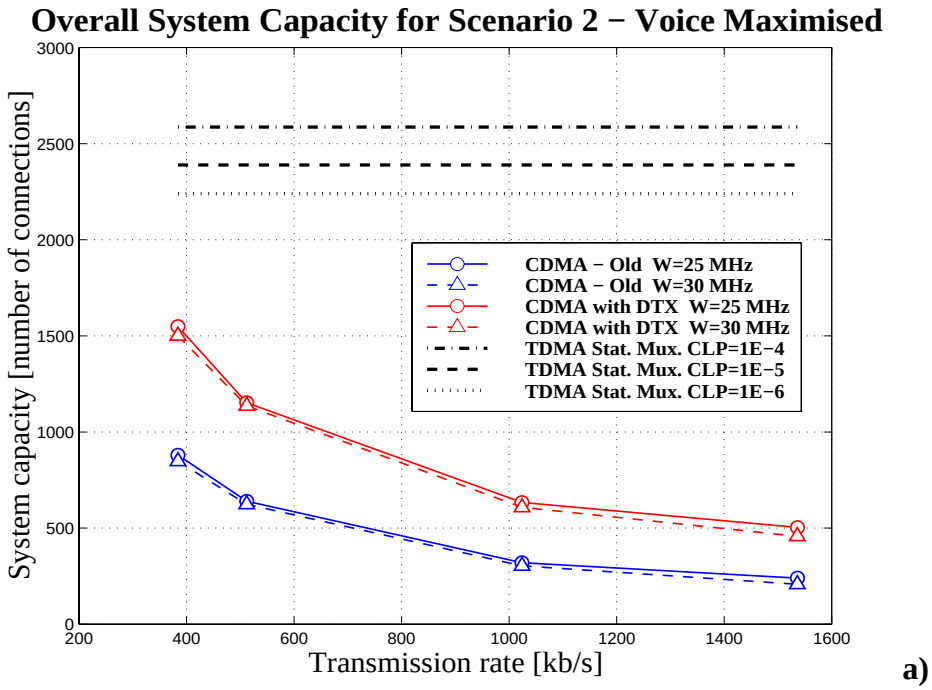


Figure 5.14 System capacity when unused capacity is allocated to only one traffic type: a) voice, b) video. Only values for channel bandwidths W_{SS} =25 and W_{SS}=30 MHz are shown., as all other channel bandwidths yield very similar results.

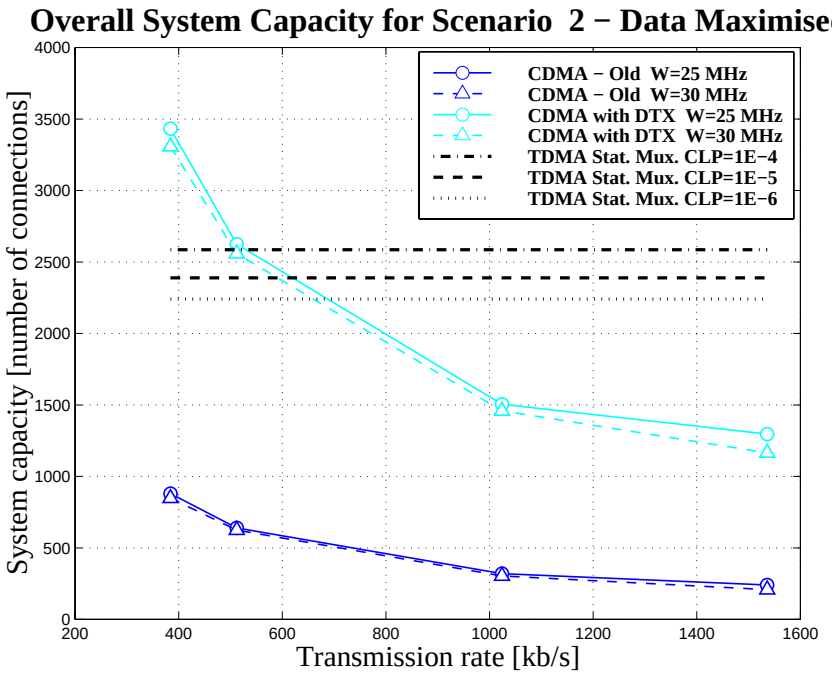


Figure 5.15 System capacity when unused capacity is allocated to: a) only data, b) equally to all traffic types. Only values for channel bandwidths W_{SS} =25 MHz and W_{SS}=30 MHz are shown., as all other channel bandwidths yield very similar results.

5.4.3.1 Relative capacity increase

Unused capacity, and consequently potential capacity increase is lower at higher transmission rates for the same loads and channel bandwidths (Figure 5.12-b and Figure 5.13). The mean unused capacity decreases with the transmission rate for a given channel bandwidth and load. This is a consequence of the lower processing gain *G* and

capacity limit C_{max}. For a fixed bandwidth, increase in transmission rate means decrease of G and hence, C_{max}. At the same time the achievable loads for higher transmission rates drop, as the *link* capacity, expressed as cell rate at which a satellite server can serve incoming traffic (C_{max} SR), increases relative to the aggregate mean arrival rate (equal to the sum of all mean rates λ, which is at most C_{max} λ_{max}, where λ_{max} is the highest mean arrival rate; as λ_{max} is constant and SR increases, the achievable load decreases). Thus to achieve even very low loads of ρ=0.1 (red curves on Figure 5.12-b and Figure 5.13) "very active" combinations of connections are required, which in turn reduces the amount of unused CDMA capacity. (E.g. a traffic combination in which video traffic is predominant, for R=1802 kb/s and W_{ss} = 25 MHz we have C_{max} = 8 and (U_1, U_2, U_3) = (1, 6, 1), yielding capacity usage of (1*0.35+6*0.75+1*0.09)/ C_{max} = 0.62 and hence unused CDMA capacity of about 0.38 or 38%). As expected, for a given channel bandwidth and transmission rate, the highest capacity increase results from the connections of the lowest value of activity factor (data traffic), and this increase drops with the traffic load and transmission rate (Figure 5.14 - Figure 5.17). This is because higher loads have less unused capacity available for further statistical multiplexing of the connections. The drop in the relative capacity increase with the transmission rate for each of the connection types follows from the results for the relative unused capacity, and the same arguments apply.

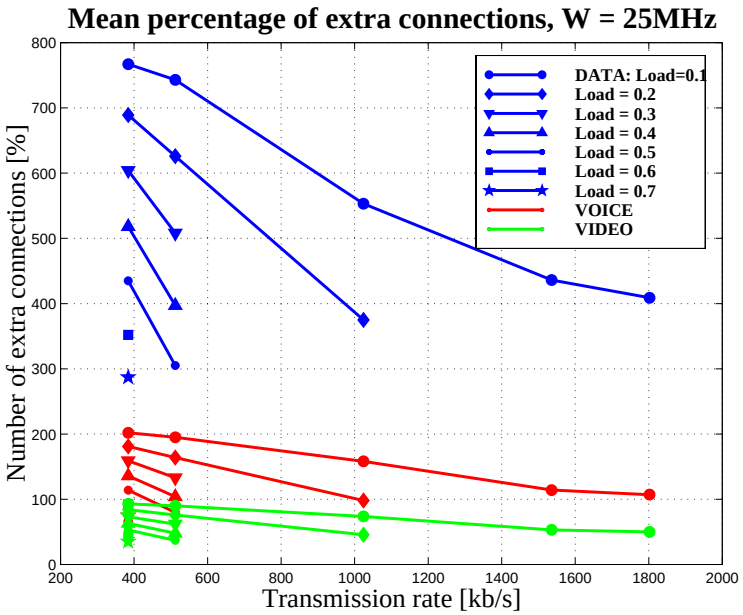


Figure 5.16 Percentage of the highest number of extra voice, video or data connections that could be accepted above the theoretical capacity C_{max} when channel bandwidth is W_{ss} = 25MHz.

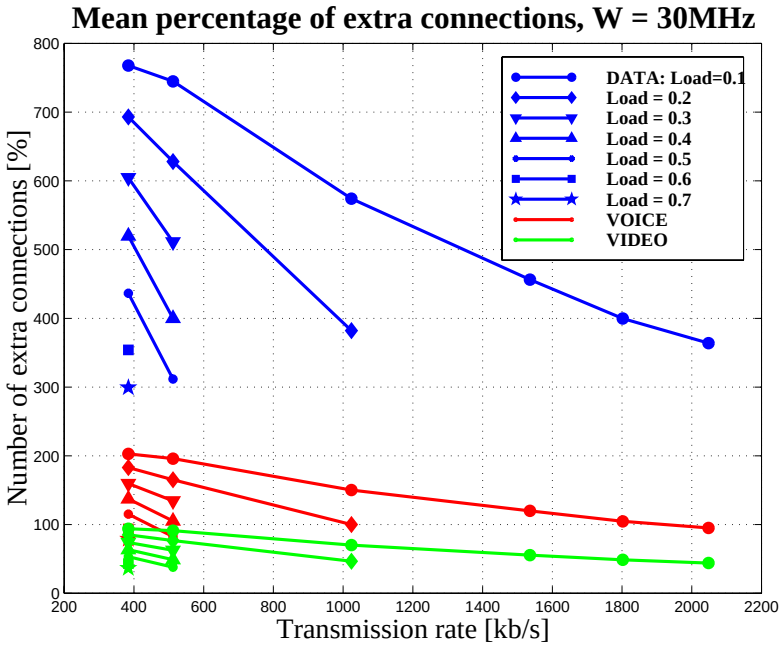


Figure 5.17 Percentage of the highest number of extra voice, video or data connections that could be accepted above the theoretical capacity C_{max} when channel bandwidth is W_{ss} = 30MHz.

5.4.4 Efficiency for Scenario 2

Efficiency for this scenario gives much higher values than Scenario 1 for all loads and transmission rates (up to 0.99 for load ρ=0.7 and 384 kb/s transmission rate, Figure 5.18). Moreover, the relative efficiency improvement is much higher than that in Scenario 1: values of 40% are obtained for 70% loads.

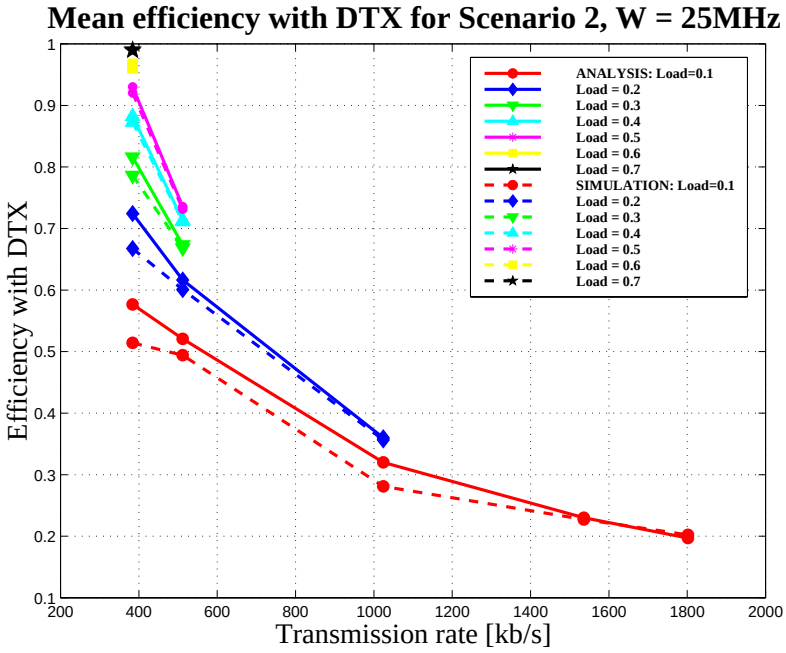


Figure 5.18 Mean efficiency when DTX is employed vs. transmission rate: comparison of analytical with simulated results for W_{ss} = 25 MHz. . All broken lines on graph are simulation results.

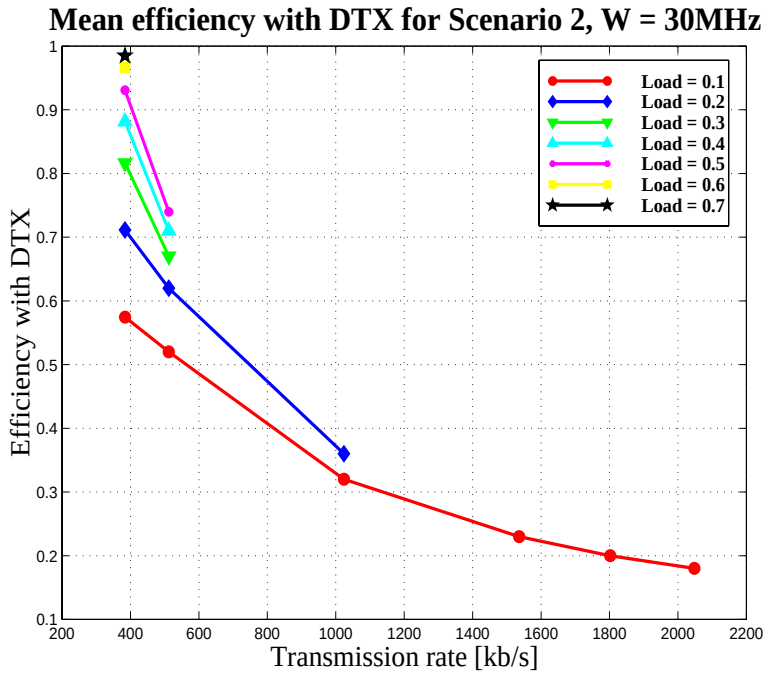


Figure 5.19 Mean efficiency when DTX is employed vs. transmission rate: analytical results for W_{ss}=30 MHz.

