REPORT ITU-R M.2038

Technology trends

(2004)

TABLE OF CONTENTS

1 Introduction

Recommendation ITU-R M.1645 defines the framework and overall objectives of future development of IMT-2000 and systems beyond IMT-2000 for the radio access network. In defining the framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000, the significant technology trends need to be considered. This Report provides further information on many of the technology trends concerning radio access network foreseen at the time of preparation of Recommendation ITU-R M.1645. Depending on their development, evolution, expected capabilities and deployment cost, each of these technologies may or may not have an impact or be used for the future development of IMT-2000 and systems beyond IMT-2000. It is expected that the research and future development of IMT-2000 and systems beyond IMT-2000 will consider these technologies and provide guidance on the applicability or influence they might have on the future of IMT-2000 and systems beyond IMT-2000.

Technologies described in this Report are collections of possible technology enablers. There is no decision implied at this stage about whether those technologies will be adopted for future mobile communication systems, and this Report does not preclude the adoption of any other excellent technologies that exist or appear in the future.

2 Scope

This Report provides information on many of the technology trends concerning radio access networks foreseen at the time of preparation of Recommendation ITU-R M.1645.

The Report addresses technology topics that appear relevant to some lesser or greater degree to the future development of IMT-2000 and systems beyond IMT-2000. The Report considers these topics in three broad categories:

– technologies which have an impact on spectrum, its utilization and/or efficiency in this context;
– technologies which relate to access networks and radio interfaces;
– technologies which relate to mobile terminals.

3 Overview of major new technologies

This section presents technology topics that appear relevant to some greater or lesser degree to the future development of IMT-2000 and systems beyond IMT-2000. Technologies having an impact on spectrum, its utilization and/or efficiency; technologies related to access networks and radio interfaces; and technologies related to user terminals are described in § 3.1, 3.2, and 3.3, respectively. Further details are provided in the related Annexes.

The demand for mobile multimedia communications has been rapidly increasing. The radio spectrum is, however, a precious and scarce resource. Therefore, novel technologies for efficient spectrum utilization to enhance the capacity of IMT-2000 and systems beyond IMT-2000 are keenly anticipated. Section 3.1 addresses new radio technologies and their impact on spectrum utilization, including technologies for improving spectrum efficiency, those using multiple antennas such as adaptive antennas and multiple-input multiple-output (MIMO), and those for handling traffic asymmetry and time division duplex (TDD).

Advanced radio resource management (RRM) algorithms and flexible frequency sharing methods are beneficial in maximizing and optimizing the frequency resource utilization. In addition, antenna and coding technologies such as smart antennas, diversity techniques, coding techniques, space-time coding, and combined technologies improve the radio link quality in multipath Rayleigh fading channels. Furthermore, efficient multiple access schemes and adaptive modulation improve the bandwidth efficiency of the systems.

Adaptive antennas improve the spectral efficiency of a radio channel, and in so doing, greatly increase the capacity and coverage of most radio transmission networks. This technology uses multiple antennas, digital processing techniques, and complex algorithms to modify the transmitted and received signals at a base station and at a user terminal. In addition, MIMO techniques can provide significant improvements in the radio-link capacity by making positive use of the complex multipath propagation channels found in certain terrestrial mobile communications. MIMO techniques are based on establishing several parallel independent communication channels through the same space and frequency channel by using multiple antenna elements at both ends of the link.

In broadband multimedia communications, asymmetric traffic is envisaged to be dominant. Due to uncertainties in future traffic asymmetry, future mobile communication systems should be adaptable to different ratios of asymmetry especially at the personal-area and the user-access levels in order to deliver the offered traffic asymmetry while simultaneously maintaining high spectrum efficiency. TDD is one of the techniques suitable to support asymmetric high data rate services while providing flexible network deployment including busy urban hotspot and indoor environments as well as wide area applications. TDD systems do not require a duplex frequency pair since both the uplink and downlink transmissions are on the same carrier within the same spectrum band. In future mobile communication systems, flexibility and integration/convergence will be key factors. In section 3.2, technologies related to IP applications and IP broadband wireless access, those related to software-defined radio (SDR), and those achieving wider coverage such as radio on fibre (RoF), multi-hop radio networks, and high altitude platform station (HAPS) are presented.

Many wireless communication systems provide users with convenient ways to access the Internet and to communicate with one another or access multimedia content. Wireless technologies are expected to progress in a direction that will allow native support of multimedia and Internet services. The technological implication of the integration of IP and wireless is more prominent in the case of mobile broadband Internet access. To support real-time or multimedia applications using end-to-end IP, all the elements, in general, of a service path must support the requirements of mobile or broadband wireless access. To support efficient IP transport over a broadband mobile environment, we essentially need a set of diverse technologies grouped around the concepts of "seamless", "broadband" and "energy-efficient".

SDR provides reconfigurable mobile communications systems that aim at providing a common platform to run software that addresses reconfigurable radio protocol stacks thereby increasing network and terminal capabilities and versatility through software modifications (downloads). Basically, SDR concerns all communication layers (from the physical layer to the application layer) of the radio interface and has an impact on both the user terminal and network side.

Radio on fibre is defined as a system that enables the transparent interconnection of a base station, or equivalent wireless system radio interface network element, to its associated transmission and reception antennas by means of an optical network. Optical fibre presents very low insertion loss to achieve long cable spans of up to several kilometres and an enormous bandwidth to transport many different RF signals over a single fibre.

Multi-hop wireless access technology utilizes multiple serial wireless connections between the target user terminal and a base station in a homogeneous system or across different systems. In a wireless system with higher frequency bands where a smaller coverage area is available, multi-hop wireless access technology may be a solution for user terminals to gain wireless connectivity to a base station.

Another solution is applying HAPS, which is a new technology based on a flying platform. The HAPS system can provide mobile cellular coverage and fixed wireless services to several regions ranging from a high-density (urban) area to low-density (rural) areas.

Flexibility and integration/convergence are also key factors for user terminals. Section 3.3 addresses technologies for achieving reconfigurable user terminals such as terminal architecture, reconfigurable processors, RF micro-electro-mechanical systems (MEMS) for achieving smaller user terminals, and user interfaces for a flexible user terminal.

Future mobile user equipment may assume characteristics of general-purpose programmable platforms by containing high-power general-purpose processors and provide a flexible, programmable platform that can be applied to an ever-increasing variety of uses. The convergence of wireless connectivity and a general-purpose programmable platform might heighten some existing concerns and raise new ones; thus, environmental factors as well as traditional technology and market drivers influence the architecture of these devices. A well-designed embedded processor with a reconfigurable unit may enable user-defined instructions being efficiently executed, since general-purpose processors such as CPUs or DSPs are not suitable for bit-level operation. This type of processor, which can handle many kinds of bit-level data processes, can be applied to various applications for mobile communication systems with efficient operation.

RF MEMS are integrated micro-devices (or systems) combining electronic and mechanical components fabricated using an integrated circuit (IC) compatible batch-processing technique. This technology can yield compact, light weight, low power, and high performance ICs to replace discrete passive RF components such as VCO, IF, RF filters, and duplexers.

Wearable computing is also a promising technology that will give birth to new ideas of man-machine interfaces applicable to user terminals. So far, many solutions are not standardized but are proprietary methods. There is also a clear need for harmonization and for open use of common open interface standards.

3.1 New radio technologies and impact on spectrum

3.1.1 Technologies for improving bandwidth efficiency

To meet the strong demand for broadband multimedia services to both nomadic and mobile users, it is necessary to increase the maximum information bit rate of systems beyond IMT-2000. To enhance the capacity of IMT-2000 and systems beyond IMT-2000, novel technologies or new concepts for improving bandwidth efficiency are indispensable. Advanced radio resource management (RRM) algorithms will be beneficial for maximizing the resource utilization. In addition, antenna and coding technologies such as smart antenna, diversity techniques, coding techniques, space time coding, and combined technologies will be necessary for systems beyond IMT-2000 to improve the wireless link quality under multipath Rayleigh fading channels. Furthermore, efficient multiple access schemes, adaptive modulation, adaptive downlink modulation, and multi-hopping technology will be needed to improve the bandwidth efficiency of the system.

Technologies for improving bandwidth efficiency which are discussed in this Recommendation include:

– bunched systems;

– ultra-wideband (UWB);

– adaptive modulation and coding (AMC);

– flexible frequency sharing;

High level descriptions of the above technologies are to be found in the following sections, whilst more detailed information is provided in Annex 1.

3.1.1.1 Summary of the technology

– *Bunched systems:* In pedestrian and indoor environments, there will be severe fluctuations in traffic demands, high user mobility and different traffic types. This highly complex environment will require advanced RRM algorithms. It could be beneficial to have a central intelligent unit that can maximize the resource utilization. This capability is provided by bunched systems.

– *WWB:* The basic concept of UWB is to develop, transmit and receive an extremely short duration burst of RF energy. The resultant waveforms are extremely broadband (typically some gigahertz).

– *AMC:* Adaptive modulation and coding schemes adapt to channel variation by varying parameters such as modulation order and code rate based on channel status information (CSI).

– *Flexible frequency sharing:* Sharing of frequency carriers between different operators is a method to optimize the use of spectrum resources.

3.1.1.2 Advantages

– *Bunched systems:* Bunched systems provide dynamic load distribution, dynamic RRM, and adaptive coverage control. Bunched systems are well suited to hotspot coverage.

– *UWB:* UWB systems provide the potential for spectrum sharing between services and more efficient use of spectrum.

– *AMC:* The advantage of AMC schemes is that the amount of spectrum utilized is based on the actual channel conditions rather than worst case channel conditions.

– *Flexible frequency sharing:* More efficient use of the spectrum resource.

3.1.1.3 Issues to be considered

– *Bunched systems:* Design issues of the radio access network (RAN) and the RRM algorithm for the bunched systems must be considered.

– *UWB:* No internationally agreed definition of UWB exists because the applications and uses to which the technology may be applied are very diverse and the devices have not been fully developed. The regulatory and interference impacts of UWB are not yet known.

– *AMC:* Delays in reporting channel conditions reduces the reliability of the channel status indicator which may cause the system to select incorrect modulation levels and coding rate.

– *Flexible frequency sharing:* The use of flexible spectrum sharing may have serious implications on the time required to scan the spectrum and locate a radio access technology (RAT) carrier after the terminal has been powered on.

3.1.2 Technology solutions to support traffic asymmetry

3.1.2.1 Background

Radio interfaces for IMT-2000 systems and systems beyond IMT-2000 may support different capabilities in the uplink and downlink with respect to traffic asymmetry. In this context asymmetry means that the basic amount of traffic and consequently the amount of needed resources may differ between the uplink and the downlink direction.

There are at least four aspects of traffic asymmetry:

– *At the personal area level:* the degree of asymmetry for traffic between devices of a personal area network (PAN).

– *At the user access level:* the degree of asymmetry for the traffic between a specific user and the network for a specific service.

– *At the cell level:* the degree of total traffic asymmetry in a specific cell.

– *At the network level:* the degree of total traffic asymmetry in the entire network.

These views differ in particular concerning the considered amount of traffic and the speed of change of the asymmetry. For individual users (i.e. at the personal area level and user access level) the degree of asymmetry may change quickly. But the degree of total asymmetry over a cell (i.e. at the cell level) and even more over the entire network (i.e. at the network level), will change much slower due to aggregation of individual services on one hand and changing mix of services on the other hand. It depends on the system design whether and how this offered changing traffic asymmetry can be delivered efficiently.

