Evolution of GSM into the Next Generation Wireless World

Prabhakar Chitrapu, Member, IEEE and Behrouz Aghili

Abstract—**GSM is the most widely deployed 2nd Generation digital cellular standard, with over 2 billion subscribers in some 213 countries and adding about 1000 new users per minute! Originally developed in the 1980s, and first deployed in 1991, GSM is a TDMA+FDMA system, providing wide area voice communications using 200 KHz carriers. Subsequently, GSM evolved into a 2.5G standard with the introduction of packet data transmission technology (GPRS) and higher data rates via higher order modulation schemes (EDGE). More recently, GERAN standards organization has been evolving further to coexist with and provide comparable services to 3G technologies. In this paper, we first provide an overview of the traditional 2G and 2.5G GSM, and then discuss the so-called "GERAN Evolution" into the 3G world. We conclude by detailing a few selected aspects of the GERAN Evolution. Overall, we aspire to demonstrate that GERAN is a vibrant, living and growing technology that exploits the latest advances in communications and signal processing.**

***Index Terms* — Land Mobile Radio Cellular Systems, Standards, Quadrature Amplitude Modulation, Diversity Methods.**

I. Introduction

GSM is the most widely deployed 2nd Generation digital cellular standard, with over 2 billion subscribers in some 213 countries and adding about 1000 new users per minute! Originally developed in the 1980s, and first deployed in 1991, GSM is a TDMA+FDMA system, providing wide area voice communications using 200 KHz carriers. Soon, the text messaging service, SMS, became hugely successful and now transmits over 1 billion messages per day. Subsequently, GSM evolved into 2.5G standards with the introduction of packet data transmission technology (GPRS) and higher data rates via higher order modulation schemes (EDGE). Together they are now referred to as GERAN, for GSM/EDGE based Radio Access Network. The theoretical peak data rate is 473.6 kbps and the practical highest data rate is approximately 200 kbps.

The advent of the new millennium saw the development of 3G technologies, based on Wideband CDMA. This standard offers much higher data rates and greater flexibility for supporting a variety of services. As 3G Networks and User Equipment is still relatively young, a number of deployment challenges remain unresolved. First, widespread commercial deployment of 3G is an expensive matter, since it entails the acquisition of new spectrum, and the development of entirely new Radio Access Networks and User Handsets. Second, because of market requirements and service compatibility issues, 3G Handsets will support 2G functionality as well. Finally, some level of continuity of services across 2G & 3G networks will be important for both the customers and operators.

Driven by the above factors, GERAN standards development organization has been engaged in a major evolution effort since 2005, in order to offer advanced radio interface capabilities in the context of emerging multi-RAT (Radio Access Technology) networks and multi-mode user equipment environment. These capabilities are expected to be mostly Mobile Station centric, with only software and minor (if any) hardware upgrades to the existing GERAN networks. This will offer cellular operators a less expensive alternative to 3G. Furthermore, since the new features are related to 3G features, implementation in Dual Mode (2G-3G) Handsets should be efficient and less expensive.

GERAN evolution is addressing both Uplink and Downlink enhancements. One of the new DL features is the usage at the Mobile Station of multiple antennas, which are exploited for Receive Diversity and interference cancellation. This is especially important as the cell sizes are becoming smaller and smaller, so that even though GERAN is a TDMA based system, interference from adjacent cells, rather than noise, is beginning to be the capacity limiting factor. Another DL enhancement feature is the allocation of multiple carriers in order to increase peak and mean data rates. Finally, Higher Order Modulation, HOM (16 and 32 QAM) and Turbo Coding are being investigated to increase the data rates and performance respectively. Similarly, HOM and Turbo Coding are being implemented in the Uplink as well.

This paper will provide a technical overview of the GERAN Evolution, expanding on the above initiatives. We will cover aspects such as link performance, spectral efficiencies, implementation complexities etc.

II. Overview of 2G&2.5G GERAN

GSM or GERAN as it came to be called in later years is a well known, well understood and well documented technology, with several textbooks dealing with the subject either in full or in part. Some useful references are [1,2,3]. Here we give a brief overview of the salient aspects.

Manuscript received March 6, 2007.

P. Chitrapu is with InterDigital Communications Corporation, 781 Third Avenue, King of Prussia, PA 19406, USA. (610-878-5730; fax: 61-878-7841; e-mail: prabhakar.chitrapu@interdigital.com).

B. Aghili is with InterDigital Communications Corporation, 2 Huntingdon Quadrangle, Melville, NY 11747. (e-mail: behrouz.aghili@interdigital.com).

A. *GSM Network Architecture*

The GSM system architecture is basically divided into two subsystems; the Network (NW) Switching System (NSS, also referred to as Core NW or CN), where the services are realized and the Base Station System (BSS, also called Radio Access NW or RAN), which gives the users access to the CN. The user terminal, which is called Mobile Station (MS), can support a range of various capabilities. Figure1 illustrates a simple example of a GSM NW, where the Base Transceiver Stations (BTS) and Base Station Controller (BSC) make up the BSS. The NSS is simply the totality of the switches and databases that exist in the NW. Here, we find the Mobile Services Switching Center (MSC), the Visitor Location Register (VLR), the Home Location Register (HLR) and the Authentication Center (AUC) as the most important nodes.

