GSM/EDGE continued evolution

Håkan Axelsson, Peter Björkén, Peter de Bruin, Stefan Eriksson and Håkan Persson

GSM/EDGE has evolved into the most widely deployed cellular telephony standard in the world, offering high-performance voice and data services. To keep GSM/EDGE competitive, even in years to come, Ericsson has proposed steps to further evolve the standard.

The authors describe the proposed technology enhancements, which improve average and peak bit rates, latency, service coverage, and spectrum efficiency. Specific solutions include dual-antenna terminals, multicarrier EDGE, mobility enhancements, reduced transmission time intervals, higher-order modulation, and new coding schemes. The combination of these solutions fulfills the requirements set by 3GPP.

Background

When introduced to the market in 1992, the Global system for mobile communication (GSM, defined by ETSI as the internationally accepted digital cellular telephony standard) was primarily a voice-only system with SMS and some low-bit-rate circuit-switched data services. Since then, however, GSM has continued to evolve unabated. General packet radio service (GPRS), for example, introduced packet-data capabilities. And the subsequent introduction of Enhanced data rates for GSM evolution (EDGE) or Enhanced GPRS has greatly extended the capabilities of packet-data services. EDGE technology, which is backward compatible with GPRS, offers up to a 200% increase in data bit rates.[1]

At present, GSM/EDGE is undeniably the most widely spread cellular technology in the world (Figure 1). Only two countries, Japan and South Korea, rely exclusively on other standards. There are currently more than 1.7 billion GSM/EDGE users worldwide, and subscriber growth continues to be high. In November 2005, the Global mobile Suppliers Association (GSA) reported that 121 GSM operators were offering commercial EDGE-based services in 70 countries, and 172 operators in 92 countries had committed to providing EDGE.[2]

Improving service performance

To keep GSM/EDGE competitive, even in years to come, Ericsson has proposed a number of enhancements that facilitate truly conversational multimedia services and improve service performance and service continuity. All major vendors and many operators have pledged their support for a continued evolution, which is being standardized in 3GPP.[3] The general objectives are to improve service performance and provide more efficient bearers. Seamless operation or service continuity between GSM/EDGE and WCDMA networks is also being addressed.

Objectives

The proposed technology enhancements are meant to improve average and peak bit rates, latency, service coverage, and spectrum efficiency. Box B (Item I) shows the targets used in the standardization.

The first objective, increased spectrum efficiency, mainly benefits operators. In many urban areas, the existing frequency spectrum is being used to the maximum extent. However, by enhancing spectrum efficiency (that is, improved ability to withstand greater levels of interference) operators can increase traffic volumes without compromising quality.

Best-effort services, such as web browsing and file downloads, typically benefit from increased peak and mean bit rates. Depending on the degree of user interactivity, these services might also considerably benefit from reduced latency.

Conversational services, such as voice over IP (VoIP) and enhanced Push to talk over Cellular (PoC), as well as online gaming services typically benefit from (or might even require) reduced latency and faster access.

BOX A, TERMS AND ABBREVIATIONS

3GPP	Third Generation Partnership Project
8-PSK	Octonary phase-shift keying
BCCH	Broadcast control channel
BSC	Base station controller
BSIC	Base station identity code
BSS	Base station subsystem
BTS	Base transceiver station
CDMA	Code-division multiple access
DTM	Dual-transfer mode
EDGE	Enhanced data rates for global evolution
EGPRS	Enhanced GPRS
ETSI	European Telecommunications Standards Institute
FTP	File transfer protocol
GAN	Generic access network
GERAN	GSM/EDGE radio access network
GMSK	Gaussian minimum-shift keying
GPRS	General packet radio service
GSM	Global system for mobile communication
MBMS	Multimedia broadcast/multicast services
MCS	Modulation and coding scheme
MMS	Multimedia messaging service
PCU	Packet control unit
PoC	Push to talk over cellular
QAM	Quadrature amplitude modulation
RLC	Radio link control
RTT	Roundtrip time
SMS	Short message service
TBF	Temporary block flow
TDMA	Time-division multiple access
TTI	Transmission time interval
VoIP	Voice over IP
WAP	Wireless application protocol
WCDMA	Wideband CDMA
WiFi	Wireless fidelity (wireless LAN)

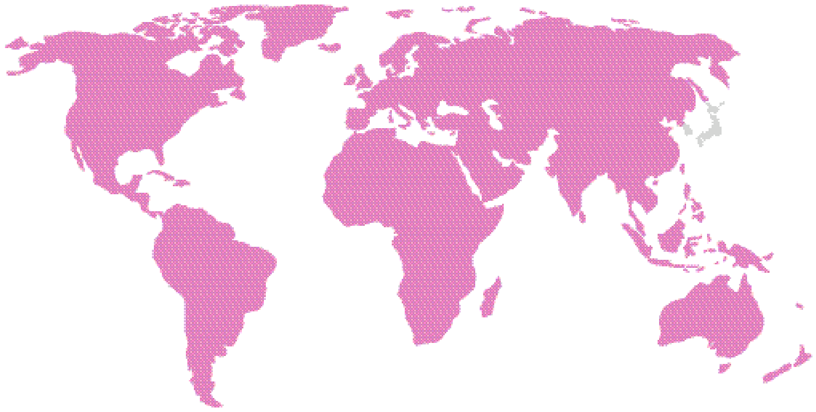


Figure 1
GSM/EDGE footprint.