3.1.2.2 Service mix in IMT-2000 systems

In IMT-2000 networks or systems beyond IMT-2000, there will be a mix of symmetric applications as well as predominately downstream (downstream direction is from base station to mobile station(s))[1] or predominately upstream (upstream direction is from mobile station(s) to base station)[1] applications using different data rates. The most recent estimates for a mix of traffic are described in Report ITU-R M.2023. An analysis of these estimates indicates that the total traffic asymmetry in a specific cell or the entire network from IMT-2000 users would have the same "down load" characteristics as in the fixed network, i.e. it is predominately downstream. However, it should be noted that the traffic characteristics and the degree of traffic asymmetry between a specific user and the network for some IMT-2000 specific services may be different. It is expected that new applications, such as picture and video clips, as well as peer-to-peer traffic, which would generate traffic from terminals or servers connected over wireless, will affect the IMT-2000 traffic mix. Due to uncertainties of the future traffic asymmetry, future radio access systems should be adaptable to different ratios of asymmetry especially at the personal area level and at the user access level to deliver the offered traffic asymmetry by maintaining at the same time high spectrum efficiency.

---

[1] Per Recommendation ITU-R F.1399 – Vocabulary of terms for wireless access.

3.1.2.3 Technical aspects

Radio interface support for asymmetric traffic can be achieved by different means:

– By asymmetric resource allocation, e.g. asymmetric frequency allocation in case of frequency division duplex (FDD) operation or asymmetric time-slot allocation in case of TDD operation.

– By symmetric uplink/downlink frequency allocation in the case of FDD or symmetric uplink/downlink time-slot allocation in the case of TDD with only partial use of the available capacity in one of the two directions.

– By applying different capacity-enhancing technologies to uplink and downlink, regardless of the resource allocation. These technologies are typically independent of the duplex scheme.

More details are given in Annex 2.

3.1.3 Advanced system innovation using TDD

TDD is well suited for asymmetric high data rate services while providing flexible low cost network deployment including busy urban, hotspot and busy indoor environments as well as wide area applications. TDD is a technique where both the uplink and downlink transmissions are on the same carrier within the same spectrum band. This means TDD technology can operate within an unpaired frequency band; i.e. no duplex frequency pair is necessary. The minimum spectrum requirement is only half the bandwidth of the FDD mode, i.e. only one 5 MHz spectrum allocation is necessary when the wideband code-division multiple acess (W-CDMA) TDD (IMT-2000 CDMA TDD) chip rate is operating at the same 3.84 Mchip/s harmonized chip rate as the W-CDMA FDD (IMT-2000 CDMA Direct Spread) mode.

Currently, within IMT-2000, TDD makes use of both CDMA and time division multiple access (TDMA) techniques to separate the various communication channels by both time slot and CDMA code. Time slots can be assigned to carry either downlink or uplink channels. The TDMA structure also permits the use of a specific algorithm by which multiple channels are jointly recognized and decoded (joint detection algorithm). This method eliminates intracell interference almost completely and helps increase system capacity. This is feasible in TDD because the transmission and reception occur at the same frequency and exhibit similar channel distortions, thus simplifying processing.

Due to the TDMA structure and the joint detection algorithm, which significantly reduces interference from other CDMA signals present in the time slot, W-CDMA TDD behaves much like a TDMA system. It does not suffer from cell breathing and the necessity to maintain sufficient operating margin to compensate for the uncertainty, nor does it require a soft hand-off capability. This is of particular value for hotspot scenarios with heavy data load and small cell sizes such as indoor and outdoor (pico- and microcells). Since time slots for uplink and downlink can be assigned separately, W-CDMA TDD is particularly suited for asymmetric traffic. The degree of asymmetry can be dynamically controlled, improving overall operating efficiency.

From the beginning, the TDD standard has been designed in anticipation of the implementation of smart antennas which can substantially improve the system capacity. Smart antennas give particular advantages in macro- and microcell scenarios where the user signals are not very scattered. Again, TDDs use of the same physical radio channel for both the uplink and downlink simplifies the processing required to shape the antenna beams. This unique characteristic, channel reciprocity, of TDD also makes it practical to implement advanced diversity and coding techniques.

Finally, TDD is cost-efficient for network deployments as it leverages the infrastructure of an FDD-only roll-out by providing scalable capacity for "hotspots". This is accomplished through a multi-tier architecture of FDD and TDD macro-, micro- and picocells.

3.1.4 Adaptive antenna concepts and key technical characteristics

3.1.4.1 Introduction, and benefits of adaptive antennas in IMT-2000

Formally, adaptive antennas may be defined[2] as "an array of antennas which is able to change its antenna pattern dynamically to adjust to noise, interference and multipath. Adaptive antennas are used to enhance received signals and may also be used to form beams for transmission".

Likewise, switched beam systems "use a number of fixed beams at an antenna site. The receiver selects the beam that provides the greatest signal enhancement and interference reduction. Switched beam systems may not offer the degree of performance improvement offered by adaptive systems, but they are much less complex and are easier to retrofit to existing wireless technologies".

Finally smart antennas are similarly defined by the same source as systems that "can include both adaptive antenna and switched beam technologies".

The reader is cautioned that there is some variation in terminologies used here; for example, non-adaptive or non-switched systems are sometimes termed smart simply due to the incorporation of masthead RF electronics, and unfortunately often the terms adaptive and beam-forming are used rather loosely.

3.1.4.2 Benefits of integrating adaptive antennas

Benefits of adaptive antennas in IMT-2000 networks

Adaptive antennas improve the spectral efficiency of a radio channel, and in so doing, greatly increase the capacity and coverage of most radio transmission networks. This technology uses multiple antennas, digital processing techniques and complex algorithms to modify the transmit and receive signals at the base station and at the user terminal. Systems in all of the existing IMT-2000 radio interfaces could enjoy significant performance improvements from the application of adaptive antenna technology.

Further improvements by including adaptive antennas in the initial design concept

While applying adaptive antenna technology to an existing radio interface can significantly improve the spectral efficiency of that radio interface, there are more significant efficiency benefits that might be derived if adaptive antenna technology is incorporated into the design of the radio

---

[2] LIBERTI and RAPPAPORT *Smart Antennas for Wireless Communications* [1999] Wiley.

interface from the outset. Many aspects of an air interface design affect the spectral efficiency gains that can be realized from the adaptive antenna technology including the following:

– duplexing methods;
– carrier bandwidth;
– modulation methods;
– signalling control: broadcast and paging methods;
– burst and frame structures;
– media access control methods.

The result of this approach can be quite significant. It can be shown that integrating adaptive antennas into the initial design concept can yield spectral efficiency increases of > 4 000% over existing 2G systems and > 400% increases over the new IMT-2000 radio interfaces.

3.1.4.3 Summary

There are a number of less commonly appreciated adaptive antenna technology advantages. For example, the inevitable redistribution of RF power amplification elements for adaptive antenna systems commonly leads to lower total amplifier cost than is likely to be the case with conventional technology. From a deployment viewpoint it is sometimes attractive to utilize adaptive antenna stations in only a proportion of the overall infrastructure in an area, and similarly the interference mitigation advantages may be particularly beneficial for such situations as cross-border coordination arrangements.

Integrating adaptive antenna systems into the design of future IMT-2000 systems and systems beyond IMT-2000, will significantly improve the spectral efficiency of these new radio systems. Spectral efficiency gains from adaptive antenna systems can be used not only to reduce the number of base stations (cells) needed to deploy an IMT-2000 network, but also to obtain significantly increased data rates within a limited amount of increasingly scarce spectrum.

3.1.5 Multiple-input multiple-output techniques

3.1.5.1 Summary of the technology

MIMO techniques can provide significant improvements in the capacity of the radio link by making a very positive use of the complex multipath propagation channels found in terrestrial mobile communications. There are many alternative solutions within this family of techniques, but they are all based on establishing several parallel independent communication channels through the same space and frequency channel by using multiple antenna elements at both ends of the link.

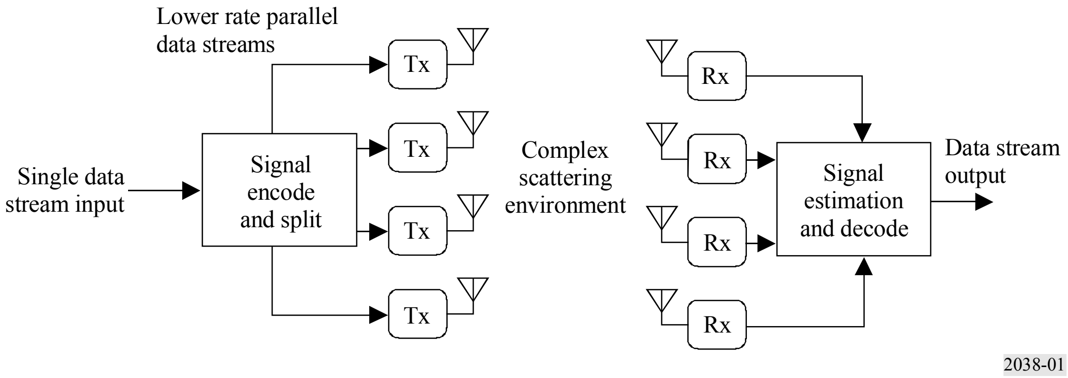
3.1.5.2 Advantages

The advantage of exploiting MIMO techniques is to increase the system throughput data rate for the same total radiated power and channel bandwidth.

In highly scattering propagation environments, the theoretical maximum data rate for MIMO algorithms increases directly in proportion to the number of antennas, rather than only being proportional to the logarithm of the number of antennas when conventional phased array beam forming methods are used. For the arrangement shown in Fig. 1 the MIMO method has a potential gain of double capacity over using conventional phased array algorithms in cellular networks.

FIGURE 1

MIMO transmitter-receiver concept



3.1.5.3 Issues to be considered

How much of these theoretical gains can be achieved in realistic deployment scenarios is the subject of ongoing research within the industry, with particular emphasis on maximizing the performance of the terminal antenna system within the restricted form factors of future terminals such as laptops, personal digital assistant PDAs and handsets, and also in minimizing the computational complexity of the signal processing algorithms.

Initial results reported in the literature have shown that most of the theoretical MIMO capacity could be exploited with appropriate terminal antenna array design. In these the antenna elements can be separated by less than a wavelength and can also make use of alternative polarization to increase the number of elements within a given size terminal. It has been demonstrated that even four elements within a terminal can give significant gains in capacity when mounted within the outline of a typical PDA.

However, for all aspects of MIMO system design a thorough characterization of the MIMO propagation channel in realistic deployment scenarios is needed and this is the subject of study in the 3GPPs and COST 259 and COST 273 research projects.

A brief review of MIMO techniques is provided in Annex 5, along with a list of references to some of the more important published work in this field.

3.2 Access network and radio interfaces

3.2.1 Software defined radios (SDRs)

3.2.1.1 General

SDR is a technology to provide reconfigurable mobile communication systems, which aim at providing a common platform for running software addressing reconfigurable radio protocol stack thereby increasing network and terminal capability and versatility by software modifications (downloads). With the proliferation of open application programming interfaces (APIs), software from different vendors can run on proprietary hardware platforms. On such platforms, the air interface protocols and applications are executed under the control of a common software environment.

