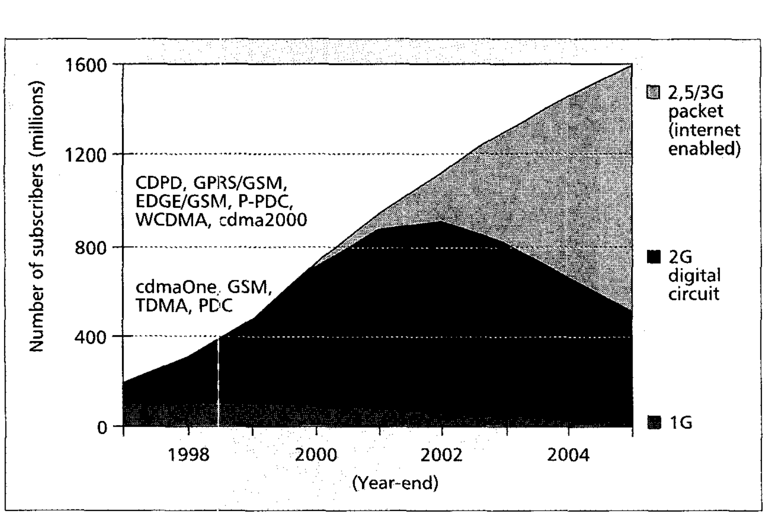
Abstract

Starting with today's 3G standards, future-generation wireless networks are discussed. Two complementing major trends are identified: seamless roaming between different air interfaces, leading to the Always Best Connected concept, and the continuous development of the current third-generation standards. The evolution of WCDMA toward high-speed downlink packet access, aiming for peak rates in the order of 8–10 Mb/s, is described as an example of air-interface evolution. Fourth-generation technologies such as *ad hoc* networking and multihop networks, still at the research level, are discussed and their impact on wireless communication systems addressed.

Future-Generation Wireless Networks

MAGNUS FRODIGH, STEFAN PARKVALL, CHRISTIAAN ROOBOL, PER JOHANSSON, AND PETER LARSSON
ERICSSON RESEARCH

The growth of mobile communication is explosive and at the end of 2000 the number of subscribers exceeded 700 million. So far mostly speech services have been offered, but with the introduction of packet-switched radio systems (GPRS) and third-generation (3G) systems starting in 2001, users will be able to select from a vast menu of mobile data services. The addition of mobility to data communication enables new services not meaningful in a fixed network, e.g., positioning-based services. However, the development of mobile services has only started and in the future we will see new application areas opening up. In this article, future-generation systems are discussed, starting with a description of current trends and a brief overview of the cellular systems of today. The focus of the article is on the access networks. We consider the evolution of current systems, exemplified by a description of possible improvements in the packet-data support in wideband CDMA (WCDMA), and more long-term 4G research such as ad hoc and multihop networking. Clearly, due to its nature a description of the long-term research on 4G is less detailed than the sections covering the evolution of the 3G systems.



■ **Figure 1.** *Cellular subscribers by standard generations.*

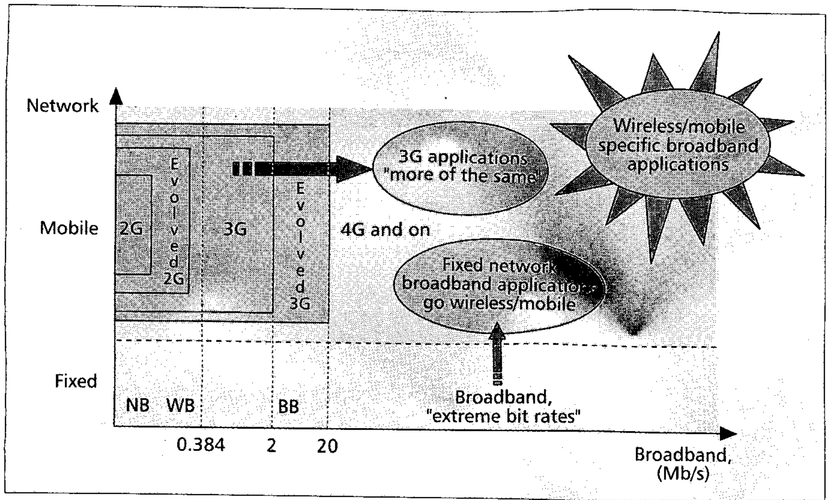
General Trends

A growing portion of cellular traffic is data traffic, and we can see the first signs of a mobile Internet. More than 8 billion SMS messages being sent during the single month of June 2000, and the fact that Mobitex grew more in 1999 than in all previous years combined, are just two examples of this trend. In Japan the i-mode service had approximately 17 million subscribers at the end of 2000, with an increase of more than 1 million subscribers per month. The number of WAP-enabled terminals in circulation exceeds 50 million, and more then 10000 WAP sites are available in 95 countries. These are just a few examples of the increasing popularity of data services and an indication that cellular services will be used for more than speech. Further, 2001 is the year for the worldwide introduction of third-generation systems and the mobile Internet with a continued increase in subscribers. At the end of 2005 we estimate there will be 1.6 billion users, of which more than 1 billion will be 3G mobile internet users, as shown in Fig. 1.

With the introduction of data services, the usage measured in min/month/subscriber is also increasing from today's figure of approximately 200 to an estimated 600 min/month/subscriber at the end of 2005. The increased usage is partly explained by traffic moving from fixed to mobile networks, but also by to the introduction of new services. These new services can be divided three main classes:

- Social communication and safety, with examples such as video-telephony, sending photos, messaging, alarm notification, and emergency location.
- Time saving and empowerment, exemplified by shopping, banking, WWW information search, news, remote control of home, and navigation.
- Fun and pleasure, for example gambling, games, sports information, and music.

Taking a broader perspective, we see the converged telecom industry being formed by adding Internet and multi-services to wireless and mobility. New applications will be developed, taking advantage of the combination of



■ **Figure 2.** *Evolution of applications toward 4G.*

high data rates and mobility, as illustrated in Fig. 2. In addition, new value chains are added by the integration of e-commerce, e-trade, e-ads, and e-entertainment.