The efficiency decreases both with load and transmission rate, but there are no significant variations with channel bandwidth (Figure 5.18 and Figure 5.19). Like in Scenario 1, at higher transmission rates more empty cells are inserted between the cell

arrivals, thus efficiency decreases with transmission rate. The trend of decreasing efficiency with the load is a consequence of the effect of the connection activity factors: to decrease the load for a given transmission rate, the traffic distribution must favour the connections with lower activity factor. That is, the load decreases when the number of video connections decreases relative to the number of voice and data connections. A consequence of the time scales of mean ON and OFF periods of the different traffic types is that the mean active periods of voice and data traffic are shorter by one or two orders of magnitude than the mean active periods of video sources. Therefore over a fixed period, the number of ON/OFF changes of voice and data traffic is greater than that of video traffic, and thus the overhead is higher. At the same time, for a single duty cycle of a source, the overhead for video is lower than that for voice, since the PCR of video traffic is higher than the PCR of voice, hence the hangover period for a video connection is shorter and there are fewer empty cells during its ON state, which results in better efficiency. With data traffic, which has the highest PCR in the ON state, the overall effect is increased overhead, since the frequency of overhead insertion increases 24 times in comparison to video traffic (ratio of mean duty cycles is 26.5/1.1 = 24), while the PCR increases and therefore the hangover decreases only 1.3 times (=512/384), resulting in the effective increase of the number of empty cells of 24/1.3 = 18 times when a single video connection is replaced by a data connection. Therefore, as the connections with lower activity factor become predominant, there are fewer useful cells and more inserted overhead and empty cells, and the efficiency decreases.

From Figure 5.18 and Figure 5.19 it can be observed that the efficiency curves are steeper for higher loads, and their slope decreases with load. This again follows from the traffic characteristics of the sources. The highest loads are achieved for R=384 kb/s and predominantly video traffic, which is because video traffic has PCR=384 kb/s, so at R=384 kb/s almost no empty cells are Inserted (there are only overhead cells). When the transmission rate increases from R=384 kb/s to R=512 kb/s, for each useful cell sent during an ON period another 512/384 = 1.3 empty cells are inserted. In relative terms, this is a huge increase in the number of empty cells transmitted. For lower loads the predominant traffic types are voice and data. For voice, when the transmission rate moves from R=384 kb/s to R=512 kb/s instead of having 384/64 = 6 empty cells between cell arrivals there are 512/64=8 empty cells, and for data no empty cells are

inserted for the two transmission rates. In relative terms the increase in empty cells is lower, which is reflected in the slope of the respective curves. As the transmission rate increases for a fixed load, the relative increase in the number of empty cells from one point to the next decreases, so the curves flatten off.

Channel bandwidth has no influence on efficiency, as efficiency is a *relative* measure of the number of useful cells transmitted. The number and size of channels have no impact on how efficiently their bandwidth is used.

In summary: capacity gained through having connections of lower activity factors has its cost in comparatively lower efficiency. Overall, Scenario 2 (Code per VCC) outperforms Scenario 1 (Code per User) with respect to efficiency.

5.4.5 Buffer sizes and delay for Scenario 2

Standard Queueing theory formulae (discrete fluid-flow analysis, [Pitts & Schormans 96, pp.82]) were used to evaluate the required buffer sizes as transmission rates were varied to give loads from ρ=0.5 to ρ=0.9 (Figure 5.20-Figure 5.22). For the choice of transmission rates studied here (above 384 kb/s), the critical traffic type is data, with the highest PCR (512 kb/s). However, as the activity factor of data traffic source is very low, the mean cell arrival rate is low, and with the mean burst size of 120 cells (T_{ON}*PCR), the buffer sizes needed for good cell loss performance are small (700 cells for CLP~10^{-11}, see Figure 5.22; buffers for bursty data traffic have often been shown to require more than 10000 cells [Winstanley 96]). For video traffic, due to the very high activity factor, any slot rate lower than the PCR produces very high cell loss, which is to be expected since video traffic is almost CBR (and for CBR traffic PCR service rate is required). In principle, the relative simplicity of the architecture for Scenario 2 makes it easier to dimension the buffers, as each buffer is fed by only one source. As transmission rates studied start from 384 kb/s, which is the PCR of video traffic, and which for data traffic produces a black line of Buffer size vs. CLP on Figure 5.22, the buffer sizes required in this scenario are smaller than that in Scenario 1, and of the order less than 1000 cells.

At rates higher than 512 kb/s this extra cell inserted before the first cell of the burst does not at all affect other cells in the burst. For R=384 kb/s, the only traffic that might experience some buffering delay is data. However due to the low data mean cell arrival

rate, load at R=384 kb/s is ρ=0.12, so queueing delay can be assumed insignificant. Delay distributions for three traffic sources are analogous to their buffer state distributions, i.e. Pr[delay=d]=Pr[state=d*SR].

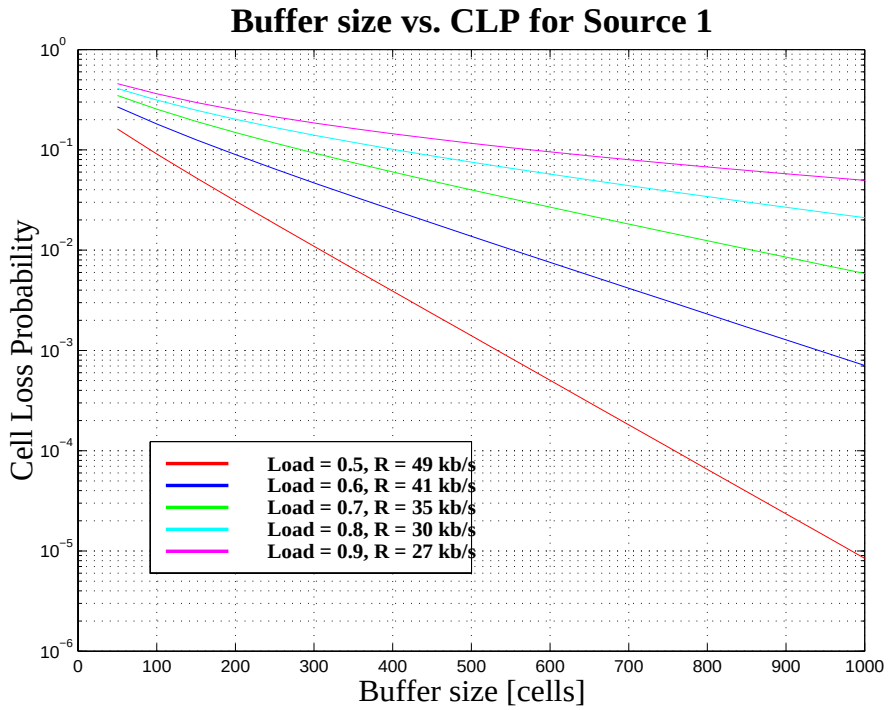


Figure 5.20 Cell loss probability (CLP) versus buffer size, for voice traffic.

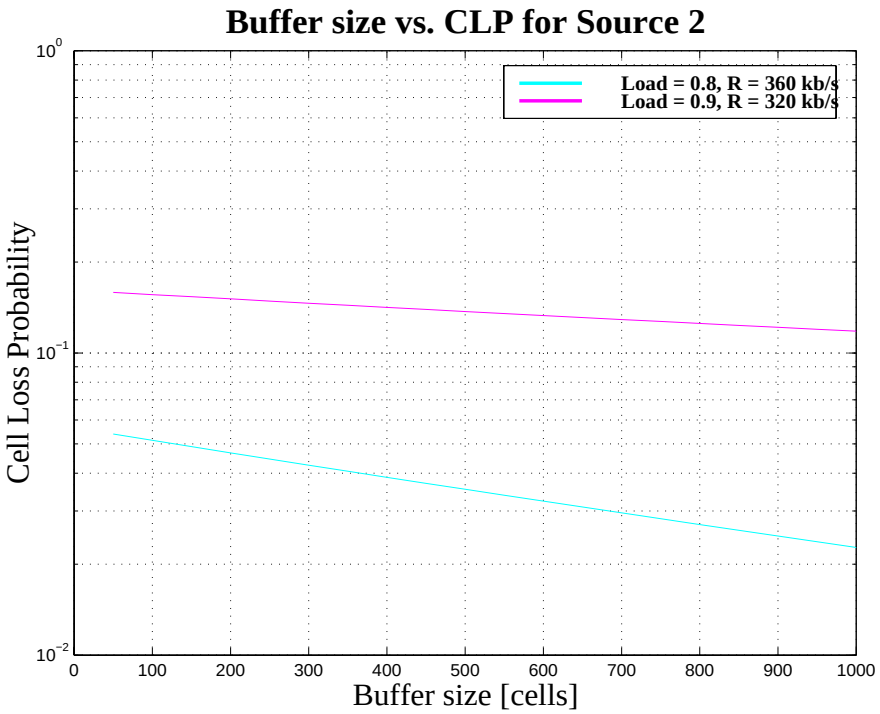


Figure 5.21 Cell loss probability (CLP) versus buffer size, for video traffic.

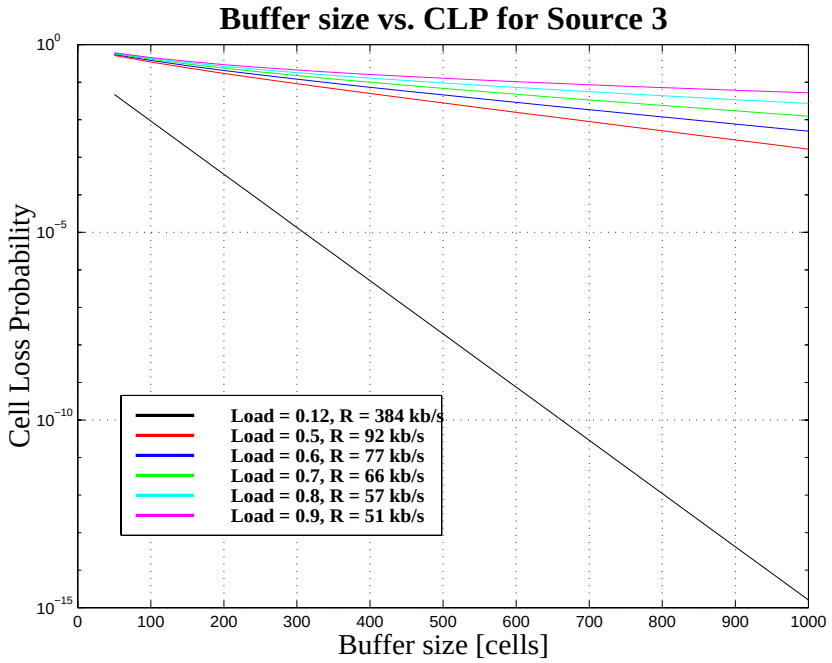


Figure 5.22 Cell loss probability (CLP) versus buffer size, for data traffic. The black line is the plot for R=384 kb/s, the only critical transmission rate of those examined for data traffic.

5.5 Scenario 3

In this scenario each connection uses a different channel dedicated only to one particular type of traffic. Transmission rates are ideally chosen to be equal (in cells/s) to the peak cell rates of respective traffic types. That means that there will be no buffering delay apart from that introduced by code re-synchronisation, that buffer sizes required are of the order of a few cells, and that empty cells arise only from the overheads (hangover and code re-synchronisation), and not from the difference between the cell arrival and cell service rates.

It was mentioned in the introduction of this chapter that it is important to compare like with like. When it comes to comparing Scenario 3 with Scenarios 1 and 2, direct comparison was not possible due to the nature of Scenario 3: there is no unique transmission rate to measure the performance against. Instead, the performance measures are plotted against:

a) processing gain, with load as a parameter, and

a) load, with processing gain as a parameter.

Figure 5.23 provides an interface graph that allows a comparison of Scenario 3 with scenarios 1 and 2. The graph shows a plot of processing gain *G* versus transmission rate

R was, with the channel bandwidth W_{ss} as a parameter. Each point $Y(G)$ on the Scenario 3 plots with processing gain as an x-axis (Figure 5.24-a, Figure 5.26-a, Figure 5.27-a and Figure 5.25-a) corresponds to the point in scenarios 1 and 2 graphs determined by the channel bandwidth W_{ss} and transmission rate R that produce the corresponding processing gain $G=W_{ss}/R$.

Before numerical results for Scenario 3 were calculated, the system was first optimised with respect to the overall system capacity. For the sake of fair comparison with the other two scenarios, the processing gain was kept constant for all channel types (as this was the case for a single system instance in Scenario 1 and 2). This means that connection signals were equally spread to their respective channel bandwidths W_{ss}. Since QoS (i.e. SNR and BER) were also assumed equal for all three traffic types, it meant that the capacity per channel was the same for all channel types (Eq. 1.4 in Chapter 1). At the same time, due to $G_i=G=const.$ the channel bandwidths W_{ss} were proportional to the sources' peak cell rates. Thus the channel type carrying the connections with highest PCR (data) have the highest channel bandwidth, and the voice channels have the lowest bandwidth. This affected the results of the optimisation: it favoured those channels with the lowest bandwidth, as all channels contribute to the total capacity by equal amounts (capacity per channel is constant). In other words, the highest system capacity required a system channel configuration with the maximum number of voice channels and minimum number of data and video channels (e.g. $[m_1\ m_2\ m_3] = [M, 1, 1]$). It is important to note that for different QoS requirements for each of the traffic types, and varying processing gains, optimisation produces system channel configurations that are not on the border of the optimisation search state space.