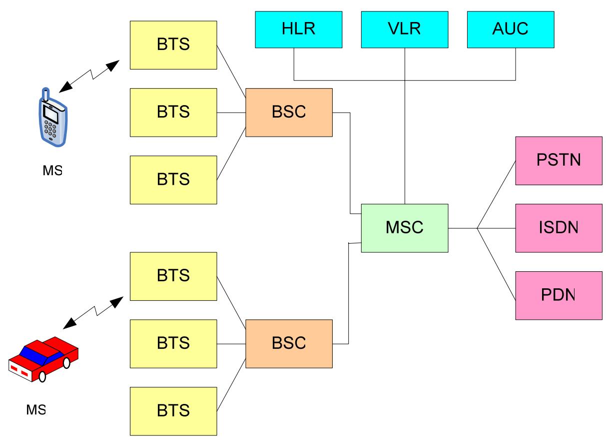

Figure-1: Physical Network Architecture [2]

B. *GSM Protocol Model*

Figure 2 below shows the protocol stack of GSM together with the entities in corresponding nodes. Focusing at the MS side, the Connection Management (CM) entity provides the signaling to setup basic voice and data calls as well as Short Message Service (SMS). The Mobility Management (MM) entity supports the means for registration and security functions such as user authentication, identification and confidentiality. However, in order for the MS to have a signaling connection with the core NW, a "dedicated" Radio Connection must be created first. The radio connection between the MS and the BSS is created by using the Radio Resource Management (RRM). All these protocols are realized as Layer 3 Signaling Protocols. The Layer 2 protocol (LAPDm), also generally recognized as the Data Link Layer, is mainly responsible for providing sequence control and error correction by means of retransmission.

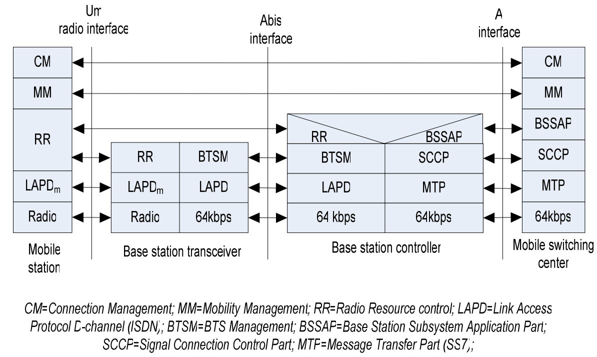

Figure-2: Protocol Architecture [2]

C. *GSM Radio Interface*

The GSM radio interface, shown in Figure-2, consists of a number of communication channels defined in terms of the radio spectrum. GSM is essentially a FDMA+TDMA system, operating with 200 KHz carriers in the frequency domain. Time is basically partitioned into a continuous synchronous sequence of 4.615 msec-long 'frames', each of which is further partitioned into 8 'timeslots'. Each time slot carries a 'burst' of modulated symbols, using the spectrally efficient GMSK modulation scheme, which carries 1 bit per symbol. The number of symbols per burst is 148 for a so-called 'normal' burst, which is used to carry 114 user data bits as well as training symbols and a small amount of signaling data. Adjacent bursts are separated by a guard time, with a duration equal to that of 8.25 symbols. This results in a symbol interval of 3.69usec and an average symbol rate of approximately 271 ksps. Starting with frames, multiframes (consisting of 26 or 51 frames), superframes and hyperframes are defined. Figure 3 depicts the various time scales described above.

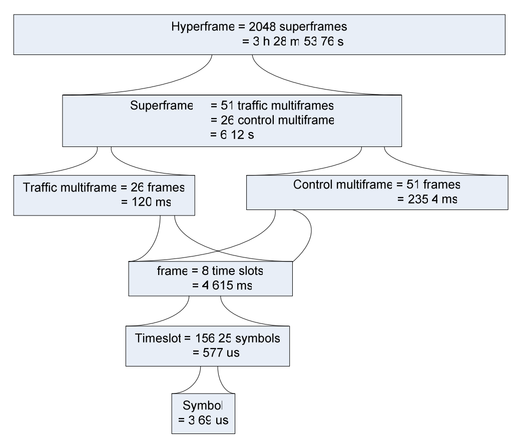

Figure-3: Various Time Scales in GSM [2]

A number of physical and logical channels are defined in terms of radio bursts. For example, a physical full rate channel consists of 1 timeslot per frame. Logical channels are characterized by the type of information carried over the actual physical channels. Some examples are Broadcast Channels (to facilitate cell selection, synchronization etc), Common Control Channels (to support Access and Assignment mechanisms), Stand-alone Dedicated Control Channels (to perform procedures such as location update etc), Traffic Channel (to transfer user data) and Associated Control Channels (to help maintain the radio connection, send handover messages, to report measurements etc).

User data, typically voice in 2G GSM systems, is protected against channel errors by a combination of Layer 1 and Layer 2 techniques. The Layer 1 techniques include Forward Error Correction Codes (mostly convolution codes) for correcting randomly distributed errors and Interleaving for randomizing bursty errors (which can then be corrected by FEC). The Layer 2 technique used is the ARQ (Automatic Repeat Request) method. The ARQ is mainly used for retransmission of entire radio bursts.

FEC Codes are characterized by the so-called channel code rate, which is the ratio of the data rate before coding to data rate after the coding. Often the coded data stream is 'punctured' by deleting some coded bits, so that the punctured data can be exactly mapped onto the assigned physical channel.

In GSM, a voice call is first source coded by the RPE-LTP (Regular Pulse Excitation – Long Term Prediction) speech codec, which compresses input speech to approximately 13 kbps, which is increased to about 22 kbps after FEC coding. This is the data rate supported by a full rate physical channel (114 data bits per 4.615 msec frame and 24 frames in a 26 frame multiframe), which means that a voice communication requires a Traffic Channel using a single full rate physical channel. The resulting voice capacity is 8 simultaneous voice calls for each 200 KHz carrier. Subsequent improvements in speech coding technology saw the introduction of the Half Rate Coder (doubling the voice capacity) and the Enhanced Full Rate Coder (producing higher perceptual speech quality without diminishing the voice capacity).

D. Packet Data Services and E-GPRS

Since GSM system was primarily designed for voice, a Circuit Switched architecture was followed even for data applications. In other words, dedicated physical channels were allocated for individual data applications. However, because the traffic of most data applications is "bursty" in nature, this was obviously not an optimized solution. Therefore, the GSM society introduced a Packet Switched transport service called General Packet Radio Service (GPRS) in 1997.

The GPRS is realized by maintaining and modifying the BSS as well as introducing a new type of CN based on IP (Internet Protocol) routing technology. The new nodes in the CN are called Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The data are routed as IP-packets within the CN and beyond to external Packet Data Networks (PDNs).

