Fifty years with mobile phones
From novelty to no. 1 consumer product

Olof Billström, Lars Cederquist, Magnus Ewerbring, Gunnar Sandegren and Jan Uddenfeldt

This year marks fifty years since Ericsson introduced the world's first automatic mobile telephone system for a few hundred subscribers in Stockholm and Göteborg (Gothenburg), Sweden. The associated mobile units were colossal 40kg blocks of radio equipment designed for installation in the baggage compartment of a car.

The extraordinary success of mobile communications, from those early pioneering efforts to today's miniature telephones and advanced services – the result of global technical labors of the highest art – reads like a page from a storybook. From the outset, Ericsson has been a major player, driving many facets of development and helping to create the standards that prevail in the industry today.

The expertise Ericsson gained from its first systems, and a steady ambition to create an optimum, automatic network, laid the groundwork for a vision that has stood the test of time. Ericsson continues to forge ahead as a technical leader in the field of mobile communications – perhaps more so now than ever before thanks to high-speed packet access (HSPA), the turbo version of the 3G network.

1956: 0G – the first networks

Enthusiasts at Ericsson and elsewhere had long nurtured the idea of creating a radio-based mobile communications system. In the mid-1950s, a handful of systems were tested in several countries around the world. For Ericsson's subsidiary, Svenska Radioaktiebolaget (SRA), this development was a natural next step – it had been selling private mobile radio systems to the police and other groups since the 1940s. Now, users would dial numbers on a telephone to call other telephones.

What set Ericsson's system apart from the others was that it was fully automatic – it did not require manual control of any kind. The system, called MTA, operated in the 160MHz band using pulsed signaling between the terminal and base station (MTB, an upgrade introduced in 1965, used dual-tone multifrequency signaling). SRA developed and supplied the base stations and mobile units (Figure 1).

Initially, mobile telephony was synonymous with car phones or voice communication over a mobile radio in a car. It did not matter, therefore, that the equipment was bulky or heavy. The equipment was also power hungry, because for signals from the mobile unit to reach the base station, the output power from the mobile unit had to match that of the base station.

The Swedish Telecommunications Administration (Telia, formerly Televerket) took the first MTA system into operation on April 25, 1956, but real system tests had begun two years earlier, in 1954. During the 1960s, Televerket simultaneously operated two systems. The two were more or less identical except for a minor difference in signaling and how users connected to the network.

Each system consisted of one transmitter and several receivers and was built using relays and radio tubes. Four duplex channels enabled the systems to handle approximately 100 users apiece. The majority of subscribers were salaried professionals, such as doctors or lawyers. The main system shortcomings were limited coverage and subscriber density. Ultimately, Televerket concluded that it would have to take a radical new approach in order to successfully create a national or multinational mobile network.

1981: 1G – the analog era

NMT (Nordic Mobile Telephone) was the first truly mobile system. Ericsson and Televerket deployed this open standard in Scandinavia in 1981. At about this same time, Ericsson also supplied a turnkey system, consisting of switches, base stations and mobile units, to Saudi Arabia.

Shortly afterward, Bell Labs in the USA developed what came to be known as AMPS (Advanced mobile phone system). The first AMPS system was deployed in 1984. One other system, TACS (Total access communication system, a modification of AMPS), was developed and inaugurated in Great Britain in 1985. Ericsson manufactured equipment for each of these standards and quickly grabbed a 40% global market share of analog systems. In fact, Ericsson was the first supplier to take a TACS system into operation (for Racal, currently Vodafone).

These early systems were all analog systems based on frequency-division multiple access (FDMA) technology, which allocates a distinct frequency channel to each user. Today, we think of these standards as first-generation (1G) mobile systems, but in their day they were considered fourth-generation systems. We would remind the reader that the systems were designed for telephony from cars – apart from the handset, all mobile subscriber equipment was installed in the baggage compartment of a car.

It took nearly ten years to develop NMT, mainly because system designers wanted to make the most of coming finesses in microelectronics: microcircuits, digital frequency synthesis (to combine radio channels in a terminal), digital signaling, digital switching methods (which were necessary for managing and controlling telephone switches and rapid change-over during handover between cells), and so on.