Mobile Generations

After the first-generation analog mobile systems, the second-generation (2G) mobile systems were introduced to the market around 1991. The shift from analog to digital and the rollout of new systems made the shift to the second generation very clear. The second-generation systems offered higher capacity and lower costs for network operators, while for the users, short messages and low-rate data services were added to speech services. Wide area roaming is another advantage, especially for GSM, which is available in almost all parts of the world. Presently, the 2G systems are GSM, TDMA, PDC, and cdmaOne, found to the left in Fig. 3. TDMA, cdmaOne, and GSM are all used in the U.S. GSM is used in most parts of the world except in Japan, were PDC is the second-generation system used.

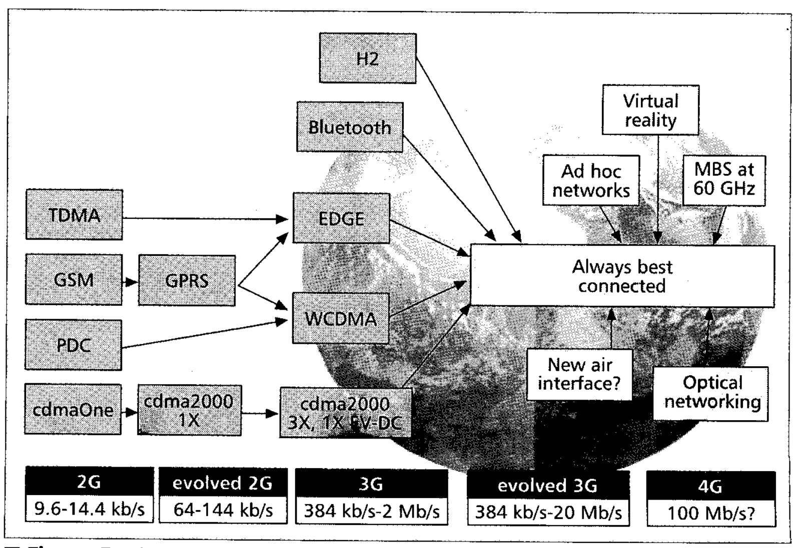
A good example of an important evolution of the 2G systems, sometimes known as 2.5G, is the ability to use packet-switched radio connections over the air. For GSM systems, the packet-switched solution is General Packet Radio Service (GPRS). The main investment for the operators lies in the new packet-switched core network, while the extensions in the radio access network mainly are software upgrades. For the users, GPRS offers the possibility to always be online and only pay for the data actually transferred. Data rates of up to 20 kb/s per used time slot will be offered, and with multiple time-slots per user in the downlink, attractive services can be offered.

The shift to third-generation in the radio access networks is presently ongoing. The worldwide introduction of WCDMA will take place in 2001 and 2002, starting in Japan and continuing in Europe. In the U.S., several 3G alternatives will be available. GSM and TDMA operators can evolve toward EDGE, with WCDMA as a possible further step, while cdmaOne operators can evolve toward cdma2000 systems.

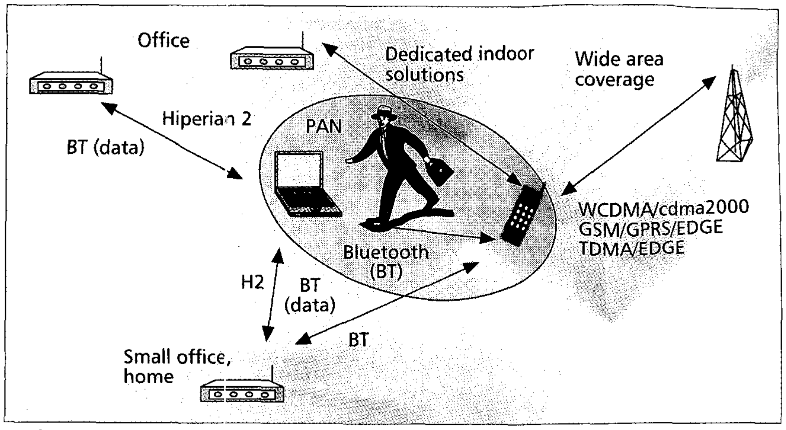
WCDMA, as specified by the third-generation partnership project (3GPP), is a 3G system operating in 5 MHz of bandwidth. Variable spreading and multicode operation is used to support a multitude of different radio access bearers. Different service classes are supported by an advanced quality-of-service (QoS) support. Data rates up to 384 kb/s for wide-area coverage and up to 2 Mb/s for local-area coverage are provided. For a system evolved as described in a later section, considerably higher peak rates are foreseen.

EDGE is an evolution of GPRS with data rates of up to 60 kb/s per time-slot together with improved spectrum efficiency. EDGE uses higher-order modulation together with link adaptation and incremental redundancy to optimize the radio bearer to the radio connection characteristics. Currently, additions in the form of a new set of radio access bearers to align EDGE toward WCDMA are being standardized within R5 of the 3GPP standards and are expected to be ready by the end of 2001. The same service classes as in WCDMA and the same interface to the core network will be used, the so-called Iu interface, shown in Fig. 5.

cdmaOne has evolved into cdma2000, and is available in two flavors, 1x and 3x. The former uses the same 1.25 MHz bandwidth as cdmaOne and supports up to approximately 600 kb/s, while the latter is a multi-carrier system using 3.75 MHz and supporting up to approximately 2 Mb/s. At the moment, the focus on 3x is very limited. As a complement to 1x, the standardization body 3GPP2 has recently specified 1xEV-DO (1x EVolution-Data Only). 1xEV-DO uses a separate 1.25 MHz carrier and supports best-effort data traffic only, using a new air interface compared to cdma2000. Voice traffic has to be carried on a separate cdma2000 carrier. The peak rate in the 1x EV-DO downlink is almost 2.5 Mb/s, excluding overhead. Phase two of the 1x evolution, known as 1x EV-DV (1x EVolution-Data and Voice), is currently being discussed within 3GPP2 and there are a number of proposals under consideration. The purpose is to specify an extension to cdma2000 1x in order to support high-rate data and voice on the same carrier.