On the radio interface, GPRS uses the same physical channels as circuit switched GSM. However, some new logical channels have been defined to enable GPRS specific procedures. An example is the PDTCH, which uses a physical channel shared among several users. The basic radio unit is the so-called "Radio Block", which consists of 4 'consecutive' bursts. By allocating a different number of radio blocks to a user, the effective data rate can be varied, depending upon the demands of the data application at hand.

Further increases in data rates were made possible in 1999 by the standardization of 8PSK Modulation, in addition to GMSK. Since 8PSK carries 3 bits per symbol, theoretically the data rates can be tripled without affecting the symbol rate and hence the bandwidth. By way of terminology, GPRS enhanced with 8PSK modulation came to be referred to as Enhanced GPRS (E-GPRS) where EDGE stands for Enhanced Data rates for GSM Evolution.

Since, data applications, unlike voice communications, require variable data rates as well as channel error protection, a number of "Modulation and Coding Schemes" (MCS) have been defined in E-GPRS. The following table shows the schemes along with the raw data rates and user data rates that, respectively, do and do not include channel coding and signaling overhead.

Table-1: Data Rates in E-GPRS

Scheme (Family)	Modulation	Channel Coding Rate	User Data Rate (kbps)
MCS-1 (C)	GMSK	0.53	8.8
MCS-2 (B)	GMSK	0.66	11.2
MCS-3 (A)	GMSK	0.85	13.6/14.8
MCS-4 (C)	GMSK	1.0	17.6
MCS-5 (B)	8PSK	0.37	22.4
MCS-6 (A)	8PSK	0.49	27.2/29.6
MCS-7 (B)	8PSK	0.76	44.8
MCS-8 (A)	8PSK	0.92	54.4
MCS-9 (A)	8PSK	1.0	59.2

The above date rates assume that a single timeslot is allocated per frame. If all the 8 timeslots were to be allocated to the same data stream, clearly the maximum theoretical data rate would be 473.6kbps! In reality, however, the maximum data rates are lower reaching up to some 200 kbps, limited by the number of timeslots that can be allocated to a user and processed by a mobile.

In addition to ARQ techniques, error protection at Layer 2 is enhanced through the introduction of Incremental Redundancy (IR) and Link Adaptation (LA). In the former, the incorrectly received packets are retransmitted using a different puncturing pattern, which are 'soft-combined' with the previously sent packet. In Link Adaptation, the Modulation and Coding

Scheme is changed depending upon the channel conditions. If the change happens within the same ‘Family’ of MCS schemes (as indicated in Table 1), then data packets (at the RLC level) are guaranteed to be commensurate, in the sense that they allow integer partitioning or combining.

The allocation of packet data resources is orchestrated by the MAC (Medium Access Control) protocol, which also ensures packet data sequence control and error correction. The MAC protocol allows for static as well as dynamic allocation on a per radio block basis. The GPRS Mobility Management (GMM) protocol mainly acts like the MM protocol and the Session Management (SM) protocol is used for IP-address allocation as well as QoS negotiation.

III. Overview of GERAN Evolution

The late 90’s and the 2000’s saw an explosive growth in mobile wireless communication technologies, especially in terms of new and more capable radio interfaces and a multitude of new usage scenarios. In the wide area coverage cellular scenarios, the ITU (International Telecommunications Union) initiated the IMT2000 project, with a hope for a truly global radio interface for the new millennium. This effort resulted in two 3rd generation radio interfaces, developed by the Standards Development Organizations (SDOs) 3GPP and 3GPP2, with the former being based on a 5 MHz WCDMA (Wideband Code Division Multiple Access) technology and the latter being based on Multi Carrier extension of the 2G 1.25 MHz CDMA radio interface. (The radio access network in 3GPP is referred to as UTRAN). Both these radio interfaces are capable of delivering data rates up to several Mbps and are well suited for data intensive applications, including streaming & interactive multimedia communications. In the scenario of local area data oriented services, IEEE developed the WLAN (Wireless Local Area Network) radio interfaces, based on OFDM (Orthogonal Frequency Division Multiplexing) technology and providing up to 54 Mbps. Similar developments took place in the area of Personal Area Wireless Services for Cable Replacement applications (e.g. Bluetooth), Wireless Gaming devices (e.g. ZigBee), Warehouse & Asset Tracking applications (RFID) etc.

In light of such unprecedented developments, one may get an impression that 2G & 2.5G radio technologies, such as GERAN, may become gradually obsolete. However, this is clearly not the case when one considers that GERAN has over 2 billion users, at the time of writing, spread all across the globe, so that the GERAN technology is expected to last for several more decades to come. Migration to newer 3G technologies is expensive from an operators’ point of view, not only because of the new networks that need to be built, but also due the spectrum licenses needed to deploy these new radio technologies. Recent history of spectrum regulation and licensing has shown that the latter aspect can be extremely expensive and therefore prohibitive to operators across the globe in general. This calls for continued enhancement of 2G & 2.5G networks.

Another important point to keep in mind is that as new 3G radio networks are being deployed, they will be coexisting with the 2G & 2.5G networks. This means that the level of service provided should remain comparable as users move across the networks, justifying the need to evolve 2G&2.5G services.

Finally, current and future handsets will most likely be multi-mode devices, with 3G capabilities. This suggests that enhancing 2G&2.5G capabilities to be on par with 3G capabilities will be feasible with only a marginal increase in the overall dual mode device complexity.

Driven by such motivations, the GERAN standardization organization (namely 3GPP, www.3gpp.org) has been continuing to enhance the capabilities of GERAN both at the Physical Layer as well as higher layers. In this section, we will give an overview of these efforts. We will see that some of these enhancements have already been standardized (in Release 6 of GERAN) and even implemented, while others are being actively standardized (Work Items for Release 7&8 of GERAN), while yet others are being studied for standardization (Study Items for Releases 7&8 of GERAN).