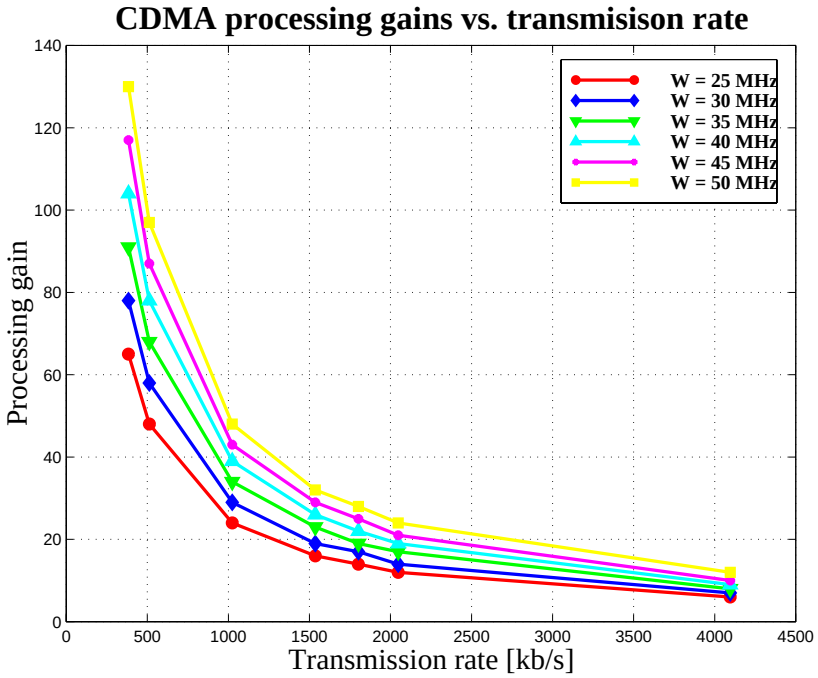


Figure 5.23 Plot of processing gains *G* versus transmission rate *R* for various channel bandwidths *WSS*.

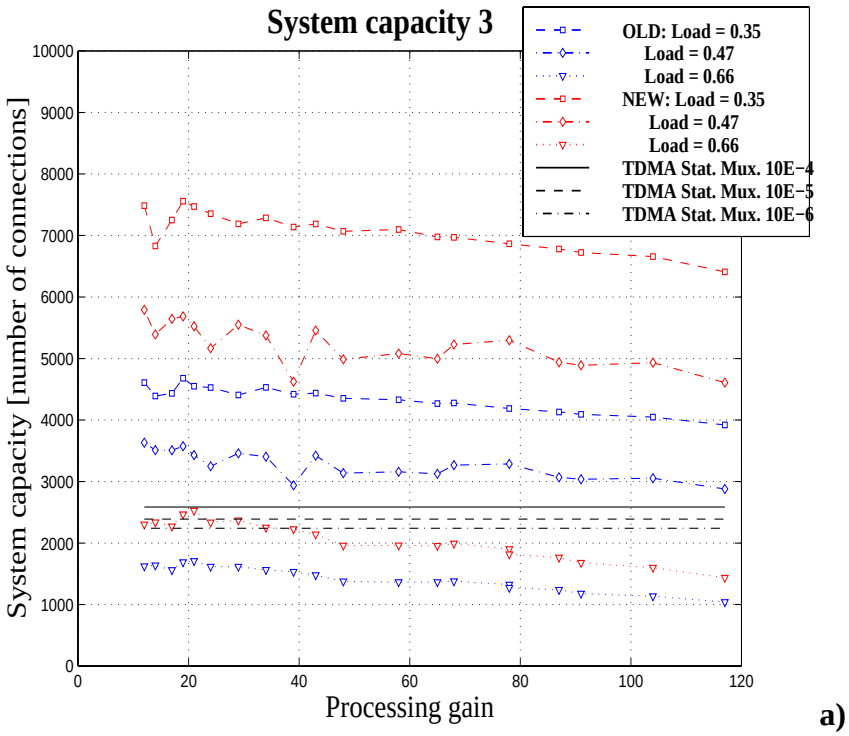
It should be noted that in Scenario 3 the load is varied by changing the channel configuration [m_1 m_2 m_3]: for different channel configurations different loads can be achieved, depending which traffic type (i.e. channel type) is favoured (Figure 5.24-b, Figure 5.26-b, Figure 5.27-b and Figure 5.25-b). That is, for the same *G*, a configuration that supports a higher number of video connections offers a higher load to the satellite server (while yielding the lower overall system capacity).

No simulation was performed for this scenario, as the main difference from the Scenario 2 is physical separation of channels (on the physical layer), which cannot be modelled on a cell level simulator. Scenario 3 can be seen as three Scenario 2 systems put together, each with a different allocated system bandwidth (corresponding to the total satellite bandwidth W_{sat} in Scenario 2). Each of these systems has a fixed transmission rate and channel bandwidth (different from the other two), and each carries only one type of traffic instead of three, as in Scenario 2. Thus from the agreement of simulated and analytic results shown for Scenario 2, follows the validity of Scenario 3 analysis presented in Chapter 3. Thus only analytic results for Scenario 3 are presented in this thesis.

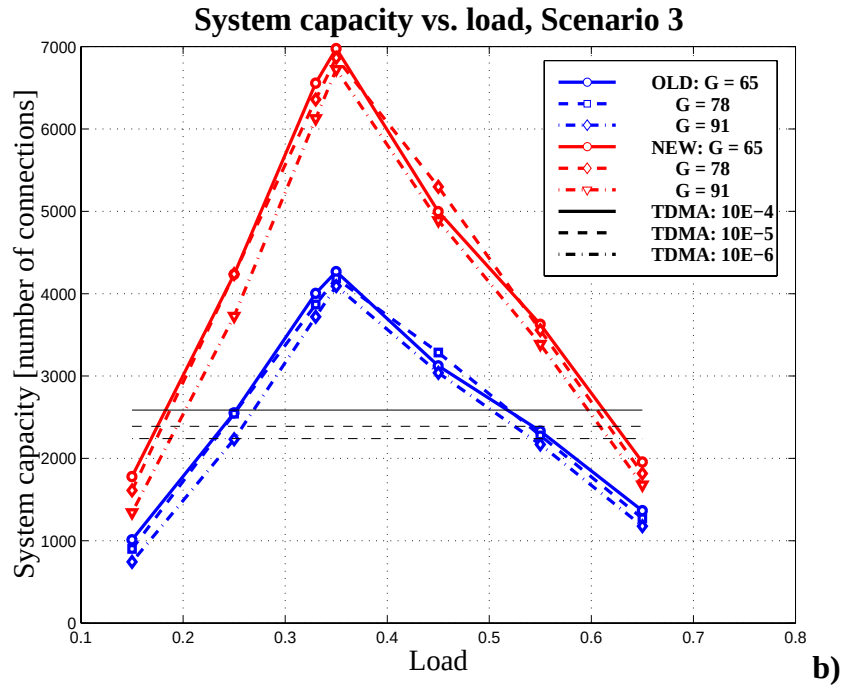
5.5.1 Capacity for Scenario 3

Total capacity

The maximum achievable system capacity is much higher than in Scenario 1 or 2, and is obtained for all but two channels being voice, i.e. a system load of 0.35. For higher and lower loads the total achievable system capacity becomes comparable to those in Scenario 2. This is clearly demonstrated by the triangular peaky shape of Figure 5.24-b. If this figure is compared to Figure 5.14-a where voice connections are maximised, it is clear that the maximum achievable capacity with Scenario 2 is comparable to the minimum possible capacity of Scenario 3 (load of 0.1) - around 1500 connections.



a)



b)

Figure 5.24 Total system capacity without DTX (blue) and with DTX (red): a) vs. processing gain and b) vs. load. Black line represents capacity arising from effective bandwidth allocation (CLP=10^{-4}) in statistical multiplexing TDMA

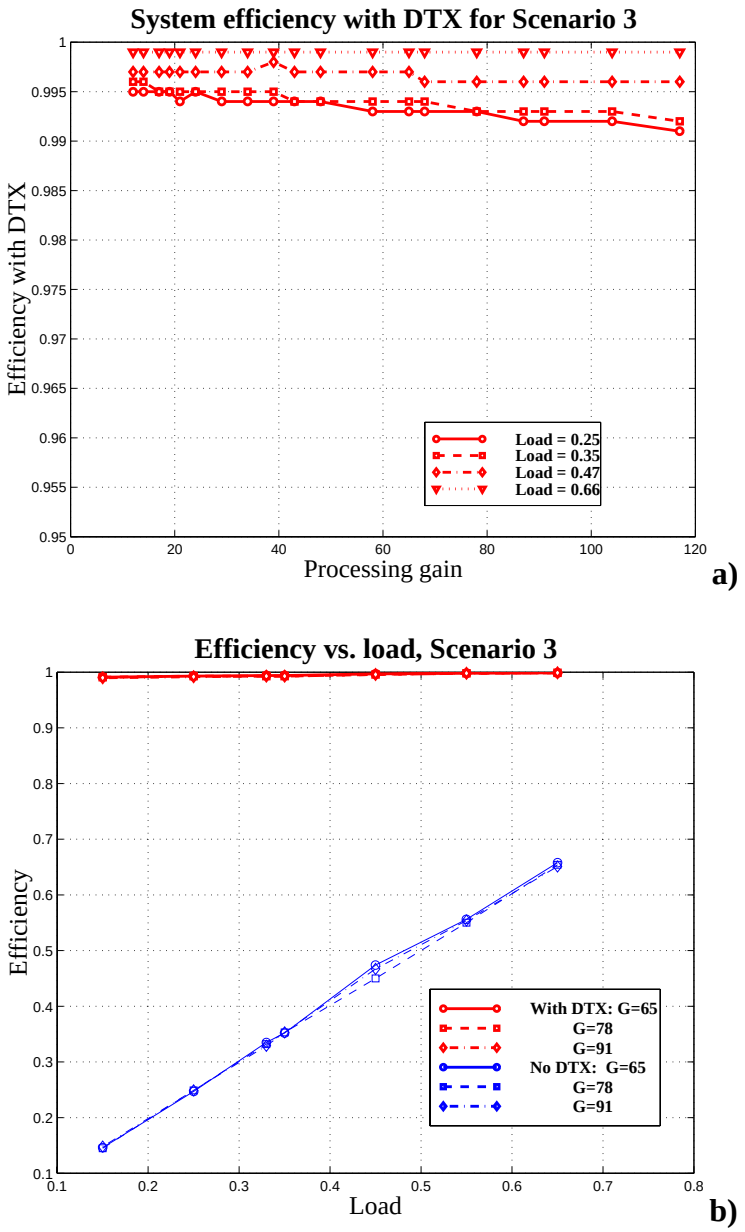


Figure 5.25 Efficiency achievable with DTX vs. a) processing gain and b) load.

When comparing to Scenario 1 it should be remembered that in Scenario 1, one user represents 3 connections. Thus, with this particular arrangement of traffic, Scenario 1 exhibits slightly better performance than Scenario 2, and the values of 3*750 connections are in the lower range of Scenario 3. Figure 5.24-a shows that the capacity does not depend on processing gain, which is expected: as the processing gain increases, the number of channels decreases as the channel bandwidth increases. Had the system

configuration in Scenario 3 been chosen to have equal number of channels for all traffic types, it would still outperform Scenarios 1 and 2, as in such a configuration the total number of connections with DTX would amount to just over 5000 (for processing gain G=65 and 8 channels of each type, with bandwidths W_{voice}=65*64 kb/s, W_{video} = 65*384 kb/s and W_{data}=65*512 kb/s).

The ripples on Figure 5.24-a are due to bandwidth quantisation.

Capacities achievable by idealised TDMA statistical multiplexing described in section *Comparison with TDMA* are shown for comparison on Figure 5.24. TDMA statistical multiplexing capacity (based on effective bandwidth) is an idealised measure of TDMA capacity, and therefore any real TDMA system would yield a lower capacity, so it can be seen from Figure 5.24 that Scenario 3 outperforms TDMA when the CDMA system is optimised (in the studied case, when the voice channels are maximised).

5.5.1.1 Relative capacity increase

Figure 5.26-a shows that relative capacity increase does not depend on the processing gain. It decreases with the increase of load, as expected. The relative capacity increase within a channel depends entirely on the activity factor of the traffic type the channel is carrying and thus within a channel the relative capacity increase is constant. Small variations with the processing gain are a result of total bandwidth channel partitioning into unequal bandwidth sizes. While the relative capacity increase is constant within a channel, varying the load is caused by altering the channel distribution [m_1 m_2 m_3] and so individual relative capacity contributions of different traffic types vary. For low loads the predominant channels are those of the data traffic type, with lowest activity factor and therefore highest relative capacity increase. As the number of voice channels increases and number of data channels decreases, the load is increased, but so is the overall activity (proportion of ‘ON’ periods) too, so the relative capacity increase drops. The ripples on the curves of Figure 5.26-a are due to bandwidth quantisation: for some values of processing gain, the amount of bandwidth that has not been taken into account in the analysis due to rounding causes the capacity to drop below the trend line.