SDR concerns therefore basically all communication layers (from the physical layer to the application layer) of the radio interface (see Fig. 19) and impacts both mobile terminal and the network side.

As key objectives, SDR shall provide means for:

– adaptation of the radio interface to varying deployment environments/radio interface standards;

– provision of possibly new applications and services;

– software updates;

– enabling full exploitation of flexible heterogeneous radio networks services.

In Annex 6 we provide more details on architecture for reconfigurable terminals and supporting networks.

3.2.1.2 General requirements for SDR

The provision of SDR poses requirements on the mobile communication system, which fall into three distinct groups:

– radio reconfiguration control;

– creation and provisioning of services over converging networks and different radio access modes;

– user environment management.

Moreover SDR has to consider and take into account appropriate security functions that allow reliable operation and avoid any potential abuse despite the high flexibility provided by SDR.

3.2.1.3 Logical SDR-architecture

The logical SDR-architecture has to support the following functions:

– management of terminal, user and service profiles in the network entities and the terminal;

– efficient download control and reconfiguration management for terminals and network entities;

– negotiation and adaptation functionalities for services and RATs (e.g. vertical handover);

– assurance of standard compliance.

These functions are logical functions, i.e. they can be implemented in different places in the network. Moreover they can be distributed within the network and between network and terminal.

An example of such a logical SDR-architecture (terminal and network aspects) is given in Annex 6.

3.2.1.4 Constraining considerations

SDR, due to its huge flexibility and due to the possibilities to change nearly all parameters of the radio interface or higher layer parameters (e.g. parameters in the transport layer) are potential subject to standardization, if mixed operations (mix of different hardware and software vendors) and open APIs between modules are required.

Related topics to be considered are for example:

– security functions for reliable and trusted software download (e.g. software download limited to manufacturer approved builds available only from a manufacturer's secure server to protect manufacturer's regulatory liability for system integrity);

– for the terminal: separation of functionalities used for applications and for radio-specific software;

– for the terminal concerning new applications and services: request user confirmation before software update to avoid incompatibility with other already installed software.

3.2.2 High data rate packet nodes (HDRPN)

Since packet data services display different characteristics than voice data services it may be possible to take advantage of the characteristics of certain packet data applications to enhance the performance of the system when it accommodates these services. One such change in architecture and structure that takes advantage of the more tolerant delay characteristics of certain classes of packet traffic is the HDRPN concept. This concept places nodes, HDRPN, close to routes that mobile subscribers are expected to traverse and when the subscribers are in close proximity to these nodes the system transfers large files at high data rates to users that have large files waiting for them. The HDRPN do not transmit sufficient power to allow the mobile terminals to receive high data rates when they are not in the proximity of one of the high rate packet nodes. This will translate into less interference across the region and may result in fewer base stations.

Future IMT-2000 systems are expected to provide high data rate packet services, see new questions, that will seriously test the practical limits of existing technology. It is anticipated that this type of packet link is likely to be asymmetrical with the downlink transfer frequently operating at a much higher data rate. Often the data is not sensitive to short delays and as much as a minute delay may be acceptable. This set of requirements are different than the original requirements for IMT-2000 which strongly emphasized voice requirements and balanced transmission paths. In the next phases of IMT-2000 it is essential for us to re-examine the basic architecture to determine if these new requirements might affect the structure of the system. In some of the new applications it will be practical to negotiate a reasonable delay value in other cases a best effort capability will suffice. Internet users have become used to a best effort category of service when they have used line modems for access. If a large file takes a minute to transfer at a rate of 144 kbit/s. The same file can be transferred in six seconds at 1 444 kbit/s. Therefore, if the delay to start the transfer is 54 s in the later case the transfer is still completed at the same time as the first case.

Mobile terminals in vehicles are generally moving quite rapidly and are, therefore, quickly changing their relationship to base stations. This is particularly true for automobiles on expressways and high-speed trains. Therefore, since the class of data described above can tolerate short delays it is logical for the highly mobile terminals to receive larger files at a high data rate when they are close to a high rate packet node. This ultimately reduces the cost of terminals/base stations and can significantly reduce the interference for other terminals/base stations. The mobile terminals can receive lower rate data over the entire region.

3.2.3 Internet technologies and support of IP applications over mobile systems

3.2.3.1 Summary of the technology

The Internet technologies and wireless technologies have to move in a direction to accommodate each other more natively. The technological implication of the integration of IP and wireless is more prominent in the case of mobile broadband access of Internet. In order to support real-time or multimedia applications using end-to-end IP, all the elements, in general, of a service path, it is necessary to support the requirements of mobile or broadband wireless access. Similarly, access networks should be equipped to support high-speed IP mobility while maintaining negotiated quality of service (QoS). For example, mobile network components should be able to monitor and evaluate the wireless channel conditions, and adjust transmission parameters accordingly to avoid severe degradation of throughput.

For supporting efficient IP transport over mobile environment, we essentially need a set of diverse technologies grouped around the keywords, “seamless”, “broadband” and “energy-efficient”. From these we can derive many Internet-related technologies, for example, QoS, routing and handoff, location management, QoS management, wireless resource management, paging/signalling protocols, terminal architecture, operating system support, adaptive system reconfiguration, etc.

In view of support of IP applications the following applications can be mentioned for example.

Less QoS-demanding services like web access, email or SMS are already being provided over current cellular systems. Among other potential and challenging IP-based applications over mobile systems, VoIP is the front-runner, currently being implemented in increasingly smaller devices, from notebook PCs to PDAs. However, issues like high bandwidth requirement, handover delay, etc. will hinder its deployment unless significant improvement is not made, for example, through efficient header compression or seamless handover by access points.

Mobile commerce/banking services are other lucrative applications where the potential of integration of Internet and mobile technologies can be exploited. In such services, specifically and in all other services in general, a reliable (continuous, non-breaking) and secure wireless environment is a prerequisite. The solution may come through improved wireless technology or Internet technology or both. A security framework to support IP applications over mobile systems is required.

3.2.3.2 Advantage of the technology

If the current trend of mobile usage continues, the integration with Internet technologies will bring about a revolution in the wireless communications industry that will affect vendors, service/application/context providers, policy makers, and users.

Current wireless systems are (and probably future systems will also be) independently designed, implemented and operated to meet different requirements on mobility, data rates, services, etc. Some, if not all, of these systems can simultaneously provide services at a specific geographic location, creating a heterogeneous wireless environment for users in overlaid service areas. Also, next generation wireless networks would be heterogeneous networks that support multiple broadband wireless access technologies and global roaming across systems constructed by individual access technologies. For the seamless integration of heterogeneous wireless systems, all-IP solution appears to be most promising.

3.2.3.3 Issues to be considered

Beginning with the network architecture there are several important issues that should be considered to realize an efficient mobile Internet environment as well as heterogeneous wireless networks. To build a sustainable system we must address issues related to security and scalability. Also, interoperability with legacy and future systems, IP addressing, IPR issues, etc., have to be considered.

3.2.4 IP broadband wireless access

For the foreseeable future, data services are and will be carried predominantly on IP networks. The overall network architecture therefore should evolve to an end-to-end architecture that, at layer 3, is a pure IP architecture. Such an architecture will enable transparent access from mobile devices to all data applications that are accessible via the Internet and corporate intranets. Provided that there is adequate system capacity and user data rates, this approach will allow the wireless data market to grow organically at the same rate as the explosive growth of e-commerce and "infotainment", which services are currently assumed to be wireline in nature. This approach also eliminates the need for duplicate programming and/or repackaging of content, i.e. only one version for both Internet/intranet and wireless applications. To allow this to happen, a network architecture needs to evolve that brings IP transparency to the network edge and utilizes IP protocols. The required network architecture is illustrated in Fig. 2, and the ramifications for the protocol stacks are shown for one embodiment in Fig. 3.

FIGURE 2

Example of a pure-IP access network architecture

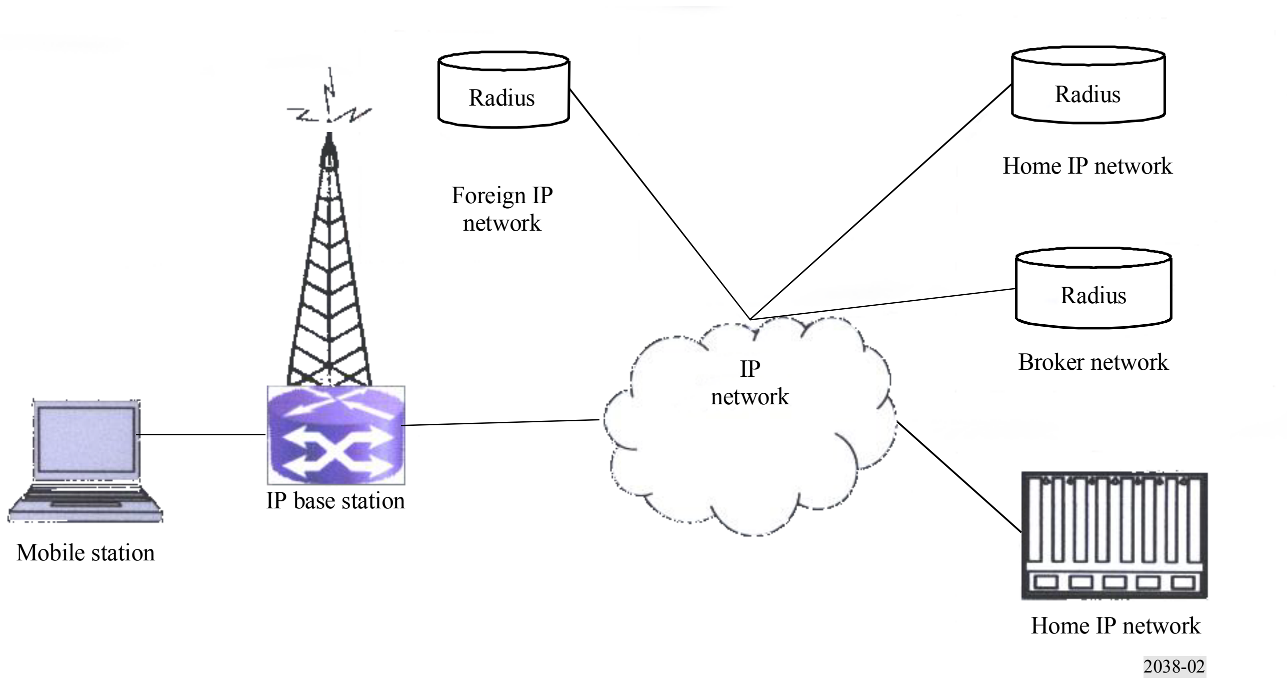
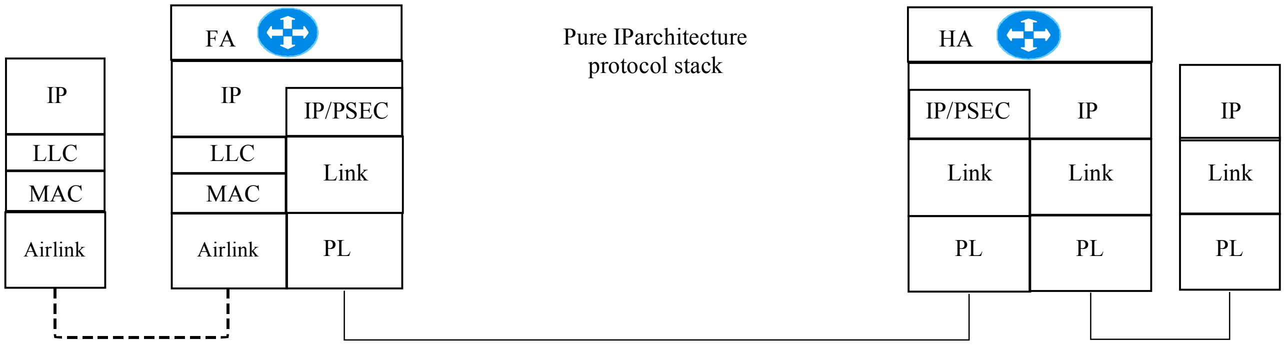


FIGURE 3

Example of a pure-IP access network protocol stack



FA: Foreign agent
IP: Internet protocol
LLC: Logical link control
PL: Physical layer
HA: Home agent
IPSEC: Internet protocol security
MAC: Media access control

2038-03

Examples of protocols required:

1 *Mobile IP* – To support mobility. Additional enhancements need to be worked to allow seamless roaming at vehicular speeds.