A. Physical Layer Enhancements

1) Flexible Layer One (FLO): In 2.5G GERAN, optimized logical channels for real-time services were standardized and specified in terms of specific channel coding, puncturing and interleaving. While these were optimal for the particular service, they were not flexible, in the sense that with the introduction of each new service, new coding schemes had to be developed and standardized. This problem was likely to grow as a number of new IP based Multimedia Services (IMS) were being defined. A solution was to design a flexible physical layer (i.e. layer-1), which provided a framework in which optimized channel coding schemes could be specified at the time of call setup. Fortunately, this problem was solved in 3GPP UTRAN via the concept of Transport Channels (TrCH), so that GERAN simplified and adapted the concept. The result was a so-called Flexible Layer One (FLO), specified in [4] and described in [3].

Essentially, with FLO, the physical layer offers Transport Channels (TrCH) instead of logical channels to the MAC layer. A transport channel is used to transmit one data flow with a given QoS over the radio interface. A number of transport channels can be active at the same time and multiplexed at the physical layer.

The configuration of a transport channel i.e. the number of input bits, channel coding, interleaving etc. is denoted by the Transport Format (TF). As in UTRAN, a number of different transport formats can be associated to one transport channel.

The configuration of the transport formats is completely controlled by the RAN and signaled to the MS at call setup.

On transport channels, transport blocks (TB) are exchanged between the MAC layer and the physical layer on a transmission time interval (TTI) basis. For each TTI a transport format is chosen and indicated through the transport format indicator (TFIN).

Only a limited number of combinations of the transport formats of the different TrCHs are allowed. A valid combination is called a Transport Format Combination (TFC). The set of valid TFCs on a basic physical channel is called the Transport Format Combination Set (TFCS).

In order to decode the received sequence the receiver needs to know the active TFC for a radio packet. This information is transmitted in the Transport Format Combination Indicator (TFCI) field. This field is basically a layer 1 header, which is decoded first and is used to decode the payload data. Figure below illustrates the concept of FLO.

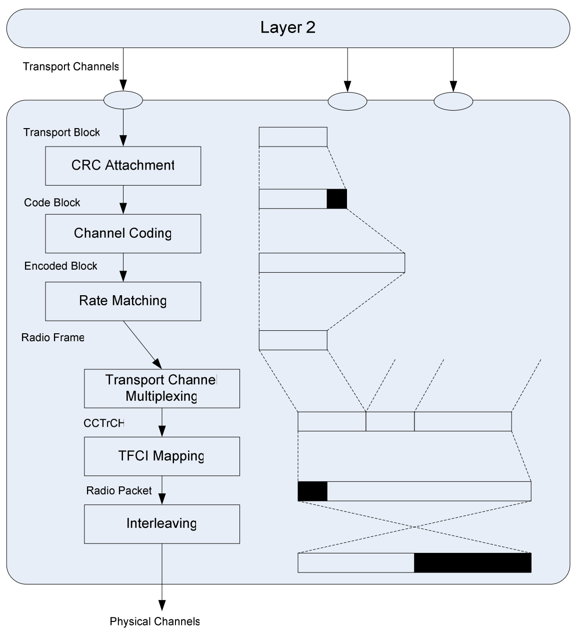

Figure-4: Concept of Flexible Layer One [3]

2) Downlink Advanced Receiver Performance (DARP): As the cell sizes decreased to accommodate increasing traffic load, co-channel interference becomes the limiting factor, as opposed to additive random noise. Especially in the downlink, MSs are affected by the interference from one or more base stations using the same frequency and timeslot. Since the characteristics of the interfering signals are known in general terms (such as modulation types, training sequences etc), it is possible to develop interference cancellation techniques, leading to a class of receiver algorithms referred to as Single Antenna Interference Cancellation (SAIC). At a high level, two approaches are possible for SAIC, namely Joint Detection (JD) algorithms and Blind Interference Cancellation (BIC) algorithms. As the name suggests, JD algorithms essentially attempt to decode both the desired signal and the interfering signal, whereas BIC algorithms attempt to cancel the interference based on general characteristics of the interfering signal. Clearly, JD algorithms require synchronization between the signals, which may not be possible in the case of un-synchronized base stations. For the BIC algorithms, a particular non-obvious property of the GMSK modulated signals is exploited. The property is that GMSK symbols, although carrying a single bit per symbol, span the 2-dimensional complex symbol space, so that the received signal can be split into two I-Q components, mimicking two separate signals from two virtual antennas [3,5]. This observation allows the application of several well known 2 antenna interference cancellation algorithms. Figure 5 illustrates the large gains that are possible with an SAIC algorithm [8]. The parameter DIR represents the ratio of the Dominant Interferer to the rest of the Interferers and Noise.

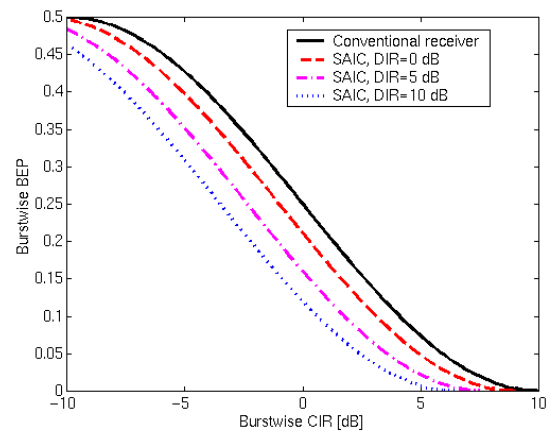

Figure-5: Illustrating the gains realizable by SAIC [8]

Due to such impressive performance, DARP performance requirements have been standardized in GERAN Release 6 [6].

3) Mobile Station Receive Diversity (MSRD) with Dual Antennas: As handset technology improves, use of multiple antennas will become a reality. As such, GERAN Evolution is currently standardizing the use of 2 receive antennas to provide diversity gains [7]. In a certain sense, it may be viewed as the Dual Antenna extension of DARP/SAIC. It is expected that this will be standardized in Release 7 of GERAN, due to be completed later in 2007. This topic will be dealt with in greater detail later in section IV.

4) Downlink Dual Carrier (DLDC) transmission: GERAN uses a relatively narrowband carrier of 200 KHz. In order to increase peak and mean data rates, the use of multiple carriers has been investigated in both uplink and downlink. However, it has been decided that only the use of 2 carriers in the downlink will be standardized, in Release 7 of GERAN [7]. This topic will be elaborated later in section V.

