Paging System Design Issues and Overview of Common Paging Protocols

Yianni Michalas and Leandros Tassiulas *
Institute for Systems Research and
Department of Electrical Engineering
University of Maryland
College Park, MD 20742, USA
Phone: (301) 405-6620
Fax: (301) 314-8586
E-mail: yianni@eng.umd.edu, leandros@eng.umd.edu

7/15/98

Abstract

Paging services constitute a large and integral part in Personal Communication Systems (PCS), traditionally providing one-way and recently two-way communication. We describe paging through its main application as a personal selective calling system. A detailed description of the paging network architecture along with the main paging signaling protocols is provided. The issue of battery savings of the user terminal is discussed and the protocol mechanisms for reducing energy consumption are identified. In addition, simulcasting, the transmission technique commonly employed in paging protocols, along with other techniques employed in paging to improve reception and reduce the effects of multi-path fading, are examined.

1 Introduction

An important function in a cellular wireless network is the precise tracking of a mobile user, in order for the user to receive calls and messages anywhere in the network. The process of locating the exact position of the mobile user is called paging. Paging services are a large and integral part of Personal Communication Services (PCS). In its original definition, paging is a one-way personal selective wireless calling system [1]. Its function is to locate and alert the mobile user without receiving any immediate acknowledgment from the user or establishing any real-time interactive

*Corresponding Author

communication between the calling and the called parties. Although this form of paging service does not provide interactive communication among mobile users, it has some advantages over the present cellular telephony (requires less radio bandwidth, inexpensive service, the pager device is smaller lighter and cheaper, and utilizes less power). However within the framework of second and third generation cellular radio systems, confirmation that correct messages have been delivered to mobile users can be achieved through two-way paging systems. Today these systems are incorporated in cellular phone network platforms such as the short message service of the Global System for Mobile Communications (GSM).

Paging issues though do appear in cellular telephony and mobile computing since paging plays a fundamental role in the conventional tracking strategy (also known as location management) of Mobile Stations (MS) within cellular wireless networks like GSM. Paging is employed in the call set-up process, before a call can be routed to the mobile user to establish a real-time connection.

The purpose of this report is to explore how paging is used as a main PCS application by describing the conventional network architectures and protocols of the main digital paging systems used today. In addition a brief description of some of the technologies they utilize to successfully deliver paging messages over the radio channel and an overview of paging in GSM networks is provided in the last section.

2 Main Types of Paging Messages

Today's paging systems use several features to alert the paging subscriber anytime and almost anywhere, as coverage can be extended globally. In addition more varied applications of paging besides alerting, such as E-mail, voice-mail and faxes are in the process of being integrated in the latest paging technologies. The traditional main types of paging messages are tone only, voice, numeric and alphanumeric.

- Tone pagers alert users with a loud beeping tone or a vibrating signal and typically a flashing indicator. A tone pager has a dedicated telephone number where usually the last 6 digits will represent the user's unique paging address. Every time the number is dialed the pager is triggered. With some paging receivers the sound of the alert tone may be changed to indicate group call, a page from a different system, or a priority page. Typically the number of alert tones is restricted to four. This type of messaging provides great savings in air-time but has the disadvantage that the alerted subscriber only knows that he has to call certain pre-arranged numbers based on the kind of alert tone received.

- Voice pagers alert users with a tone, and then the voice message is received. Voice paging provides detailed information and message urgency. A typical receiver may store up to 12 minutes of voice messaging and some voice pagers have memory capability enabling the user to store voice messages for later reference.

- Numeric pagers alert users with either a tone or a vibration followed by the signaling receipt of a numeric message, typically a telephone number. They use small LED or LCD screens to display the digit string to provide the user with an immediate notification to contact a specific individual. This is the most common type of paging messaging.

- Alphanumeric pagers alert users with either a tone or a vibration followed by a signaling receipt of an alphanumeric message. The text message can provide detailed and complete information. Messages may be sent through a telephone's keypads, a telex or a modem. Pagers may store several tens of messages and a total number of characters in excess of 1000. Mechanisms also exist for 'protecting' certain messages. Alphanumeric messaging typically eliminates the need to return telephone calls, saving time and increasing efficiency.

3 Typical Paging Network Architecture

The main components of the traditional one-way paging system consist of the input source and delivery system, the paging terminal, the base station controller, the base station and finally the pager (figure 1).

The input source or user terminal equipment, i.e. telephone, generates the messages and the corresponding delivery system i.e. the Public Switched Telephone Network (PSTN typically handles alert tone and numeric messages) or the Public Switched Data Network (PSDN handles alphanumeric messages), delivers and distributes the messages to the paging network usually through T1 (up to 24 subscriber lines) or E1 (up to 30 subscriber lines) trunks. Paging requests are sent to the paging terminals through the user access interface by using telephone trunk protocols or specialized data input protocols such as the Telocator Alphanumeric Protocol (TAP) and the Telocator Data Protocol (TDP), that can handle messages of digital data from portable computer devices [1].

The paging terminal is responsible for receiving, processing, sorting, billing and forwarding information from the caller to the base station controller or to other paging terminals within a multi-city paging network. Paging terminals are connected to each other via an internetwork interface. Over the years several manufacturers developed their own proprietary protocols for inter-terminal