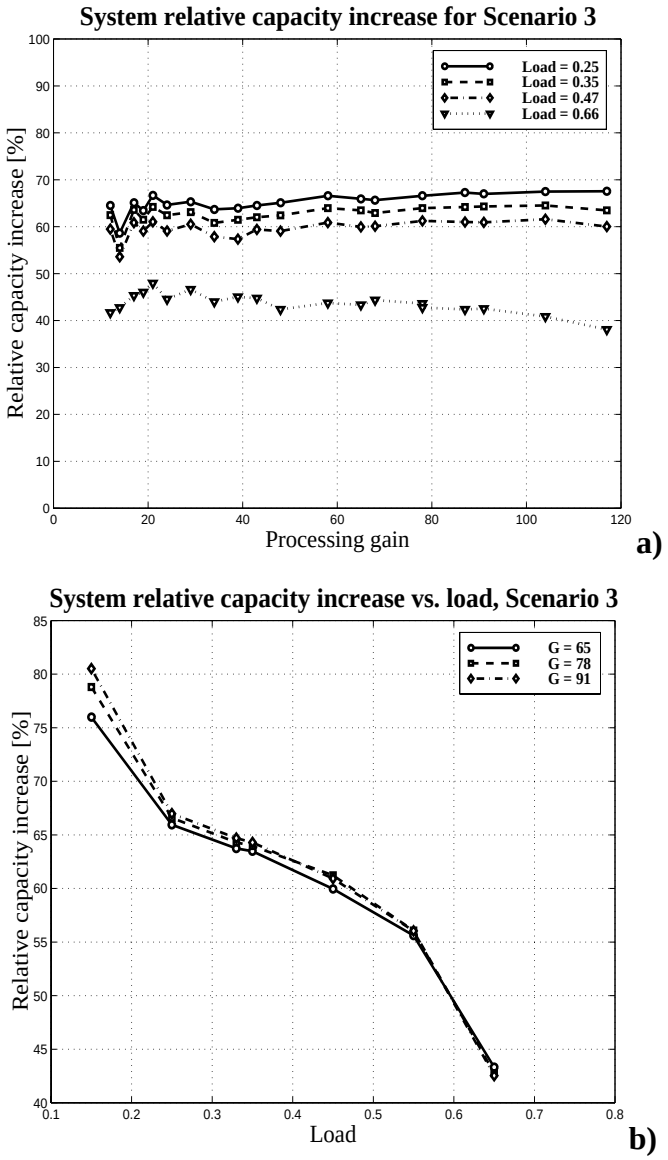


Figure 5.26 Relative capacity increase vs. a) processing gain and b) load.

5.5.2 Efficiency for Scenario 3

Scenario 3 can achieve the highest efficiency of all three scenarios. Efficiency does not change with the processing gain, since efficiency itself is a relative measure of the number of useful cells transmitted. Change of the processing gain changes the number of channels, but not the proportions within a channel. Efficiency is more-or-less constant with load, the curves on Figure 5.25-a for different loads are shifted by at most 0.5%, which cannot be seen on the coarser scale of Figure 5.25-b. This slight increase in efficiency with load is a consequence of the traffic characteristics of the majority traffic in each different configuration. As was explained earlier, higher loads are produced

when the majority traffic is video, with longest duty cycles. To decrease the load, the amount of video traffic must be decreased and the amount of voice and data traffic increased. That means more overhead, due to the shorter duty cycles and more frequent ON/OFF state changes.

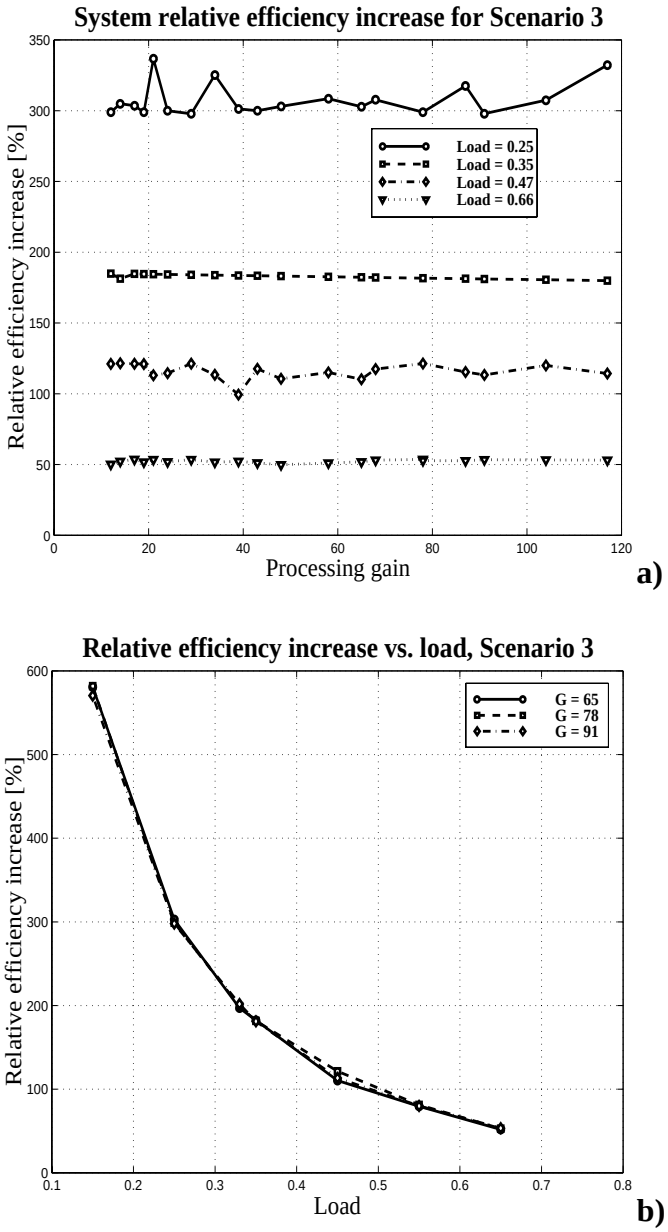


Figure 5.27 Relative efficiency increase obtained with DTX vs. a) processing gain and b) load.

Relative efficiency increase drops exponentially with load on Figure 5.27-b, and the same trend can be observed from Figure 5.25-b, where efficiency without DTX (i.e. load) was plotted (blue line). The ripples on the parallel lines of relative efficiency increase vs. processing gain are again computational “noise”.

5.5.3 Buffer sizes and delay for Scenario 3

As all traffic is being transmitted at their peak cell rate, minimal buffering is required (only a few cells). Virtually zero queueing delay is being introduced, and the code re-synchronisation overhead is of the order of one cell (or less); thus all connections are being delayed by about 1 cell slot, which is insignificant in this context.

6. DISCUSSION

The motivation for this thesis was the need to devise a medium access scheme based on CDMA for broadband communications via satellite that would allow transmission of heterogeneous traffic while using the satellite bandwidth efficiently. The concept of statistical multiplexing present in fixed broadband communication networks was brought into the context of CDMA-based satellite access networks. The aim of the research was to establish how this statistical multiplexing in the code domain could be used to improve multirate (hence, broadband) CDMA system performance on the network (capacity) and ATM layers (efficiency), to evaluate the relative performance improvements, and to establish if there is a potential in its use as a medium access scheme. To achieve this, two main objectives were set:

1. to study the *performance* of three different multirate CDMA system configurations that employ statistical multiplexing in the code domain, in order to:
1. understand how the process of statistical multiplexing in the code domain is affected by a traffic mix and the choice of two system design parameters (transmission rate and channel bandwidth), and
1. identify the best system configuration;
1. to define and evaluate a *methodology* which would allow quick and easy multirate CDMA system analysis and dimensioning when a form of statistical multiplexing in the code domain is used. This methodology could be used with any similar multirate CDMA system to assess the system performance and help choose the optimal design parameters for the given choice of traffic types.

Definition of the basic system configurations was necessary in order to assess quantitatively how the system performance (capacity, efficiency) depended on traffic characteristics and how it changes with both traffic distribution and variation of the system design parameters. The study was not concerned with the issues of source modelling; the main issue was the relative difference in the characteristics of the three chosen traffic sources, particularly in the parameters that affect medium access control (granularity of mean ON and OFF periods, mean and peak bit or cell rates). It is this relative difference in the traffic characteristics of the sources that dictates the trends in the system performance according to the traffic mix. The focus of the study was to

understand the relations between the changing traffic mix on one side, and the system performance on the other. This understanding allows more general conclusions about the *trends* in the system performance depending on traffic characteristics, transmission rates and channel bandwidths, that can be applied to any combination of traffic sources in the same or similar system configurations.

The study proposed a methodology that allows accurate numerical assessment (and therefore prediction) of the performance depending on many system variables, and can be used to eliminate the need for heavy simulations. The methodology comprises of applying mathematical analysis described in Chapter 3 to a range of values of the system design parameters that need to be optimised: channel rates and channel bandwidths, given the characteristics of the expected traffic and their QoS requirements in terms of CLR i.e. BER. Optimisation is based on capacity in terms of the number of connections or the number of users. By applying numerical analysis, optimal system parameters can be found for any of the described configurations. To establish the accuracy of the numerical method, it was necessary to run simulations or carry out alternative mathematical analysis and compare the results obtained by different methods.

Comparison with TDMA was by referring both presented CDMA medium access scheme and other potential medium access schemes in TDMA to the results of an idealised statistical multiplexing scheme. The capacity of this idealised measure can be used as a reference point to which the capacity of any other broadband medium access scheme could be compared. By establishing the relative performance difference with respect to the idealised statistical multiplexing based on effective bandwidth, other medium access schemes can now be compared to the statistical multiplexing in the code domain presented in this thesis.

6.1 Validity of Methodology

The methodology consisted of:

- definition of traffic types to be accommodated by the system (traffic characteristics and QoS requirements);
- definition of the values of interest for the physical layer system parameters (channel bandwidths, transmission rates, total available bandwidth) in the given system configurations;

- evaluation of required overheads (hangover period, code acquisition time);
- evaluation of the performance at the cell level as system parameters were varied, to establish how the system performance depends on the traffic carried;
- identification of the optimum set of physical layer system parameters for the best ATM and network layer performance;
- optimisation of the number of each type of system channel for the maximum capacity in Scenario 3, using simulated annealing or a genetic algorithm.

The methodology was verified by:

- cell-level simulation (Scenario 1), where no appropriate queueing results were available for validation.
- hybrid simulation/analysis method (Scenario 2), where cell-level simulation was used as an input to further analysis, since the full-scale cell-level simulation was not appropriate.

Scenario 3 is a refinement of Scenario 2 and as such did not need additional verification by simulation or analysis further to what was done for Scenario 2, as the validity of Scenario 2 implies the validity of Scenario 3.

Both simulation and the hybrid simulation/analytical method showed good agreement with the results obtained by the numerical analysis at the cell level. It was demonstrated that the methodology accurately predicted the system behaviour under arbitrary traffic mix for loads lower than 0.75, and that it can be used to analyse how different traffic types and loads affect overall system performance. Since a real system never operates at loads higher than 0.75, the analysis is applicable to real systems.

The methodology can therefore be used to study traffic mixes of larger number of traffic types in similar system configurations. While this study assumed equal BER requirements for the CDMA channel, and equal transmitted signal powers, this assumption did not affect either the generality of the results or validity of the methodology.

6.2 Further work

The analysis presented in this thesis did not take into account the queueing effects at high loads in Scenario 1. A further problem for queueing theorists could be to generalise the results for capacity by taking into account the queueing effects when the multiplexer load tends to 1.

A real system would require an optimisation that takes into account the relative load distribution among different traffic types. That is, the optimisation criterion would be capacity, but with the constraint that a particular traffic distribution is supported by the optimal partitioning of system bandwidth into different channels. With the constraints imposing how the total traffic load is distributed among different traffic types, one of the parameters that would need to be optimised is processing gain, that is, channel bandwidth itself (since the optimal transmission rate for a given channel i.e. traffic type, from the point of view of efficiency and delay is the peak cell rate of the carried traffic). In such a case a variation of the simulated annealing (or genetic) algorithm with more complex constraints and objective function is required. In this thesis, the channel bandwidths and processing gains were given as constants, i.e. input parameters to the optimisation. Modifying the optimisation algorithm to accommodate a greater number of variables is not difficult, and the only implication will be longer algorithm search times due to a more complex and larger state space. The complexity of the state space would result from the non-linear constraints that relate capacity and processing gain of channels (Eq. 1.4). Simulated annealing is an algorithm capable of dealing with non-linear constraints.

The research reported in this thesis is the first step towards the study of link and network layer protocols in a multirate CDMA-based system. Future work can address the statistical multiplexing based CDMA scheme for non-real time data.

The data link layer provides mechanisms for error detection and correction: automatic repeat request (ARQ) and forward error correction coding (FECC). On a satellite link, where propagation delays are long, ARQ mechanisms are characterised by low throughput. An adequate high-performance ARQ error control protocol is required for broadband satellite access network. FECC allows higher throughput at the expense of higher bandwidth requirement and lower bandwidth utilisation. In ATM, FECC is provided within both the cell header (ATM layer) and the cell payload (ATM adaptation

layers, AAL), although this is not sufficient for a satellite (wireless) channel. More recently, [Cain & McGregor 97] have shown that suitable error correction mechanisms for wireless ATM exist. Additional coding can significantly improve the BER and consequently CLR performance on satellite links. [Cain & McGregor 97] have addressed coding for random errors, coding for avoiding burst errors on the satellite link, as well as techniques for randomisation of burst errors introduced by convolutional decoders that are used to overcome burst errors naturally present in satellite links (e.g. cell interleaving). The bandwidth required for error correction coding can be accommodated by the analysis presented in this thesis through the consideration of the mean duration of sources' ON and OFF periods.

The work reported herein did not investigate the effect of the scheme on the performance of ARQ protocols, i.e. how the variable BER due to the statistical multiplexing in the code domain may affect the required number of re-transmissions and therefore, the performance of the ARQ protocols. In the case of an access network employing an ARQ error control mechanism, the packet throughput and end-to-end delay would need to be studied within a context of statistical multiplexing in CDMA.