The system was composed of mobile stations, base stations, and telephone switch equipment. The mobile stations, essentially small radio stations with keypads, were often packaged in a module that could be mounted in the cockpit or dashboard of a car. The telephone switches consisted of a subsystem (called MTS) of AXE. The base stations could thus connect to the switch via ordinary

Figure 1
The mobile unit for MTA, the world's first automatic mobile telephone system.

Figure 2
The size of mobile handsets quickly diminished between 1981 and 1998 thanks to rapid advances in microelectronics and digital tecchnology. In particular, GSM gave rise to pocket-sized mobile phones.

telephone lines. And thanks to their modular design, the AXE switches could coordinate mobile and public transit traffic. Initially, the AXE switches were considered to be too powerful (and expensive) for mobile systems, but rapid network growth soon validated this design choice.

To provide coverage to large areas, the NMT network was built from a series of cells (a technique developed by Bell Labs) with one base station per cell. Topography and the anticipated number of subscribers determined the actual number of cells and location of base stations. The base stations associated with a given telephone switch constituted a service area.

To prevent interference, neighboring base stations transmitted on separate frequencies. One channel in each base station was designated a calling channel; the other channels were used for voice traffic. In small NMT base stations, the calling and traffic channels could be combined. A base station's range depended on mast height and topography. The range was often about 20km.

The system employed duplex channels: one for transmission (TX) and one for reception (RX) with 4.5MHz of bandwidth in each direction and 25kHz of separation between adjacent channels. This gave the system a total of 180 channels. Duplex transmission and reception were separated by 10MHz.

Handover and roaming are two techniques at the very heart of mobility. Handover entails passing off calls between base stations. Roaming entails keeping track of users when they are not connected so the system can know where to place a call. Mobile radio solutions had previously employed DTMF signaling to manage handover and roaming, but Ericsson's solution (adopted by the industry for the NMT standard) introduced digital signaling, which was faster and less complex. The system also supported international roaming.

Two versions of the NMT standard were developed:

- NMT 450 for the 450MHz frequency band; and
- NMT 900 (in 1986), which used smaller cells for the 900MHz frequency band. These new frequencies put new technical requirements on radio network planning. The NMT 900 extension primarily responded to a need for more channels, to accommodate more subscribers.

1991: 2G – transition to digital technology

As the 1970s were winding down (and before the analog systems had been taken into operation) the industry foresaw the need for solutions with much greater capacity, paving the way for smaller, more energy-efficient base stations and mobile terminals. In addition, Ericsson foresaw that digital mobile telephony would improve spectrum efficiency, compared with analog FM transmission, and allow for the introduction of microcells. New frequencies, two blocks of 25MHz, were reserved in Europe for mobile telephony in the 900MHz band.

In 1982, a Scandinavian and Dutch initiative in CEPT (European Conference of Postal and Telecommunications Administration) started a *Group Speciale Mobile* (GSM) to define and specify a pan-European mobile system for these new frequencies. At the same time, Ericsson and Televerket began working on the digital system that would become GSM. Ericsson completed its first experimental system in 1984. With a 16kbps voice coder, the system proved that digital technology yielded considerably more capacity than analog systems. The work gave rise to very advanced voice coding, channel coding and modulation techniques in 1985. The GSM group was transferred to ETSI (European Telecommunications Standards Institute) when participating operators, manufacturers and other organizations formed it in 1988.

GSM was to have digital radio transmission and open, clearly defined interfaces, so that operators could select equipment from different manufacturers. Initially, GSM was intended to be a pan-European mobile network, but it became a world standard when Telstra (Australia) signed a memorandum of understanding in 1992.

A main theme of Ericsson's was that GSM would facilitate pocket-sized battery-driven telephones. And, sure enough, GSM was a breakthrough for handheld mobile phones, and thus the first global mobile system to enjoy really substantial growth (Figure 2).

As its access technology, GSM combined time-division multiple access (TDMA) with FDMA, which had been used in analog systems. TDMA divides the digitally modulated carrier into a number of timeslots for transmitting compressed information. This approach proved to have several advantages, including greater traffic capacity, frequency-hopping capability to combat fading, and signal strength measurements made in terminals to facilitate handover. The decision to go with TDMA also greatly reduced costs as well as the weight and volume of handsets.