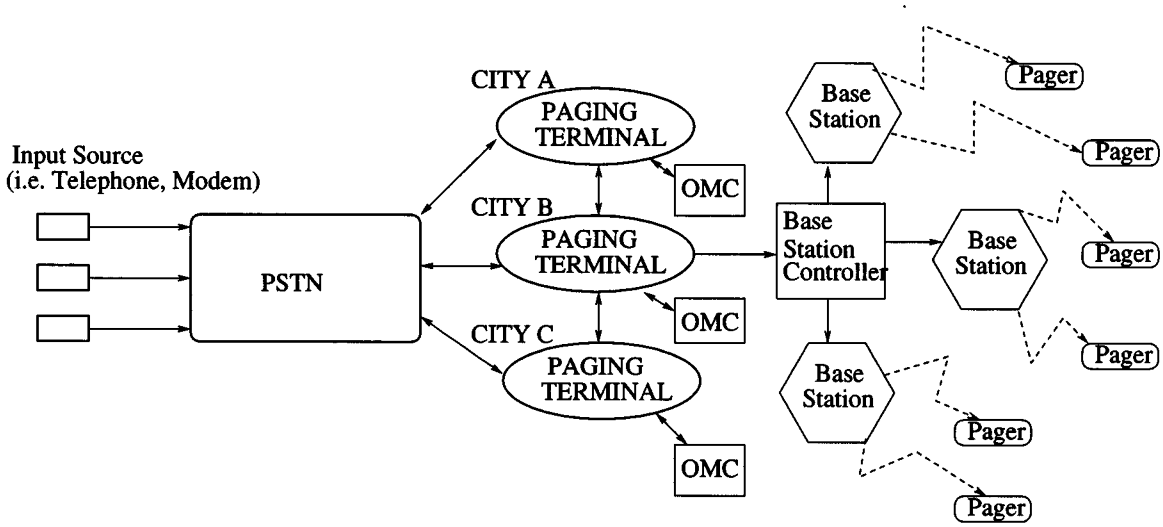

Figure 1: Typical Paging Network Architecture

communication until a standard point-to-point protocol, known as the Telocator Network Paging Protocol (TNPP) was introduced, to move radio page requests within networks and between different carrier networks. TNPP assures reliable delivery in a point-to-point communications environment. Each TNPP packet contains a destination address but it needs to work on top of a routing algorithm to fully utilize the network capabilities. TNPP can also be used in a satellite network environment, where page requests can be broadcast over satellite to a large number of paging terminals in a one-way transmission mode since the remote paging terminals do not have the ability to acknowledge proper receipt. The protocol supports this type of network by allowing multiple transmissions of the data without creating duplicate information at the destination mode. Recently, the Telocator Interswitch Paging Protocol (TIPP) for inter-terminal paging communications was developed to increase throughput, message size, and provide features beyond those provided by TNPP [2].

Existing high capacity terminals may support more than 1 million pagers through 1900 T1 and E1 input lines [1]. Each terminal maintains a large customer database with the necessary information to alert the pager. It validates the type of call, determines the authenticity of the subscriber by looking up a database of pager addresses known as cap codes, specifies the type of message to be used, and converts voice messages to alphanumeric text or stores them in a mailbox for feature reference. Most important though, it encodes the address and message into the appropriate paging signaling protocol. The signaling protocol, that will be covered extensively later, is the root of all paging systems since it dictates the capacity, the latency and signaling speed, the pager battery life and the data integrity of the paging system.

All maintenance, administration and operation functions of the paging network are performed by the operation and maintenance center (OMC) that is able to access the customer database of each terminal, add, delete or update customer records, collect billing information etc.

The base station controller is used to distribute the paging signals to the appropriate base stations accordingly, when the services of a subscriber are limited to a smaller geographical area than the one covered by the paging terminal. The paging signal may be sent to multiple paging terminals for simulcast paging or sequential paging. Transmitter sequencing incurs a higher delay but may be necessary when the pager is in a region where simulcasting can not be used effectively. Since the radio channel is shared by different paging service providers, the base station controllers must also be coordinated with each other.

The base station transmitters accepts the data from the paging terminals via leased telephone lines, PSDN (X.25), RF links or satellite-based networks. Satellite communications technology is typically used between paging terminals and highly dispersed base stations or in some advanced paging services as base stations to broadcast messages directly to the pagers.

Base stations are primarily designed for one-way paging but recent designs allowing two-way messaging have also been implemented. Page codes are transmitted on an assigned RF using dedicated paging channels, FM radio or subcarriers of TV stations having the paging signal superimposed on the normal TV or radio channel in both the VHF and UHF band.

A wide area paging allows the transmission of messages via the base station to receivers within a defined area of the base station. The distance varies considerably between 5 and 20 km from the base station depending upon the location, power and aerial system of the base station and the location of the receiver [3]. The signal is attenuated by buildings, trees, foliage, and even human bodies especially at high frequencies. Tests have shown that frequencies between 80MHz and 460 MHz are the most suitable for paging in densely built areas. In practise frequencies commonly used for paging are the VHF low 40 MHz and high 150 MHz bands, the UHF 450 MHZ and 900 MHz. In some countries 280 MHz is also available. Attenuation of the signal levels caused by penetrating buildings ranges from 14 to 22 dB at 150 MHz, 18 dB at 250 MHz and 12 to 18dB at 400 MHz [3].

The base stations need high transmitting power (kwatts) and to be even more effective, high antennas to cover large areas. Using large-scale propagation models associated with the paging systems, it was shown that in most situations, doubling the transmitter power results in only a 1.4

times increase in field strength at the receiving point, while doubling antenna height nearly doubles this field strength [4]. Omni-directional antennas with equal gain in all directions and directional antennas providing higher gain in a particular direction, are combined to provide a coverage pattern so as to minimize interference with other communication systems that may use frequencies similar to that used by the paging system. As in all radio mobile communication systems, the received power is predicted by the physics of the three basic propagation mechanisms of reflection, diffraction and scattering, that also describe the small-scale fading and multipath propagation [5]. The coverage of a base station system is measured experimentally and strategically designed so as to minimize the path loss within its covered area [6].

The pager device consists of four basic components.: the receiver, the decoder, the controller logic and the display [7]. The receiver demodulates the data from the RF carrier and passes the data to the decoding section. The receiving section consists of a small built-in ferrite rod and coil antenna, which has a loss factor of approximately 16 dB relative to a half wave dipole, and several other components [3]. A typical page receiver will consist of: An oscillator, an IF filter using a 7th-order low-pass Butterworth filter with a limiting frequency of 10 kHz, a limiter and discriminator that derives the estimate of the input signal and passes it to a VCO, and a post filter for the output of the discriminator using a 3d-order-low-pass Bessel filter with a limiting frequency of 2.9 kHz [8].

The decoder corrects errors, looks at the binary information and identifies the code for the pager while rejecting the messages for other pagers. Pagers may share the same code to be used for group paging or may be assigned different page codes (typically up to four) for different paging functions.

The control logic provides service features and operation functions such as duplicate message detection, message screening, message freeze, multiple alerting codes and so on. Finally, a Liquid Crystal Display (LCD) or a Liquid Emitting Diode (LED) screen are used to display the message and other information

4 Paging Signal Protocols

In a paging system, the paging terminal, after accepting an incoming page and validating it, will encode the pager address and message into the appropriate paging signaling protocol. The signaling protocols implement different encoding formats to successfully carry pager specific address information along with message data and error correction-detection data, to groups or individual pagers. Digital transmission techniques are used to send the binary coded formats, such as 2 and 4-level Frequency Shift Keying (FSK) and PAM/FM. Most paging formats are manufacturer-specific and often proprietary, but there are a few common paging protocols that have been put on public domain so that different manufacturers may produce compatible pagers. Presently, the widely used paging protocols typically support one-way paging, but can be incorporated in two-way paging systems as well. The three most common protocols are POCSAG, described in [9], FLEX [4] and HERMES [10], while two-way paging is supported by ReFLEX and InFLEXion [4].

4.1 One Way Paging-POCSAG

4.1.1 General

The Post Office Code Standardization Advisory Group (POCSAG) is the oldest of the three but is still widely used. POCSAG was developed in 1981 to support tone only, numeric and alphanumeric messages. Today POCSAG operates at 512, 1200 and 2400 bits per second(bps) and can handle up to two million addresses. POCSAG uses 2-level FSK modulation with a deviation difference of +/- 4800 HZ.