■ **Figure 3.** *The path to future generation mobile systems.*



■ **Figure 4.** *Personal area network and always best connected.*

At the same time as 3G standards are being introduced, other air interfaces have been developed and standardized. First of all, Bluetooth is already available, enabling devices to communicate over short distances. The strength of Bluetooth is low power consumption and a design enabling low-cost implementations. Bluetooth will be integrated into mobile phones, laptop computers, PDAs, and so forth. The first version of Bluetooth will offer up to 700 kb/s, but higher data rates, up to approximately 10 Mb/s, are currently being standardized for later releases.

Wireless local area networks (WLANs) based on the different versions of the IEEE 802.11 standard have been around for some years in the 2.4 GHz ISM band. Data rates up to 11 Mb/s with reasonable indoor coverage have been offered. The next step in the WLAN evolution is the new systems developed for the 5 GHz band. Products based on two different standards, Hiperlan 2 (H2) and IEEE 802.11a, will be available starting in early 2002. The physical layers of the two are more or less identical, with a carrier spacing of 20 MHz, OFDM modulation, and data rates up to 54 Mb/s. The difference is the MAC protocol, where Hiperlan 2 has a more advanced protocol supporting QoS and mobility in a consistent way.

Having defined WCDMA, EDGE, and cdma2000 as 3G technologies, and having recognized the presence and importance of Bluetooth and WLAN, we can start discussing future-generation mobile systems [1]. Here we foresee two trends complementing each other. First, seamless roaming and handoff possibilities between the above mentioned air interfaces, leading to the Always Best Connected (ABC) Concept. The ABC is further described in a later section. The second trend is the continuous development of the existing air interfaces. WCDMA, Hiperlan 2, and Bluetooth are all just in their beginning stages, with the first versions released in 2001–2002. Hence, the potential for improvements is large and many exiting new features will be standardized and available in products in the medium time frame, 2003-2005. In this article we will focus on some aspects of the evolution of WCDMA, further described later.

Beyond 3G evolution are 4G mobile communication systems. We believe that 4G will build on 3G and evolved 3G systems. At the moment different issues are being discussed in the research community and it is still unclear what will characterize 4G systems. One possible application driving the need of 4G systems could be augmented (virtual) reality applications [2, 3], requiring high bitrates both in the radio interface and in the fixed network. Air-interface issues, ad hoc networking, and multihop networks are used as examples of 4G research and are further elaborated upon in a later section. Finally, we present a summary and some general conclusions.

Always Best Connected and PAN

Picture a user carrying a number of personal electronic devices such as mobile phones, laptop computers, PDAs, and so on. Today these devices are used independently of each other. Imagine that the devices could interact with each other, for example sharing documents between laptops at a meeting or delivering e-mails to the PDA instead of the stationary computer. If the devices were equipped with wireless communication, for example a Bluetooth connection, they could form a personal area network (PAN) with a typical scenario illustrated in Fig. 4. In the scenario envisaged, a user will at least carry a cellular terminal and a PC or PDA, but other devices such as printers and cameras could also be integrated into the PAN. The mobile terminal will be used to connect to a wide-area coverage radio access network, for example, a WCDMA or a GSM network, or both.

In addition to the wide-area coverage connection of the terminal, a PC can also be connected to the cellular network by means of the Bluetooth connection between the PC and the terminal. Thus, wide-area coverage is provided to a PC by means of Bluetooth and 2G/3G systems. When the user moves into an office, indoor coverage is obtained by the wide-area cellular infrastructure as well, or by dedicated indoor systems based on either the outdoor cellular technology or WLAN standards.

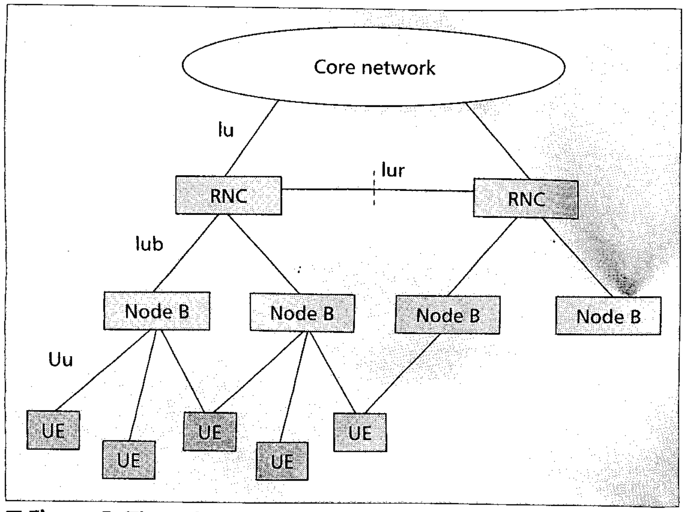
Bluetooth access points can be used for medium data rates. Currently, Bluetooth supports user rates up to 700 kb/s, but this figure will be increased in future releases. Bluetooth will work well for connections over limited ranges, on the order of 10 meters. For high data rates, WLAN radio access networks such as Hiperlan 2 can be used. Hiperlan 2 supports data rates from about 8 Mb/s up to 54 Mb/s, depending on the location of the user with respect to the access port. Additionally, Bluetooth data access could be housed in Hiperlan 2 WLAN access ports at a low additional cost and supporting an even wider range of devices. In small offices or in home environments, Bluetooth access ports alone can be an inexpensive alternative to connect to the network. To the user, the switching between the different air-interface technologies will be invisible. Hence, the user can always be assured of having the best possible connection to the data services requested.