2 *Radius and AAA* (authentication, authorization, accounting)– For support of security features and accounting.

3 *SIP* – For support of end-to-end service control.

To allow optimal end-to-end performance the wireless access networks itself also needs to be IP-aware. Current wireless access networks also assume that the predominant data traffic is highly asymmetrical, the typical model being web-surfing. Though this is true in many cases, more symmetrical data traffic applications are rapidly emerging. Such applications include videoconferencing and various enterprise data applications.

3.2.4.1 Summary of the technology

Table 1 shows the distinguishing characteristics between existing 3G interfaces and the emphasis of an IP access network. Work on such a mobile broadband wireless access network is currently ongoing in IEEE 802.20.

3.2.5 Radio on fibre (RoF)

3.2.5.1 Summary of the technology

In this Report, radio on fibre (RoF) is defined as a system which enables the transparent interconnection of a base station, or equivalent wireless system radio interface network element, with its associated transmission and receiving antennas by means of an optical network. The signals propagating through the optical network are a replica of the signals at the base transmitter station (BTS) radio interface.

The definition can be generalized to include in it not one, but several BTS repeaters, as long as all of them are housed in the same room and share one common basic infrastructure, i.e., power, air conditioning, etc.

Such RoF systems are described in Annex 10.

TABLE 1

Subject	Broadband IP access network	3G
End-user	Fully mobile, high throughput data user Asymmetric or symmetric data services End-user devices initially PC card enabled data devices Full support of low-latency data services	Voice user requiring data services Highly asymmetric data services End-user devices initially data enabled handsets Support for low latency services still an issue
Service provider	Wireless data service provider – Greenfield start or evolving cellular carrier Global mobility and roaming support	Cellular voice service provider evolving to data support Global mobility and roaming support
Technology	New PHY and MAC optimized for packet data and adaptive antennas Licensed bands below 3.5 GHz Packet oriented architecture Channelization and control for mobile multimedia services. Mobile-IP based High efficiency data uplinks and downlinks Low latency data architecture	See Recommendation ITU-R M.1457 Licensed bands below 2.7 GHz Circuit oriented architecture – evolving to packet on the downlink Channelization and control optimized for mobile voice services. MAP/SS7 based Medium efficiency data downlinks, low efficiency uplinks Latency still an issue

3.2.5.2 Advantages of the technology

RoF systems are applicable when the distance between BTS and antennas is so large that it becomes impractical to connect them through coax cable, even with the use of in-line repeaters. Optical fibre presents very low insertion losses, which allows unrepeatered fibre cable spans of up to several kilometres, and enormous bandwidth: many different RF signals can be transported over one single fibre. RoF systems use simple analogue modulation of lightwave signals without modulation and demodulation of the RF signal. The RF signal channels can be inserted or extracted by straightforward optoelectronic circuitry. RoF is also immune to electromagnetic interference and grounding problems; the fibre cross-section is very low, which allows several tens of fibres to be bundled in one single optical cable; optical cable is rugged, it can be placed in ducts, hung on poles or directly buried; with a small extra cost it can even be lined with a rodent protective steel cover.

In microcellular scenarios where different wireless systems are co-sited, RoF enables the use of centralized processing, i.e., the system RF heads are placed in close proximity to the antennas, and the wireless systems processing equipment is housed in a centralized room, usually under controlled environmental conditions. Depending on the deployment scenarios, centralized processing might bring the following benefits. It allows a dense deployment of repeaters in urban environments, diminishes the number of required building rooftop housing installations, lowers the need for costly high-power RF amplifiers, and improves the spatial distribution of BTS capacity.

These benefits, though applicable to all wireless systems, are especially significant in IMT-2000 – and beyond – cellular networks for the following reasons:

– IMT-2000 frequency bands are higher than those assigned to 2G. Similarly, it is reasonable to expect that IMT-2000 systems beyond frequency bands will be higher than those assigned to 2G, with correspondingly higher propagation losses. This would make it more difficult to provide adequate coverage with macrocells only, so IMT-2000 and systems beyond would favour more comprehensive microcell deployment.

– Because of their higher capacity compared to 2G, IMT-2000 and systems beyond may require a larger number of cells to cover a given geographical area. Since it is becoming increasingly difficult to commission new sites, solutions like RoF allow BTS equipment concentration to simplify radio network rollout.

– Compared to 2G systems, the capacity of an IMT-2000 and systems beyond carrier is quite large. This favours radio coverage solutions, like RoF, which allow tailoring the spatial radio distribution of the carrier capacity to the specific coverage area, or coverage volume, requirements.

3.2.5.3 Issues to be considered

When an RoF system is used to radiate the same carrier, or carriers, from different antennas, the handover among the cells under the same BTS is not required. However, some interference may occur in the overlapping area between cells because of the multipath routing through different antennas.

When such an RoF system has antennas widely spaced apart from the BTS, the spatial accuracy of any location system based on the wireless system cannot be better than the distance between the antenna and BTS. This might lower the precision of an IMT-2000 based location system, whose accuracy can be few tens of metres if it uses differential time of arrival procedures to determine mobile terminals relative position to different BTS.

Furthermore, key devices such as the optical modulator and low noise amplifier would be developed later.

3.2.6 Multi-hop radio networks

3.2.6.1 Summary

Multi-hop radio networks are mobile radio communication networks characterized by the existence of radio nodes (extension points) that provide retransmission capabilities. These radio nodes can be mobile terminals with special relay functionality ("ad hoc multi-hop networks") or fix installed extension points, that operate exclusively as relays ("structured multi-hop networks").

TABLE 2

Ad hoc multi-hop networks	Structured multi-hop networks
– No fixed installed additional infrastructure – Coverage depends on existence of further relaying mobile terminals in the area and cannot be reliably planned – May provide interconnectivity between mobile terminals in an area and also to access points – Mobile terminals must provide some kind of network functionality	– Extension points are installed as additional infrastructure – Coverage extension is guaranteed and is planned – Provides extended connectivity to the access point and relay capability for local communication

Multi-hop radio networks are able to use multiple subsequent wireless connections between a user terminal and a base station. Thereby other user terminals or fixed extension points enhance the coverage of a base station. The multiple subsequent wireless connections can be established within a homogenous system (e.g. cellular or RLAN system) or across different systems (e.g. some hops via cellular, some hops via RLAN systems).

3.2.6.2 Advantages of multi-hop radio network technology

Recommendation ITU-R M.1645 specifies as a target for systems "beyond IMT-2000" peak useful data rates of 100 Mbit/s (high mobility) and up to 1 Gbit/s (low mobility). Such data rates require large carrier bandwidths, which are most likely available only above 3 GHz. Both the large transmission bandwidths and the used frequency bands above 3 GHz will result in radio ranges that will be about one order lower than those of IMT-2000 systems.

This small range has a coverage and a capacity implication:

– In conventional single-hop radio networks small cell sizes would require correspondingly high numbers of base stations to achieve an ubiquitous coverage. This would increase the infrastructure costs.

– Small cell radii and the high data rates per cell lead to very high traffic capacities per area. This offered traffic capacity will very likely exceed significantly the average traffic demand per area. This leads to uneconomic network deployments.

Multi-hop radio network technology provide means to expand the coverage per base station and allow scalability of the radio network to match offered traffic capacity and demanded traffic capacity. Therefore this technology leverages fast deployment of wireless networks with low cost.

3.2.7 High altitude platform station (HAPS)

HAPS are stations located on an object at an altitude of 20 to 50 km and at a specified, nominal, fixed point relative to the Earth. In Regions 1 and 3, the bands 1 885-1 980 MHz, 2 010-2 025 MHz and 2 110-2 170 MHz and, in Region 2, the bands 1 885-1 980 MHz and 2 110-2 160 MHz may be used by HAPS as base stations to provide IMT-2000.

One proposed realization of a HAPS platform may consist of an extremely strong, lightweight, multilayer skin containing buoyant helium, a station keeping system consisting of GPS and an advanced propulsion system, a telecommunications payload, thin film amorphous silicon solar panels for daytime power, and regenerative fuel cells for night time power. The enabling technologies are high efficiency solar cells and fuel cells that are both lightweight and durable, high strength ultra thin fibre and helium impermeable seal, thermal and pressure control/management techniques, as well as advanced phased antenna array and MMIC (microwave monolithic integrated circuit) technologies.

A HAPS is designed with a lifespan of five to ten years. Service beyond this term is limited by the gradual degradation of solar and fuel cells, structural fatigue and the decomposition of gas-storage modules. Ongoing advances in high strength, lightweight, UV-resistant composite materials, fuel cells, solar cells, and compact, high-speed semiconductor device will likely extend the lifespan of second generation HAPSs.

An IMT-2000 terrestrial system utilizing HAPS consists of communication equipment on one or more HAPSs located by means of station-keeping technology at nominally fixed points in the stratosphere (at about 20 km altitude), one or more ground switching/control stations, and a large number of fixed and mobile subscriber access terminals. The system uses radio transmission technologies (RTTs) that satisfy IMT-2000 requirements to offer high density and high-speed communications capacity to fixed and mobile stations. The HAPS architecture is in concept much like a very tall terrestrial tower that is sectorized into hundreds of cells.

The HAPS telecommunications payload consists of multibeam light-weight reflector or phased-array antennas, transmit/receive antennas for gateway links with ground switching stations, and a very large bank of processors that handle receiving, multiplexing, switching and transmitting functions. The payload can utilize various multiple-access techniques and standards (e.g., TDMA, CDMA) that meet IMT-2000 requirements. The HAPS telecommunications payload can be designed to serve as the sole station in a stand-alone infrastructure (essentially, replacing the tower base station network with a "base station network in the sky") or can be integrated into a system that employs traditional terrestrial base station towers, satellites, and HAPSs.

A HAPS system will provide mobile cellular coverage and fixed wireless services to several regions ranging from a high-density (urban) area to low-density (rural) areas. The high gain transmit/receive antennas used on the HAPS project a large number of cells onto the ground in a pattern similar to that created by a traditional cellular system. The HAPS cellular coverage will likely include three regions:

– high-density (urban);

– moderate density (suburban);

– low-density.