All services (voice as well as packet-data) stand to benefit from improved coverage and from having terminals that are always connected to the most appropriate base station.

Most services can benefit from faster initial response from the network. This is especially true for conversational services and applications with limited system interactivity prior to uploading or downloading application data (for example, WAP, MMS and FTP). Although not presently pursued in 3GPP, solutions have been proposed that would decrease access time (when the mobile terminal is known in the packet control unit, PCU) to less than 500ms.

Given that the installed base of GSM/EDGE equipment is very large, Ericsson has tried to minimize the effects of the proposed technology enhancements on base station hardware. As a consequence, the proposed solutions have no impact on transceiver units and other base station hardware. What is more, they assume the same (current) network architecture (Figure 2). In general, only radio network software and terminals will need to be updated.

Proposed solutions

Ericsson has proposed solutions that fulfill the objectives outlined above, some of which coincide with proposals from other partici-

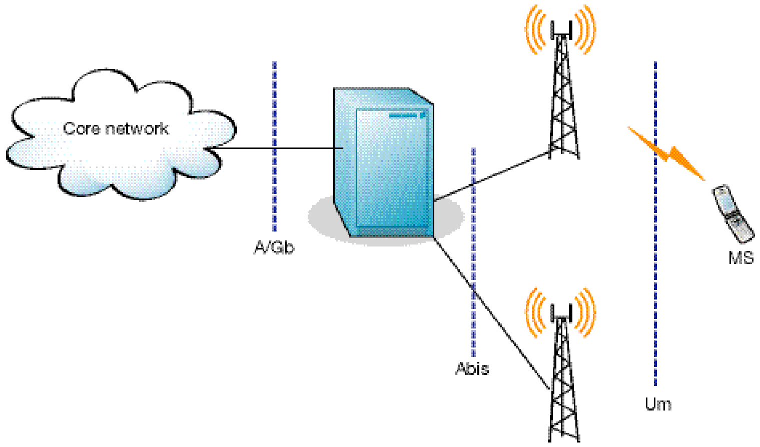


Figure 2
GSM/EDGE system architecture.

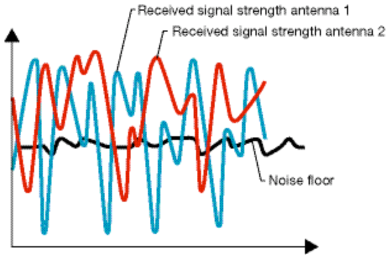
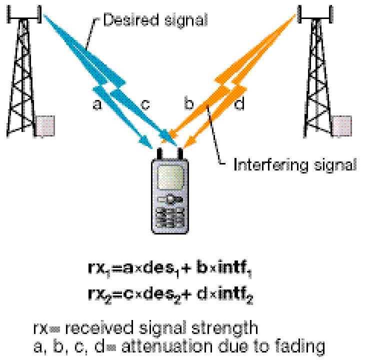


Figure 3
Received signal strength on two antennas in a mobile terminal.

Figure 4
Dual receive antennas can be used for interference cancellation. The received signal on each antenna is the sum of the desired signal and interference, weighted with different attenuation factors due to fading (denoted a, b, c and d).



pants in the standardization work. Most of the proposals are currently being discussed in 3GPP, either in the Future GERAN Evolution Work Item or other 3GPP Work Items. Some of Ericsson's proposals address several of the desired performance enhancements, whereas others address only one specific area.

Dual-antenna terminals

A challenge associated with radio communication is that the strength of received radio signals varies rapidly relative to the position of the transmitter, the receiver, and other objects that scatter signals. This phenomenon, known as fading, can make a signal too weak to be captured. One other challenge is that other transmitters (terminals or base stations) in the vicinity can cause interference. Various techniques, such as channel coding, frequency hopping, and selective retransmission schemes, are used to combat fading and interference.

Dual-receive-antenna systems are an efficient weapon against fading. To date, however, these have only been used on base station receivers. Figure 3 shows how two antennas mounted on a terminal, separated in space, polarization, or both, can receive two signals of the same transmission with different fading characteristics. This increases the probability that at least one of the signals will be strong enough (with some margin above the receiver noise floor) to be captured. Moreover, by combining two signals, one can sometimes capture a transmission that would otherwise have been altogether too weak.

Dual-antenna solutions can also be used to efficiently handle interference. Obviously, the desired signal as well as all other received interference is subject to fading on the two antennas. But by combining the signals, one can cancel out interference by taking into account the instantaneous attenuation of the different signals due to fading (denoted *a*, *b*, *c* and *d*, Figure 4). This technique is known as interference cancellation.

Experiments and computer simulations show that dual-antenna solutions in GSM terminals yield substantial improvement. In situations with limited coverage (that is, when the signal is too weak to be received correctly), dual-antenna terminals can cope with signal levels 6dB below (or one-fourth) that of single-antenna terminals. Moreover, dual-antenna terminals can handle almost 10dB (or ten times) more interference.