WCDMA Evolution

Wideband CDMA (WCDMA) is rapidly emerging as the leading global 3G (IMT-2000) standard, and products based on the first release of the standard, R99, have just started to appear on the market. However, the standard will continue to evolve for several years and in this section two possibilities for this evolution are described: high-speed downlink packet access [4] and service negotiation.

High-Speed Downlink Packet-Data Access

The current WCDMA specification fully satisfies the IMT-2000 requirements, including support of data rates up to 2 Mb/s in indoor/small-cell-outdoor environments and up 384 kb/s with wide-area coverage, as well as support for both high-rate packet data and high-rate circuit-switched data. However, to satisfy the future demands for packet-data services outlined in the previous sections, there is a need for a substantial increase in this figure, especially in the downlink, and high-



■ **Figure 5.** *The architecture of a WCDMA system.*

speed downlink packet access (HSDPA) is currently a work item for release 5 within 3GPP. The main goal of HSDPA is to allow WCDMA to support downlink peak data rates in the range of approximately 8-10 Mb/s for best-effort packet-data services, that is, far beyond the IMT-2000 requirement of 2 Mb/s. Furthermore, HSDPA should also enhance the WCDMA packet-data capabilities in terms of lower delay and improved capacity.

Many new technologies are currently being considered for this first step of WCDMA development, including higher-order modulation and fast link adaptation, fast hybrid ARQ, fast scheduling, fast cell selection, and multiple-input-multiple-output (MIMO) antenna solutions.

Higher-order modulation, e.g., 16-QAM and 64-QAM, provides higher spectral efficiency in terms of bit/s/Hz compared to QPSK and can thus be used to provide peak data rates on the order of 10 Mb/s within the current 5 MHz WCDMA bandwidth. However, higher-order modulation schemes are significantly less robust to noise, interference, and other channel impairments. Hence, higher-order modulation should be combined with fast link adaptation where the coding and modulation scheme employed is rapidly adapted to the instantaneous channel conditions. By using fast link adaptation, users experiencing favorable channel conditions, for example, close to the cell site, can be assigned higher-order modulation and high-rate coding, e.g., 16-QAM and rate 3/4, thus achieving higher peak rates. Similarly, users with less favorable conditions, for example, users close to the cell border or users experiencing a fading dip, need to use robust QPSK modulation and low-rate coding.

In the case of packet-data services, the receiver typically detects and requests a retransmission of erroneously received data units. Combining the soft information from both the original transmission and any subsequent retransmissions prior to a decoding attempt, generally known as hybrid ARQ type-II, will improve performance and also add robustness against link-adaptation errors. Hence, a fast hybrid ARQ scheme tightly coupled to the link adaptation mechanism is beneficial. The link adaptation serves the task of selecting a good initial estimate of the amount of redundancy needed in order to avoid an excessive number of retransmissions. The hybrid ARQ mechanism serves the purpose of fine-tuning the effective code-rate (implicit link adaptation) and compensates for any errors in the channel-quality estimates used for the link-adaptation.

Fast scheduling is the mechanism to determine which user to transmit to in a given time interval. It is a key element in the design of a packet-data system, because to a large extent it determines the overall behavior of the system. Maximum system throughput is obtained by assigning all available radio resources to the user with the currently best radio-channel conditions, while a practical scheduler should include some degree of fairness. By selecting different scheduling algorithms, the operators can tailor the behavior of the system to suit their needs.

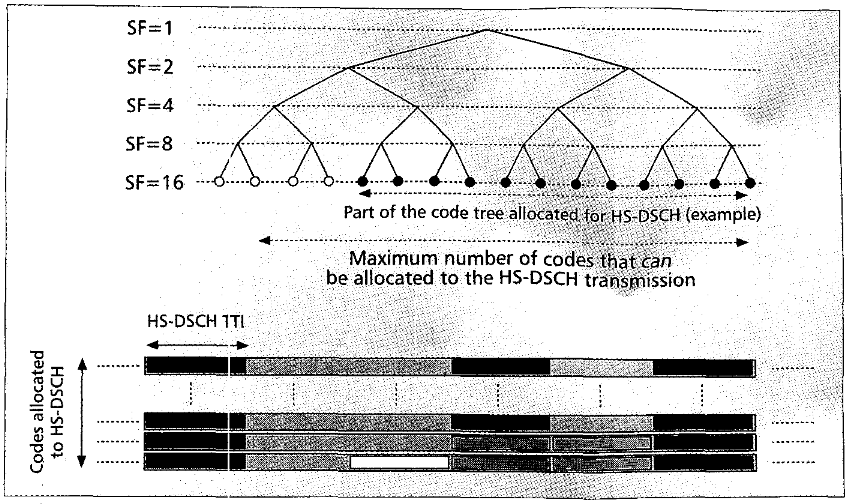
Fast cell selection, where the UE rapidly selects the cell with the most favorable radio conditions, is one possibility to increase the utilization of the shared radio resources, and can be used as an alternative to soft handover. Fast cell selection can be seen as a complement to fast scheduling, providing a spatial dimension in the scheduling in addition to the temporal dimension.

MIMO, multiple-input multiple-output antenna systems, is a technique for increasing the data rates in rich scattering environments through the exploitation of multiple independent channels, implemented by means of multiple transmit antennas and multiple receive antennas.

With the basic principles above, there is a possibility for unequal service provision, offering higher data rates to users in favorable conditions. It can be noted that some of the above technologies have been applied to GSM for the evolution toward EDGE. Thus, to some extent HSDPA can be seen as an EDGE-like evolution of WCDMA, similar to EDGE being an evolution of GSM toward 2.5G and 3G. In the following sections, it will be shown how the key technologies discussed above can be added to WCDMA in order to facilitate a smooth evolution.