The system dynamically reassigns capacity among the cells on a minute-by-minute basis in order to focus the capacity where it is most needed at any given time. For instance, the HAPS can direct additional capacity toward automobile traffic during rush hour and then shift it to a stadium during an evening sports event or performance. This gives the HAPS greater flexibility than traditional systems and can be used along or in concert with traditional terrestrial systems to prevent system overload in hotspots.

The above describes one system approach to HAPS. Alternate realization approaches are also feasible.

3.3 Mobile terminals

3.3.1 Terminal architecture

From the user's perspective, IMT-2000 and systems beyond IMT-2000 represent a fundamental change in expectation. Rather than merely expecting a "new and improved" but "static" collection of applications and services, the user will have an expectation of a dynamic, continuing stream of new applications, capabilities and services; a "Moore's Law" rate of advancement of new applications and capabilities.

Such a continuing stream will flow from a healthy ecosystem of general-purpose programmable platforms supported by a large, robust, and vibrant developer community.

New mobile user equipment (UE) are assuming characteristics of general-purpose programmable platforms by:

– containing high power general-purpose processors that follow Moore's law of dramatically increasing price/performance;

– providing a flexible, programmable platform that can be used for an ever-increasing variety of uses.

The convergence of wireless connectivity and a general-purpose programmable platform heightens some existing concerns and raises new ones, so that environmental factors as well as traditional technology and market drivers will influence the architecture of these devices.

Some important environmental factors are economic, security, and privacy.

Combining with the environmental factors we have traditional market and technology drivers: user value pull, security requirements pull, and technology enablers.

To maintain network and user space integrity, communications software will be "decoupled" and executed in parallel with user applications being written to a general-purpose processor running in a general-purpose execution environment. This partitioning maximizes the economic viability by allowing application development to evolve independent from communication standards, as well as enhancing security by providing autonomous network and user spaces.

Creating coexistent autonomy for the radio subsystem, application subsystem, and memory subsystems portions is evolving as a means to solve the triple environmental requirements of enabling economically viable products and services; while maintaining network and corporate security, and user sovereignty over application space and data privacy. Put anecdotally, "good fences make good neighbours".

3.3.2 RF micro-electro-mechanical systems (MEMS)

Future personal communication systems will require very lightweight, low power consumption, and small size. The requirements of IMT-2000 terminal such as small size, multifrequency bands, multi-mode and functional complexity demand the use of highly integrated RF front-ends and a compact system on chip solution. Despite many years of research, widely used discrete passive components based on electronic solutions cannot easily satisfy the above requirements of the future IMT-2000 terminal.

RF MEMS are integrated micro devices (or systems) combining electronic and mechanical components fabricated using an integrated circuit compatible batch-processing technique. This technology can yield small size, light weight, low power and high performance to replace discrete passive RF components such as VCO, IF, RF filters and duplexer. The system on chip using this technology can reduce the actual implementation size by 1/10.

As the users of the future wireless communication systems continually push handset manufacturers to add more functionalities, the manufacturers are confronted with trade-offs among cost, size, power and packaging constraints. It is anticipated that RF MEMS will emerge as a breakthrough technology to satisfy these constraints of future terminals. The commercialization of RF MEMS for future terminals will be within the next five years.

3.3.3 New innovative user interfaces

How the user experiences new telecommunications technology, depends on the services offered but also on the usability, design and quality of the terminals. Wearable computing is a popular study item at universities worldwide, giving new ideas of man-machine interfaces applicable also for mobile terminals.

Text messaging is the killer data application of today and a very frequency efficient way of communicating compared e.g. to a voice call. Multimedia messaging is expected to be the next boom, requiring a large display. Combining a practical text input method and a large enough screen on a single small terminal is a challenge.

So far, many of the solutions offered, e.g. for text entry, are not open standards but proprietary methods including IPRs. Proposed physical keyboards often tend to add features and/or buttons to the conventional dialling keypad instead of decreasing the number of keys that could instead be the goal in order to minimize the space required.

There is also a clear need for harmonization and for recommended use of common open interface standards in this area. For example, if a user gets used to one type of keyboard and becomes a committed and skilled user of it, she or he will get frustrated if the next phone, new version or another brand, has a different or slightly different user interface solution and the learning curve must be restarted.

3.3.3.1 One example of a new physical interface

Annex 13, as an example, describes a proposed method for combining text entry and a large display on a single compact mobile terminal. The presentation of the global keyboard optimized for small wireless devices (GKOS) back panel keyboard demonstrates that completely new types of physical user interfaces can still be found, and hopefully encourages manufacturers to study this issue more and maybe further refine the proposed concept to obtain a common standard for this kind of solution. The concept is an open standard and was first published on 5 October 2000. For more detailed information on GKOS, check also http://gkos.com/.

3.3.4 Reconfigurable processors, terminals and networks

3.3.4.1 Summary of the technology

Since bit-level data processing is needed in such areas as interleaving, error correction and detection, ciphering, and scrambling in mobile communication systems, particularly in baseband digital processing of mobile terminals as well as base stations, a processor which can perform high-speed bit-level data processing is required. However, general purpose processors, such as CPUs or DSPs, are not suitable for bit-level operation, and hence, a well-designed embedded processor with a reconfigurable unit is required so that user defined instructions are efficiently executed. It has a special execution unit for user custom instructions, other than normal execution units such as an integer execution unit. The special execution unit can be designed to be suitable for bit-level data processing, and it is realized with reconfigurable circuitry because custom instructions are different from application to application.

As shown in Annex 14, a typical reconfigurable processor has configuration information which defines connections between circuit elements in the execution unit and functions of those circuit elements, where the configuration information is supplied from configuration memory. It also has the capability that various custom instructions are executed by changing a corresponding address of the configuration memory. The configuration information can be updated in one clock cycle while the configuration memory can hold a set of configuration information. They can be rewritten in run-time, which allows users to define instructions other than the predefined ones. In a typical prototype, as shown in Annex 14, the configuration information is 256-bit length, the configuration memory can hold 32 sets of configuration information, and the size of reconfigurable circuit is about 50 K gates including a configuration memory, occupying several per cent of the entire size of the processor. This technology is also applicable to base stations.

3.3.4.2 Advantages of the technology

Since this type of processor has a reconfigurable unit that can handle many kinds of bit-level data processing, it can be applicable to various applications for mobile communication systems with efficient operation. For instance, the performance of the processor with the reconfigurable unit for processing a data encryption standard (DES) algorithm is more than six times higher than that of a processor without a reconfigurable unit. The processor can also be applicable to wireless communication systems, for instance, to Bluetooth digital baseband processing such as forward error correction (FEC), cyclic redundancy check (CRC), or scrambling at a speed a few times faster than conventional processors.

3.3.4.3 Issues to be considered

The reconfigurable unit is designed aiming for bit-level data processing and is suitable for various bit-level data processing tasks. It is applicable to some processing necessary for wireless communication systems. On the other hand, consideration needs to be given to another type of digital baseband processing which handles byte-level or word-level (non-bit level) data. For those types of data processing, especially for enhanced IMT-2000 or systems beyond IMT-2000, it might be necessary to either extend the reconfigurable unit or to adopt a different type of reconfigurable unit. This approach may result in the inevitable implementation of silicon devices of a huge scale, and hence, careful consideration needs to be given in terms of comprehensive efficiency, gate size, power consumption, and applicability.

4 Conclusions

This Report provides useful information on some of the technology enablers which are foreseen, such as the spread of IP-based technologies, increasing signal processing power in semiconductors and the enlargement of transport capacity in networks. Those technology enablers are in different areas, such as new radio technologies having an impact on spectrum utilization, access network and radio interfaces, mobile terminals, and system-related technologies.

It is expected that those technologies will be considered in the research and development of, but not necessarily used for, the future development of IMT-2000 and systems beyond IMT-2000. While this Report does not give an exhaustive list of potential technologies for the future development of IMT-2000 and systems beyond IMT-2000, it should be noted that other newly emerging technologies that are not covered in this Report would be taken into consideration as well.

5 Terminology, abbreviations

Table 3 provides an explanation of the terminology used for the current and enhanced IMT-2000 terrestrial technologies, and may prove useful in understanding the background to some of the topics presented in this Report.

TABLE 3

IMT-2000 terrestrial radio interfaces

Full name	Common names
IMT-2000 CDMA Direct Spread	UTRA FDD WCDMA UMTS
IMT-2000 CDMA Multi-Carrier	CDMA2000 1X and 3X CDMA2000 1xEV-DO CDMA2000 1xEV-DV
IMT-2000 CDMA TDD (Time-Code)	UTRA TDD 3.84 Mchip/s high chip rate UTRA TDD 1.28 Mchip/s low chip rate (TD-SCDMA) UMTS
IMT-2000 TDMA Single-Carrier	UWC-136 EDGE
IMT-2000 FDMA/TDMA (Frequency-Time)	DECT

The following listing of abbreviations and their meaning may similarly prove useful.

AA	Adaptive antennas
AAA	Authentication, authorization, accounting
ALU	Arithmetic-and-logic unit
AMC	Adaptive modulation and coding
API	Application programming interface
ARPU	Average revenue per user
BAC	Basic access component
BAN	Basic access network
BASM	Basic access signalling manager
BMM	Bandwidth management module
BS	Base station
BSI	Base station interface
CCN	Common core network
C/I	Carrier-to-interference ratio
CMM	Configuration management module
C/N	Carrier-to-noise ratio
CoMM	Cooperative mode monitoring
CRC	Cyclic redundancy check
CSI	Channel status information
CU	Central unit
DES	Data encryption standard
FDD	Frequency division duplex
FEC	Forward error correction
GKOS	Global keyboard optimized for small wireless devices
HAPS	High altitude platform station
H-ARQ	Hybrid ARQ
HDRPN	High data rate packet nodes
HRM	Home reconfiguration manager
IP	Internet Protocol
IMSI	International mobile subscriber identity
LMM	Local mobility management
LOC	Locator component
LRM	Local resource manager

MCS	Modulation and coding scheme
MEMS	Micro-electro-mechanical systems
MIMM	Mode identification & monitoring module
MIMO	Multiple-input multiple-output
MNSM	Mode negotiation and switching module
MUT	Multiservice user terminal
NI	Network interface
PAN	Personal area network
PDA	Personal digital assistant
PRM	Proxy reconfiguration manager
RAN	Radio access network
RAT	Radio access technology
RAU	Remote antenna units
RHAL	Radio hardware abstraction layer
RoF	Radio on fibre
RRM	Radio resource management
RSSI	Received signal strength indication
SDP	Session description protocol
SDR	Software defined radio
SDRC	Software download and reconfiguration controller
SDR-CF	SDR core framework layer
SHO	Soft hand-off
SIP	Session initiation protocol
S/N	Signal-to-noise ratio
SPRE	Software download and profile repository
SRM	Serving reconfiguration manager
SWD	Switched diversity
TDD	Time division duplex
TRSA	Terminal reconfiguration serving area
UE	User equipment
UWB	Ultra-wideband
WAP	Wireless application protocol

Annex 1

Technologies for improving bandwidth efficiency

1 Bunched systems

1.1 Introduction

In pedestrian and indoor environments, there will be severe fluctuations in traffic demands, high user mobility and different traffic types. This highly complex environment will require advanced RRM algorithms. It will be beneficial to have a central intelligent unit that can maximize the resource utilization.