The basic signaling pattern used, is a sequence of coded binary data following a synchronous format, that allows pages to be transmitted in a single batch structure. POCSAG provides a battery-saving capability and a fairly high code capacity. The fundamental unit of transmission is a "Cycle", which has a variable time duration depending upon the paging messages queued to be sent. Each cycle consists of a preamble section of 576 alternating 0's and 1's followed by an unlimited number of batches of complete codewords. Each batch begins with a 32-bit frame synchronization codeword (SC) and contains eight 64-bit frames of two 32-bit codewords. There are three types of codewords: address, message and idle codewords. Any two combination of the three types can be used to form a frame. The frame synchronization code marks the start of the batch of codewords. A format of the signal is shown in figure 2.

4.1.2 Preamble

Each transmission starts with a preamble of 576 bits, which serves as a "wake-up" call for all the pagers in the network. The pager's decoder uses the preamble to determine if the data received is a POCSAG signal and to attain bit synchronization and thus acquire word and batch synchronization. POCSAG guarantees that a pager will not miss the start of a Cycle by requiring the pager receivers to monitor the channel at least twice during the length of a preamble.

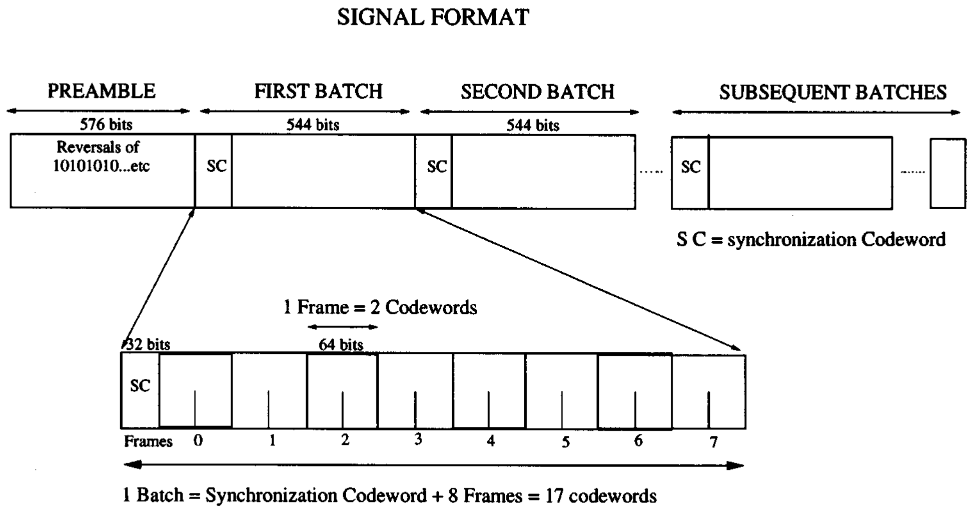

Figure 2: POCSAG transmitted paging signal

4.1.3 Batch Structure

A batch consists of 544 bits, broken down to the frame synchronization codeword (SC) followed by 8 frames each containing 2 address or message or idle codewords, or any combination of the three, for a total of 17 codewords. The frame synchronization is a unique reserved codeword recognized by the receivers as the initialization of a new batch. Frames are numbered from 0 to 7 and the pager population is similarly divided into 8 groups. Thus each pager is allocated to one of the 8 frames according to the last 3 bits of its 23 bit identity, i.e. 000=frame 0, 001=frame 1, 111=frame 7, and only needs to examine the address codewords in that frame. This frame structure within a batch multiplies the address possibilities of each codeword by 8 and most importantly provides the pager receiver with a simple battery saving technique.

The data message always begins with an address codeword. Message codewords flow immediately after and may be transmitted in any number of consecutive frames directly after the address codeword. The message can use any number of codewords, embracing one or more batches but without displacing the synchronization codeword. Message termination is indicated by a new address codeword or an idle codeword and not by a priori length information. This way full channel utilization is achieved, to be directed to a particular subscriber if necessary, even though it may block addressing in some of the succeeding frames. Wherever there is no meaningful codeword to be transmitted in a batch, an idle codeword is used.

At a rate of 512 bps each batch is only 1.0625 seconds in duration and transmission can then be terminated allowing for time-slotted multi-transmitter systems to be engineered easily.

4.1.4 Codewords

All codewords contain 32 bits which are transmitted with the most significant bit first.

- The synchronization codeword has a very low correlation with the preamble. Its first 31 bits are a pseudo random sequence which can be generated from a 5-stage feedback shift register.
- An address codeword has its first bit (the flag bit) as a zero to distinguish it from a message codeword. The bits 2-19 are address bits corresponding to the 18 most significant bits of the 23 bit identity assigned to the pager (recall that the last 3 bits correspond to the frame assignment). The pager looks at these bits to find its unique address. Bits 20 and 21 are the two functions bits which are used to select the required address from a total of four assigned to the pager in each frame. The total number of addresses is $2^{20} * 8$ (number of frames) ≈ 8 million. Bits 22 to 31 are the parity chick bits and the final bit (bit 32) is chosen to give even parity.
- The message codeword always starts with a 1 and the whole message always follows directly after the address codeword. The framing rules of the code format do not apply until the message is terminated, however although messages may continue to the next batch, the normal batch structure is maintained (16 codewords and the sync). After the end of the message any waiting address-codewords get transmitted with the first appropriately assigned to the first frame or half frame. Message codewords have 20 bits, bits 2 to 21, followed by the parity check bits and even parity bit.
- The idle codeword is a unique reserved codeword used to complete a frame in the absence of an address or message codeword.

4.1.5 Generation of Codewords

POCSAG uses a (31:21) BCH + Parity, cyclic block code. The 21 bits of each codeword correspond to the coefficients of a polynomial having terms from x^{30} down to x^{10}. The polynomial is divided, modulo-2, by the generating polynomial $x^{10} + x^9 + x^8 + x^6 + x^5 + x^3 + 1$. The check bits correspond to the coefficients of the terms from x^9 to x^0 in the remainder polynomial found at the completion

			Function bits	Function bits		Even Parity Bit
Bit number	1	2 to 19	20	21	22 to 31	32
Address Codeword	0	Address Bits			Parity check Bits	
Message Codeword	1	Message Bits			Parity check Bits	

Figure 3: Codeword format

of this division. The complete block of information and check bits, corresponds to the coefficients of a polynomial which is integrally divisible in modulo-2 fashion by the generating polynomial. An additional bit is added to the 31 bits to provide the even bit parity check of the whole codeword. This codeword generation method results in a Hamming distance of 6. Several error control deciding algorithms can be used by the receiver:

Using "hard decision" decoding:

- Any 5 bit errors or any one error burst of length not to exceed 11 bits, can be detected.
- Any single bit error can be corrected and any 4 bit errors or any error burst of length not to exceed 7 bits, can be detected.
- Any 2 errors can be corrected and any 3 errors detected.
- Every single burst of length not exceeding 4 bits can be corrected.