5) Higher Order Modulation (HOM): A well known way to increase data rates is to increase the modulation order. As it is done in UTRAN, GERAN Evolution is currently standardizing 16 QAM and 32 QAM for Release 7 [7]. This topic will be elaborated later in section VI.

6) Higher Symbol Rate (HSR): Through feasibility studies, it has been determined that the symbol rate of the GERAN waveform can be increased by a factor of 1.2, with the spectrum still satisfying the spectral mask requirements [6,7]. This effectively increases the peak data rates by 20%.

7) Turbo Coding: It is well known that Turbo Codes are near-optimal codes approaching Shannon capacity limits and that it has been standardized in 3G UTRAN. Accordingly, GERAN is adopting the same Turbo Coding scheme [7]. This introduces additional MCS classes to those shown in Table-1. Since Turbo Codes perform best for longer data blocks, Turbo Coding is expected to perform well with HOM. This feature will be standardized in Release 7 of GERAN.

B. Higher Layer Enhancements

1) Latency Reduction for Conversational PS (VOIP): From the very origin of GSM, the basic time unit in delivery of voice samples was chosen to be 20ms. The same tact was also used when GPRS was introduced by sending the basic piece of info, called "Radio Blocks", in four consecutive TDMA frames using one time slot. As Packet Switched networks offer more capacity and flexibility for the operators, the support of Voice over IP (VoIP) became an attractive solution. This, on the other hand, calls for reduction of latency, which is somewhat natural in PS networks. Accordingly, GERAN is standardizing various techniques to reduce the latency in Conversational Packet real-time Services, such as voice and video.

The two latency reduction techniques that are being considered are the Reduced Transmission Time Interval (RTTI) and Fast Ack/Nack Reporting (FANR). The RTTI is realized by sending the information in 10ms, instead of 20ms, using two timeslots on the same carrier frequency in two consecutive frames.

In order to guarantee the delivery of data packets, it has always been customary (in GPRS) that the receiver sends acknowledgements for the correctly decoded blocks. In the DL direction, where the MS is the receiver, due to the fact that several mobiles can be multiplexed on the same physical resource, the NW decides when a certain mobile should send acknowledgement by simply polling the MS. When being polled, the MS sends a so called Ack/Nack report using a Control Block. It has to be mentioned that Control Blocks in (E)GPRS cannot contain any user data. This mechanism works properly for applications that can tolerate delays. A real time service, on the other hand, demands rapid responses from the receiver. Therefore, the idea of FANR is that the receiver, when sending data, will piggyback an Ack/Nack report. Currently, there are two different approaches being discussed, one time based and another one based on the sequence number.

2) Generic (WLAN) Radio Access Networks (GAN): Due to the increasing prevalence of Wireless LAN, the idea of utilizing WLAN coverage became very appealing for the operators. This is particularly important as it frees up radio spectrum hence increasing the capacity. The whole idea is also recognized as Fixed Mobile Convergence (FMC). GERAN standards introduced a new node, called GAN Controller (GANC), interfacing the CN by using the exact same protocols and procedures defined in GERAN, to enable this idea in Release 6. This new GANC also communicates with the mobile using WLAN (IEEE 802.11) Access Points and rules. Obviously, handover of voice calls, being the most important service, was defined between GERAN and GAN from the day of introduction, and extensions to PS Handovers is under current standardization.

IV. "Mobile Station Receive Diversity" with Dual Antennas

The use of multiple antennas is a technique that is being explored in many wireless communication systems, since it promises great improvements for capacity (and hence data rates) as well as performance (by providing transmit and receive diversity). In GERAN Evolution, the use of 2 antennas at the Mobile Station is being explored in order to provide receive diversity, which can in turn be exploited to improve the performance in the presence of channel impairments, interference and noise. We mention that the use of multiple antennas at the Base Station has been common in GERAN. They have been used for Receive Diversity as well as Transmit Diversity. So, although the use of multiple transmit antennas at the Base Station coupled with multiple receive antennas can in principle enable full fledged MIMO in the downlink, the GERAN standardization efforts have not yet embraced this aspect.

The basic concept is shown in Figure-6 where the MS has two receive antennas, which receive two versions of the transmitted signal from a desired BS transmitter, as well as interference from one or more other BSs. The channel coefficients $\{h_{ij}\}$ are complex numbers in general [7].

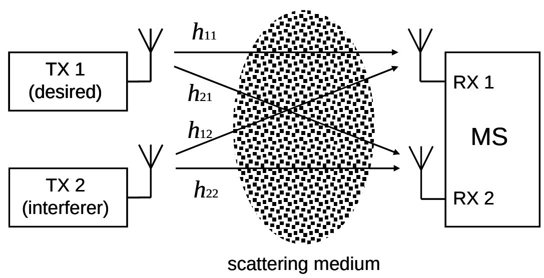

Figure-6: Basic Concept of MSRD [7]

It is assumed that the channel coefficients h_{mn} are superpositions of L multipath components (MPC), each of which interacting with the scattering medium through a different path. Each MPC is described by its angle of departure (AOD) and its Angle Of Arrival (AOA) and its complex amplitude. A MPC may arise due to single scattering, multiple scattering, or line-of-sight transmission.

The signals received by the two antennas may be correlated, in general, depending upon the scattering multipath channel environment, physical placement of the two antennas as well as the usage of the device by the user. Note that the correlation may be different for signals from different Base Stations. Figure below shows correlation values for signals received from 7 BTSs measured in an outdoor setting, at highway driving speeds [7]. (BTS A&G are from the same site as are BTS C&F).