Architecture – The current UTRAN (Universal Terrestrial Radio Access Network) architecture is illustrated in Fig. 5. A number of RNCs (radio network controllers) are connected to the core network. Each RNC controls one or several node Bs (base stations), which in turn communicates with the UEs (user equipment). The basic technologies for HSDPA described in the previous section rely on rapid adaptation of the transmission parameters to the time-varying channel conditions, and the corresponding functionality should therefore be placed close to the air interface, preferable in the node B. This is in contrast to the current WCDMA architecture, where, for example, scheduling and transport format selection is performed in the RNC. Thus, for HSDPA, parts of this functionality will be placed in the Node B as part of a new entity, MAC-HSDPA. However, the current RNC entities are advantageously kept. Ciphering and in-order delivery are examples of functions provided by the RNC. Furthermore, in soft handover scenarios between two different Node Bs, the RNC can guarantee no loss of data if the hybrid ARQ mechanism in the Node B fails, for example due to a limited number of retries. Hence, with this background the added features of the Node B should not be seen as a replacement of the RNC, but rather as a complement, providing, from the RNC perspective, a highly reliable channel supporting high data rates.

General Channel Structure – HSDPA introduces a new transport channel, HS-DSCH (High Speed Downlink Shared Channel), primarily to be used for best-effort packet data and supporting the technologies discussed above. Similar to the current DSCH, the HS-DSCH corresponds to a common channelization code resource shared among several users, primarily in the time domain. In Fig. 6 the principal structure is illustrated with a part of the code tree allocated to HS-DSCH.



■ **Figure 6.** *The channel structure for the HS-DSCH.*

Allocation of the HS-DSCH to different users is done on an HS-DSCH TTI (transmission time interval) basis. HS-DSCH is proposed to use a shorter TTI than the current minimum WCDMA TTI of 10 ms (one frame length). A short HS-DSCH TTI has several advantages. First, it reduces delays, which is especially important if one or several retransmissions are required before the data is correctly received. Second, it allows for a finer granularity in the scheduling process, facilitating better tracking of the channel variations. It should also be noted that there are no major interleaver gains associated with longer TTIs, as the time diversity offered by channel variations are exploited by the link adaptation, hybrid ARQ, and scheduler instead of the decoder. Hence, a short TTI is preferred, and an HS-DSCH TTI equal to 2 ms is selected. Within each TTI, the HS-DSCH code resource is preferably allocated to a single user at a time, that is, the primary way of sharing the HS-DSCH code resource should be in the time domain. Additional support of HS-DSCH code multiplex implies that data to two or more users could be transmitted in parallel within the same TTI, using distinct parts of the set of all channelization codes allocated for HS-DSCH.

In addition to the transmission of user data on HS-DSCH, there is also a need for associated control signaling to the UE. First, the UE (UEs in the case of code multiplex) scheduled for the HS-DSCH in the upcoming HS-DSCH TTI must be notified. There is also a need for additional lower-layer control information, e.g., the transport format to be used, including modulation and coding scheme, and hybrid ARQ-related information. This control information is only relevant for the UE(s) for which there is data on the HS-DSCH and can therefore be transmitted on a shared control channel. Similarly to the current DSCH, each UE to which data can be transmitted on the HS-DSCH has an associated dedicated physical channel, DPCH. The associated DPCH carries an indicator telling the UE when to read the shared control channel and the HS-DSCH. In addition to this indicator, the associated DPCH is used to carry power control commands for the associated uplink DPCH and, if needed, other services, e.g., circuit-switched voice.

Adaptive Modulation and Coding – As discussed previously, higher-order modulation in conjunction with link adaptation provides a tool for maximizing the instantaneous utilization of the fading radio channel. By transmitting the HS-DSCH at a (in principle) constant power, that is, no fast downlink power control, the modulation and coding scheme can be selected to maximize the downlink user throughput given the instantaneous channel conditions. Note that the use of link adaptation instead of fast power control does not mean that the HS-DSCH power cannot vary for other reasons, for example due to the variations in the power used by other downlink channels.

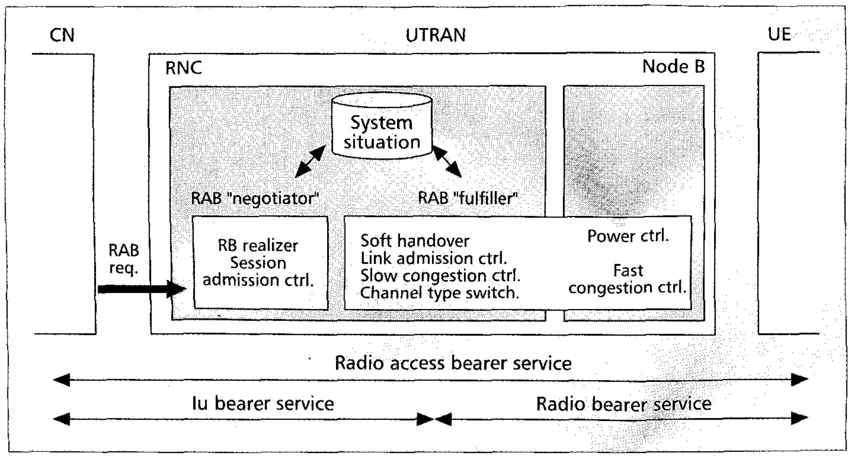
A number of transport blocks, originating from the RNC, are concatenated, followed by the addition of a CRC for error detection, turbo encoding, modulation, and multicode spreading. The use of a single CRC for all transport blocks in the TTI reduces the overhead compared to using a CRC per transport block. Furthermore, in most cases when a TTI is received in error, most of the transport blocks are erroneous and must be retransmitted. Hence, there are no significant gains associated with being able to identify erroneous transport blocks separately.

Selecting the instantaneous modulation and coding scheme can be done in several ways. A scheme where the UE controls the link adaptation is possible, but would prevent the possibility of taking into account traffic aspects such as the amount of data to transmit. It is therefore preferred to let the Node B control the link adaptation. Explicit measurement reports from the UE based on the signal-to-interference ratio of the common pilot channel (CPICH) is one possibility upon which to base the selection of modulation and coding scheme. Another possibility is to base the selection on the transmit power of the power-controlled associated downlink DPCH.