On the network layer, the study has provided the initial steps towards the development of a CAC suitable for a multirate CDMA network access (see Chapter 3, Scenario 2). For Scenario 3 the CAC mechanism would be even simpler as channels for different traffic types are separated, although it would be governed by the same principle of exploiting the traffic activity factors. CAC study would need to take into account a greater range of traffic source models, e.g. Poisson sources, multi-state generally modulated Markov or deterministic processes, etc.

Two assumptions were made regarding the power of the sources: first, that perfect power control was used, and second, that the transmissions are switched off completely during source silent periods. The study could be extended by investigation of the cases when non-ideal power control is used, and when the transmission continues during sources' silent periods but at a lower transmitted power level.

In the system configuration of Scenario 3 it was assumed that the system was first optimised off-line. However, future systems will require flexibility to quickly adapt to changing traffic demands. Thus a form of on-line system re-dimensioning in Scenario 3 would be desirable. This in turn could imply the changes on the physical layer, and

would therefore be dictated by the advancement of the technology that enables hardware flexibility. Provided hardware flexibility is possible, faster optimisation algorithms for system on-line dimensioning should be developed.

7. CONCLUSIONS

The research has been carried out that assesses the system capacity and efficiency performance achievable through the use of a medium access scheme that employs statistical multiplexing of ATM traffic in the code domain of DS-CDMA [Timotijevic & Schormans 96, Timotijevic & Schormans 97, Timotijevic & Schormans 98a, Timotijevic & Schormans 98b]. Varying mixes of three different traffic types have been considered in the assessment of the impact of the traffic characteristics on system performance. Methodology for numerical assessment of system performance and system parameter optimisation has been presented. Understanding how varying traffic mixes influence system capacity is the first step towards the development of traffic control functions for an ATM satellite system employing DS-CDMA. In that context, a feasible CAC algorithm has been proposed that could be used in an ATM DS-CDMA system using statistical multiplexing in the code domain described in this thesis. The conclusions drawn from the results can be generalised for a greater number of ON-OFF traffic types.

7.1 Comparison of system scenarios

Three system configurations or scenarios were investigated:

1. Code per User scenario with *R=const*. (Scenario 1), i.e. with equal channels and uniform transmission rate across the system. All connections are first multiplexed prior to spectrum spreading by a PN code.
1. Code per VCC scenario *R=const.* (Scenario 2), also with equal channels and uniform transmission rate across the system. However the connections are not multiplexed before spectrum spreading, instead each connection has its own CDMA PN code.
1. Code per VCC scenario *R<>const.* (Scenario 3), with unequal channels, different number of channels of each type, and different transmission rate for each channel.

Overall Scenario 3 showed best performance in all criteria observed (capacity, efficiency, buffer requirements), while Scenario 2 has comparable performance to that of Scenario 1 in terms of capacity, but better in terms of efficiency and buffer requirements.

7.1.1 Capacity

The relative capacity increase in Scenarios 2 and 3 can be measured only in terms of one particular connection type. The study takes into account the values of relative capacity increase in terms of each connection type. For the same loads, transmission rates and channel bandwidths, Scenario 2 is comparable to Scenario 1 in terms of the number of connections. However, the relative capacity increase that can be achieved by employing DTX is much higher in Scenario 2 (between 100% and 800%, Figure 5.15 in Chapter 5) than in Scenario 1 (15%, Figure 5.3). Overall, Scenario 3 performs best: for equal distribution of traffic types highest capacity can be achieved (around 5000 as opposed to around 2000 as in Scenario 1 and 2). An advantage of Scenario 2 over Scenario 1 is the fact that the distribution of traffic can be more easily controlled, as codes are allocated to connections and not multiplexed streams (users or terminals). In Scenario 1 the total capacity may not be well utilised as not all three connections may be active at the same time i.e. during one session, which was the assumption underlying the analysis. Thus better utilisation of resources is possible with Scenario 2 than Scenario 1.

7.1.2 Efficiency

For optimally dimensioned systems, Scenario 3 outperforms Scenario 2, which is better than Scenario 1, for.

Overall, given the comparison of capacity performance, it can be said that Scenario 3 is the best although it is also the most complex to dimension. It requires optimisation for maximum capacity, which depends on the traffic that is to be accommodated by the network and its QoS requirements. Scenario 2 is better than Scenario 1, and offers simpler buffer dimensioning (and smaller buffers). In general, existence of traffic sources with longer of ON and OFF periods increases system efficiency, while traffic sources whose ON and OFF periods are shorter cause lower efficiency, due to more frequent overhead insertion. Thus for a different traffic mix than the one chosen for this thesis, the methodology can be used to determine and compare the relative difference in efficiency performance of Scenarios 1 and 2.

The comparison of efficiency performance between Scenario 1 and 2 is not always straightforward, and will depend on the traffic types present in the system. The choice of transmission rate in Scenario 2 will be dictated by the existence and characteristics of

the most CBR-like traffic type: if this traffic type has a PCR higher than other traffic types (as well as high activity factor), transmission rate in Scenario 2 will have to be high enough to ensure good cell loss performance of this traffic type, causing channels carrying other traffic types to experience low efficiency due to insertion of empty cells. For the same traffic mix, the number of inserted empty cells in Scenario 1 would be relatively lower, as the traffic streams would be multiplexed together (assuming that the chosen transmission rate for Scenario 1 is between the aggregate peak cell rate and aggregate mean cell rate). Thus for a case when the most CBR-like traffic type has higher PCR than other traffic types, Scenario 1 seems like a better system configuration.

Scenario 1 requires simulation to determine the required buffer dimensions, particularly if the system is expected to operate at higher loads, since no mathematical models exist to predict the queueing behaviour of a mix of limited number of heterogeneous ON-OFF sources.

7.2 System parameters

For Scenario 1, an optimal transmission rate is the one that allows loads close to but below 70%, as that operating point yields good delay, capacity and efficiency performance.

For Scenario 2, a transmission rate as low as possible, but close to the PCR of the traffic type most similar to CBR, is optimal. This yields good efficiency and capacity, while giving satisfactory buffer CLP due to buffering for the most CBR-like traffic type.

For Scenario 3, transmission rates equal to the sources' PCRs were chosen, and when optimised Scenario 3 outperformed Scenarios 1 and 2. The choice of PCRs for channel transmission rates was made in order to minimise queueing delays in the buffers.

REFERENCES AND BIBLIOGRAPHY

[Aart & Korst 89] E. Aarts and J. Korst: *Simulated Annealing and Boltzmann Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computing*. John Wiley & Sons Ltd., 1989

[Abramson 96] Norman Abramson: *Wireless Random Access for the Last Mile*. ALOHA Networks, Inc., invited paper, Conference on Multiaccess, Mobility and Teletraffic for Personal Communications, Paris 1996

[Andermo & Brismark 94] P.G Andermo,. and G. Brismark: *CODIT, a testbed project evaluating DS-CDMA for UMTS/FPLMTS*. Proc. of the IEEE 44th Vehicular Technology Conference "Creating tomorrow's mobile systems", 8-10 June 1994, Stockholm, Sweden, Ch. 389, pp. 21-25

[Andermo 95] P.G.Andermo: *Code Division Testbed project – CODIT*. Mobile and Personal Communications, 1995, pp. 23-29

[AT&T 95] *Application of AT&T Corporation to Federal Communications Commission for authority to construct, launch and operate the VoiceSpan System of twelve satellites in the Ka-band domestic and international fixed-satellite service and to construct four partial Ka-band satellite group spare*. Before the FCC, Washington, D.C. 20554, File numbers 156-SAT-P/LA-95 through 162-SAT-P/LA-95; September 29, 1995

[Akyldiz & Jeong 97] I.F. Akyildiz and S.H. Jeong: *Satellite ATM networks: A survey*. IEEE Communications Magazine, July 1997, Vol.35, No.7, pp.30-43

[Baier 93] A. Baier: *Multi-rate DS-CDMA: A promising access technique for third generation mobile radio systems*. Proc. of PIMRC '93, Yokohama, Japan, September 1993, pp.114-118

[Baier & Panzer 93] A.Baier and H.Panzer: *Multi-rate DS-CDMA radio interference for third-generation cellular systems*. Proc. of the 7th IEE European Conference on Mobile and Personal Communications, 13-15 December 1993, Brighton, England, Ch. 46, Conf. Pub. No. 387, pp.255-259

[Baiocchi 91] A. Baiocchi, M. Carosi, M. Listanti and A. Roveri: *Modeling of a distributed access protocol for an ATM satellite system: an algorithmic approach*. IEEE JSAC, January 1991, Vol. 9, No. 1, pp. 65-75

[BAF2 92] BAF Deliverable 2: *Specification of Access Control Functions*. RACE 2024, December 1992, Reference: R2024_QMW_W21_DS_P_002_b1

[Bella 96] Private communication with Luigi Bella, Head of Transmission Systems Section in European Space Agency Technical Centre (ESTEC), November 1996

[Bottcher 92] A.Bottcher: *Modelling and Performance of Packet Synchronisation in a Data Communications System*. International Journal on Satellite Communications, 1992, Vol. 10, pp. 343-347

[Brand & Aghvami 96] A.E. Brand and H.Aghvami: *Performance of a Joint CDMA/PRMA Protocol for Mixed Voice/Data Transmission for Third Generation Mobile Communication*, IEEE JSAC, December 1996, Vol. 14, No. 9, pp.1698-1707

[Bruneel & Kim 93] Herwig Bruneel and Byung G. Kim: *Discrete-Time Models for Communication Systems Including ATM*, Kluwer Academic Publishers, Dordrecht, 1993

[Cain & McGregor 97] J.B. Cain and D.N. McGregor: *A recommended error control architecture for ATM networks with wireless links*. IEEE JSAC, January 1997, Vol. 15, No. 1, pp.16-28

[Chitre *et al*. 94] D. M. Chitre, D. S. Gokhale, J. L. Lunsford and N. Mathews: *Asynchronous Transfer Mode (ATM) operation via satellite: issues, challenges and resolutions*. International Journal on Satellite Communications, 1994, Vol. 12, pp. 211-222

[Coddington 96] P. Coddington: *Lecture notes on Monte Carlo simulation for statistical physics*. WWW page: http://www.npac.syr.edu/users/paulc/lectures/montecarlo/

[Coddington 97] Correspondance with Paul Coddington via email (paulc@cs.adelaide.edu.au)

[CODIT 95] *CODIT Final Review Report*, Editor: P.G. Andermo (Ericsson Radio Systems AB). Issue 2.0, 21 November 1995. Project Title: UMTS Code Division Testbed (CODIT). Project Number: R2020.

[Delli Priscoli & Sestini 96] F.Delli Priscoli and F. Sestini: *Effects of imperfect power control and user mobility in a CDMA cellular network.* IEEE JSAC, December 1996, Vol. 14, No. 9, pp. 1809-1817

[Del Re *et al*. 95] E. Del Re, R. Fantacci, G. Giambene: *Efficient dynamic channel allocation techniques with handover queueing for mobile satellite networks*. IEEE JSAC, February 1995, Vol.13, No. 2, pp. 397-405

[Drakos 97] Nikos Drakos's Web pages on: http://cbl.leeds.ac.uk/nikos/pail/ml/section3.7.html, University of Leeds, Computer Based Learning Laboratory

[Eng & Milstein 94] T.Eng, L.B. Milstein: *Comparison of hybrid FDMA/CDMA systems in frequency selective Rayleigh fading*. IEEE JSAC, June 1994, Vol. 12, No. 5, pp. 938-951

[Elhakeem *et al*. 94] A.K. Elhakeem, R. Di Girolamo, I.B. Bdira and M. Talla: *Delay and throughput characteristics of TH, CDMA, TDMA and hybrid networks for multipath faded data transmission channels*. IEEE JSAC, May 1994, Vol. 12, No. 4, pp.622-637

[Elia 96] Private communication with Carlo Elia, Transmission Systems Section of the European Space Agency Technical Centre, November 1996

[Email 1] Electronic communication with Kari Kärkkäinen of Oulu University, dated 6 August 1996. Subject: CDMA simulation.