Multicarrier EDGE

The GSM radio interface is based on time-division multiple access (TDMA) with 200kHz carrier bandwidth. To increase bit rates, today's GPRS and EDGE terminals can use multiple timeslots for transmission, reception, or both (Figure 5, left). The GSM standard allows for up to eight timeslots in each direction (uplink and downlink), giving a theoretical peak bit rate of close to 480kbps. However, from a design and complexity point of view, it is best to avoid simultaneous transmission and reception. Therefore, in practice, today's terminals typcially receive on a maximum of five timeslots because they must also transmit (on at least one timeslot) as well as measure the signal strength of neighboring cells. The actual number of timeslots a terminal uses for transmission is also limited.

To increase bit rates, one can introduce multiple carriers for the downlink and uplink. This straightforward enhancement increases peak and mean bit rates in proportion to the number of carriers employed (Figure 5, right). For example, given four carriers with eight timeslots each, the peak bit rate would be close to 2Mbps. The limiting factor in this case is the complexity and cost of the terminal, which must have either multiple transmitters and receivers or a wideband transmitter and receiver. The use of multiple carriers has only a minor impact on base transceiver stations.

Mobility enhancements

The functionality needed to handle user mobility is a fundamental component in any cellular network. Services must be maintained as users move from one cell to another, and a good or adequate radio environment must exist for the establishment of new sessions. Seen in terms of spectrum efficiency, each active user should always be served by the most appropriate (closest) base station.

In advanced GSM/EDGE networks that employ tight frequency reuse, being connected to the most appropriate base station is essential for achieving good spectrum efficiency. In tight frequency-reuse scenarios, user mobility can compromise service quality, particularly in cases that involve an inhomogeneous cell plan, fast-moving users, or tight reuse of the broadcast control channel (BCCH). User mobility is managed in three general steps (in both active and idle modes for voice and data but with different processes and requirements in each case):

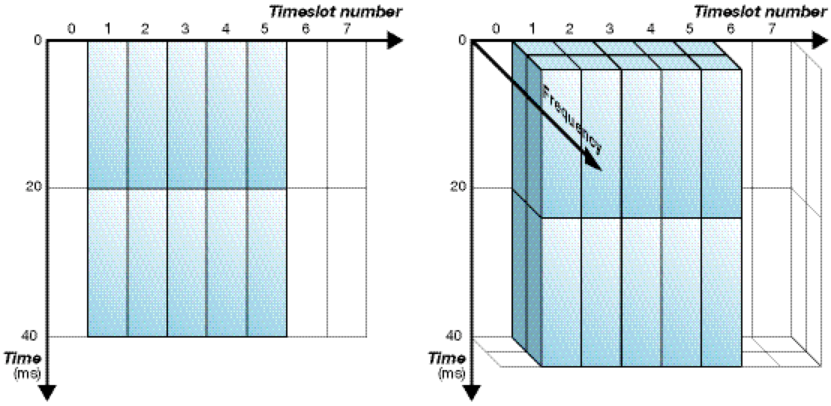


Figure 5
Left: Radio blocks are transmitted on five parallel timeslots on one carrier.
Right: Radio blocks are transmitted on ten parallel timeslots on two carriers.

measurements are taken to determine whether or not cell change is necessary; processing of the measurements (to reach this decision); and the actual cell change.

In idle mode, the main limitation of managing user mobility is that the measurements of neighboring and serving cells are not stringent enough for environments with tight frequency reuse. Consequently, sessions are sometimes initiated in a sub-optimum cell, creating extra interference and leading to service interruptions (dropped calls). Putting more stringent requirements on measurements of the received signal strength would eliminate this limitation.

In active mode, faster and more accurate measurements of neighboring cells would improve performance. At present, mobile terminals identify neighboring cells separately and infrequently by decoding the base station identity code (BSIC). Furthermore, they solely measure the total strength of the received signal. This means that if several nearby base stations use the same frequency, the contribution from each base station cannot be resolved. Ericsson proposes to introduce simultaneous identification of neighboring cells (BSIC information) and measurements of signal strength on all bursts over the BCCH carrier.[4] This would significantly reduce the risk of incorrect

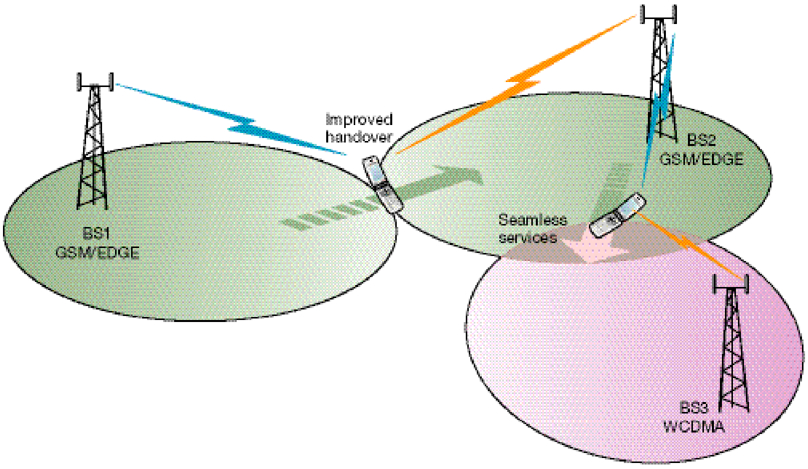


Figure 6
Mobility enhancements.