The next choice, in 1987, was between broadband TDMA (2MHz) and narrowband TDMA (200 to 300kHz). A decisive factor was that narrowband TDMA facilitated true pocket-sized telephones whereas broadband TDMA would require the terminals to have considerably greater output power, which in turn, would require cooling of some kind and much larger batteries.

Ericsson participated in a comprehensive field trial in Paris in December 1986, using, in contrast to a number of its competitors, the narrowband TDMA system that later came to serve as the basis for the GSM standard (Figure 3). Once the standard was set, Ericsson had a clear head start over its competitors, supplying its first GSM network to Mannesmann (currently Vodafone) in Germany, in 1991, and handheld mobile phones, also to Mannesmann, in 1992.

The GSM system parameters were set as follows:
- 200kHz channel width; and
- digital transmission at 22.8kbps – 13kbps was to be used for voice and the rest for channel coding.

In addition, the system would employ slow frequency hops and time multiplexing with eight timeslots. Half-rate voice coders were later introduced with close to 8kbps for voice, making it possible to double the number of voice channels over the same bandwidth.

Two additional digital systems were developed after GSM:
- Digital AMPS (D-AMPS, later called TDMA). The system, which was compatible with AMPS, was first taken into operation in 1992; and
- PDC (first called JDC). The system was developed and deployed in Japan in 1993.

These systems, which were based on TDMA, had three timeslots but used different channel bandwidths.

At the same time, another standard was being developed in the USA: IS95. This standard borrowed its radio technology, code-division multiple access (CDMA), from the military. In essence, CDMA differentiates between users by assigning distinct codes to each. IS95, which used a 1.25MHz carrier, later became known as narrowband CDMA (compared with wideband CDMA, which has a 5MHz carrier).

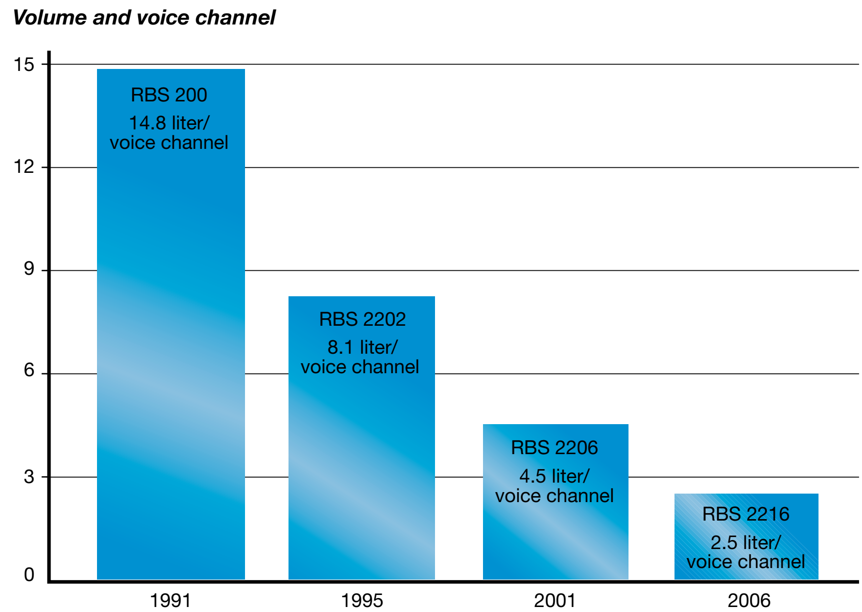
Ericsson played a decisive role in the struggle between TDMA and FDMA for D-AMPS. There was not any available frequency band in the USA for D-AMPS. Instead, the digital system would have to be integrated into the existing analog AMPS network. Ericsson proposed a TDMA system with a channel width of 30kHz (the same as for analog AMPS). By means of live demonstrations of the TDMA solution in Los Angeles, Ericsson convinced TIA of the system's advantages. The first version of TDMA shared a control channel with AMPS. However, to extend battery life, a digital control channel was introduced in later generations.

The GSM system architecture differed somewhat from that of NMT. It consisted of a mobile switching system (MSS) with mobile switch and home location register (HLR) and visitor location register (VLR), a base station subsystem (BSS) with groups of base stations connected to a base station controller (BSC), and an operations support system (OSS). And then there were the mobile terminals, of course, which were now either handheld or installed in cars.