Using "soft decision" decoding:

- Check and correct any burst of poor quality bits up to 11 bits.
- Check and correct any 5 poor quality bits.

4.1.6 Message Formats

Numeric-only messages use 4 bits per character to represent the decimal numerals and 6 other special characters such as spaces, hyphens etc. The bits are transmitted in numerical order and there are usually 5 characters per message codeword. The address for numeric-only messages has its function bits set to 00.

Alphanumeric or general data format can be used for messages requiring a greater range of characters. They use 7 bits per character using the CCITT alphabet No 5 while the complete message is partitioned into 20 bit blocks. The function bits of the addresses for these types of messages are set to 11.

4.2 Motorola's FLEX Protocol

4.2.1 General

Motorola developed and released in 1993 a family of multi-high speed paging protocols transmitting in speeds of 1600, 3200, and 6400 bps. FLEX pagers can operate at any of the three speeds and can automatically detect the correct speed based on information from the paging signal.

The higher speeds enable FLEX to accommodate a higher number of users than POCSAG, while providing even longer battery saving times. The FLEX format supports the delivery of tone only, numeric, alphanumeric and binary data to remote receivers. Paging data is transmitted in fixed size batches which utilize a data interleaving scheme, providing good protection against burst errors. Error protection is also enhanced by embedded checksum information within the transmitted data. Consecutive interleaved blocks are grouped to form a frame and sequences of frames are transmitted cyclically each hour. Unlike POCSAG, FLEX is completely time synchronized, transmitting frames at very specific time intervals.

FLEX uses 2-level FSK as POCSAG does, for the slowest transmission of 1600 bps, while the 6400 bps uses 4-level FSK where each two-bit sequence (symbol) is represented with a different frequency level at -/+1600 and -/+4800 Hz.

4.2.2 FLEX Cycle

The FLEX protocol is a synchronous code that assigns pagers to frames. These frames occur at a rate of 128 every 4 minutes. A single FLEX cycle consists of 128 frames, so there are 15 FLEX cycles/hour. Each pager is assigned a particular frame (or frames) within a FLEX cycle. After "power on" the pager determines the time of its pre-assigned frame and only turns on its receiving circuitry during that time, without the need of a preamble detection as done in POCSAG. Frame assignments are configured within the pager.

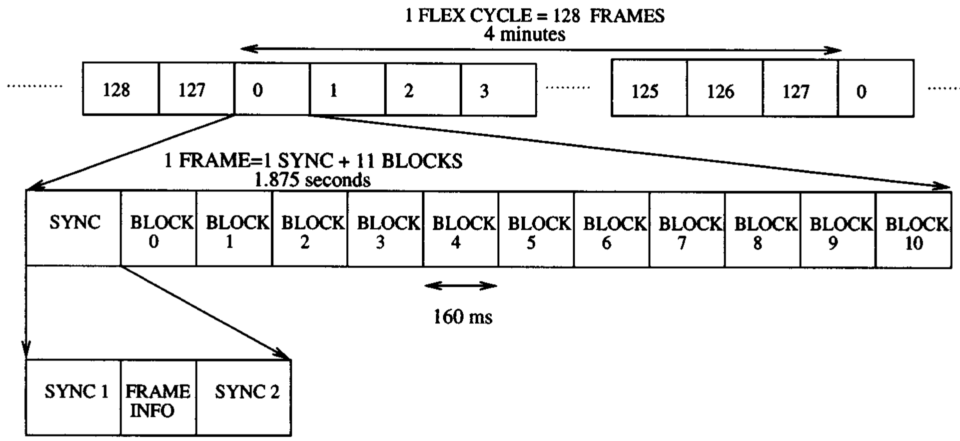

Figure 4: FLEX Paging Protocol

4.2.3 FLEX Frame

A FLEX frame is 1.875 seconds long and consists of a sync field and 11 blocks, each of time length of 160ms. The sync information takes 115 ms to transmit and is composed of three parts: sync 1, frame information and sync 2. The first two parts are always transmitted at 1600 bps using 2-level FSK. Sync 1 starts each frame to provide initial timing synchronization and to indicate the speed of the rest of the frame. The frame information contains the frame number (0-127), the cycle number (0-14), a roaming support bit, a repeat paging bit, a light traffic indication, and a 4 bit checksum. Sync 2 provides synchronization at the frame's block speed to allow the rest of the frame to be decoded correctly from the receiver [11]. The information being transmitted within a FLEX frame is divided into fields as shown in figure 5. There is the block information field(BI), the address field(AF), the vector field(VF), the message field(MF) and the idle blocks(IB). As shown in the figure the fields are not constrained by block boundaries and can arbitrarily cover any number or fraction of blocks. The BI field is typically one word long and indicates where in the frame the AF and the VF start, how many priority ("urgent") addresses will be placed at the beginning of the field and other information regarding the frame. The address field contains the addresses of the pagers to be paged during the frame time. The vector field has a one to one correlation with the AF so the pager knows where to look for the vector. The vector points to the start word of the message, in order for the pager to know where the designated message begins. The use of the vectors decouples the message from the address allowing subscriber units to share the same message, without having to repeat the message and thus increasing the effective channel bandwidth. The message field simply contains the messages and the unused blocks are filled with bits representing idle-alternating ones and zeros.