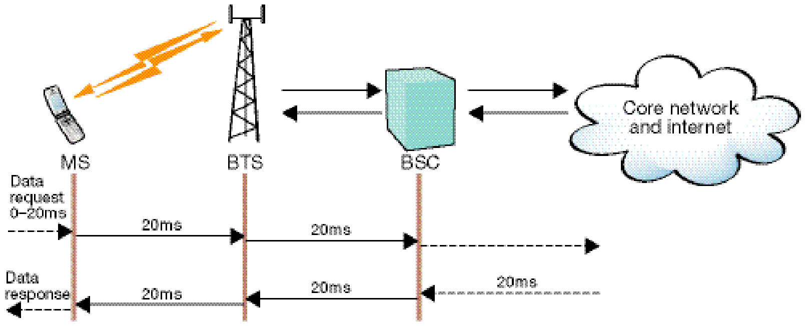


Figure 7
Data flow diagram with TTI = 20ms.

handover, resulting in substantially fewer dropped calls. It would also speed up the process of detecting new, strong, neighboring cells.

Mobility enhancements apply equally well to voice and data services (although the detailed effects differ). They also improve service continuity, for example, by means of faster and more accurate handover between GSM/EDGE and WCDMA networks (Figure 6) and enhanced handover functionality in dual-transfer mode (DTM).

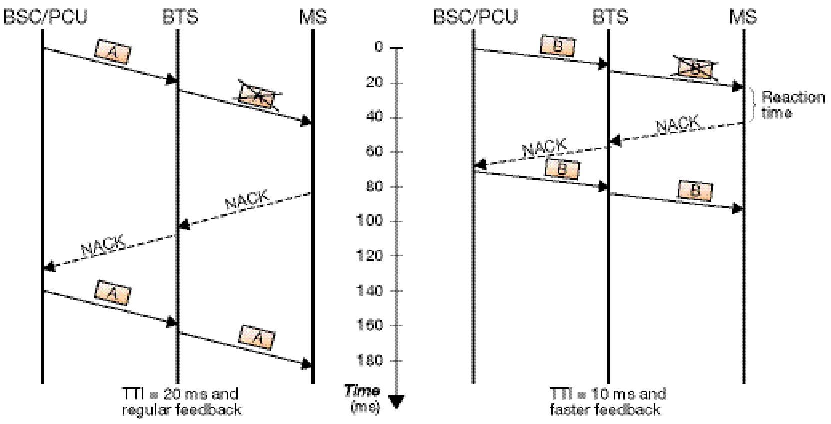
Reduced TTI

Latency has a major influence on user experience. In particular, conversational services, such as VoIP and video telephony, require low latency. Other services that benefit from low delay are gaming and applications with extensive *handshaking*, such as e-mail.

It is difficult to substantially improve latency without reducing the transmission time interval (TTI). The roundtrip time (RTT) in advanced GSM/EDGE networks with an Ericsson base station subsystem (BSS) is 150ms. This figure includes network delays but not retransmissions over the radio interface. Radio blocks are transmitted over four consecutive bursts on one timeslot using a 20ms TTI. Figure 7 illustrates the data flow with a 20ms TTI.

Reducing the TTI improves latency substantially and immediately. To reduce TTI, one can either use fewer than four bursts (smaller radio blocks) or transmit all four bursts on more than one timeslot (for example, parallel timeslots on two carriers). Ericsson estimates that reducing the TTI from 20ms to 10ms will reduce the roundtrip time from 150ms to 100ms.

Figure 8
Example of transfer using reduced TTI and faster feedback enhancements.



Faster feedback

To help the transmitter better understand the radio environment, feedback information is sent via the radio link control (RLC) protocol over the air. The RLC protocol typically runs in acknowledged mode, which requires the retransmission of lost radio blocks. Although feedback information is crucial for efficient transmission over the radio interface, it is also time-consuming. The procedure requires the receiver to periodically send (on request)

- acknowledgements of radio transmissions; and
- information about the current radio environment.

Faster feedback enables the transmitter to retransmit lost data earlier and makes radio transmission more efficient. By putting more stringent requirements on reaction times and by introducing support for immediate response to unsuccessful radio transmissions, one can ensure that lost radio blocks are retransmitted much earlier, which reduces latency. One can reduce latency even further by combining faster feedback with reduced TTI (Figure 8).

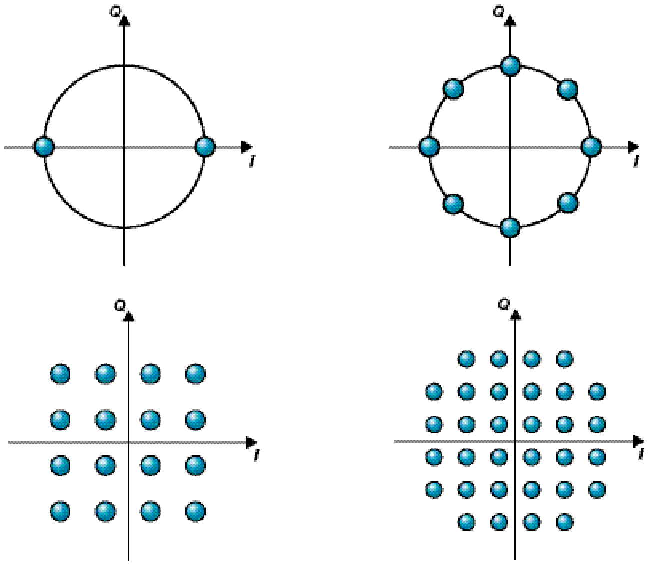


Figure 9
Top left: Gaussian minimum-shift keying (GMSK). Top right: Octonary phase shift keying (8-PSK). Bottom left: 16QAM (16-level constellation). Bottom right: 32QAM (32-level constellation).