The bunched system consists of a limited number of RAUs that are connected to a functional entity named central unit (CU). All intelligence as well as significant parts of the signal processing are located in the CU. The RAUs are simple antenna units capable of transmitting and receiving user signals. The local centralization at the CU level permits the use of near optimal algorithms for resource management because the CU has complete knowledge of all allocated resources at any time. This results in very efficient resource utilization within the bunched system. Furthermore, the bunched system can be enhanced to allow the RAN to detect changes, make intelligent decisions, and implement appropriate actions, either minimizing or maximizing the effect of the changes.

With a major shift from voice to high-data rate services for systems beyond IMT-2000, it is necessary to increase the system capacity. Bunched systems are well suited for hotspot applications. The coverage of bunched systems can be extended easily and has any desired geometrical shape. The move towards smaller cells will also make RAN planning process intrinsically more difficult and expensive. The bunched system can coexist with pre-existing microcell and cooperates with other bunched systems when it organizes the wireless network. Design issues of the RAN architecture and the RRM algorithms for the bunched systems must be addressed.

1.2 System characteristics

System characteristics and benefits of bunched systems are described and compared with those of conventional system.

– *Locally-centralized architecture*

  Bunched systems consist of a CU and multiple RAUs, and these RAUs are connected to the CU. Because all transceivers are installed in the CU, all the channel resources are controlled by the CU. Therefore, bunched systems have a locally centralized architecture where a CU controls all the RAUs. This system is different from the conventional cellular system by having one or multiple antennas at a cell site.

– *Hierarchical cell structures of RAN*

  Bunched systems can be regarded as the lowest cell layer, which covers dense traffic areas in a hierarchical cell structure. This system can coexist with existing macrocells as well as other bunched systems. Bunched systems are suitable for high speed data services in urban and indoor areas.

– *Dynamic load distribution*

Bunched systems can disperse system load concentrated on small regions. When traffic load is concentrated on one of the RAUs, the CU maintains service QoS by allocating more channel resources on that RAU.

– *Dynamic RRM*

Bunched systems can manage channel resources dynamically, since the CU can synchronize all the RAUs. By having transceivers in the CU, the CU can control all the channel resources. Borrowing of channel resources between RAUs can be controlled easily, making RRM dynamic and adaptive in an environment where large traffic change occurs.

– *Adaptive coverage control*

In bunched systems, coverage can be extended easily and the network can be planned as expected. This is achieved by deploying additional RAUs where coverage extension is needed. Moreover, dead spot problems can be solved by installing RAUs flexibly according to the radio network circumstances.

1.3 Required system control algorithm

Required algorithms are presented for bunched systems operation and construction. There are two types of algorithms, one is for control within a bunch (i.e. intra-bunch) and the other is for control between adjacent bunches (i.e. inter-bunch).

– *Intra-bunch*

  – Dynamic multicasting

  A dynamic multicasting scheme is used to improve the system performance. An RAU selecting scheme can be applied as a dynamic multicasting technique. With the selecting scheme, the interference is reduced by transmitting the signal through selected RAUs. Thus, an optimized selecting algorithm is needed.

  – Adaptive RAU coverage control

  Coverage control can disperse heavy traffic load from an RAU to another. This prevents QoS degradation and high-speed data service can be provided easily. Traffic load in each RAU must be measured, and an algorithm for adjusting RAU coverage according to load variation is needed.

  – Handover between RAUs within a bunch

  When a user moves into an adjacent RAU in the system with dynamic multicasting, handover between RAUs is needed. However, handover is not required in a system without dynamic multicasting, i.e. when a user signal is broadcast within a bunch. Serving RAU must be switched to another if only one RAU is selected for transmitting at any time. If the signal is transmitted through multiple RAUs, a group of the selected RAUs must be updated.

– Macro diversity techniques

Various multipaths of radio propagation are generated between multiple RAUs and a mobile terminal. Path diversity can be utilized, because signals are received through different paths of different RAUs. In this case, an algorithm is required to combine the received signals to get a macro diversity gain.

– Radio on fibre (RoF) technologies for RAU and CU links

RoF technologies are required to transfer signals between RAUs and the CU. RoF technologies can be used for a transmission scheme in which the radio signals are modulated to the optical signal. All the RAUs and the CU will require devices and algorithms for successful conversion of optically modulated signals into radio-modulated signals.

– *Inter-bunch*

– Dynamic resource assignment

Dynamic resource assignment can utilize radio resources effectively. Dynamic resource assignment schemes such as dynamic frequency allocation and power control can reduce interference between bunches, and improve system performance. Network status must be measured continuously to dynamically manage resources.

– Adaptive bunch coverage control

Coverage control can disperse heavy traffic load between bunches. Load dispersion by coverage control allows high speed data service. Therefore, like RAU coverage control, a coverage control algorithm according to load variation is required as well.

– Handover between bunches, and between bunches and macro BSs

Handover between bunches is needed when a user passes through bunches controlled by different CUs. When a user moves from a bunch service area to a macro BS area, a handover between different system layers is also required.

1.4 Illustrative capacity analysis of bunched systems

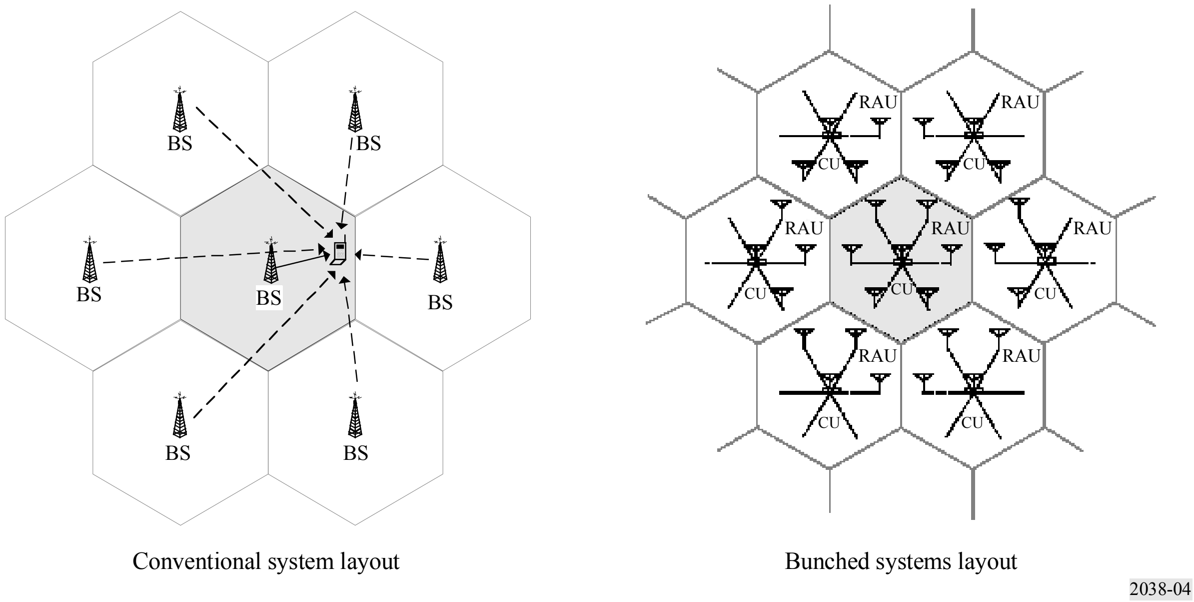
The bunched systems' capacity is evaluated in this section by applying the technology to CDMA radio access networks. Most studies on bunched systems have been focused on improving trunking efficiency in bandwidth limited systems (e.g. FDMA or TDMA systems) [Ariyavisitakul *et al.,* 1996]. Another benefit of bunched systems is reduction of path loss between terminal and RAUs, leading to the capacity improvement in interference-limited systems such as CDMA networks. When applying bunched systems to CDMA radio access networks, interference power on uplink can be decreased, increasing the capacity [Spilling *et al.*, 1999]. The next generation system requires efficient utilization of radio resource on downlink more than on the uplink, due to the traffic asymmetry. In this section, the downlink capacity gain of bunched systems over the conventional cellular system is derived and presented.

1.4.1 Model

Conventional BS systems and bunched systems in CDMA RANs are considered. Bunched systems consist of a CU and multiple RAUs. A cell consists of one BS in the conventional system, but multiple RAUs with one CU in the bunched systems. Both systems are modelled by locating BSs and CUs, respectively, at the centres of a hexagonal grid pattern as shown in Fig. 4.

FIGURE 4

System layout



Performance measure includes outage probability and BS transmission power. In a conventional system, received E_b/N_0 on downlink can be represented as follows:

$$\left(\frac{E_b}{N_0}\right)_{i,j} = \frac{P_{i,j} / L(d_{i,j})}{Isc + Ioc + NoW} \cdot \frac{W}{R_i} \qquad (1)$$

where:

i: index of terminal

j: index of BS

$P_{i,j}$: power allocated to i-th terminal at j-th BS

W: spreading bandwidth

R_i: data rate

N_0W: thermal noise power

$L(d_{i,j})$: path loss between j-th BS and i-th terminal (including antenna gain).

In bunched systems, the received E_b/N_0 can also be represented as follows:

$$\left(\frac{E_b}{N_0}\right)_{i,j} = \sum_{k=0}^{Number\ of\ RAUs}\left(\frac{P_{i,(k,j)}\ / L(d_{i,(k,j)})}{Isc,sr + Isc,or + Ioc + NoW} \cdot \frac{W}{R_i}\right) \tag{2}$$

where:

i: index of terminal

k: index of RAU

j: index of CU

$P_{i,(k,j)}$: power allocated to i-th terminal at k-th RAU of j-th CU

$P_{(k,j)}$: total transmission power of k-th RAU of j-th CU

$L(d_{i,(k,j)})$: path loss between i-th terminal and k-th RAU of j-th CU (including antenna gain)

$Isc,sr = (1 - \rho_{sr})\left(P_{(k,j)} - P_{i,(k,j)}\right) / L(d_{i,(k,j)})$: interference level from same RAU within a bunch

$Isc,or = (1 - \rho_{or}) \sum_{t \neq k}^{Number\ of\ RAU} P_{(t,j)} / L\,(d_{i,(t,j)})$: interference level from different RAUs within a bunch.

In bunched systems, the maximal ratio combining scheme utilizing macro diversity between RAUs can be applied to improve the received E_b/N_0. An ideal maximal ratio combining scheme is assumed in the above equation.

In the conventional cellular system, the received interference can be divided into same cell and other cell interference as shown in equation (1). In bunched systems, same cell interference can be regarded as the sum of two different interferences. One is a received interference level from the same RAU within a bunch, i.e. $I_{sc,sr}$, the other is an interference level from different RAUs within a bunch i.e. $I_{sc,or}$. The received signals from different RAUs along other paths can be less orthogonal to the desired signals than those from the same RAU. This reduction of orthogonality is considered in equation (2).

Outage is defined as an event where the BS (or RAU in bunched systems) has insufficient power to maintain the received E_b/N_0 at the required E_b/N_0. Hence, outage probability can be defined as follows:

$$P_{outage} = P\left(\left(\frac{E_b}{N_0}\right)_i < \left(\frac{E_b}{N_0}\right)_{req} \mid P_{TX} \geq \text{Maximum BS or RAU transmission power}\right) \tag{3}$$

1.4.2 Results

A computer simulation is developed for the performance evaluation. A WCDMA FDD system is selected as the system model. System parameters and the default values set for the analysis are listed in Table 4.