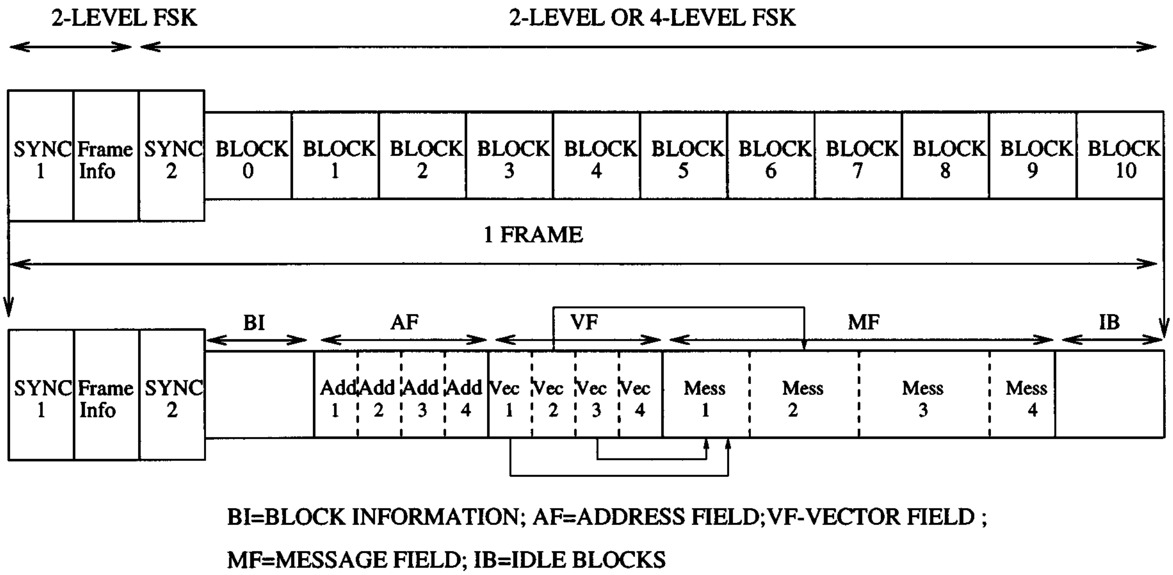

Figure 5: FLEX Frame Information

4.2.4 FLEX Data Integrity

Since the paging channel is heavily subjected to fading and time-variant multipath propagation due to simulcasting, a block interleaver for the coded data is necessary to effectively deal with bursty errors. A block interleaver formats the bits into rows of 32 bits(one codeword); 8 rows for 1600 bps (figure 6), 16 rows for 3200 bps and 32 rows for 6400. The bits are read column-wise and transmitted over the channel. At the receiver, the de-interleaver stores the data in the same rectangular array format, and reads them row-wise to reconstruct the message. Codewords are constructed using the BCH (31,21) format with even parity having a 2-bit error correction capability. Thus, the protocol itself can can be used to correct any burst error for 10ms. (16 bits for 1600bps, 32 for 3200bps and 64 for 6400bps).

4.2.5 FLEX Flexibility

FLEX provides a range of operating modes which enable the paging terminal to optimally schedule the transmitting pages on queue, based upon traffic loading conditions. Long messages may be

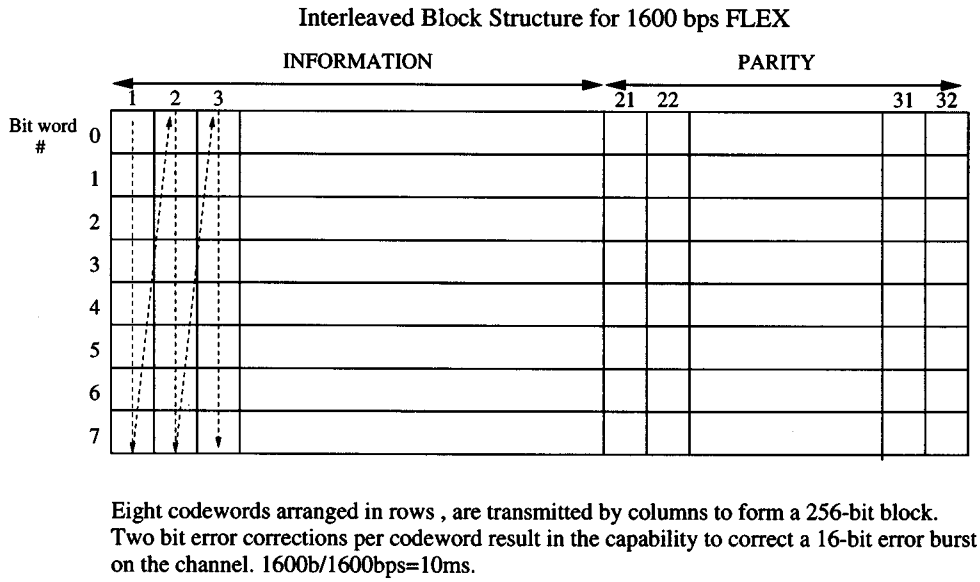

Figure 6: FLEX's block interleave structure

partitioned into shorter segments to be reassembled at the receiver. Thus shorter messages can be transmitted in between larger messages, utilizing the channel more effectively. For the channel rate of 6400 bps, (in 4-level FSK 3200 symbols per second) users are grouped into one of four "phases", a, b, c, d with each user having knowledge of its phase. The phase of a codeword is given by its row number modulo 4.

4.3 ERMES

4.3.1 General

The European Radio Message System (ERMES) is a paging protocol defined by the ETSI as the pan-European paging standard. ERMES uses a common frequency range in all European countries, allowing subscribers to roam throughout Europe and receive pages. ERMES uses 4-level PAM/FM modulation to transmit data at 6250 bps. In addition, the ERMES air interface I1 operates with 16 25 kHz channels in the radio band between 169.4125-169.8125 MHz with the pagers being able to scan all 16 channels for the best frequency for optimum reception. This allows the network operator to change system operating frequencies as necessary to meet capacity requirements without having to modify pagers.

4.3.2 ERMES signal structure

ERMES partitions every hour into 60 cycles, as shown in figure 7. Every cycle is partitioned into five sequences of 12, and every subsequence is divided into 16 batches, labeled A through P. The pager population is divided into 16 groups, each being allocated to one of the 16 batches. The batch number, subsequence number, and cycle number of each transmission is encoded into the system information partition of each batch. Over the 16 different frequencies, the first batch to be sent in each subsequence is a different batch number. Batch A is the first batch of Channel 1, the sixteenth batch of Channel 2, the second batch of Channel 3, etc. This allows the pager to step through all the channels without losing any messages.

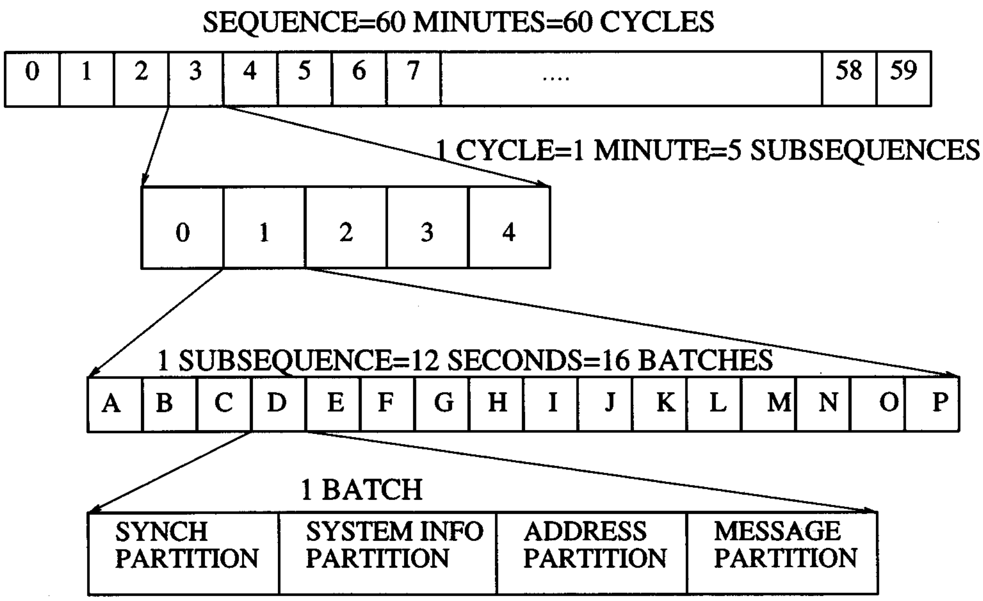

Figure 7: ERMES PAGING SIGNAL

Each pager is specified by a 35 bit address known as Radio Identity Code (RIC). It consists of 13 bits, specific to the "home system" database, and a 22 bit local address. A batch contains separation partitions of information known as the synchronization partition, system information partition, address partition and message partition.