BOX B, 3GPP

I. MAIN 3GPP OBJECTIVES AND TARGETS

Improved spectrum efficiency (or capacity) in an interference-limited scenario
Target: 50% increase (measured in kbps/MHz/cell for data and Erl/MHz/cell for voice)

Increased peak data rates
Target: 100% increase in downlink and uplink

Improved coverage, voice and data (noise-limited scenario)
Target: Sensitivity increase of 3dB in downlink

Improved service availability (when cells are planned for voice)
Target: 50% increase in mean bit rate for uplink and downlink at cell edges

Reduced latency, initial access (“no TBF assigned”)
Target: Round-trip time (RTT) of less than 450ms (in non-ideal radio conditions over the radio interface, counting from the mobile terminal up to *Gi* interface and vice versa)

Reduced latency, after initial access
Target: RTT of less than 100ms (in non-ideal radio conditions over the radio interface, counting from the mobile terminal up to *Gi* interface and vice versa)

Compatibility
Coexistence with legacy frequency planning
Coexistence with legacy mobile terminals
Minimum impact on existing BTS, BSC and core-network hardware

II. RECENT AND CURRENT 3GPP ACTIVITIES

MBMS (Rel-6)
Multimedia broadcast/multicast services (MBMS) enable a point-to-multipoint radio bearer for broadcast or multicast services (for example, mobile TV), allowing many users to receive the same information or media stream. MBMS enables efficient radio and network distribution of content, because only one bearer needs to be initiated in any given cell for use by every MBMS user in that cell.[8]

Packet-switched handover (Rel-6)
Packet-switched handover procedures in 3GPP Rel-6 enhance the support for real-time packet-based services, such as VoIP, as well as support for sharing the load of packet-data services between GSM/EDGE and WCDMA. This feature significantly reduces interruptions of packet-switched data flow when a terminal changes cells or systems. Delay-sensitive services, such as VoIP and gaming, can thus enjoy full mobility within GSM/EDGE and between GSM/EDGE and WCDMA. This also means that operators can steer user traffic between cells and between GSM/EDGE and WCDMA with very little impact on the service.

GAN (Rel-6)
The Generic Access Network (GAN) standard enables operators to use other types of access networks, such as Bluetooth and WiFi, to access GSM services. When users leave the Bluetooth or WiFi coverage area, they are handed over to GSM cells.[9]

Handover of shared and dedicated resources in DTM (Rel-7)
Handover of shared packet-switched and dedicated circuit-switched resources in dual-transfer mode (DTM) will further reduce interruptions and enhance service.

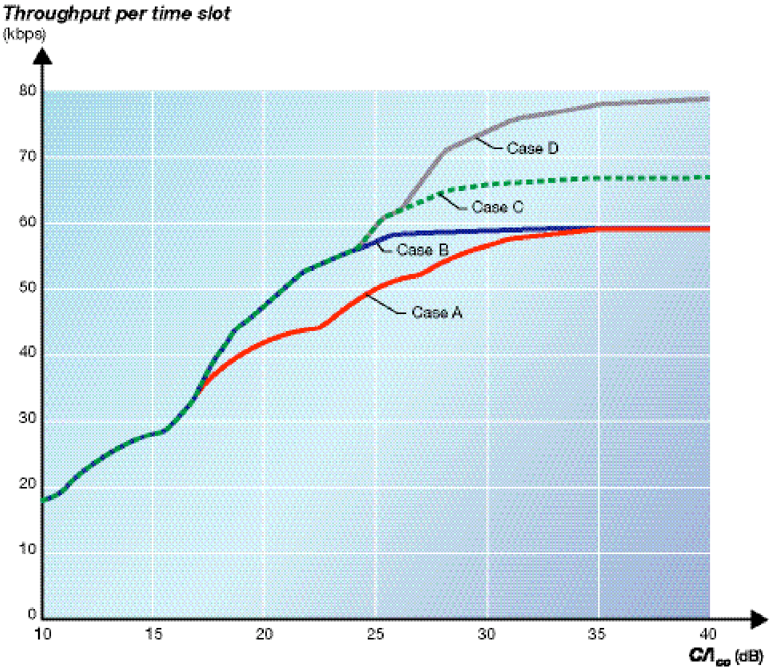


Figure 10
Bit rate per timeslot versus carrier-to-interference level for different modulation and coding techniques.

Higher-order modulation: 16QAM and 32QAM

GSM uses Gaussian minimum-shift keying modulation. GMSK is a binary modulation, which means one bit can be transmitted over the radio interface per symbol (Figure 9, top left). EDGE technology introduced a new modulation constellation: octonary phase shift keying (8-PSK), a non-binary modulation that allows three bits to be transmitted per symbol, increasing the peak bit rate from approximately 20kbps to around 60kbps per timeslot (Figure 9, top right).