The capacity of the GSM system could be increased with smaller cells and more base stations. This quickly led to the allocation of new frequency bands at 1800MHz and 1900MHz for small cells. The base stations were initially nearly two meters tall and weighed several hundred kilograms. But rapid advances in technology led to small stations that were easy to install and maintain (Figure 4). Microstations were also developed for indoor and outdoor usage. These could transmit with low power and have a range of a few hundred meters (compared with several kilometers for a macrostation).

Figure 3
Two versions of the first GSM mobile phone: prototype and commercial handset.

Figure 4
Four generations of GSM base stations.



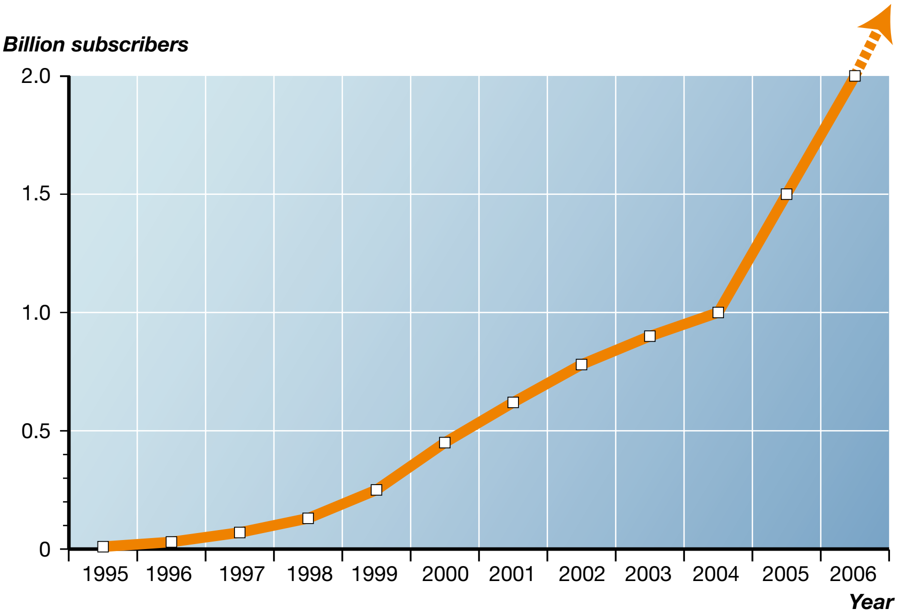
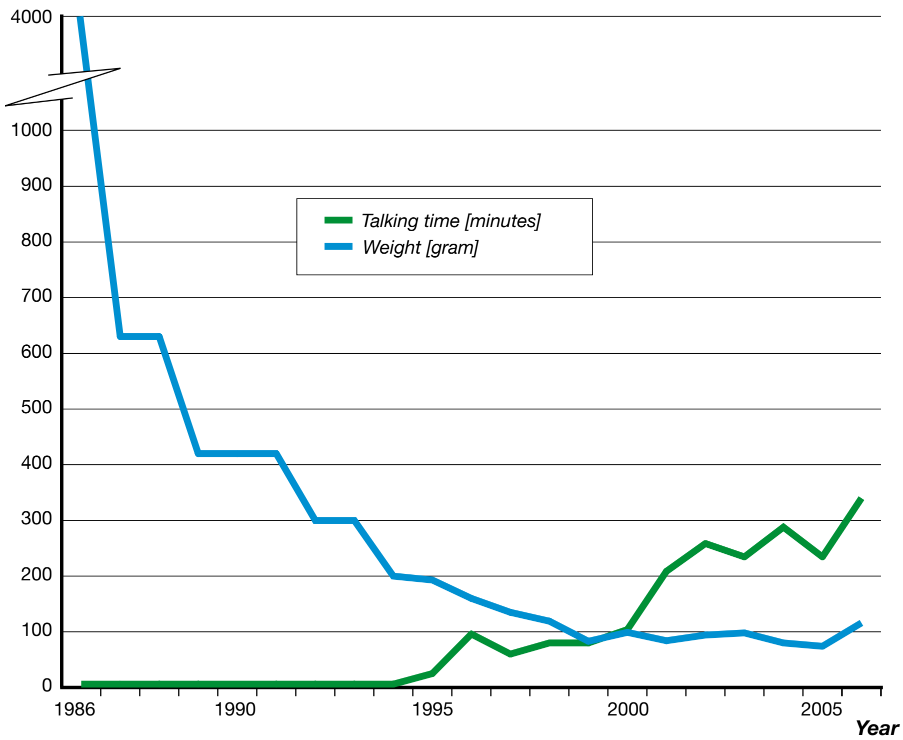