The bunched systems have one CU and seven RAUs. Each RAU is located at the 2/3 radius from the centre. The Monte Carlo simulation model consists of two tiers, and data from the centre cell is collected for statistics. The maximum transmission power of the BS is 20 W both in the bunched systems and the conventional system. Then each RAU in the bunched systems is considered to have a transmission power limitation of 20/7 W.

TABLE 4

Applied system parameters

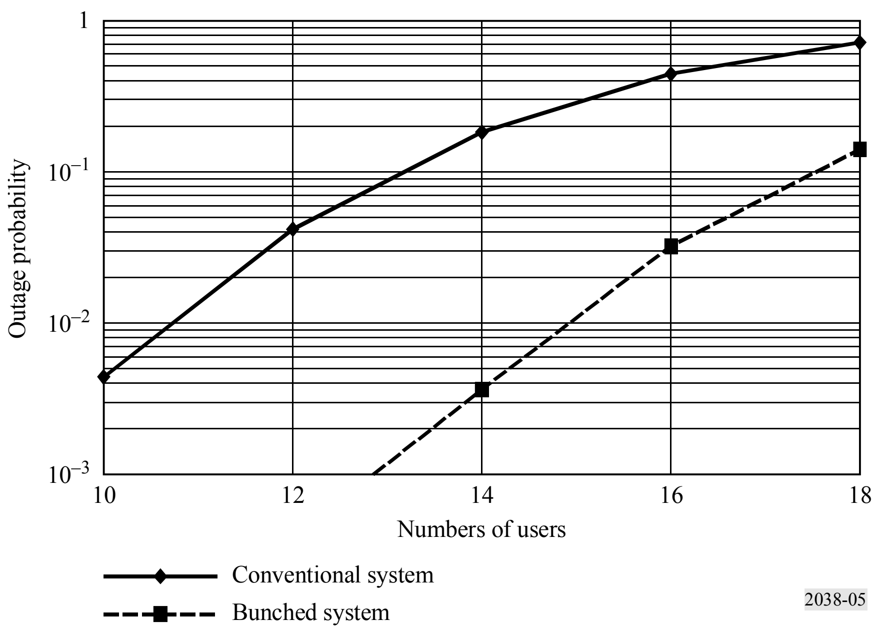
Parameter	Value
Cell radius	1 km
Chip rate	3.84 Mchip/s
Shadowing STD	8 dB
Overhead Channel power ratio	0.15
Antenna gain	15 dBi
Data rate	64 kbit/s
Required E_b/N_0 in downlink	5.4 dB
User distribution	Uniform distribution
Orthogonality within a RAU	0.6
Orthogonality between different RAUs	0.3
Cell TX power	20 W (43 dBm)
Power control	Perfect

1.4.2.1 Outage probability

Outage probabilities with various number of users are shown in Fig. 5. The results show that CDMA bunched systems can accommodate more users than the conventional system. It can be seen that the downlink capacity can be increased about 40% with the QoS goal of 1% outage.

FIGURE 5

Outrage versus number of users (64 kbit/s)

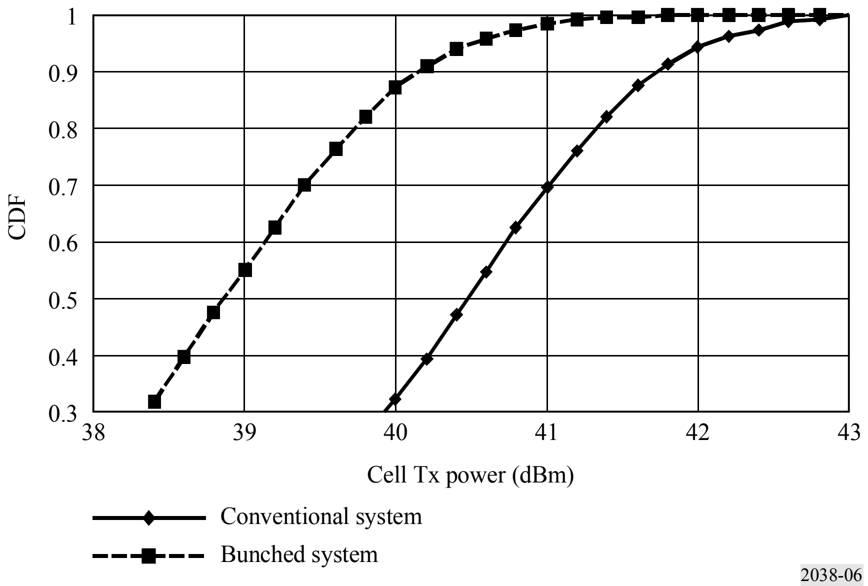


1.4.2.2 Total BS (or CU) transmission power

Figure 6 shows the cumulative distribution functions (CDF) of the BS and CU transmission powers. The result shows that CDMA bunched systems transmit lower power than the conventional system. The improvement is about 2 dB at median value. In the bunched systems, signal attenuation between user and RAUs is smaller than that of conventional cellular system, because distributed antennas are closer to the user. As the required transmission power is reduced, interference power decreases and more users can be accommodated.

FIGURE 6

Distribution of cell transmission power (64 kbit/s, 10 users)



2 Ultra-wideband technology

There are further possible access techniques under development like UWB technology. To date no internationally agreed definition of UWB exists because the applications and uses of these devices may be put to, for communications and other uses, are very diverse and have not been fully developed. Because many UWB devices and applications may be developed that have different technical and operational characteristics, the regulatory and interference impacts of UWB devices are not known yet. One administration has, however, adopted rules including technical standards and spectrum restrictions for low power UWB operations in an attempt to ensure that existing and planned radio services are adequately protected[3]. The applicability of UWB technology depends on the setting of appropriate interference limits and limitations on spectrum allowed for operation. The basic concept of UWB is to develop, transmit and receive an extremely short duration burst of RF energy. The resultant waveforms are extremely broadband (typically some gigahertz).

---

[3] The United States Federal Communications Commission adopted a First Report and Order, on 14 February 2002, on UWB transmission systems. See In the Matter of Revision of Part 15 of the Commission's Rules, Regarding Ultra-Wideband Transmission Systems, ET Docket 98-153, First Report and Order, 67 FR 34872, May 16, 2002. The document is available on the Internet at: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-02-48A1.pdf.

A time modulated ultra-wideband (TM-UWB) transmitter emits specially formed ultra-short pulses (or, monocycles) with tightly controlled pulse-to-pulse intervals. This can result in low average power, noise-like, CW-like, or pulse-like signals that can transmit data, voice and video communications or can be used as a personal radar, or as a positioning and tracking device. Single RF monocycles can be transmitted through a broadband antenna, and, by using a matched receiver, can be recovered again.

Known TM-UWB systems usually use pulse position modulation on a pulse-by-pulse basis. The receivers use a cross-correlator that gives the receiver the ability to detect and recover the signal. A single bit of information is generally spread over multiple monocycles. The receiver coherently sums the proper number of pulses to recover the transmitted information. This greatly increases the processing gain.

The extremely short pulse duration allows for a large number of transmit time slots. By shifting each monocycle's actual transmission time over a large time-frame in accordance with a suitable code, one can channelize pulse trains. In a multiple access system, each user would have its own code sequence. Only a correlation receiver operating with the same template waveform (e.g., code sequence) can decode the transmitted signal.

Presently, the effects of interference to digital communications systems from narrow pulses have not been widely studied, nor is there operational experience. In particular, CDMA cellular systems rely for their operation on fast power control. This is usually signalled over the radio interface using a few data bits (which are not interleaved because of the need for a fast response time). This means that digital communications systems (and cellular systems in particular) may be susceptible to interference that consistently corrupts particular bits within the transmitted data. For this reason, careful attention has to be spent on the compatibility of UWB systems with CDMA cellular systems, until the characteristics of the UWB emissions and their effect on digital communications are better understood.

More generally, some administrations have begun examining UWB technology and are engaged in extensive analyses of technical and national regulatory aspects of implementing UWB devices, including the potential for harmful interference from UWB devices into other systems, particularly safety-of-life services.

3 AMC (adaptive modulation and coding) and hybrid ARQ (H-ARQ)

IMT-2000 and systems beyond IMT-2000 are considering supporting a wide range of services, including high rate multimedia services. Such a growing demand for capacity makes it important to maximize the spectral efficiency. Various techniques have been proposed to make wireless communications systems spectrally efficient. One of the important research areas regarding this issue is AMC (adaptive modulation and coding). AMC can be used to increase transmission rates over fading channels. AMC schemes adapt to channel variation by varying parameters such as modulation order and code rate. The basic principle of AMC is to change the MCS based on the channel status information (CSI). Therefore, the scheduler has to know about CSI in order to select the appropriate modulation and coding scheme. Errors in the channel estimation, however, may cause the scheduler to select the wrong MCS level. Delay in reporting channel information also reduces the reliability of the estimated CSI due to the continuously varying mobile channel.

H-ARQ (hybrid ARQ) can be combined with AMC to increase overall performance. It enables the implementation of AMC by reducing the number of required MCS levels and the sensitivity to measurement error and feedback delay. Two well-known methods for H-ARQ are chase combining and incremental redundancy (IR). The chase combining method involves the retransmission by the transmitter of the coded data packet. The decoder at the receiver combines these multiple copies of the transmitted packet according to the received *S*/*N*. Thus, diversity gain is obtained. IR is another way for the H-ARQ technique wherein, instead of sending simple repeats of the entire coded packet, additional redundant information is incrementally transmitted if the decoding fails on the first attempt.

3.1 Adaptive downlink modulation based on user location, received *C*/*I* level, and required transmission rate

3.1.1 Introduction

Many modulation schemes such as single carrier (SC), direct spread (DS)-CDMA, orthogonal frequency division multiplexing (OFDM), multicarrier (MC)-CDMA have been proposed for mobile systems, nomadic wireless access and fixed wireless access. The selection of radio interface depends on the specifications of the system. OFDM is an attractive modulation scheme for its high immunity to multipath fading and its capability of offering high transmission rate. However, the link quality of the OFDM system is degraded when the co-channel interference signal level from adjacent cells is increased. On the other hand, the spread spectrum technique has tolerance to co-channel interference but it is difficult to enhance its transmission rate per user by restriction of allocated bandwidth.

OFDM and CDMA combined modulation schemes such as MC-CDMA and MC-DS/CDMA are attractive techniques that increase the processing gains in the frequency domain and time domain, respectively. In addition, the OFDM and CDMA combined scheme offers high transmission rate under multipath fading environments and mitigates co-channel interference from adjacent cells. However, bit error rate performance deteriorates when orthogonality between sub-carriers and spreading codes are degraded by large delay spread, frequency offset and other factors.

3.1.2 Technology concept

As explained above, each modulation scheme has distinct physical features. Namely, schemes have advantages and disadvantages in accordance with channel conditions such as *C*/*N*, *C*/*I*, delay spread, and other parameters.

Adaptive downlink modulation based on user location, received *C*/*I* level, and required transmission rate will be a candidate for enhancing the system capacity that allocates the preferable modulation scheme to each time slot per user. The features of the adaptive downlink modulation are summarized as follows.