[Fong *et al*. 96] M.-H. Fong, V.K. Bhargava, Q. Wang: *Concatenated orthogonal/PN spreading sequences and their application to cellular DS-CDMA systems*, IEEE JSAC, April 1996, Vol. 14, No. 3, pp. 547-558

[GAA 97] Genetic Algorithms Archive on the Web: http://www.aic.nrl.navy.mil:80/galist/

[Genesis 97] *Genetic Algorithms Archive on the Web* - source code of Genesis 5 by John J. Grefenstette: http://www.aic.nrl.navy.mil:80/galist/src

[Geraniotis et al. 94] E. Geraniotis, Y.-W. Chang, W.-B. Yang: *Multi-Media Integration in CDMA Networks*, Proc. IEEE Third International Symposium on Satellite Telecommunications and Applications, 4-6 July 1994, pp. 88-97

[Geraniotis *et al*. 95] E. Geraniotis, M. Soroushnejad, W.-B.Yang: *A Multi-Access Scheme for Voice/Data Integration in Hybrid Satellite/Terrestrial Packet Radio Networks*, IEEE Trans. On Communications, February/March/April 1995, Vol. 43, No. 2/3/4, pp. 1756-1767

[Gilhousen *et al*. 91] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, Jr., C.E. Wheatley: *On the capacity of a cellular CDMA system*. IEEE Trans. Vehicular Technology, May 199,1Vol.40, No.2, pp. 303-312

[Gilderson & Cherkaoui 97] Jim Gilderson, Jafaar Cherkaoui: *Onboard Switching for ATM via Satellite*, IEEE Communications, July 1997, Vol. 35, No. 7, pp.66-70

[Hart 97] W.E. Hart: *Global Optimization Survey*. Scandia National Laboratories WWW page: *http://www.cs.sandia.gov/opt/survey*, last updated 1997

[Holmes 82] J.K. Holmes: *Coherent Spread Spectrum Systems*. John Wiley & Sons Inc., 1982

[Hui 84] J.Y.N. Hui: *Throughput analysis for code division multiple accessing of the spread spectrum channel.* IEEE JSAC, July 1984, Vol. SAC-2, No. 4, pp. 482-486

[Hung *et al.* 96] A. Hung, M.J. Montpetit, G. Kesidis, P. Takats: *A framework for ATM via satellite*. GLOBECOM 1996, November 1996, pp.1020-1025

[I.356] ITU-T Recommendation I.356: *B-ISDN ATM Layer Cell Transfer Performance*, Geneva, May 1996

[I.371] ITU-T Recommendation I.371: *Traffic Control and Congestion Control in B-ISDN*, Geneva, July 1995

[IEE 96] *Proceedings of the IEE Colloquium on ATM over Satellite*, Savoy Place, London, November 1996, Reference number: 1996/224

[Jalali & Mermelstein 94] A. Jalali and P. Mermelstein: *Effects of diversity, power control and bandwidth on the capacity of microcellular CDMA systems.* IEEE JSAC, June 1994, Vol. 12, No. 5, pp. 952-961

[Kaltenschnee & Ramseier 95] T. Kaltenschnee, S. Ramseier: *Impact of Burst Errors on ATM over Satellite - Analysis and Experimental Results*. Proc. of the 10th International Conference on Digital Satellite Communications, May 1995, Vol. 1, pp. 236-243

[Kärkkäinen *et al.* 94] K.H.A. Kärkkäinen, M.J. Laukkanen, H.K. Tarnanen: *Performance of an asynchronous DS-CDMA system with long and short spreading codes - a simulation study.* MILCOM 94: IEEE Military Communications Conference/AFCCEA Fort Monmouth Chapter Symposium, Fort Monmouth, New Jersey, October 2-5, 1994, pp. 780-784

[Kärkkäinen 96] Kari Kärkkäinen: *Code families and their performance measures for CDMA and military spread-spectrum systems*. Ph.D. Thesis, 1996, , Acta

Universitatis Ouluensis Technica C 89, University of Oulu, Department of Electrical Engineering, 90570 Oulu, Finland

[Kou & Leib 96] C.F. Kou and H. Leib: *Power imbalance effects on packet CDMA*. IEEE JSAC Vol. 14, No. 9, December 1996, pp. 1830-1840

[Laarhoven & Aart 87] P.J.M. Van Laarhoven and E.H.L. Aarts: *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Company, Dordrecht, Holland, 1987

[Lee 91] W.C.Y. Lee: *Overview of Cellular CDMA*. IEEE Trans. Vehicular Technology, May 1991, Vol.40, No.2, pp. 291-302

[Lucantoni & Reilly 97] D.M Lucantoni, P.L. Reilly: *Supporting ATM on a Low-Earth Orbit Satellite System*, IsoQuantic Technologies LLC, http://www.isoquantic.com/ATMsatellites-1.html

[Lunsford *et al.* 95] J. Lunsford, S. Narayanaswamy, D. Chitre, M. Neibert: *Link Enhancement for ATM over Satellite Links*. Proc. of the 10th International Conference on Digital Satellite Communications, May 1995, Vol. 1, pp. 129-136

[Makrakis *et al.* 96] D. Makrakis, R. S. Mander, P. Papantoni-Kazakos: *A New Medium Access Control Protocol for Integrated Traffic Personal Communication Networks*. Proc. of the Multiaccess, Mobility and Teletraffic for Personal Communications, Paris 1996

[Maral 93] G. Maral, M. Bousquet: *Satellite Communications Systems*. 2nd Edition, John Wiley and Sons, 1993

[Mauger & Rosenberg 97] R. Mauger, C. Rosenberg: *QoS guarantees for multimedia Services on a TDMA-Based Satellite Network*, IEEE Communications, July 1997, Vol. 35, No. 7, pp.56-65

[Mazzini & Tralli 98] G.Mazzini and V.Tralli: *Performance evaluation of DS-CDMA systems on multipath propagation channels with multilevel spreading sequences and RAKE receivers*. IEEE Trans. on Communications, February 1998, Vol. 46, No. 2, pp.244-257

[McTiffin *et al.* 94] M.J. McTiffin, A.P. Hulbert, T.J. Ketsoglou, W. Heimsch, G. Crisp: *Mobile access to an ATM network using a CDMA interface*. IEEE JSAC, June 1994 Vol. 12, No. 5, pp. 900-908

[Monsen 95] P. Monsen: *Multiple-access capacity in mobile user satellite systems*. IEEE JSAC, February 1995, Vol. 13, No. 2, pp. 222-231

[MSIM 97] *Microsim User Guide and Reference Manual*. Report number ELE01RPTK7/007/03. Department of Electronic Engineering, Queen Mary and Westfield College, London, 1997

[Nemhauser & Wolsey 88] G.L. Nemhauser and L.A. Wolsey: *Integer and Combinatorial Optimization*. John Wiley & Sons Inc., 1988

[Newson & Heath 94] P. Newson and M.R. Heath: *The capacity of a spread spectrum CDMA system for cellular mobile radio with consideration of system imperfections*. IEEE JSAC, May 1994, Vol. 12, No.4, pp. 673-684

[Noneaker & Pursley 94] D.L. Noneaker and M.B. Pursley: *On the chip rate of CDMA systems with doubly selective fading and RAKE reception.* IEEE JSAC, June 1994, Vol. 12, No. 5, pp. 853-861

[Nazari & Ziemer 86] N. Nazari and R.E. Ziemer: *Probability of error for DS-SS multiple access systems.* 24th Annual Allerton Conference on Communications, Control and Computing, October 1986, pp.528-529

[Nazari *et al.* 87a] N. Nazari, R. Ziemer and J. Liebetreu: *The effects of the code period on the performance of asynchronous direct-sequence multiple-access spread-spectrum systems.* GLOBECOM '87, November 1987, Tokyo, Japan, pp.625-629

[Nazari *et al.* 87b]] N. Nazari, R. Ziemer and J. Liebetreu: *Performance of direct-sequence spread-spectrum systems in the presence of multiple-access interference and jamming.* Proc. IEEE Military Communications Conference 1987, pp. 915-921

[Omura 85] J.K. Omura, M.K. Simon, R.A. Scholtz, B.K. Levitt*: Spread Spectrum Communications*. Vols. 1-3, Computer Science Press, Inc., 1985

[Ors *et al.* 98a] T. Ors, Z. Sun, B.G. Evans: *MAC protocol for ATM over satellite*. IEE Conference Publication, 1998, No.451, pp.185-189

[Ors *et al*. 98b] T. Ors, Z. Sun, B.G. Evans: *An adaptive random-reservation MAC protocol to guarantee QoS for ATM over satellite.* Proc. IFIP TC6/ WG6.2 Fourth International Conference on Broadband Communications, 1st-3rd April 1998, Stuttgart, Germany

[Ors *et al*. 97] T. Ors, Z. Sun, B.G. Evans: *Meshed VSAT satellite network architecture using an on-board ATM switch*. IEEE International Performance, Computing & Communications Conference, Proceedings, 1997, pp.208-214

[Osama & Hussein 97] Osama Kubbar and Hussein T. Mouftah: *Multiple Access Control Protocols for Wireless ATM: Problems Definition and Design Objectives*, IEEE Communications Magazine, November 1997, pp. 93-99

[Papadimitriou & Steiglitz 82] C.H. Papadimitriou and K. Steiglitz: *Combinatorial Optimization: algorithms and complexity*. Prentice Hall, 1982

[Passas *et al.* 97] N. Passas, S. Paskalis, D. Vali, L. Merakos: *Quality-of-service-oriented medium access control for wireless ATM networks*. IEEE Communications Magazine, November 1997, Vol.35, No.11, pp.42-50

[Pickholtz *et. al.* 87] R.L. Pickholtz, D.L. Schilling, L.B. Milstein: *Theory of spread-spectrum communications - a tutorial*. IEEE Trans. Communications, May 1992, Vol. COM-30, No. 5, pp. 855-884

[Pitts & Schormans 96] J.M. Pitts and J.A. Schormans: *Introduction to ATM design and performance*. John Wiley and Sons Inc., Chichester, 1996

[Press 89] W.H. Press: *Numerical recipes in Pascal*. Cambridge University Press, 1989

[Pursley 77] M.B. Pursley: *Performance analysis for phase-coded spread-spectrum multiple-access communications -Part I: System analysis*. IEEE Transactions on Communications, August 1977, Vol. COM-25, pp. 795-799

[Pursley & Roefs 79] M.B. Pursley and H.F.A. Roefs: *Numerical evaluation of correlation parameters for optimal phases of binary shift-register sequences.* IEEE Trans. on Communications, October 1979, Vol. COM-27, No. 10, pp. 1597-1605

[Ramalho 96] M.F.N. Ramalho: *Application of an automatically designed fuzzy logic decision support system to connections admission control in ATM networks*. Ph.D. Thesis, December 1996, Dept. of Electronic Engineering, Queen Mary and Westfield College, University of London, London, U.K.

[Ramseier & Kaltenschnee 94] S. Ramseier, T. Kaltenschnee*: ATM over Satellite, Part II: Review, Analysis, Experiments*. Deliverable for the COST226/FE228 project, December 1994, Ascom Tech Ltd.

[Sampath *et al.* 95] A. Sampath, P. Sarath Kumar, J.M. Holtzman: *Power Control and Resource Management for a Multimedia CDMA Wireless System*, WINLAB, Rutgers University, Report No. TR-104, August 1995

[Sampath *et al.* 96] A. Sampath, N.B. Mandayam, J.M. Holtzman: *Analysis of an Access Control Mechanism for Data Traffic in an Integrated Voice/Data Wireless CDMA System*, Proc. 46th Vehicular Technology Conference, 28 April-1 May 1996, Atlanta, Georgia, USA, pp. 1448-1452

[Sampath & Holtzman 97] A. Sampath and J.M. Holtzman: *Access Control of Data Integrated Voice/Data CDMA Systems: Benefits and Tradeoffs*, IEEE JSAC, October 1997, Vol. 15, No. 8, pp.1511-1525

[Sarwate *et al.* 84] D.V. Sarwate, M.B. Pursley and T.U. Basar: *Partial correlation effects in direct sequence spread-spectrum multiple access communication systems*. IEEE JSAC, May 1984, Vol. COM-32, No. 5, pp. 567-573

[Schormans *et al.* 94] J.A. Schormans, J.M. Pitts, K. Williams and L.G. Cuthbert: *Equivalent capacity for on/off sources in ATM*, Electronics Letters, October 1994, Vol. 30, No. 20, pp. 1740-1741

[Schormans *et al*. 94a] J.A. Schormans, J.M. Pitts and L.G. Cuthbert: *Exact fluid-flow analysis of a single on/off source feeding an ATM buffer*. Electronics Letters, July 1994, Vol.30, No.14, pp.1116-1117

[Schormans *et al*. 96] J.A. Schormans, J.M. Pitts, B.R. Clements and E.M. Sharf : *Approximation to M/D/1 for ATM CAC, buffer dimensioning and cell loss performance.* Electronics Letters, February 1996. Vol.32, no.3, pp 164-165.

[Schormans 97] Private communication with J.A. Schormans, senior lecturer at Queen Mary and Westfield College, University of London, London

[Sklar 88] B. Sklar: *Digital Communications*. Prentice Hall, Inc. 1985

[Soroushnejad & Geraniotis 95] M. Soroushnejad and E. Geraniotis*: Multi-Access Strategies for an Integrated Voice/Data CDMA Packet Radio Network.* IEEE Transactions on Communications, February/March/April 1995 Vol. 43, No. 2/3/4

[Strange *et al.* 94] A. Strange, S.L. Arambepola, A.J. Flavin, A. Stevenson: *Carriage of SDH over satellite systems*, International Journal on Satellite Communications, 1994, Vol. 12, pp. 239-247

[Timotijevic & Schormans 96] T. Timotijevic and J.A. Schormans: *ATM over DS-CDMA in a satellite communications system*, UK Teletraffic Symposium, Manchester University, March 1997

[Timotijevic & Schormans 97] T. Timotijevic and J.A. Schormans: *A traffic analysis of ATM over DS-CDMA over a satellite link*, UK Teletraffic Symposium, University of Durham, 23-25 March 1998

[Timotijevic & Schormans 98a] T. Timotijevic and J.A. Schormans: *Performance analysis of ATM transmission over a DS-CDMA channel*, Proc. on International Conference on Broadband Communications (BC'98), IFIP-TC6-WG6.2, 1-3 April 1998, Stuttgart, Germany, pp. 95-106

[Timotijevic & Schormans 98b] T. Timotijevic and J.A. Schormans: *ATM-level performance analysis on a DS-CDMA satellite link using DTX*, accepted for publication in IEE Proceedings on Communications, 1 September 1998

[Turin 84] G.L. Turin: *The effects of multipath and fading on the performance of direct-sequence CDMA systems*. IEEE JSAC, July 1984, Vol. SAC-2, No. 4, pp. 597-603

[Van Nee et al. 95] R. Van Nee, R.N. Van Wolfwinkel and R. Prasad: *Slotted ALOHA and code division multiple access techniques for land-mobile satellite personal communications*. IEEE JSAC, February 1995, Vol.13, No. 2, pp. 382-388