Figure 5
GSM and UMTS subscriber growth.

Figure 6
Evolution of mobile handsets, size (weight) and talking time.



From the outset, GSM had been specified to support international roaming, and as a consequence it became a world standard. This, in turn, was a major breakthrough for the pocket-sized handheld telephone. Today, in 2006, GSM has been deployed in more than 100 countries and is used by more than two billion subscribers (Figure 5). Moreover, more than 800 million new GSM phones are sold each year.

Although the first version of GSM was optimized for voice, it could support circuit-switched data transmission at rates of up to 9.6kbps. A modest rate by today's standards, but this was the starting point for mobile data communication and access to the internet.

2000: 2.5G – from circuit-switched to packet data

GSM continued to evolve, yielding greater capacity and facilitating new services. One method of increasing data transmission rates was to use multiple timeslots in circuit-switched transmissions. With the introduction in 1998 of high-speed circuit-switched data (HSCSD) GSM could offer data transmission rates of up to 57kbps.

But the real breakthrough came in 2000, with the transition to packet-data technology: general packet radio service (GPRS). This solution reused as much of the circuit-switched GSM network as possible – frequencies, base stations and BSC – while complementing it with new base station software and two new nodes for packet data:

- a packet-data switch called serving GPRS support node (SGSN), which corresponds to the mobile switch in circuit-switched networks; and
- a gateway GPRS support node (GGSN) for accessing the internet.

Ericsson anticipated that IP networks would be used for transport. Accordingly, its mobile phones were also designed for internet access.

GPRS technology reserves at least one timeslot for packet-data traffic but can use as many as four or five. Therefore, if more timeslots are available, GPRS can yield data transmission rates of up to 100kbps.

In 1993, primarily to improve GSM data rates, Ericsson began researching a new modulation method called enhanced data rates for global evolution (EDGE). The regular approach to modulating the radio carrier, Gaussian minimum-shift keying (GMSK), was based on phase and amplitude modulation using fixed amplitude and two phase modes, so that each mode corresponded to one bit. The new modulation scheme, called eight-phase shift keying (8PSK), used eight modes and three bits per symbol, yielding a three-fold increase in the bit rate.

The GSM core network was also updated to support multimedia, IP, and streaming video. Likewise, its interface toward the BSS was modified to serve as the basis for UMTS, the 3G standard that would constitute the next phase of evolution.

2001: 3G – wideband coding technology

The transition to 3G and new technologies in the mobile network had been driven by user demand for access to the internet and data services, such as e-mail and multimedia. Networks were thus being optimized to deliver packet data instead of for circuit-switched transmission.

The main approach to 3G was wideband (5MHz) code-division multiple access (WCDMA) in the 2GHz frequency band. In hindsight, one can see that WCDMA (the technology that Ericsson, among others, has promoted, and the path selected in Europe to realize 3G) delivers the best performance of any alternative. Simply put, to reach really high data rates, one needs a broader carrier. At the same time, EDGE was developed for GSM, and in the USA, narrowband cdmaOne (1.25MHz) evolved into CDMA2000.

As with GSM, Ericsson was first out with WCDMA, in large part because it had begun fundamental research on WCDMA even before GSM had been taken into operation. In 1995, Ericsson had both a test system and a prototype terminal that was the same size as the GSM prototype had been (Figure 3). In other words, it too could be reduced to a pocket-sized phone. Shortly thereafter, Ericsson entered into strategic collaboration with NTT DoCoMO in Japan. This led to delivery, in 1997, of a pre-commercial system that would also be used for comprehensive field trials.