The bit rates can be increased even further by introducing higher-order modulation, such as quadrature amplitude modulation (QAM). Figure 9 (botton left and right) shows QAM constellations with 16 levels (peak rates of up to 80kbps per timeslot) and 32 levels (peak rates of up to 100kbps per timeslot).

Given these constellations, we see that as more bits are transmitted per symbol, the signal points move closer together, making the signal more susceptible to interference. On the other hand, the higher bit rates permit more robust channel coding to be used, which more than compensates for increased susceptibility to interference.

Figure 11
Average user bit rates for 10th, 50th and 90th percentiles versus system load.

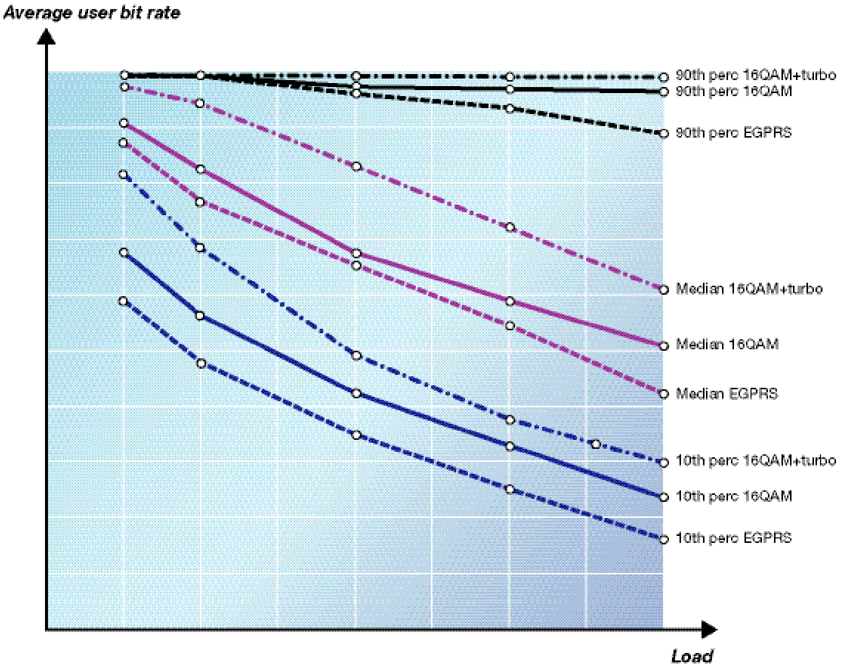


Figure 10 shows the relationship between bit rates and the carrier-to-interference ratio (C/I). *Case A* shows bit rate per timeslot for EDGE using 8-PSK modulation. When some of the modulation and coding schemes (MCS) are replaced with 16QAM equivalents, the bit rate increases (*Case B*). In this example, 16QAM has solely been used to increase robustness against interference, not to increase peak data bit rates. The technique (16QAM) enables the system to handle more than three times as much interference (5dB). The introduction of MCSs with 32QAM further increases the data bit rates, as indicated by *Case C* (one new 32QAM MCS with a peak rate of 67.2 kbps per timeslot) and *Case D* (one more, new, 32QAM MCS with a peak rate of 79.2 kbps per timeslot).[5]

The introduction of 16QAM has no effect on base transceiver station (BTS) hardware. The introduction of 32QAM, however, could be a limiting factor. Figure 11 shows system performance in terms of average user bit rates when 16QAM is used (instead of 8-PSK) for MCS-7, MCS-8 and MCS-9 (solid lines).

New coding schemes: turbo codes

Channel coding is used to increase robustness against interference. GSM and EDGE use what are called convolutional codes, which offer good flexibility (they can easily be adapted to any bit rate and code rate) and low encoding/decoding complexity.

In terms of error correction, more sophisticated classes of channel coding schemes outperform convolutional codes. One such class of coding schemes, called turbo codes, is used in other cellular systems such as WCDMA.[6] The process of decoding turbo codes is more complex than that of decoding convolutional codes, but because future terminals will generally support GSM/EDGE as well as WCMDA, the decoding circuitry for turbo codes will already exist in terminals and can thus be reused for GSM/EDGE.

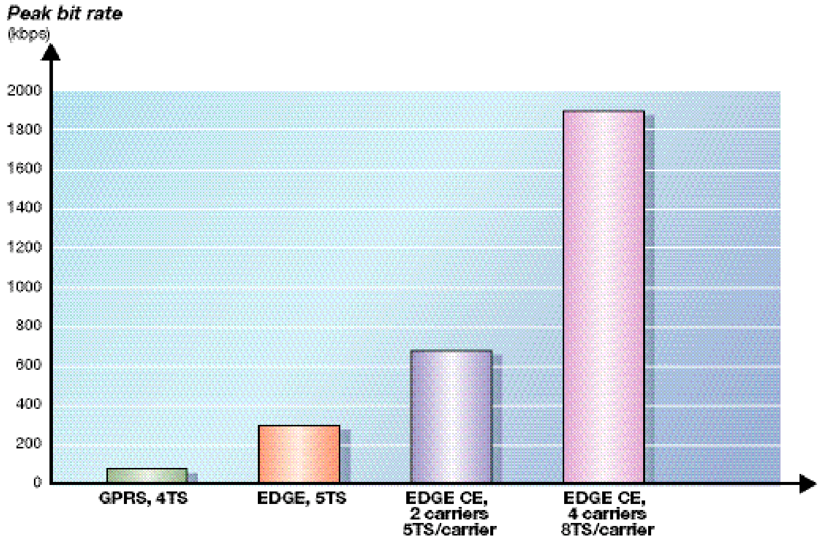


Figure 12
Peak bit rate enhancements.