Messages follow directly after the address partition, and are preceded by a 36 bit header. The header contains the 22 bit local pager address, information about what type of message (alphanumeric, tone-only, numeric or data page), and a message "ID" number. If the pager finds that the next message number is not in order it can notify the user that a message has been missed. In this case, ERMES provides a message retrieving feature by remote calling.

Like the FLEX protocol the bits of the system, address and message partitions in a batch, are encoded into block codewords. However contrary to FLEX, only the message partition gets transmitted in an interleaving manner. Every nine message codewords are grouped into a codebook and transmitted in an interleaving fashion. Each message codeword is 30 bits long, including 18 information bits and 12 parity bits allowing for similar error correction capabilities like FLEX.

4.4 Two Way Paging Protocols-MOTOROLA'S ReFLEX and InFLEXion

The most recent paging technology advancement is the introduction of two way paging technologies that became available after the FCC's approval of a 50 kHz spectrum for PCS applications. The pager and the base station function as transceivers. The pager receiver is enhanced with a battery efficient small transmitter, capable of sending acknowledgments to the base station's receiver to be forwarded anywhere in the network. Some of the two way protocols support autonomous registration on power-on or when the pager enters the coverage region of the whole paging network in a similar fashion to the GSM structure. This way the paging control system can determine when and where to send any waiting messages to the pager as soon as it becomes aware of its presence.

ReFLEX is a strictly proprietary, two way messaging protocol designed by Motorola and using FM for message transmission. It pairs the FLEX one-way channel with a reverse channel allowing pagers to transmit their responses on different frequencies than the ones used in the forward channel. The ReFLEX protocol operates on a 25 or 50 kHz outbound channel. The wide channels provide higher data rates and customer capacity at higher bandwidth. ReFLEX is a completely time synchronized protocol to ensure proper operation. However the time response periods may be dynamically changed or assigned based upon current loading conditions of the channel. There are two versions of ReFLEX available today; ReFLEX 25 and ReFLEX 50. ReFLEX 50 combines four 12.5 kHz forward FLEX channels (one master and three slave channels) to offer a combined data rate of 25.6 kbps. The ReFLEX 25 protocol allows on the forward 50 kHz channel two RF carriers transmitting at 250 Watts out of the base stations. This design is more robust than the ReFLEX 50 because it allows guard bands between the carriers. The inbound channel is 12.5 kHz and can support speeds from 800 bps up to 9,600 bps. The speed is a function of the number of receivers deployed in each transmitter's coverage area. ReFLEX uses 4-level FSK modulation with Reed Solomon (RS) coding, digitally engineered simulcast techniques and linear base stations to reduce inter-modulation products at varying frequencies and high output power. The pager and the base stations can use re-transmission schemes as well as store and forward mechanisms.

A new protocol, InFLEXion, is being developed by Motorola to create a voice and data messaging network with frequency reuse capabilities similar to those of cellular telephony. The protocol is based on ReFLEX's capability to request the pager to respond to a location request combined with directed transmission of pages within specific geographic areas. Voice messaging is then sent to the nearest transmitter(s) to the pager over the 50 kHz channels through variable compression techniques allowing 24 times more subscriber capacity of today's 25kHz channels. Simultaneously, other voice messages may be delivered on the same frequency to other pagers that are geographically separated to avoid channel interference. With InFLEXion data signaling has a throughput on the 50 KHz channel up to 112 Kbps. It is expected that Motorola will be using both voice compression and a complex modulation scheme like Quadrature Amplitude Modulation (QAM), to sustain the high bit rates specified.

5 Battery Savings achieved by the Signal Protocols

Extended battery life is very crucial to the development of the paging industry, since maximizing battery life, means lowering operating costs, fewer battery changes and availability of smaller pagers for the subscriber.

Pagers are powered by a simple AA or AAA battery. For the battery life to be approximately 2 months the average load must not exceed 4ma. In order to achieve this low power design several technologies are applied in the signal processing circuits of the pager device. However the most crucial and effective battery savings are achieved simply by the actual implementation of the paging signaling protocol. The pager employs battery-saving techniques to conserve power by periodically switching the pager into low power mode. Appropriate circuitry to "wake-up" the pager is used and depending on the paging signaling protocol, the pager only looks for paging information during a particular point in time (Motorola-FLEX and HERMES), or the actual signal being detected, is what "wakes up" the pager(POCSAG). Achieving longer battery lives is a crucial requirement that every signaling protocol aims to satisfy [12].

POCSAG provides a battery saving capability through the partition of its fundamental unit of transmission, the "Cycle", into the preamble and the frames. Subscriber units have their receivers turned off during battery saving mode and are required periodically to turn on and listen for a few milliseconds for the preamble tone. The pagers uses the preamble as a "wake-up"call and in case the data received is not a POCSAG signal they go back to battery saving mode for about 1 second. Operation between the subscriber unit and the paging terminal is initially asynchronous until the detection of the preamble signal and synchronous thereafter. Recall that each pager device has a pre-assigned frame determined by the last bits of its 23 bit identity. After detecting the preamble the receiver needs to be turned on only during the synchronization codeword and its particular frame for a total of 3 codewords per batch . Since there are a total of 17 codewords in a single batch, the energy requirement is reduced to about $\frac{3}{17}$ of that of constant reception, with the receiver being off 82% of the batch time. Motorola holds a patent which uses a two-bit error correction scheme to shut down the receiver after only 8 bits of its particular frame rather than 32, which increases the battery-save time of the receiver even more [12].

The FLEX protocol is completely synchronized allowing the paging receivers to have pre-assigned frames within the FLEX cycle to receive data destined for them. All addresses are grouped together in the beginning of the frame so subscriber units are able to quickly determine if there are any messages pending for them by checking only the address field of the frame, otherwise they enter the battery-save mode. A special feature of FLEX, called "Frame Collapse", is utilized to increase the sleep time of the pager receiver, by ignoring some of the bits within a the frame and still capturing all the information. A frame collapse of two will half-size the channel bandwidth and allow scarce paging channels to be multiplexed with other paging protocols, such as ERMES.

ERMES, similarly to FLEX, uses batch transmission techniques to increase the battery life of the pager by assigning pagers to one of 16 batches. To increase battery saving even more, pagers may not only be assigned a specific batch, but can be programmed to be paged in a particular subsequence or even cycle. Again the synchronized nature of ERMES allows the receiver to know when the particular batch is being transmitted to look for any data. In addition , within the address partition of the batch, each address is sorted in decreasing order according to the first 18 bit of the address codeword. This important innovation of ERMES allows the pagers to quickly determine if there unique address is included in the batch or not. If the initial address is larger than their own they can immediately enter battery saving mode.