Two other factors further strengthened Ericsson's leading position in 3G:

- Ericsson initiated a new international forum for 3G, the Third Generation Partnership Project (3GPP), which united ETSI in Europe with ARIB in Japan and gave the industry a timely push. By managing standardization issues and in-house product development in parallel, Ericsson could save considerable time.
- Ericsson opted to base its radio nodes (base stations and RNCs) on a new platform (CPP) that was completely packet-based, flexible, and could handle very high data rates. CPP laid the groundwork for ATM and IP transport. Today, CPP serves as a solid and stable platform for very advanced 3G products, such as turbo 3G (HSPA), and the future, long-term evolution of 3G (LTE).

WCDMA uses a 5MHz carrier (compared with 200kHz for GSM), which in its first incarnation, provided data rates of up to 384kbps. This is enough to support video telephony and internet browsing. The broader carrier also eliminates the problem of fading (a weakness of narrowband code-division technology) by distributing interference throughout the entire band.

WCDMA technology is based on code-division techniques (as opposed to frequency division in time) and makes very effective use of spectrum. Code-division techniques enable a large number of calls to be sent and mixed over the same 5MHz channel without noticeable interference. Each call is differentiated by means of a distinct code, which the recipient decodes and filters. Moreover, because all calls are on the same frequency, operators need not plan or optimize network frequency.

One technical challenge with CDMA is power control. Simply put, signals from mobile terminals at different distances from the base station must all have the same strength at the antenna. The solution is very fast control algorithms.

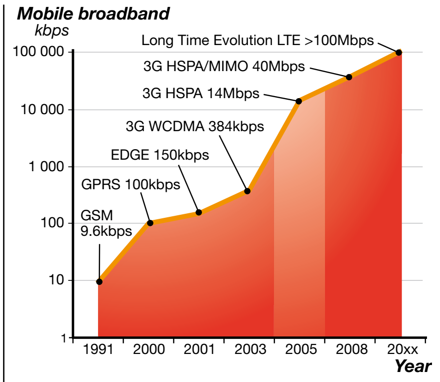


Figure 7
Evolution of mobile data bit rates from GSM to LTE.

One other complication is time dispersion, a phenomenon that arises when signals are reflected on their way to the base station. Here the solution calls for complex Rake receivers.

In addition, optimized signal management is needed in the base station to reduce power and heat.

Other important pieces of the puzzle include a good voice coder and a robust video coder that can handle everything from slow sequences to rapid exchanges.

Accordingly, the new WCDMA concept called for entirely new base station architecture. Previously, the main focus was on channels; that is, the base stations mainly consisted of several transceivers (transmission and receiver units). By contrast, WCDMA employs a function-oriented architecture that manages all processor resources in a

BOX A, A PHONE IN EVERYBODY'S POCKET

Two factors have played a decisive role for the success of mobile telephony: the size and price of mobile phones. Ericsson introduced the world's first pocket-sized mobile phone for NMT 900 in 1987. And during the next ten years it would shrink the size of its mobile phones by 80% (Figure 6). Performance, on the other hand, increased one hundred fold, opening the door to a variety of advanced new services (Figure 7). This evolution is still far from over. The first terminals were telephones with enough intelligence to connect calls, but today's terminals are actually powerful handheld computers.

Compare the NH 72 (1989) with the Sony Ericsson W850i (2006):
The Ericsson NH 72 was a 420g mobile terminal for NMT voice service. A unique feature in its day enabled users to adjust the volume (loudness) of the handset. The Sony Ericsson W850i Walkman (Figure 8), by contrast, is a 3G terminal that weighs 116g, works on five frequency bands and two systems and supports GPRS, Bluetooth, WAP browsing, e-mail, SMS, MMS, Java, FM radio, games, music, video, voice memo, a camera, and much, much more.

Figure 8
The W850i Walkman from Sony Ericsson.

pool, to enable dynamic bandwidth and allocation of capacity (from only a few kilobits per second to 14Mbps depending on service), and to combine circuit-switched data with packet data. The capacity of Ericsson's first WCDMA base station was more than three times that of contemporary GSM base stations.