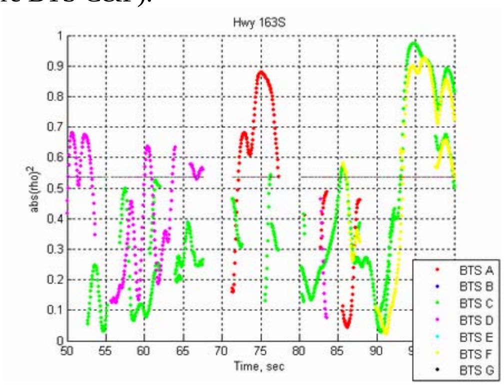

Figure-7: Measured Correlation Values [7]

In addition to correlation of the signals received by the two antennas, the gain experienced by the two antennas may also be different, due again to physical design of antennas, scattering medium and factors involving user interaction. Practical values in related scenarios range from 0-13 dB [7 and references therein].

At the receiver, the two received signals are processed separately at the RF level and combined at the baseband level, in order to maximize the SNR and SIR. Shown in Figure 8 is an example of the improvement in the radio link performance due to dual antenna diversity. This figure depicts uncoded BER vs. CIR (Carrier to Interference Ratio) for the case of single antenna conventional receiver and a dual antenna advanced receiver. The simulation parameters include: GMSK modulation, 2 Interferers 10 & 20 dB below the desired signal, Typical Urban Channel, pedestrian speed 3 kmph.

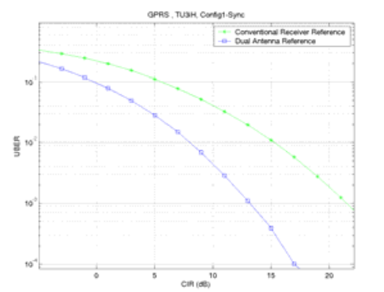

Figure 8: Uncoded BER vs. CIR for single & dual antenna receivers [7]

System level performance improvement can be demonstrated by considering the Figure 9, which shows the percentage of satisfied users (defined as users whose calls are not blocked and whose calls have less than 2% FER for the duration of the call) as a function of the load of the carrier (which controls the blocking probability and the amount of interference). The figure also compares the performance of a Conventional Receiver (CR-purple cuve), Single Antenna Interference Cancellation Receiver (DARP – blue curves) and Dual Antenna MSRD Receivers (DARP-MSRD – red & black curves). Flat fading and Typical Urban (TU) fading as well as different values of correlation & gain imbalance are addressed. The figure shows dramatic improvement with MSRD.

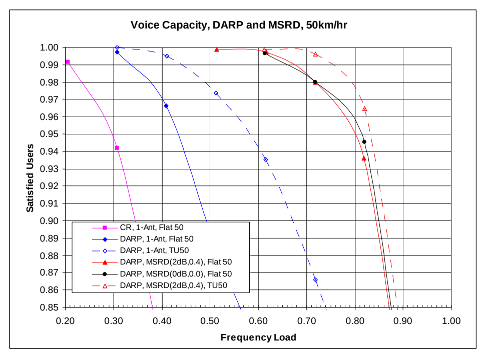

Figure 9: System Performance Improvement with MSRD [7]

V. "DOWNLINK DUAL CARRIER" TRANSMISSION

The maximum data rates are ultimately limited by the 200 KHz carriers that GERAN uses. Currently, the theoretical maximum instantaneous rate is about 473kbps. In the real world, bit rates of the order of 100-200kbps are feasible on 2-4 timeslots. Therefore, GERAN Evolution is considering the usage of multiple carriers to increases the data rates in both the uplink and downlink.

At the time of writing this paper, this feature is being standardized only in the downlink with 2 carriers, whereas the uplink counterpart is still in the feasibility stage. The two carriers need not be adjacent but will probably be constrained to a maximum separation.

It is worth noting that even dual carriers in the downlink will indirectly increase the data rates in the uplink also, because a smaller number of timeslots is used for downlink transmission, leaving more timeslots for uplink usage.

Downlink dual carrier usage has another indirect advantage for the uplink transmissions. In GERAN, carriers are typically assigned in pairs, with one carrier for uplink and the other for downlink. Assuming that dual carriers are used for simultaneous data transmission only in the downlink, the uplink data transmission can be dynamically allocated between the two uplink carriers. For example, the dynamic allocation may be optimized based upon the uplink channel conditions.

In addition to increasing peak data rates, using multiple carriers also reduces latencies. For example, using 2 carriers and 1 timeslot per frame per carrier will halve the latency involved in transmitting a radio block (i.e. 4 bursts) to 10 msecs.

Error protection can also be improved with dual carriers, because, for example, interleaving of the data can be done across both carriers, providing frequency diversity in addition to time diversity provided by interleaving. Figure 10 illustrates the potential gains that can be realized for different MCS schemes with or without Frequency Hopping for the cases of single or dual carriers [7]. It is seen that about 1 dB gain is achievable.

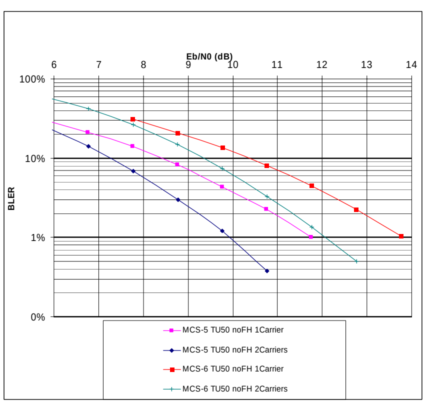

Figure 10: Channel Coding Gains due to 2 carrier transmission [7]

It is also possible that advanced channel coding methods (essentially of a joint time-frequency type) can be developed for joint coding across the two carriers.

Although the primary motivation for dual carriers is increased data rates, the two carriers may also be used to provide diversity, by transmitting same data (or jointly coded data) across both the carriers. Such techniques will improve the coverage probabilities, especially at the cell edges, where SNR & SIR are typically low.

Next, we mention that dual carriers in the downlink can facilitate simultaneous reception of MBMS (Multicast and Broadcast Multimedia Services) as well as transmission and reception of PS and CS services. Presently, if a MS is receiving MBMS data, it cannot engage in point to point PS/CS data transmissions/receptions.