[Vatalaro *et al.* 95] F. Vatalaro, G.E. Corazza, C. Caini and C. Ferrarelli: *Analysis of LEO, MEO and GEO global mobile satellite systems in the presence of interference and fading.* IEEE JSAC, February 1995, Vol. 13, No. 2, pp. 291-300

[Vaquero 96] Private communication with Salvador V. Vaquero, a graduate trainee in ESTEC, held in ESTEC on 17th April 1996, and simulation results of delay vs. time of coverage for two earth stations for Waves-Jocos produced by S. V. Vaquero

[Vojcic *et al.* 94] B.R. Vojcic, R.L. Pickholtz and L.B. Milstein: *Performance of DS-CDMA with imperfect power control operating over LEO satellite link.* IEEE JSAC, May 1994, Vol. 12, No. 4, pp. 560-567

[Vojcic *et al.* 95] B.R. Vojcic, L.B. Milstein and R.L. Pickholtz: *Total capacity in a shared CDMA LEOS environment*. IEEE JSAC, February 1995, Vol. 13, No. 2, pp.232-244

[Yang & Geraniotis 94] W.-B. Yang and E. Geraniotis: *Admission Policies for Integrated Voice and Data Traffic in CDMA Packet Radio Networks*, IEEE JSAC, May 1994, Vol.12, No. 4, pp.654-664

[WEB/Alcatel] *http://www.alcatel.com/our_bus/telecom/products/space/text_a1a.htm*

[WEB/Esa] *http://esapub.esrin.esa.it/pointtotest/test299.html*

[WEB/Lloyd] Lloyd Wood's WWW page: *http://www.ee.surrey.ac.uk/Personal/L.Wood/constellations/*

[WEB/Ors 96] Tolga Ors's WWW page: *http://www.ee.surrey.ac.uk/Personal/T.Ors/atmsat/*

[WEB/Tor] Tor Wisloff's WWW page: *http://www.idt.unit.no/~torwi*

[Winstanley 96] Private communication with Steven Winstanley, ATM consultant at Roke Manor Research (now part of Siemens), Romsey

[Wisloff 96] T. Wisloff: *On satellite path diversity for GSM rate compatible TDMA big LEOs*. Multiaccess, Mobility and Teletraffic for Personal Communications, Edited by Bijan Jabbari, Philippe Godlewski and Xavier Lagrange, Kluwer Academic Publishers, May 1996

[Wittig 97] M. Wittig: "Large-Capacity Multimedia Satellite Systems", IEEE Communications Magazine, July 1997, Vol. 35, No. 7, pp.44-49

[Wu *et al.* 94] W.W. Wu, E.F. Miller, W.L. Pritchard, R.L. Pickholtz: *Mobile Satellite Communications*. Proc. IEEE, September 1994, Vol. 82, No. 9

APPENDIX A: MEAN BUSY AND IDLE TIMES OF THE MULTIPLEXER

Variables:

K	number of traffic types (in the numerical analysis K=3)
α_k, β_k	state transition probabilities for the kth connection (applying to the ends of time slots)
α_{MUX}, β_{MUX}	state transition probabilities for the ON-OFF process at the output of the multiplexer (applying to the ends of time slots)
λ	mean aggregate arrival rate at the multiplexer input
λ_k	mean arrival rate of the kth connection
SR, SR_k	slot rate expressed in [cells/s]; it is equal to the transmission rate R divided by cell size: $SR=R/424$
PCR, PCR_k	peak cell rate of a source during busy (ON) period
B, B'	random variable representing duration of the busy period with and without overhead
I, I'	random variable representing duration of the idle period with and without overhead
T_B , T_{Bk}	mean duration of a busy period of a multiplexer including overhead transmission time
$T_{B'}$, $T_{Bk'}$	mean duration of a busy period of a multiplexer without overhead
T_I , T_{Ik}	mean duration of an idle period of a multiplexer including overhead transmission time
$T_{I'}$, $T_{Ik'}$	mean duration of an idle period of a multiplexer without overhead
$T_{hgo'}$, $T_{hgo,k}$,	mean hangover period
T_{acq}	mean acquisition period
$p_{ON,k}$, $p_{OFF,k}$	probability that the connection k is ON/OFF,
P_{ON} , P_{OFF}	probability that the multiplexer (process at its output) is ON/OFF
$T_{ON,k}$, $T_{OFF,k}$	mean ON and OFF periods of the k^{th} connection,
$E[.]$	expected (mean) value

Duration of an OFF state in terms of number of cells or cell slots is found as an expectation of a geometrically distributed random variable [Pitts & Schormans 96]:

$$E(OFF) = \sum_{j=1}^{\infty} j \cdot (1-\alpha)^{j-1} \cdot \alpha = \frac{1}{\alpha} \quad [\text{cell slots}] \qquad \text{A.1}$$

$$E(ON) = \sum_{j=1}^{\infty} j \cdot (1-\beta)^{j-1} \cdot \beta = \frac{1}{\beta} \quad [\text{cells}] \qquad \text{A.2}$$

From equations $E(OFF) = \sum_{j=1}^{\infty} j \cdot (1-\alpha)^{j-1} \cdot \alpha = \frac{1}{\alpha}$ [cell slots] and A.2, the parameters α_k and β_k of a source of traffic type *k* can be found when the duration of the mean ON and OFF periods of a source are known together with its peak cell rate (PCR_k) during the ON period, and a slot rate of a multiplexer (*SR*):

$$\alpha_k = \frac{1}{\overline{T_{OFF,k}} \cdot SR} \qquad \text{A.3}$$

$$\beta_k = \frac{1}{\overline{T_{ON,k}} \cdot PCR_k} \qquad \text{A.4}$$

In order to characterise the process at the output of a multiplexer it is necessary to determine the mean busy and idle periods. This can be done using the same steps as above:

$$E[I'] = \frac{1}{\alpha_{MUX}} \quad [\text{slots}] \qquad \text{A.5}$$

$$E[B'] = \frac{1}{\beta_{MUX}} \quad [\text{cells}] \qquad \text{A.6}$$

The process at the output of a multiplexer is taken from the idle to the active state when at least one of the sources goes from the idle to the active state. Thus we have:

$$\alpha_{MUX} = 1 - \prod_{k=1}^{K} (1-\alpha_k) \qquad \text{A.7}$$

The probability that the multiplexer is idle is a product of the probabilities of individual sources being idle:

$$P_{OFF} = \prod_{k=1}^{K} p_{OFF,k} \qquad \text{A.8}$$

and also:

$$P_{OFF} = \frac{E[I']}{E[B'] + E[I']} = \frac{T_{I'}}{T_{I'} + T_{B'}} \qquad \text{A.9}$$

yielding the expression for the mean multiplexer busy period duration:

$$E[B'] = E[I'] \cdot \frac{P_{ON}}{P_{OFF}} = E[I'] \cdot \frac{1 - P_{OFF}}{P_{OFF}} \quad \text{[slots].} \qquad \text{A.10}$$

The expression for *E[B']* is in terms of the number of multiplexer service time slots To express the duration of the busy and idle periods in seconds, *E[B]* and *E[I]* have to be multiplied by the service time, i.e. divided by the slot rate:

$$T_{B'} = \frac{E[B']}{SR} \qquad \text{A.11}$$

$$T_{I'} = \frac{E[I']}{SR} \qquad \text{A.12}$$

To conclude: to determine the mean duration of the ON and OFF periods of the process at the output of a multiplexer, we find α_{MUX} from A.12, insert it into A.10 to find *E[I]*, find P_{OFF} from A.13, and find *E[B]* from A.15.

APPENDIX B: SCENARIO 3 SYSTEM PARAMETER OPTIMISATION BY SIMULATED ANNEALING

This section is intended to give a general overview of the concept of simulated annealing, and of the associated implementation issues. A detailed coverage of the theory of simulated annealing can be found in [Laarhoven & Aarts, 87] and [Aarts & Korst 89].

B.1 The concept

The idea of simulated annealing was derived from the analogy between solving complex combinatorial optimisation problems and the physical process of annealing of solids [Laarhoven & Aarts, 87]. In the process of annealing, solids are first heated up to a very high temperature at which all particles of the solid randomly arrange themselves in a form of liquid, followed by a slow process of cooling down, in which the particles re-arrange themselves to form crystals with minimum potential energy. This perfect state of minimum-energy crystal structure can only be achieved if the starting temperature is sufficiently high to ensure that all the particles are excited to completely random positions within the liquid, and the cooling process is sufficiently slow for the particles to achieve the required thermal equilibrium at each stage in the cooling process (that is, the state of minimum energy for a given temperature). The process of cooling to a solid is governed by an exponential physical law that relates the temperature of the solid and its energy in thermal equilibrium at that temperature. In thermal equilibrium, the probability that a solid has energy E at temperature T is described by the Boltzmann distribution [Laarhoven & Aarts 87]:

$$\Pr(\mathbf{E} = E) = \frac{1}{Z(T)} \cdot e^{-\frac{E}{k_B \cdot T}} \qquad \text{B.1}$$

where k_B is the Boltzmann constant and $Z(T)$ is a normalisation factor over all energy states j [Aarts & Korst 89]:

$$Z(T) = \sum_j e^{-\frac{E_j}{k_B \cdot T}} \qquad \text{B.2}$$

Thermal (quasi) equilibrium is described by the law:

$$E = k_B \cdot T \qquad \text{B.3}$$

which gives the minimum achievable energy for given temperature T.

The analogy with large scale combinatorial optimisation is in the following:

- The solid configurations correspond to the solutions i.e. the points of the search state space of a combinatorial optimisation problem.
- The energy of the solid corresponds to the objective cost function.
- Temperature is translated into the control parameter, which regulates the cooling process i.e. the speed and rate of the search.

Starting from an arbitrary configuration, the search generates new configurations from the old ones by some generation mechanism. This is equivalent to the particles' reconfiguration. The new solution (configuration of the solid) is located in the neighbourhood of the original state. The energy (cost) of the new configuration (solution) is compared to the energy of the previous one. If the new configuration has energy lower than the previous one, the transition is accepted. If it is higher, the transition is accepted with probability that follows Boltzmann's distribution.

In this way, the process (search) is approaching the thermal equilibrium at a given temperature value - that is, the optimal solution for a given control parameter. This algorithm is called Metropolis iteration. Simulated annealing algorithms are a sequence of Metropolis iterations with gradually decreasing values of the control parameter - that is, temperature T - whereby the near-optimal solution is reached for some very low value of the control parameter for which no significant further improvement in the cost function (energy) is possible. The value of a "significant improvement" is quantified by some stop criterion.

The speed and convergence of the algorithm is determined by the number of the Metropolis iterations for a single value of the control parameter, and the initial value and the mechanism that governs the decrease of value of the control parameter.

B.2 Implementation

To implementat the algorithm, the following have to be defined:

1. the states, i.e. configurations (description of a state space)
2. the generation (reconfiguration) mechanism
3. the cost function *E*
4. the initial value of the control parameter *T*
5. the function that governs the decrease of the control parameter *T*
6. final value of *T*, i.e. the stop criterion
7. the number of reconfigurations i.e. Metropolis iterations at each value of *T*.

The parameters 1-3 are problem-specific. The parameters 4-7 define the convergence of the algorithm and the accuracy of the final solution. The rules given in [Aarts & Korst 89] for specifying these parameters were followed in the implementation of the algorithm used in this thesis. These rules result in a polynomial-time cooling schedule.

B.2.1 Control Parameter

Initial value of temperature has to be sufficiently high to ensure that all the particles are randomly distributed in the physical state of a liquid. This means that the energy of the particles is so high that all the particles' perturbations would lead to the decrease of energy. By analogy, the control parameter *T* has to be high enough to ensure that all reconfigurations are accepted in the search process [Laarhoven & Aart 87]:

$$\forall i, j: \quad e^{-\frac{\Delta E_{i,j}}{T_0}} \approx 1 \qquad \text{B.4}$$

The algorithm implemented followed the rule for the initial value of the control parameter given in [Aarts & Korst 89, pp.60]:

$$T_0 = \frac{\Delta E^+}{\ln\left|\frac{m_2}{m_2 \cdot \chi - m_1(1-\chi)}\right|} \qquad \text{B.5}$$

where:

m_1 is number of transitions that resulted in the cost decrease

m_2 is the number of transitions that resulted in cost increase

ΔE^+ is the average increase in cost (summed over the cost-increasing transitions),

χ is the acceptance ratio, that is the ratio of transitions that have been accepted divided by the total number of generated transitions.

B.2.1.1 Decrementing the control parameter

The following rule for decrementing the control parameters was used [Aarts & Korst 89, pp.63]:

$$T_{k+1} = \frac{T_k}{1 + \frac{T_k \cdot \ln(1+\delta)}{3\sigma(T_k)}}, \quad k = 0, 1, \ldots \qquad \text{B.6}$$

where

T_k, T_{k+1} are the two consecutive values of the control parameter;

$\sigma(T_k)$ is the standard deviation of the values of the cost function for the control parameter T_k,

found from [Aarts & Korst 89, pp. 25]:

$$\sigma(T) = \left|\frac{1}{L}\sum_{i=1}^{L}\left(E_i(T) - \overline{E(T)}\right)^2\right|^{\frac{1}{2}}$$

δ is the distance parameter whose value is close to 1, typically between 0.85 and 0.99. Small values of distance parameter lead to small decrements, large values to large decrements.