Hybrid ARQ – Hybrid ARQ, where the soft information from retransmissions requested by the UE is combined with soft information from the original transmission prior to decoding, greatly improves performance and adds robustness to link adaptation errors.

Retransmission requests can be made either per transport block or per HS-DSCH TTI (or combinations thereof), but as previously stated, the advantages of performing retransmissions individually on transport blocks are limited since, in most cases, either most of the transport blocks transmitted within a TTI are erroneous or all of them are correctly decoded. Hence, retransmission requests per HS-DSCH TTI simplify the required uplink signaling. If all data in a TTI is correctly decoded, an acknowledgment is sent to the Node B, using the associated uplink dedicated physical control channel. On the other hand, if the data in a TTI is erroneously decoded, a retransmission is requested within a few TTIs. Once the retransmission occurs, the UE can perform soft combining of the previously received versions of the data with the retransmission before decoding, thereby greatly increasing the probability of successful decoding. Retransmissions are requested until the data is either correctly decoded or a predetermined number of attempts have occurred. Furthermore, as the hybrid ARQ mechanism resides in the Node B, retransmissions can be rapidly requested, allowing for the small delays required at the high data rates envisioned.

The utilization of soft information from both the original transmission and the retransmission attempts can be performed in different ways, e.g., Chase combining or incremental redundancy (IR). Chase combining, where identically encoded data is transmitted in response to a retransmission request, is the simpler of the two and the receiver only has to form a weighted sum of the multiple copies (maximum ratio



■ **Figure 7.** *Model of service management.*

combining) before decoding. Incremental redundancy, on the other hand, implies that additional parity bits are transmitted in response to a retransmission request. The additional parity bits can be used in conjunction with the previously transmitted ones, effectively forming a lower rate code.

For the hybrid ARQ scheme to work properly, the erroneous retransmission entity must be identified in a robust fashion. One possibility is to assign a number to each transmitted entity, transmit this number to the UE separately from the user data, and include the number in any retransmission requests. This allows the implementation of a conventional selective-repeat protocol at the cost of signaling the identity. Another possibility is to use a stop-and-wait protocol, which only requires a single bit acknowledgement in the reverse signaling. To be able to fully utilize the channel for a single UE, N staggered stop-and-wait protocols can be operated in parallel. The TTI number can be used to identify which of the N channels is currently receiving information. By imposing a strict requirement on the timing of the acknowledgment relative to the data transmission, there is no need for explicitly signaling which of the N channels is acknowledged as this can be deducted from the TTI number.

Fast Scheduling – The scheduler is a key element in the design as it controls the allocation of the channel among the users and to a large extent determines the overall behavior of the system. The main objective of the scheduler is to obtain a high throughput in the system by transmitting to users with instantaneously favorable channel conditions, while at the same time maintaining a certain degree of fairness between users. As mentioned in the introduction, different schedulers can be used for different scenarios as there is no need to standardize the scheduling algorithm used, but instead different vendors can choose different criteria. Information upon which the scheduler can base its decisions includes, but is not limited to, the predicted channel quality, as used by the link adaptation, the current load of the cell, and different traffic priority classes.

Other Technologies – Some of the possibilities for evolving WCDMA toward higher-speed packet-data support were outlined and briefly discussed in the previous sections. Additional technologies, such as multiple antennas, fast cell selection, and interference-suppressing receivers, could be considered as well and used in conjunction with the previously discussed methods. Multiple antennas at both the transmitter and receiver is an interesting technique, which can be considered as an optional feature in later phases of the WCDMA evolution to increase data rates even further, at the cost of an increased complexity. Interference suppressing receivers in the UE can, in most cases, be introduced without affecting the overall design. One interesting aspect of the evolution described is that the end user directly will benefit from a more advanced receiver and get higher data rates, contrary to traditional systems, where the performance gain mainly is transferred from the UE to the network by the power control mechanism.

Service Negotiation over Iu – In the first release of the WCDMA standard, R99, QoS requirements are described in UTRAN by means of a set of radio attributes. There are four different traffic classes defined: conversational, streaming, interactive, and background. Each of these traffic classes has different characteristics. For example, for the conversational traffic class the delay requirements are very stringent to support real-time services. The requirements on the delay can be set per class. In addition to the delay, there are a number of other parameters that can be set, such as SDU (service data unit) error rate, residual bit error rate, maximum SDU size, maximum bit rate, and guaranteed bit rate, just to mention a few. During setup of a service, these parameters are set by the core network based on the type of service requested. The parameters are then sent to the UTRAN, as is depicted in Fig. 7, and based on these parameters the UTRAN will configure the different protocols. If, for example, the delay parameter is set very loosely and the SDU error rate is set very low, the UTRAN might set up the radio access bearer (RAB) using the acknowledged mode of the RLC protocol, which means that retransmissions will be used. On the other hand, if the delay requirements are stringent and the sizes of the packets to be transmitted are predefined, the transparent mode RLC can be used, meaning that no retransmissions will be used and no RLC overhead will be added to the packets.

One important input when realizing certain RABs is the actual load situation in the system. If the load is very high, there is a possibility that an incoming RAB request cannot be realized. Currently, the UTRAN can only inform the core network that the requested QoS cannot be met. No negotiation of the parameters is possible. In future systems, however, it is very likely that an application could still work reasonably well, even if the first QoS request could not be met. Thus, negotiation of the QoS parameters would be advantageous. For example, if a certain application requests a high-rate RAB, but the application would function reasonably well with a lower bit rate, the request for the RAB would not necessarily be rejected in the case of a highly loaded system, but the RAB would rather be set up with the lower bit rate.

Which parameters should be negotiable is still an open question. It is clear that some parameters, for example the traffic class, are non-negotiable. A conversational service cannot become a background class service. On the other hand, parameters such as bit error rate, SDU error rate, and transfer delay, could most likely be negotiable.