Finally, note that dual carrier usage will have some impact on the MS design. For example, with dual carriers in the downlink, the RF front end may be designed with two separate Rx filters or a single wideband filter. Similarly, if dual carriers are used in the uplink, the peak-to-average ratio of the modulated signals will increase. This in turn requires a larger power amplifier backoff. Figure 11 and 12 compare the peak-to-average ratios for 8PSK modulation with single and dual carriers. Note that the PAPR increases to a maximum of 6 dB with dual carriers.

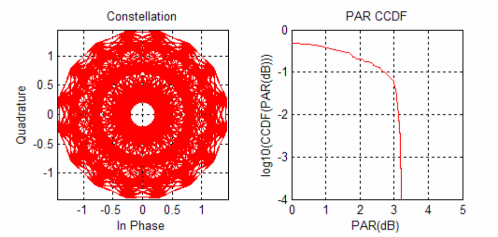

Figure 11: Single-carrier 8PSK constellation and PAR CDF [7]

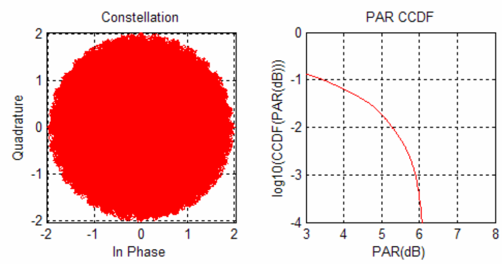

Figure 12: Dual-carrier 8PSK constellation and PAR CDF [7]

VI. HIGHER ORDER MODULATION

Recall that GERAN presently uses GMSK and 8PSK modulation schemes, carrying 1 and 3 bits per modulated symbol respectively. GERAN Evolution is standardizing 16 QAM & 32QAM, carrying 4 & 5 bits per symbol respectively. These extra bits can be used to increase the data rates (while keeping the channel coding rate the same) or to improve the robustness due to channel errors (by decreasing the coding rate while keeping the data rate unaffected).

One of the first design issues when introducing a new modulation scheme is the constellation, which affects several aspects such as Peak-To-Average Ratio (which affects the power amplifier design), bit-to-symbol mapping and Minimum Inter-symbol distance (which affects the BER performance). Shown below are some examples that are being considered for standardization – square, concentric [7].

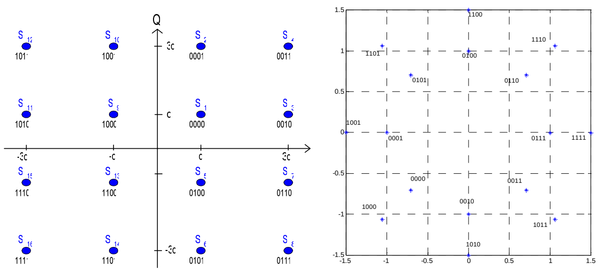

Figure 13: Example constellations for 16-QAM [7]

The peak-to-average ratio is important because the power amplifier has to be linear over the dynamic range of the signal. The PAPR of higher order modulations will naturally be higher than for legacy modulation schemes, but can be reduced by introducing inter-symbol phase rotations. For example, the PAPR for the square 16-QAM with pi/4 rotation is about 5.3 dB, compared to 3.3 dB for the legacy 8PSK with 3pi/8 phase rotation [7].

When groups of bits are mapped to higher order constellations (for example 4 bits to a 16-QAM symbol), some bits have higher reliabilities compared to the others, depending upon bit-to-symbol mapping scheme used. For example, in the mapping shown in the square 16-QAM constellation in Figure 13, the first two bits have higher reliability than the last two bits. Figure 14 shows that there is about a 3 dB greater reliability for the high reliability bits [7].

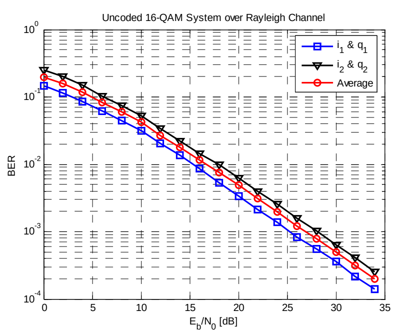

Figure 14: Differing reliabilities of 16QAM bits [7]

If the data bits themselves have differing requirements for reliability, then the bit-to-symbol mapping can be optimized by matching the reliability of bits to the reliability of the symbols. For example, if the data to be mapped is Turbo Coded data, then the systematic bits can be mapped to higher reliability positions and parity bits to the lower reliability positions.

The last property, namely inter-symbol distance, affects the raw uncoded BER. For a given average power, the average minimum distance for Higher Order Modulations will decrease compared to legacy modulation schemes, tending to increase the raw BER. However, the increased data rates with Higher Order Modulations can be exploited to improve the channel coding, so that the coded BER may actually be lower for HOMs than for legacy modulations. In order to understand this, we consider various modulation and coding combinations (MCS classes). The table below shows the 9 MCSs defined in the existing GERAN based on GMSK and 8PSK. The table additionally shows 5 new MCS schemes based on 16QAM. Of these, two new MCS classes are defined (MCS-8-16QAM & MCS-9-16QAM) that improve the code rate and three other MCS classes (MCS-10-16QAM & MCS-11-16QAM & MCS-11-32-QAM) which increase the data rates [7].

Table 2: New MCS classes made possible by Higher Order Modulation [7]

Modulation and coding scheme	Modulation	User Data Rate (kbps)	Payload Coding Rate
MCS-1	GMSK	8.8	0.53
MCS-2	GMSK	11.2	0.66
MCS-3	GMSK	14.8	0.85
MCS-4	GMSK	17.6	1.00
MCS-5	8PSK	22.4	0.37
MCS-6	8PSK	29.6	0.49
MCS-7	8PSK	44.8	0.76
MCS-8	8PSK	54.4	0.92
MCS-9	8PSK	59.2	1.00
MCS-8-16QAM	16QAM	54.4	0.67
MCS-9-16QAM	16QAM	59.2	0.73
MCS-10-16QAM	16QAM	67.2	0.83
MCS-10-32QAM	32QAM	67.2	0.65
MCS-11-32QAM	32QAM	81.6	0.79

Shown below is an example link coded BER performance for legacy MCS-8 and MCS-8-16QAM [7].