This rule follows from the fact that in a slow cooling schedule quasi-equilibrium is maintained throughout the annealing process, and by analogy, the algorithm must ensure

that the steady-state probability distributions for the two consecutive values of the control parameter are sufficiently close. That means [Aarts & Korst 89, pp. 61]:

$$\forall k \geq 0 \therefore \quad \left\| \mathbf{p}(T_k) - \mathbf{p}(T_{k+1}) \right\| < \varepsilon$$

where

k	is the kth configuration
$\boldsymbol{p}(T_k), \boldsymbol{p}(T_{k+1})$	are the vectors representing the steady-state probability distributions at temperatures T_k and T_{k+1}.

B.2.1.2 Final value of the control parameter

The algorithm is terminated if for some very small number ε_{stop} the following holds [Aarts & Korst 89, pp. 64]:

$$\frac{T_k}{E_{T_0}} \cdot \left. \frac{\partial E_T}{\partial T} \right|_{T=T_k} < \varepsilon_{stop}$$

where:

E_{To}	is the mean value of the cost function at the initial value of the control parameter
E_T	is the mean value of the cost function for control parameter T.

This is effectively the extrapolation of the mean or expected cost as the control parameter decreases.

When further improvements in the value of the expected cost cannot be achieved, or the only improvements are so minimal that they are negligible, the algorithm is terminated with the near-optimum solution.

B.2.2 Length of Metropolis chain

The number of iterations for a single value of the control parameter must ensure that the thermal or quasi equilibrium at that value of T is achieved. That means that it must perform a sufficient number of transitions so that there is [Aarts & Korst 89]:

"... a sufficiently large probability of visiting the at least a major part of the neighbourhood of a given solution."

Aarts and Korst proposed that the number of iterations be equal to the size of the solution neighbourhood [Aarts & Korst 89, pp. 64-65]. However, in large scale combinatorial problems the size of the state space may be prohibitively large. According to [Coddington 97] the length of the iteration chains (Markov chains) is in practical applications determined by experimentation. The length of chains is gradually increased for as long as the optimal value of the cost function achievable in a chain improves. The chain length is fixed at the value at which its further increase no longer produces improved values of the objective function.

B.3 Comparison of results

Genetic algorithm *Genesis 5* (GENEtic Search Implementation System) was used to verify the results obtained by simulated annealing. Genesis 5 outputs the 20 best system configurations (solutions), their cost function (i.e. evaluation) and the trial and generation counters of their first occurrence. In order to compare the results of the simulated annealing and the genetic algorithm, and assess how good are the "good" results, it is necessary to know what solutions are bad, and what is the capacity i.e. the cost function of these bad solutions (configurations). The simulated annealing algorithm developed was adjusted to produce both the 50 best and 50 worst solutions. All configurations found by both simulated annealing and the genetic algorithm were compared with the best solution found by the simulated annealing. As the optimum solution depends on the QoS and processing gain, the algorithm was run for 5 different combinations of the input system parameters: processing gain and QoS (SNR and E_b/N_0 of each traffic type). An optimal solution would favour the traffic type that had highest capacity per channel and lowest channel bandwidth. Thus the optimisation algorithm in the case of equal or similar processing gains and QoS requirements will push the result towards one of the boundary solutions, where the number of narrow bandwidth, high capacity channels will be maximised and the number of other channels will be minimised (i.e. their number will

equal to 1). In other words, under these circumstances the optimum solution was one with the maximum number of small channels and a small number of larger channels.

To push the optimisation algorithm away from the boundary solution, sets of input parameters were chosen such that one type of traffic channel would have the advantage of low QoS requirements (resulting in large channel capacity), but a different traffic type would have the advantage of lower processing gain (higher processing gains yield wider channel bandwidths, so fewer such channels can be accommodated within the fixed bandwidth of a satellite transponder).

As the capacity function in the system scenarios is a function of three variables, the four-tuple [x, y, z, $c=f(x,y,z)$] defines a 4-dimensional vector space, making the full graphical plot of the function $c=f(x,y,z)$ impossible. So a number of different plots were produced to allow graphical comparison of the two algorithms, and allow visual assessment of the regions of good and bad solutions. The plots are:

1. 3-dimensional plot showing points [x, y, $f(x,y,z)$] for the obtained configurations. This shows the regions of good and bad solutions with respect to their cost function evaluations, and the number of channels (x, y) for the first two traffic types.
2. 3-dimensional plot showing the vectors originating in the best (optimal) solution [x_{max}, y_{max}, z_{max}] and ending in all other generated solutions (x, y, z). This pictures the 3-dimensional region of possible good and bad configurations, their distribution and position in comparison to the best obtained configuration.
3. 2-dimensional "target" plot showing the circles of the radius equivalent to the Hamming distance between the points [x_{max}, y_{max}, z_{max}, $f(x_{max}, y_{max}, z_{max})$] and [$x$, y, z, $f(z,y,z)$], where the centre of the target represents the best solution i.e. point [x_{max}, y_{max}, z_{max}, $f(x_{max}, y_{max}, z_{max})$].

The cost function values of best solutions obtained by the simulated annealing and those obtained by Genesis 5 differ by no more than 5%. Depending on the system input parameters, the worst and best solutions can be very close or far away. In fact, the proximity of the good solutions to each other, and the proximity of cost function values for the wide range of the top 50 good solutions indicate that the cost function is very flat.

That means that for particular sets of traffic and system parameters the optimisation may not be necessary, as it would not yield significant improvement of system capacity. This is an important conclusion; however, it is conditional on the particular set of system parameters like QoS (SNR and E_b/N_0) required for each traffic type, processing gains, and peak cell rates i.e. transmission rates used for each of the traffic types supported. In general it would not be known a priory whether optimisation is vital or not.

Figures B.1-B.3 illustrate results of Table B.2, for the parameters listed in Table B.1.

	Voice	Video	Data
Transmission rate [kb/s]	64	384	512
Channel bandwidth [MHz]	7.0	10.4	6.6
Processing gain	100	27	13
Channel capacity [number of connections]	10	19	7
E_b/N_0 [dB]	9	12	12
SNR [dB]	5.5	4.5	5.5
Worst configuration:[1, 1, 71]	**Capacity** = 526		Load = 0.11
Best configuration: [2, 46, 1]	**Capacity** = 901		Load = 0.75

Table B.1 System parameters for test 1 and results for simulated annealing optimisation.

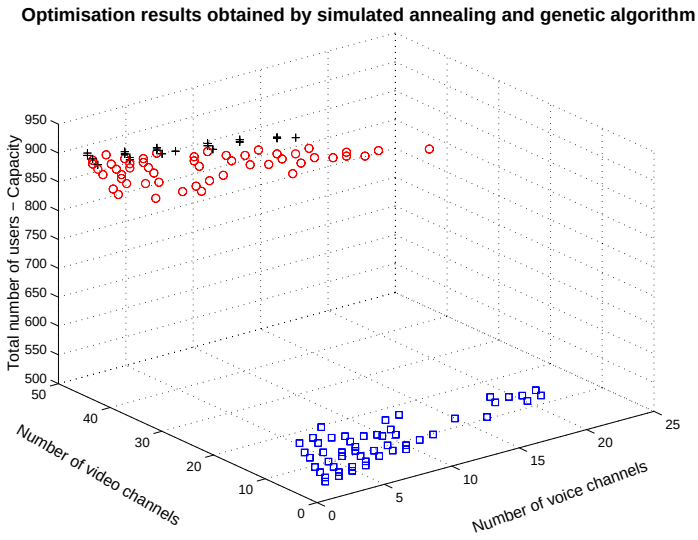


Figure B.1 Best and worst solutions of comparative optimisation tests - Cartesian 3D plot: squares- 50 worst solutions by SA; o - 50 best solutions by SA; + - 20 best solutions by Genesis 5.

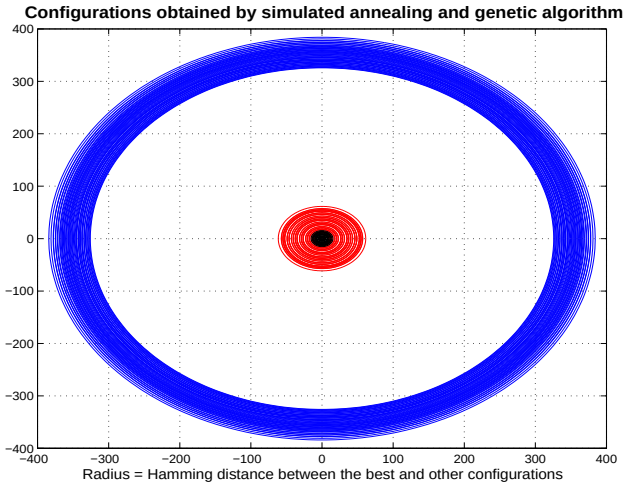


Figure B.2 Best and worst solutions of comparative optimisation tests - Target plot: centre is the best solution. Radii are equivalent to the Hamming distance between the best and other solutions.

Table B.1 shows that the effect of optimisation on the system capacity and achievable load, for the chosen system parameters, is significant. Tables B.3-B.10 give system parameters and results for the 5 tests carried out to validate the simulated annealing optimisation algorithm against Genesis 5, and assess the effect of optimisation.

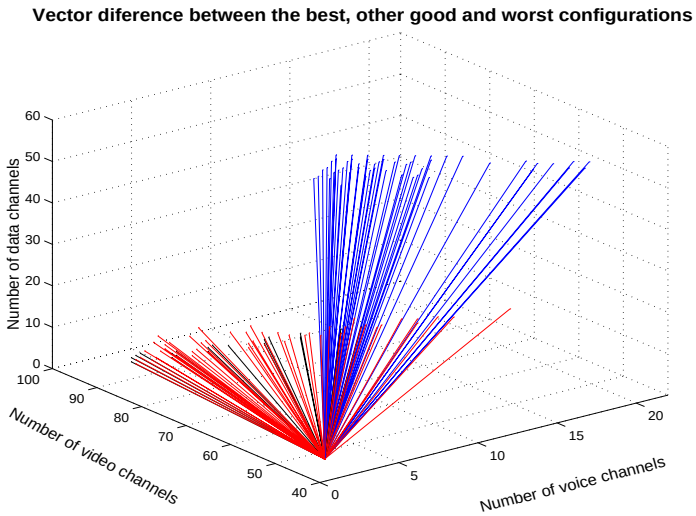


Figure B.3 50 Vector difference between the best configuration [1,46,2] and other 50 best configurations (red), 20 best configurations obtained by Genesis (black) and worst configurations

Test 1	**Solution [m_1,m_2, m_3]**	**Cost i.e. Capacity**	**Comment**
Best SA	2 46 1	901	Load = 0.75, boundary solution, video dominates
Best GA	1 47 2	906	close to SA, boundary solution
Worst SA	1 1 71	526	Load = 0.11, data dominates

Table B.2 Results for optimisation test 1.

Test 2	**Solution [m_1,m_2, m_3]**	**Cost i.e. Capacity**	**Comment**
Best SA	2 48 1	890	Load = 0.75, boundary solution
Best GA	8 44 1	869	close to SA, video dominates
Worst SA	2 1 76	494	Load = 0.11

Table B.3 Results for optimisation test 2.

Parameters for Test 2	Voice	Video	Data
Transmission rate [kb/s]	64	384	512
Channel bandwidth [MHz]	7.0	10.0	6.1

Processing gain	100	26	12
Channel capacity [number of connections]	10	18	6
E_b/N_0 [dB]	9	12	12
SNR [dB]	5.5	4.5	5.5

Table B.4 System parameters for optimisation test 2.

Test 3	**Solution $[m_1,m_2,m_3]$**	**Cost i.e. Capacity**	**Comment**
Best SA	50 14 1	703	Load=0.61
Best GA	43 19 1	699	flat cost function: for similar capacity GA and SA yield different solutions
Worst SA	2 1 76	566	Load=0.11

Table B.5 Results for optimisation test 3.

Parameters for Test 3	Voice	Video	Data
Transmission rate [kb/s]	64	384	512
Channel bandwidth [MHz]	7.0	10.0	6.1
Processing gain	100	26	12
Channel capacity [number of connections]	10	14	7
E_b/N_0 [dB]	9	11.5	12
SNR [dB]	5.5	5	5

Table B.6 System parameters for optimisation test 3.

Test 4	**Solution $[m_1,m_2,m_3]$**	**Cost i.e. Capacity**	**Comment**
Best SA	66 2 2	772	Load=0.71, voice dominates
Best GA	66 3 1	766	close to SA; very flat cost function as worst case is close to the best
Worst SA	1 46 4	691	Load=0.38, video dominates

Table B.7 Results for optimisation test 4.

Parameters for Test 4	Voice	Video	Data
Transmission rate [kb/s]	64	384	512
Channel bandwidth [MHz]	7.0	10.0	6.1
Processing gain	100	26	12
Channel capacity [number of connections]	11	14	9
E_b/N_0 [dB]	8.5	11.5	12.5
SNR [dB]	5	5	4.5

Table B.8 System parameters for optimisation test 4.

Test 5	**Solution [m_1,m_2,m_3]**	**Cost i.e. Capacity**	**Comment**
Best SA	73 2 1	765	Load=0.39, voice dominates, boundary solution
Best GA	74 1 1	764	close to SA, boundary solution
Worst SA	2 2 75	721	Load=0.11, very flat cost function

Table B.9 Results for optimisation test 5.

Parameters for Test 5	Voice	Video	Data
Transmission rate [kb/s]	64	384	512
Channel bandwidth [MHz]	7.0	10.0	6.1
Processing gain	92	23	12
Channel capacity [number of connections]	11	13	9
E_b/N_0 [dB]	8.5	11.5	12.5
SNR [dB]	5	5	4.5

Table B.10 System parameters for optimisation test 5.