WCDMA also requires a transport network that can deal with variable bandwidth. To begin with, asynchronous transfer mode (ATM) technology was used between the switch and base station. ATM, a circuit-switched variable packet link of sorts (that is, a cross between pure circuit-switched technology and pure packet-data transmission), gives operators a very flexible and economic way of transmitting data. Today, however, most networks are based on or evolving toward an IP paradigm.

Ericsson delivered the first WCDMA systems to *3* in Italy and to J-Phone in Japan. Ericsson was also first, in 2005, to deliver HSPA (the turbo variation of 3G) to Cingular in the USA and Vodafone in Japan.

Turbo 3G and LTE

The evolution did not end with 3G. High-speed 3G or high-speed packet access (HSPA), which is mainly a software upgrade of WCDMA, further increases radio spectrum efficiency 200% to 300%. Compared with first-generation WCDMA, HSPA improves mobile network data rates to a level of true broadband. At present (2006), HSDPA delivers 4Mbps and this figure will continue to improve.

Operators around the world are now rapidly installing HSPA technology in their 3G networks. The idea behind HSPA is to exploit current radio conditions (in intervals measured in microseconds) for each mobile terminal in a given area, by sending data to the mobile terminal that enjoys the best reception. Optimum data bit rates are selected using link-adaptation technology.

The objective is to make optimum use of available power by transmitting as much data with as little power as possible. This way, HSPA supports up to four times more traffic per cell in the network than the first release of WCDMA. HSPA also considerably reduces network latency. If GPRS had a delay of approximately 1s (early version), then by comparison, WCDMA would be close to 150ms, and HSPA, 70ms. This increase in performance sets users up for a completely new experience, putting mobile broadband on par with fixed-line ADSL (at least).

And the introduction of multiple input, multiple output (MIMO) antenna technology in base stations and mobile phones will bring about further exponential increases in data bit rates. Ericsson has already begun testing MIMO in the field – as usual, long before the technology is ready for commercial use.

The idea is to push HSPA technology to its limits, and tests show that one can reach data bit rates of 40Mbps with 5MHz of radio bandwidth. At the same time, Ericsson foresees an evolution to a 20MHz carrier, which will yield transmission bit rates of up to 200Mbps and flexible use of available bandwidth. Work on the long-term evolution of 3G (LTE) is already underway (Figure 6). And as always, Ericsson is conducting early system tests to gain first-hand knowledge of this new technology.

Operators and suppliers are currently discussing the systems, their very high peak data rates, and large bandwidths in relevant standardization bodies.

Mobile communications have evolved very rapidly since the introduction, in 1956, of a car-borne, walkie-talkie-like mobile system for a few hundred subscribers. At that time, nobody had any idea how fast this concept would catch on or evolve. Even the most optimistic predictions fell far short of the truth.

TERMS AND ABBREVIATIONS

3GPP	Third Generation Partnership Project
8PSK	Eight-phase shift keying
AMPS	Advanced mobile phone system
ARIB	Association of Radio Industries and Businesses
ATM	Asynchronous transfer mode
BSC	Base station controller
BSS	Base station subsystem
CDMA	Code-division multiple access
CEPT	European Conference of Postal and Telecommunications Administration
D-AMPS	Digital AMPS
DTMF	Dual-tone multifrequency
EDGE	Enhanced data rates for global evolution
ETSI	European Telecommunications Standards Institute
FDMA	Frequency-division multiple access
GGSN	Gateway GSN
GPRS	General packet radio service
GMSK	Gaussian minimum-shift keying
GSM	Group Speciale Mobile
GSM	Global system for mobile communication
GSN	GPRS support node
HLR	Home location register
HSCSD	High-speed circuit-switched data
HSPA	High-speed packet access
IP	Internet protocol
LTE	3GPP long-term evolution of 3G
MIMO	Multiple input, multiple output
MoU	Memorandum of understanding
MSS	Mobile switching system
NMT	Nordic mobile telephone
OSS	Operations support system
PDC	Personal digital communication (first called JDC)
RNC	Radio network controller
RX	Receiver
SGSN	Serving GSN
SRA	Svenska Radioaktiebolaget
TACS	Total access communication system
TDMA	Time-division multiple access
TX	Transmitter
UMTS	Universal mobile telecommunications system
VLR	Visitor location register
WCDMA	Wideband CDMA