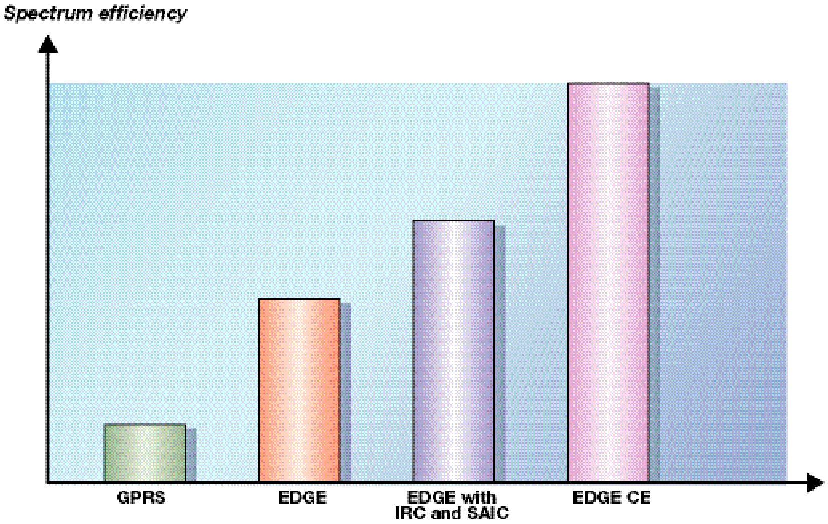


Figure 13
Spectrum efficiency enhancements.

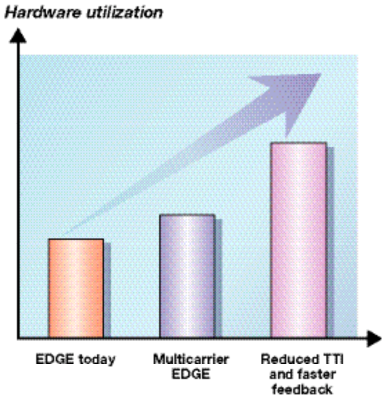


Figure 14
VoIP hardware utilization per cell, assuming multicarrier EDGE, reduced TTI, and faster feedback techniques.

Turbo codes perform well only when the code block size is sufficiently large, which means they are most suitable for high-bit-rate EDGE channels when used in combination with 8-PSK modulation or higher-order modulations.

Figure 11 also compares the average user bit rates when turbo coding is used in combination with 16QAM (dash-dotted lines), with normal EDGE and 16QAM without turbo coding. Compared with regular EDGE with 8-PSK, the combination of 16QAM and turbo coding improves the user bit rates 30% to 40% for the median user.

Summary of potential gains

The various enhancements we have discussed thus far have little or no effect on the GERAN architecture or radio network equipment. Major benefits can be derived by upgrading the network software but without the need for new hardware (Box C). Moreover, the enhancements are backward-compatible, which means legacy mobile terminals can coexist with new mobile terminals, sharing the same radio channels and making efficient use of radio resources. The enhancements also benefit services, especially where conversational services (for example, VoIP) are concerned.

Reduced latency and higher-bit-rate bearers yield efficient voice bearers on the packet side of the GSM network. User sessions are multiplexed more efficiently on the same bearer, and the mouth-to-ear delay reaches acceptable levels of around 200 to 300ms in a realistic radio environment.[7] In addition, web browsing and data downloads and uploads are faster, which further improves the user experience and is expected to benefit the use of radio resources. Figures 12 and 13 show the peak bit rate enhancements and spectrum efficiency gains.

BOX C, BENEFITS OF UPGRADING NETWORK SOFTWARE AND TERMINALS

A summary of the benefits of each individual enhancement follows below:

- Dual-antenna terminals improve capacity and mean bit rates substantially, especially when used to cancel interference in the downlink. Furthermore, receiver sensitivity is vastly improved, which improves coverage (as defined in the noise-limited case). Individual users benefit immediately. System gains are dependent on terminal penetration.
- Multicarrier EDGE yields increased bandwidth. This is manifested by increased peak and mean bit rates both over the uplink and downlink. Operators can retain the current channel structure, introducing complexity on a step-by-step basis as demand for bandwidth grows.
- Mobility enhancements increase mean bit rates and spectrum efficiency because users are always served by the most appropriate base station. Service continuity is also improved thanks to faster and more accurate handovers within GSM/EDGE networks and between GSM/EDGE and WCDMA networks.
- Reduced TTI and faster feedback reduce latency and improve overall service performance. Sessions that involve interactive data transfers will on average take less time. The enhancements also significantly reduce the mouth-to-ear delay and improve hardware efficiency for conversational services (for example, VoIP).
- Higher-order modulation used in combination with turbo coding substantially increases peak and mean bit rates. Service coverage is also improved thanks to increased robustness (when, for example, MCS-8 and MCS-9 are realized with 16QAM instead of 8-PSK, which is used in the existing EDGE standards).