6 Fading and Simulcast

6.1 Small-scale fading and multipath

As in every radio propagation model, the transmission path between the transmitter and the receiver experiences small scale fading because of the multipath nature of the signal due to reflection, diffraction and scattering, and the characteristics of the narrowband noisy channel. Fading occurs simply because of the interference between two or more versions of the transmitted signal that arrive at the receiver at different times. Depending on the signal parameters (such as signal bandwidth, symbol period, etc.) and the channel parameters (such as channel bandwidth,rms delay and doppler spread) there are two main fading types. The first is based on multipath time delay spread and the other based on the doppler spread [5].

As explained in [13] the mobile radio channel can be modeled as a linear time varying channel, where the channel changes with time and relative to the distance between transmitter and receiver. The fading effects related to the coherence bandwidth are based on the multipath time delay spread and cause the transmitted signal to experience either flat or frequency selective fading. Flat fading occurs if the bandwidth of the signal is much smaller than the bandwidth of the channel and the symbol period is greater than the rms delay spread (a statistic that describes the variations of channel time delay values). In flat fading the spectral characteristics are preserved at the receiver, but the strength of the received signal will still vary with time since the channel gain varies in time due to the multipath. Flat fading channels are also known as amplitude varying channels with an amplitude distribution usually characterized by a Rayleigh distribution.

If the channel has a constant gain and linear phase response over a bandwidth smaller than the signal bandwidth, then the received signal experiences frequency selective fading. In this case the received signal constitutes of multiple attenuated versions delayed in time. This happens because in the frequency domain the gain is different for different frequency components. Frequency selective fading is caused by multipath delays which approach or exceed the transmitted symbol period.

Fading effects due to doppler spread occur because of the time variations in the channel's response. They are realized through movements of the mobile and variations in the radio channel environment. Depending on how fast the transmitted signal changes compared to the rate of change of the channel, there can be fast fading and slow fading. Fast fading only deals with the rate of change of the channel due to motion. In this case the channel impulse response changes rapidly within the transmitted symbol duration which causes frequency dispersion due to the doppler spreading, leading to signal distortion. In slow fading the channel impulse response changes at a much slower rate than the signal. The doppler spread of the channel is much less since the channel can be considered to be static.

6.2 Simulcast for wide-area paging

In order for the paging system to achieve a wide-area coverage and compensate for the receiver's small and inefficient antenna, relatively high field strengths must be produced from an appreciable number of widely separated transmitters. In addition paging systems use the VHF and UHF bands that are highly susceptible to obstruction losses due to hills, trees and buildings which create shadowing. The method used to overcome these problems is known as quasi synchronous operation or simulcast. Simulcast is a reliable method to achieve wide-area coverage by transmitting the same paging signal from multiple paging transmitters at approximately equal times. In simulcasting, the relative carrier frequencies, the relative audio frequency phases and the relative audio levels of the various paging transmitters should be kept within tight tolerances. Ultra-stable oscillators should keep the difference in the carrier frequencies below specified limits and be regularly re-synchronized for proper simulcast operation. Transmitters are separated by different distances so when the signal is sent out from the base controller to multiple transmitters it is necessary to add additional delays to the transmission path to ensure proper synchronization. Audio equalizers which are variable time delay elements are used at the base stations to perform the synchronization.

Although the paging system can be appropriately designed to artificially add delays to compensate for the difference in land lines and radio links feeding the transmitter as well as differences due to equipment delays, it can not compensate for propagation differences experienced at the receiver due to the position and mobility of the pager. Furthermore most often no line of-sight exists between transmitters and pagers, so reception relies purely on reflections. Consequently the several paths of propagation from a unique transmitter and others participating in the simulcast, can be regarded as a multipath environment producing the familiar fading effects at the receiver. In the simulcast paging environment fades occur between delayed versions of the same signal that arrive within roughly 10dB to 15dB of each other and may be as much as 20dB or even 40 dB weaker than the local mean [14]. In addition, for the Rayleigh fading channel the error rate decreases only inversely with the SNR, in contrast to the non-fading channel which decreases exponentially. This means that transmitters must transmit a very large amount of power to assure a low probability of error, however a large amount of power is not possible for technical and economical reasons. To deal with the undesirable effects of multipath nature the paging system implements FSK modulation with non-coherent detection at the receiver since it not possible to establish a phase reference for the multipath signals. FSK operation receivers are less sensitive to the resulting strength errors since digital rate signals(especially at lower rates) are much easier to synchronize than direct analogue signals. Provided the equalizers and synchronizers of the receiver are optimally designed under the low power constraint, several alternative solutions can deal with the undesirable effects of fading, by means of code interleaving, frequency offset [15] and time diversity techniques [16], [17].

Code interleaving was described earlier as a reliable transmission method used in FLEX and ERMES, that effectively deals with bursty errors due to fading. By interleaving the coded data, the bursty channel is transformed into a channel having independent errors, since the waveform corresponding to the bits or symbols of the codeword fade independently. In FLEX the interleave period allows for error correction capability for error bursts up to 10ms and in ERMES up to 2.88ms [14].

Decoders of simulcast systems at the receiver generally can tolerate a misalignment between signals of equal energy and frequency of about one-fourth of the symbol time. For POCSAG at 2400 bit/sec this corresponds to 104 μs and for FLEX at 3200 bits/s 2-level FSK it is 78 μs while for the 4-level FSK it is 156 μs. Notice that going from a 2-level FSK to a 4-level FSK modulation scheme, for a given rate, doubles the symbol duration width of the transmitted signal and consequently the allowable delay spread. Finally for the 4-level FLEX at 6400 bits/s this will be 78 μs as in the 3200 2-level FLEX [18]. These values have been determined experimentally through performance mapping using signal analyzers with built-in GPS receivers [19] and are typical for doppler frequencies of $f_d = 1Hz$. Increasing the doppler frequency will reduce the maximum allowable delay.