The details about how the negotiation is performed remains to be decided. One possibility is to send two or more lists of parameters from the core network to the UTRAN, each with a priority. If the highest priority QoS request cannot be satisfied, the remaining QoS requests specified in the list are tried in priority order until either the request can be fulfilled or all options have been tried. Another method of negotiation is to let UTRAN make a proposal of a new set of parameter values if the requested service cannot be realized. The difficulty with this approach is that UTRAN does not know which application will use the realization and thus an estimate of the values for the new set of QoS parameters can be difficult. These issues will be resolved as the standard evolves.

4G Research

Besides speech and short-message services, third-generation systems will provide additional streaming and conversational services. Streaming video combined with audio and conversational video were envisaged when designing the third-generation systems. Thus, from an application perspective, future generations of applications should provide more advanced types of services. At this stage it is difficult to predict exactly which services will become popular in the future. On one hand, a number of services that are available in the third generation will probably also be available in the fourth generation, however improved. For example, video may be improved through better resolution and shorter delays. On the other hand, there may be services that are available in wired systems, but currently still are not desirable in wireless systems because of, for example, excessive delays. These services may become available in future systems. Furthermore, there will be a number of services currently not available at all. An interesting candidate of such a service could be an augmented reality or virtual reality service combined with a strong sense of personalization. Through personal profiles, the system knows the specific preferences of a certain user and presents data in a 3D manner to the user when he or she moves into an area where the data is of interest. Other applications of virtual reality are virtual presence. When being at different places people can still have meetings in a common virtual meeting place.

For the services listed above to become available, even higher data rates need to be supported at the physical layer. One of the issues that should be addressed in this context is whether extremely high data rates are necessary everywhere. Perhaps there is a need for extremely high data rates in small hot spots only. These hot spots can be connected to a backbone network providing reasonable rates in a wider area, but these hot spots could also be isolated. In the first case the system may know where the user is moving and schedules the data in such a way that when the user moves into a hot spot area, large bulks of data are sent to the user. In the latter case, a high-data-rate hot spot is just connected to one or a number of servers, but not connected to a backbone network.

One of the questions to be asked is whether high-data-rate systems can be designed on the basis of existing systems such as Hiperlan 2 and IEEE 802.11. Can the current 5 GHz band be used for these systems, or is another band more suitable, e.g., 17 GHz, 40 GHz, or 60 GHz? The disadvantage with the higher carrier frequencies is the high loss of free space. In addition, people, walls, and furniture will block the propagation of radio signals. This will lead to very limited coverage areas. On the other hand, the higher the frequencies, the smaller the antennas become. This makes the use of adaptive antennas in the terminals more attractive.

Besides having hot spots with very high data rates, it may be desirable to extend the coverage area for currently available data rates (i.e., WLAN) so that the services available in 3G systems will be improved. An interesting topic to study in this respect is multihop systems. For example, the coverage area of a WLAN system may be extended through seeds placed at different locations on the outside of a building. These seeds, located at fixed positions, behave as forwarding nodes, forwarding the data to WLAN access ports. But even user equipment can act as forwarding nodes. In this way the coverage area can be extended even further, however variably, depending on the movements of the users. In the following section, ad hoc networking and multihop networks will be discussed as two examples of 4G research.

Ad Hoc Networks

Ad hoc networks [5] are networks formed without any central administration or infrastructure. The PAN described earlier and the Independent BSS mode of IEEE 802.11 are simple examples of ad hoc networks. As most of the wireless research over the years has focused on systems with centralized control, it is believed that ad hoc networking is still a research field in its mere infancy. Consequently, ad hoc networking will evolve in the future both in terms of technology and in the applications it will support. It may then act as a complement to other infrastructure-based 4G technologies. In the short time span, it is likely that one will see simple applications such as file transfer or gaming. Personal devices such as PDAs, laptop computers, and cell phones can form an ad hoc PAN using Bluetooth, as illustrated in Fig. 4. This is in part possible today, but performance will improve over the forthcoming years. Some futuristic and somewhat far-reaching applications that have been suggested for ad hoc networks are autonomous networks, sensor networks (smart dust), and amorphous computing. Autonomous networks may, for example, interconnect home robots that clean, do the dishes, perform security surveillance, and so on. Sensor networks could, for instance, be used to forecast water pollution or to provide early warning of an approaching Tsunami. Amorphous computing is a futuristic research discipline that studies the concept of programmable materials built with a myriad of tiny communicating entities that react to each other and to their environment. This field is not tightly connected to what we see as 4G today, but to ad hoc networking research as such.

In the short term, research activities pertaining to ad hoc networking need to address the issues that primarily differentiate it from its infrastructure-oriented cousin. Typical differences and operational characteristics for ad hoc networks are distributed operation, dynamic network topology, fluctuating link capacity, and low-power devices. For instance, in terms of distributed operation, a node in an ad hoc network cannot rely on a network in the background to support security and routing functions. Instead, these functions must be designed so that they can operate efficiently under distributed conditions. Security in an ad hoc setting, isolated from third-party certifications, calls for new ways of distributing trust among the participants of the ad-hoc network. Billing is another aspect, related to authorization and authentication, that differs from traditional, centralized networks. How will a user entering the network pay for the services used? Sometimes it will also be desirable to enable distributed discovery of special services provided by nodes in the network. Examples of such services might be printing hardware, application software, or just pure computational power.

Given the operating characteristics, what can the user expect from an ad hoc network? The support of multimedia services will most likely be required within and throughout the ad hoc network, which calls for QoS support. However, the inherent variable communications quality makes it difficult to offer fixed QoS guarantees. Nevertheless, when communication conditions are stable, the ad hoc network should provide the same QoS as has been defined for the access network. To further improve user perception of the service, user applications that run over an ad hoc network could be made to adapt to sudden changes in communication quality. QoS together with security, routing, and mobility are areas requiring more research.

Multihop Networks

Multihop support is an independent feature that can be adapted to various network settings, but it always implies routing capabilities. This is exemplified by the use of multihop in ad hoc networks as well as recent proposals introducing multihopping in the framework of cellular systems, then denoted cellular multihop. Whereas multihop as such is not new, the use of multihop in commercial systems has not yet been deployed on a large scale. A number of interesting features may well motivate multihop to be used as a component in future wireless systems.