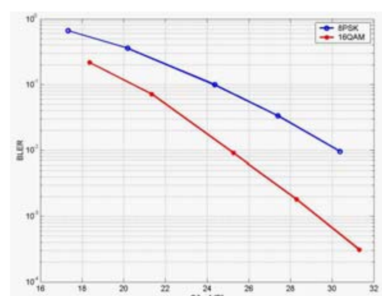

Figure 15: Coded BER performance improvement with 16QAM [7]

It is well known that GERAN uses link adaptation techniques to optimize the link performance over a wide range of channel conditions. The adaptation is done by switching among the various MCS classes, shown in Table . The figure below shows the improvement in link performance.

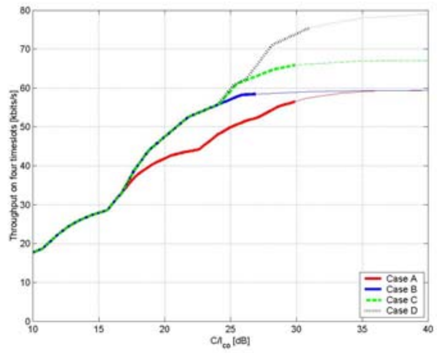

Figure 16: Link Throughput for legacy MCS classes (Red curve) and enhanced MCS classes [7]

The red curve (Case A) refers to the set of legacy MCS classes, with the highest data rate being 59.2 kbps per timeslot. The blue curve (Case B) replaces the legacy MCS8&9 with 16QAM, which keeps the data rates unchanged but improves the channel coding, resulting in improved performance at lower SNRs. The green curve (Case C) adds a new MCS-10 with a peak data rate of 67.2 kbps, improving the maximum data rates supported by the link. Finally, the black curve (Case D) adds a new MCS-11 with a peak data rate of 79.2 kbps, further increasing the maximum link rates.

VII. Conclusion

This paper attempts to demonstrate that the most widely deployed 2nd generation digital cellular system introduced in the early 1990s, namely GSM, continues to evolve taking advantage of latest advances in radio communication technologies and handset implementation technologies. It is expected that evolved GERAN will keep pace with the numerous new radio interfaces, including 3G Cellular as well as local area radio interfaces and continue to serve humanity for decades to come!

References

[1] M. Mouly, M-B Pautet, “The GSM System for Mobile Communications”, 1992

[2] D. Goodman, “Wireless Personal Communication Systems”, Addison Wesley, 1997

[3] T. Halonen, J. Romero, J. Melero, “GSM, GPRS and EDGE Performance”, 2nd Edition, John Wiley, 2003

[4] 3GPP TR 45.902 V6.8.0 (2005-01), “Flexible Layer One (FLO)” www.3gpp.org/ftp/Specs/archive/45_series/45.902/

[5] H.Trigui, D. Slock, “Cochannel Interference Cancellation within the current GSM standard”, Proc. IEEE 1998 Intl. Conf. on Universal Personal Communications, Oct 5-9, 1998, Florence, Italy.

[6] 3GPP TS 45.005 V6.0.0 (2002-11), “Radio transmission and reception” (Release 6). www.3gpp.org/ftp/Specs/archive/45_series/45.005/

[7] 3GPP TR 45.912 V7.1.0 (2006-11), Technical Report, “Feasibility study for evolved GSM/EDGE Radio Access Network (GERAN)”, (Release 7). www.3gpp.org/ftp/Specs/archive/45_series/45.912/

[8] 3GPP TR 45.903 V6.0.1 (2004-11), “Feasibility Study on Single Antenna Interference Cancellation (SAIC) for GSM networks”, (Release 6). www.3gpp.org/ftp/Specs/archive/45_series/45.903/

Prabhakar Chitrapu currently works as a Principal Engineer in the Office of the CTO at InterDigital Communications Corporation in King of Prussia, PA (since 2000). In this position, he has been involved in development of advanced wireless communication technologies for cellular and WLAN networks. He is also an Adjunct Professor at Villanova University, where he teaches graduate courses in Wireless Communications and Signal Processing. Earlier, he worked in the area of Mobile Satellite Communications at Lockheed Martin Global Telecommunications, in the area of Interactive Computer Telephony at Dialogic Corporation (now part of Intel Corporation), in the Electrical & Computer Engineering Department at Drexel University as an Assistant Professor, in the Department of Electrical Engineering at University of Pennsylvania as an Adjunct Professor, and in Digital Communication Development at Bharat Electronics Ltd, India. He has been and continues to be involved in Standards Organizations, such as TIA, ETSI and 3GPP. Prabhakar has published a book “Wideband TDD: WCDMA for the Unpaired Spectrum” (Wiley 2005) and over 50 papers in technical journals and conferences. He has over 20 patents, one of which received the highest award at InterDigital (Chairman of the Board of Directors’ award). Prabhakar holds a PhD degree in EE from Delft University of Technology, Delft, The Netherlands (1985), MS degree in EE from Philips International Institute, Eindhoven, The Netherlands (1980) and B.Tech. in EE from Indian Institute of Technology, Madras, India (1976). He is a member of IEEE.

Behrouz Aghili Originally born in Iran, Behrouz moved to Stockholm (Sweden) in 1984 where he graduated with a Master of Science degree in Engineering Physics at the Royal Institute of Technology. During the years of studying, he also started a teaching career, at the same university, as an Assistant Professor in mathematics. Behrouz taught math for about seven years and received numerous awards and recognitions. His telecom carrier started with a position in Ericsson Radio Systems where he dealt with the Operation & Maintenance of the Ericsson MSC. After two years with Ericsson, he started working for a training company in Sweden, called APIS Technical Training, dealing with GSM signaling protocols. He stayed seven years with this company where he conducted and developed training and consultancy in the area of GSM, GPRS, EDGE and UMTS. Between November 2004 and August 2006, he worked at Award Solutions, an American training company where he continued working with GERAN and UTRAN protocols as well as HSDPA. As of August 2006, he has been an employee of InterDigital, mainly dealing with the standardization of GERAN.