Diversity techniques are based on the notion that errors occur in the receiver when the channel attenuation is large during deep fades. By supplying the receiver with a replica of the transmitted signal over independently fading channels the probability that all the signal components will simultaneously fade is reduced. If p is the probability that only one of the K independently fading replicas fades below some critical value, then p^K is the probability that they all fade. One method currently used in Japan incorporated in the traditional FLEX protocol, is FLEX-TD(Time Diversity FLEX) that transmits the modified FLEX paging signal in K different time slots, where $K = 1$ to 4, and the separation between successive time slots is greater or equal to the rms delay spread of the channel. FLEX-TD shares many of the characteristics of the FLEX protocol regarding data transmission rate and signal format. The only modification is that within the frame structure (see section 4.2.3) blocks are organized in K subframes, where K is the number of replicas. However within a frame, the K subframes are not replicas of the same signal, but instead an interleave scheme is employed in terms of C frames to randomize even more the burst errors, i.e. for $K = 4$, during one cycle, the original signal will be transmitted in one of the subframes in frame 0, the first replica will be transmitted in a subframe of frame C, the second replica in a subframe of frame $2C$ and so on while the last replica will be in frame $4C$. This way the receiver also knows in which frame to look for the replicas after receiving the original signal, eliminating the need for transmission of the replica's position. Low correlation of the two fading signals is necessary to ensure diversity improvement and the minimum value of C is the duration of a frame , 1.875 s. Thus time diversion is effective even in slow fading environment where bit interleaving is not effective. Two methods are available for receiving the repeatedly FLEX-TD signals, one is called "Bit Combining Time Diversity"(BC-TD) and the other is "Code Word Combining Time Diversity"(CWC-TD). BC-TD combines the demodulator's output signals bit by bit, while CWC-TD combines the decision circuit output bit stream, code-word by code-word. In BC-TD the bit combination and construction can be done by one of two methods. One is equal combining and the other is weighted combining which assigns a weight to each received signal to secure the condition that each received codeword has identical average SNR. The receiver may provide the appropriate bias term through the weights after estimating the average signal power and evaluating the actual received power from the RSSI (Received Signal Strength Indicator)

$$S_{equal}(i) = \sum_{j=1}^{m} b_{ij} \qquad S_{weighted}(i) = \sum_{j=1}^{m} w_{ij} b_{ij} \tag{1}$$

where b_{ij} is the signal to be combined and w_{ij} is the weight. Simulations [16], [17],have shown that in terms of Bit Error Rate (BER) BC-TD outperforms CWC-TD for almost all SNR levels (especially the weighted combining), but requires more pager memory. Both Time-Diversity techniques have a superior performance in terms of BER for a given SNR compared to POCSAG 2400.

Finally another method that is implemented in the simulcast transmission environment is frequency offset transmission. As seen, simulcast creates overlapping areas where equal energy transmitted signals having the same frequency and variations in phase may cause the received power to deteriorate. To avoid the simulcast delay spread the signals can be transmitted by the different base stations in slightly variant frequencies having a small frequency offset among them. Different deviation offsets should be programmed by more than two transmitters because large frequency offsets can have a negative impact on the system due to zero beating. Zero beating occurs as a phase cancelation between only two transmitted signals when the frequency offset Δf is equal to some peculiar value and these signal cancelations occur periodically at a rate of $\frac{1}{\Delta f}$. Therefore the larger the offset the deeper and faster the fades are but with smaller duration. When the interval $\frac{1}{\Delta f}$ corresponds to the multiple of the interleave period of the transmitting code, the errors caused by the fading are concentrated for a specific codeword. These errors are impossible to correct by the BCH correction scheme, therefore the frequency offset should be very carefully designed for different transmission rates of for each protocol. For instance simulations and experiments have found that for FLEX-TD at 6400 bps frequency offsets below the ± 150 Hz should be specified while for POCSAG at 2400 bps offsets can range from 200 to 1000 Hz [15], [18].

7 Conclusion

The demand for low-cost, efficient wireless messaging services, continues to increase. Digital pagers in their two most popular forms, Numeric Display, and Alphanumeric Display, continue to be the most cost-effective way to communicate with people on the move. Main reasons are that:

- A pager costs less
- Better battery life
- Simulcasting is superior technology
- Better penetration into buildings
- Fewer missed messages
- Efficient use of radio resources
- Smaller, lighter weight unit

Pagers are now commonly used as screening devices in complement to cellular, while they will help extend the PCN's handset location capability beyond larger metropolitan areas. With the expansion of mobile users and pagers, signaling messages from/to the base stations (two-way pagers) will increase and so will the demand for mobile services. In the future paging will become a typical example of computer telephony integration. Several packages are already available to connect the internet to the paging networks and pagers using the two way technology will become sophisticated personal digital assistants. The paging network will become a platform for mobile computing. New designed efficient protocols should aim to increase the capacity of users, inter-connect large areas for global paging, transmit and receive at high speeds, incorporate sophisticated services such as two-way paging and minimize battery life while keeping the overall costs low.

References

[1] Y.b. Lin. Paging systems:network architectures and interfaces. *IEEE Network*, pages 56–61, July-August 1997.

[2] J. Moskowitz. Paging protocols. *http://www.rts-inc.com/rts/protocol.htm*, 1995.

[3] Gordon Whyte. *Mobile Radio Technology*. Newnes, UK, 1994.

[4] Motorola Inc. An explanation of paging. *http://www.mot.com/MIMS/MSPG/FLEX/*, 1996.

[5] T.S. Rappaport. *Wireless Communications. Principles and Practise.* Prentice Hall, NJ, 1996.

[6] L.I. Williams. System integration of the flex paging protocol.part 2:system design recommendations. *Mobile Radio Technology*, 14(7):12–18, July 1996.

[7] L. Stewart. Pager development, testing. *Mobile Radio Technology*, pages 22–28, May 1996.

[8] G. Stuber D. Kim and N. Hightower. Performance of simulcast systems in mobile radio environments. Phoenix, AZ, 1997. IEEE/VTC 1997.

[9] B. Dye. Recommendation 584 standard codes and formats for international radiopaging annex 1. radiopaging code no. 1. *http://village.ios.com/ braddye/rpc1.htm*, 1996.

[10] ETSI. Ermes-a guide to the ermes international paging standard. *ERMES MoU Sterering Group-The Guide*, 1996.

[11] L. Stewart. Paging technology: Flex, ermes paging codes. *Mobile Radio Technology*, pages 23–30, April 1996.

[12] B. Mangione-Smith. Low power communications protocols:paging and beyond. San Jose, CA, Oct 1995. IEEE Symposium on Low Power Electronics.

[13] John Proakis. *Digital Communications.* McGraw-Hill, 3d edition, 1995.

[14] R. Read. Radiopaging technology and systems for wide-area use. *Electrotechnology*, 5(6):20–24, Dec-Jan 1995.

[15] S. Souissi T. Sexton and C. Hill. Performance of high data rate paging in simulcast. Boulder, CO, 1997. Wireless Communications Conference.

[16] T. Sakai S. Ito, S. Ohkubo and Y. Yamao. Time diversity improvements in flex-td new generation paging systems. pages 1385–1389, Atlanta,GA, April 1996. Proc. VTC 96.

[17] M. Hasegawa K. Abe, M. Mimura and H. Katayama. Technologies for the particular propagation characteristics in the 4-level fsk paging system. Chiba, Japan, 1996. Proceedings of ISAP '96.

[18] L.I. Williams. System integration of the flex paging protocol.i system design constraints. *Mobile Radio Technology*, 14(6):23–30, June 1996.

[19] S. McFarland. Performance mapping identifies simulcast problems. *Mobile Radio Technology*, 14(5):58–68, May 1996.