There could be several reasons for employing multihopping, including: providing basic connectivity for nodes out of communication range; significantly reducing energy consumption; improving overall network performance, e.g., throughput; or simply enabling low interceptability by generating least-possible interference. For devices with short communication range, say 1–100 m, the advantage of multihopping is mainly as a coverage extender, that is, nodes being too distant from each other for direct communication exploit multihopping to enable communication. For larger communication ranges, i.e., 100–10000 m, it may be of greater concern to exploit multihopping as a method either to enable higher data rates due to improved SIR values at reception, or under the assumption of small node size, to reduce the transmit power. Multihopping for coverage extension will, to some extent, appear before 4G systems, for example with Bluetooth Scatternets or routing protocols developed by IETF's MANET working group potentially used in conjunction with 802.11 systems. Another envisioned future scenario appears as a result of ongoing broadband discussions, namely, to provide a multihop-oriented extension of Hiperlan 2 applied in, for example, rural areas and used for fixed wireless access. Hiperlan 2 devices may be placed on roofs, walls, or even indoors, and traffic is then relayed via those devices to a central access point enabling, for example, Internet access. In cellular networks, however, multihopping could provide the means for rate extension for nodes placed at the cell boarder, but also improve coverage for areas with poor coverage. Multihopping may also be used for line-of-sight connections in future high-frequency systems operating on 60 GHz, where the propagation characteristics are similar to visible light and coverage will be poor.

Apart from classical multihop or ad hoc network research issues, an area of particular interest is to ensure improved interaction between the radio medium, the channel access scheme, and the routing method. In contrast to classical routing, basic properties of the radio medium such as inherent broadcasting characteristics and channel fluctuations should be considered as well as exploited when routing is performed. This may violate traditional layering principles, but it may also enable important system improvements. Support of end-to-end QoS is another vital research area that should be emphasized. Moreover, unconventional aspects requiring more research will be found when new areas are explored, e.g., cellular multihop. The successive concentration of traffic load around a base station acting as both sink and source is one such aspect that may affect the choice of routing approach. In conclusion, multihopping as such is not new, but the setting in which it may be deployed is new, as will be the constraints dictating its function. Some interesting features of multihop systems motivate a closer examination of it as a component for future-generation wireless systems, yet much research remains to be done in this area.

Conclusions

In the previous sections aspects of future mobile communication systems were discussed. Two major trends complementing each other were foreseen, the first one being seamless roaming and handover between different systems, leading to the ABC concept. PANs were used as one example of this trend, where a variety of different devices are interconnected, for example, by the use of Bluetooth. The other major trend is the continuing development of the current 3G systems. One example of this trend, evolving WCDMA toward high-rate best-effort packet data, was outlined in some detail.

Fourth-generation systems are to a large extent in the research stage and there is not yet any clear picture of what will constitute fourth-generation systems. However, it is expected that these systems will build upon the evolved third-generation systems and be launched sometime after 2010. As examples of research topics for fourth-generation systems, ad hoc networks and multihop systems were given. These and other technologies will make possible applications and ideas previously considered to be science fiction.

References

[1] L. M. Correia and R. Prasad, "An Overview of Wireless Broadband Communication," *IEEE Commun. Mag.*, Jan. 1997, pp. 28–33.

[2] J. Vince and R. Earnshaw, "Virtual Worlds on the Internet," *IEEE Comp. Soc.*, 1998.

[3] J. Isdale, "Virtual Reality Information Resources," http://www.isx.com/~jisdale/WhatIsVr.html

[4] S. Parkvall *et al.*, "The Evolution of WCDMA Towards Higher Speed Downlink Packet Data Access," *IEEE Vehic. Tec. Conf. 2001*, spring.

[5] M. Frodigh *et al.*, "Wireless Ad Hoc Networking: The Art of Networking Without a Network," *Ericsson Rev.*, no. 4, 2000, http://www.ericsson.com/review

Biographies

MAGNUS FRODIGH is an expert in the field of new radio systems, with a focus on radio and network resource management at Ericsson Research. Since joining Ericsson in 1994 he has concentrated on radio resource management for TDMA, GSM, and WCDMA systems. More recently he has also been involved in research on short-range systems, such as Bluetooth and wireless LAN. He holds an M.S. in electrical engineering from Linköping University and a Ph.D. in radio communication systems from the Royal Institute of Technology, Stockholm.

STEFAN PARKVALL (stefan.parkvall@era.ericsson.se) works at Ericsson Research focusing on CDMA radio access. Since joining Ericsson in 1999 he has worked with cdma2000 and WCDMA and is currently involved in research and standardization of high-speed packet-data access in WCDMA. He received his Ph.D. in communications from the Royal Institute of Technology in 1996. Prior to joining Ericsson he was an assistant professor in communication theory at the Royal Institute of Technology, Stockholm, Sweden and a visiting researcher at the University of California, San Diego, USA.

CHRISTIAAN ROOBOL received his M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1992 and the Ph.D. degree from the Royal Institute of Technology, Stockholm, Sweden, in 1997. Since 1997 he has been with Ericsson. His research interests are in the area of radio protocols, future radio networks, and QoS aspects with regard to the wireless internet.

PER JOHANSSON is a researcher at Ericsson SwitchLab, an applied research laboratory within Ericsson Research. He holds master of science and licentiate degrees in electrical engineering. He joined Ericsson, in 1992 to work on traffic management and performance analysis of ATM networks, but later moved into research on wireless systems. Since 1998 his research has focused on ad hoc networks, in particular on Bluetooth ad hoc networking.

PETER LARSSON is a researcher at the Ericsson Research air interface architecture group. He holds an M.S. degree in physics engineering from the Royal Institute of Technology, Stockholm. He joined Ericsson in 1995 and has been working on GSM system development, WLAN systems, and ad hoc (multihop) networking.