Feature	Mean data rate	Peak data rate	Latency	Coverage	Spectrum efficiency
Dual-antenna terminals	x	-	-	x	x
Multicarrier EDGE	x	x	-	-	x
Mobility Enhancements	x	-	-	-	x
Reduced TTI and faster feedback	-	-	x	-	-
Higher-order modulation and turbo coding	x	x	-	-	x

Example – Conversational VoIP enhancements

Reduced TTI and faster feedback help relax the requirements put on radio transmission for delay-sensitive services by providing sufficient time for retransmitting radio blocks. In other words, the techniques make it possible to reduce the safety margins associated with the transmission of radio blocks. This thereby reduces bandwidth requirements, which increases spectrum efficiency and improves service coverage. Multicarrier EDGE makes it possible to share resources between more timeslots. Therefore, given a large pool of resources, trunking efficiency will improve. One example of such a service, with strict requirements put on delay, is conversational IP telephony, often referred to as Voice over IP or VoIP.

Figure 14 shows the normalized hardware utilization for VoIP at a given frequency load, assuming gradual introduction of

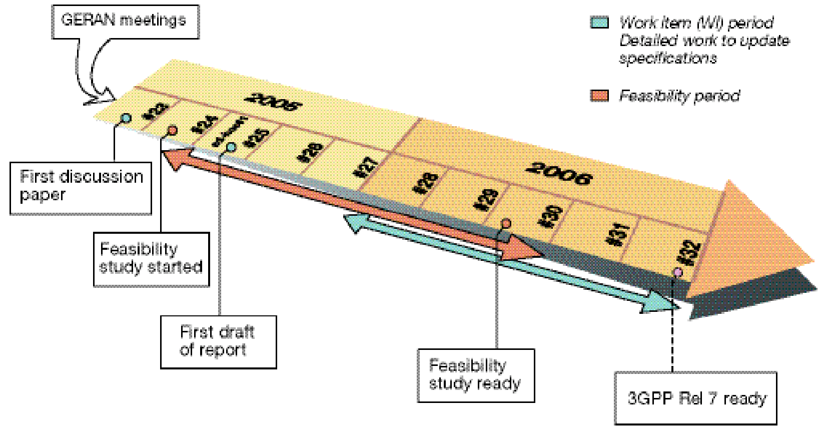


Figure 15
3GPP GSM/EDGE evolution time plan.

multicarrier EDGE, reduced TTI, and faster feedback techniques.

Standardization

All major vendors and many operators have agreed to conduct a feasibility study regarding the evolution of GERAN in 3GPP, the standardization body for GERAN. The 3GPP officially started the study in April 2005. Since then, the participants have agreed on various goals, for example, performance gains and implementation considerations relative to existing GSM/EDGE equipment (Box B, Item I). Several proposed enhancements are currently being evaluated. Apart from the mobility enhancements, nearly all the proposed enhancements described in this article are part of the current 3GPP study. Successful candidates will target Rel-7 or an earlier Work Item (Figure 15). Box B (Item II) lists other GSM/EDGE enhancements currently handled in 3GPP standardization.

Conclusion

GSM/EDGE has evolved into the most widely deployed cellular telephony standard in the world, offering high-performance voice and data services. To keep GSM/EDGE competitive, even in years to come, Ericsson has proposed steps to further evolve the standard. The combination of solutions fulfills the requirements set by 3GPP:

- Spectrum efficiency is improved, allowing more voice and data traffic in the same spectrum.
- Higher peak and mean data bit rates enable faster and better services.
- Reduced latency improves service performance, primarily for delay-sensitive services, such as VoIP.
- Coverage is improved and higher bit rates can be offered.
- Service continuity is enhanced within GSM/EDGE and between GSM/EDGE and WCDMA networks.

REFERENCES

1. Furuskär, A., Näslund, J. and Olofsson, H.: EDGE – Enhanced data rates for GSM and TDMA/136 evolution. Ericsson Review, Vol. 76(1999):1, pp. 28-37
2. GSA– the Global mobile Suppliers Association, November 3, 2005, www.gsacom.com
3. 3GPP TR 45.912 Feasibility study for evolved GSM/EDGE Radio Access Network (GERAN), (Release 7)
4. Tdoc GP-041348, 3GPP TSG GERAN, Enhancements of Handover Performance
5. Eriksson, M., Turina, D., Arslan, H., Balachandran, K., Cheng, J.: System Performance with Higher Level Modulation in the GSM/EDGE Radio Access Network, Proceedings of IEEE Globecom 2001
6. Berrou, C., Glavieux, A. and Thitimajshima, P.: Near Shannon limit error-correcting coding and decoding: turbo codes, Proceedings of IEEE International Communications Conference '93, pp. 1064-1070, May 1993.
7. ITU-T G.114
8. Bakhuizen, M. and Horn, U.: Mobile broadcast/multicast in mobile networks. Ericsson Review, Vol. 82(2005):1, pp. 6-13
9. Bäckström, M., Havdrup, A., Nylander, T., Vikberg, J. and Öhman, P.: Mobile@Home—GSM services over wireless LAN. Ericsson Review, Vol. 82(2005):2, pp. 92-99