– The base station (BS) allocates the preferable modulation scheme to each user per each time slot in accordance with link conditions such as propagation distance, received signal strength indication (RSSI) level, interference signal strength, delay spread, and required transmission rate of each user.

– Each time slot should be occupied by a single modulation scheme to avoid interference between different radio interfaces. Timing synchronization between multiple cells will be accomplished using GPS without any difficulty.

– The transmission power of each time slot is set to a different value in association with the modulation scheme. Transmission power control technique can also be applied to each modulation scheme with a different control algorithm.

– The channel frequency can be reused in every radio cell by allocating a different modulation scheme according to user location.

3.1.3 Example configuration

Figure 7 presents an example configuration of a cellular system adopting the adaptive downlink modulation scheme. OFDM and MC-CDMA are selected as the example modulation schemes for the adaptive downlink modulation scheme. The frame structure of the cellular system is arranged as shown in Fig. 8. In this Figure, a frame is divided into multiple slots. Some slots are allocated for OFDM and others are for MC-CDMA. Time synchronization of the frame and slot will be established between adjacent BS by GPS or other methods. Interference between OFDM and MC-CDMA can be avoided by allocating the time slots for each system independently.

FIGURE 7

Example configuration of a cellular system using adaptive downlink modulation

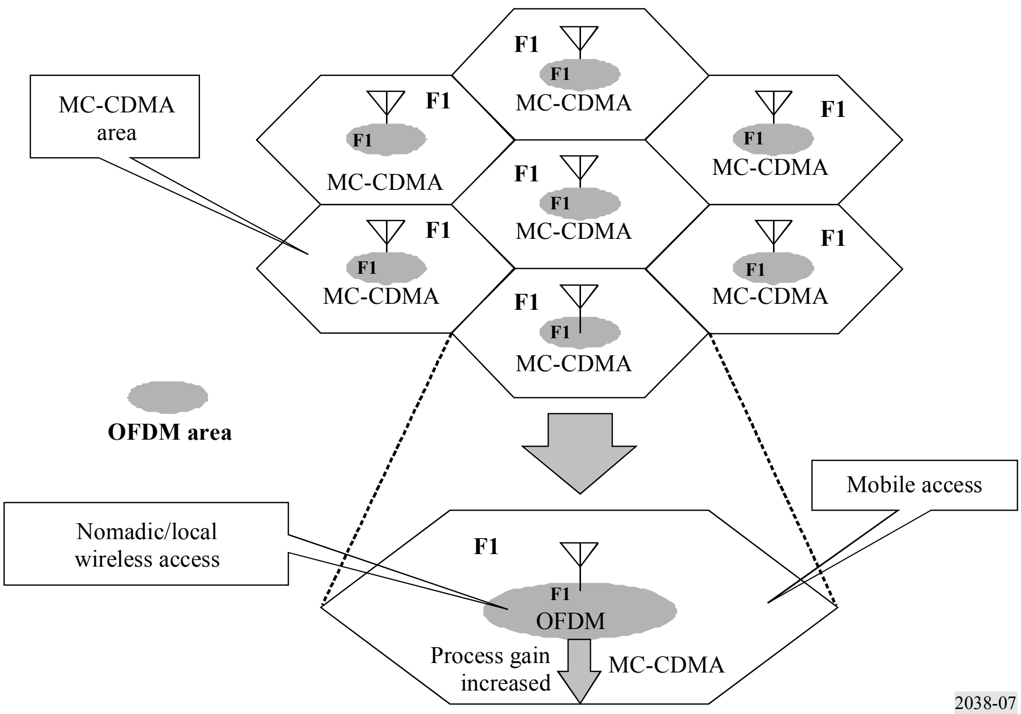
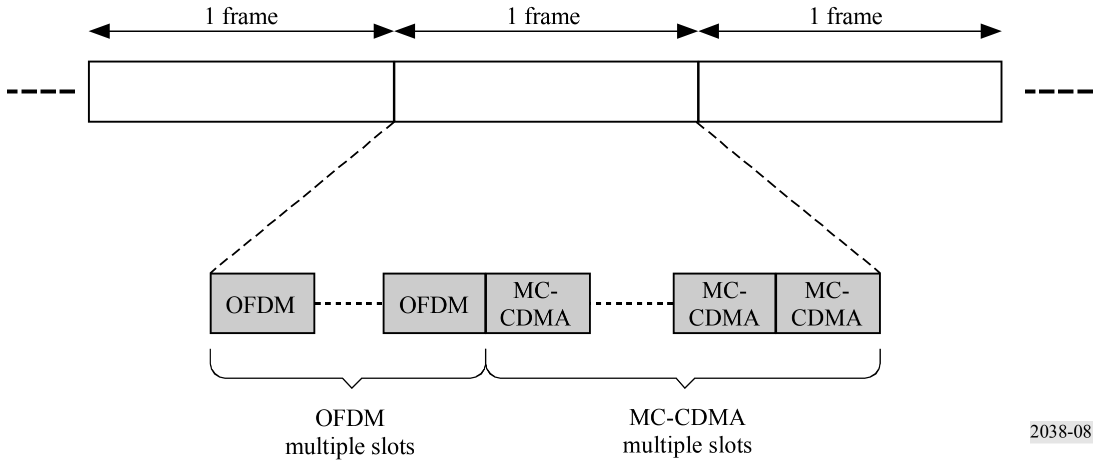


FIGURE 8

Example frame structure



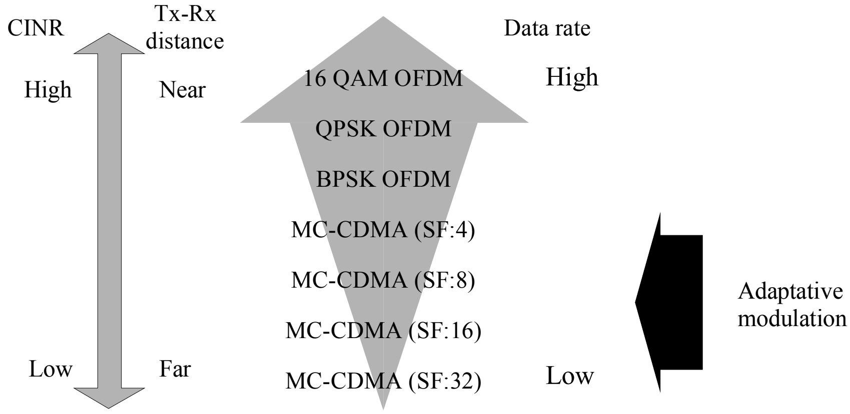
The transmission power for OFDM slots are set lower than that of MC-CDMA slots to avoid the co-channel interference between adjacent cells. In this case, service areas of OFDM slots are restricted around the BS and do not overlap as shown in Fig. 7. Therefore, the same channel frequency can be allocated in every cell, which will enhance the efficiency of channel utilization. This service area will be adequate for nomadic/local wireless access services with high data rate.

On the other hand, the transmission power of MC-CDMA slots is set higher than that of OFDM slots. The service areas of MC-CDMA are extended and overlapped as shown in Fig. 7. Co-channel interference between adjacent cells is mitigated using spreading code in the frequency domain. The selection of spreading code per user should take into account the orthogonality between the other codes used in the same cell and adjacent cells. As the same service areas of MC-CDMA signals are deployed as the current cellular systems, users will be able to establish their communication link under high mobility environments.

Figure 9 presents an example selection algorithm for the modulation scheme. When the CINR of the channel is high, the distance of the wireless link is short and the required transmission rate of the user is high, the BS assigns an OFDM slot with high rate modulation such as 16-QAM with high coding rate. If the CINR is very low and the distance of the wireless link is long, the BS allocates an MC-CDMA slot with high spreading factor and low coding rate to maintain the communication link. The concept of this algorithm is based on the combination of adaptively allocating the radio interface and adaptive selection of its parameters. As a consequence, the adaptive downlink modulation scheme will enhance the system capacity of wireless communications systems by allocating a preferable modulation scheme to each time slot per user.

FIGURE 9

Selection algorithm of modulation scheme



(CINR: Carrier to tnterference noise ratio)
(SF: Spreading factor)

2038-09

3.1.4 Conclusion

The adaptive downlink modulation scheme will be a promising technology for enhancing the system capacity of wireless communications systems by allocating the preferable modulation scheme to each time slot per user. Using this technique, the same channel frequency will be reused in every radio cell by allocating a different modulation scheme according to user location. Furthermore, co-channel interference from different modulation schemes will be avoided by separating the slot timing for each modulation scheme on time-basis. Transmission power of each radio interface is changeable to form different beam coverage. This feature will offer different service quality to different users in accordance with their requirements and channel conditions. By allocating the modulation scheme and its parameters (mapping pattern, coding rate, spreading factor, etc.) adaptively, users will be able to maintain their communications even in severe wireless environments.

The adaptive selection algorithm for the modulation scheme requires further consideration in conjunction with the specifications required for systems beyond IMT-2000. The number of slots per frame, slot length, frame length, and frame format for the system require examination. In addition, modulation type, coding rate, transmission power and other parameters for radio interface require appropriate selection for the adaptive downlink modulation technology to achieve high bandwidth efficiency. The measurement method for C/I, C/N, distance between the BS and MS, and delay spread have to be considered. The technologies combined with adaptive array antenna, diversity, space time coding, MIMO require further study to optimize the improvements.

4 Flexible spectrum sharing

4.1 A method for flexible inter-operator spectrum sharing

The radio-frequency spectrum resource is scarce and expensive. Spectrum efficiency of the cellular systems must, therefore, be optimized. This is critical especially in the presence of several competing operators in the same frequency band. Spectrum resources available to each operator should be dynamically adjusted to its needs. Sharing of frequency carriers between different operators is a method to optimize the use of these resources.

The spectrum allocation process is nowadays not very flexible. Indeed, operators are allocated a fixed amount of spectrum and they are committed to *a priori* some objectives in terms of amount of traffic and coverage. If they do not meet these commitments, part of their spectrum can be reallocated to other operators in need. However, this method may take a long time and is not flexible enough. Besides, it is not suitable if the under-utilization of spectrum occurs in peaks. A dynamic way of reallocating spectrum according to traffic needs is necessary.

This Report proposes a method to enable sharing of frequency carriers between different operators by dynamically adjusting a set of thresholds without any need of load information exchange between the operators, thereby prioritizing spectrum efficiency.

4.1.1 Description

4.1.1.1 Preamble

Frequency carriers can be shared between several operators to improve spectrum utilization. For example, in the following, frequency spectrum is used by two operators A and B, F_A and F_B are "proprietary frequencies", only network A and B respectively can use them. F_{shared} are two frequency carriers shared among networks A and B. That implies that mobile users of both networks may be assigned to these shared frequencies, depending on the load conditions. In this case, each network can potentially use up to four frequency carriers. Radio resource can therefore be adapted to the traffic needs, resulting in improved spectrum utilization.

FIGURE 10

Spectrum alocation with frequency sharing

F_A	F_A	F_{shared}	F_{shared}	F_B	F_B

2038-10

The objective of this technology trend is to propose a method for the management of these shared frequency carriers, bearing in mind the four principles mentioned above, by defining:

– admission rules of mobile users, either by direct access or hand-over;

– load control;

– priority management between operators for the use of shared carriers.

4.1.1.2 Terminology

Four thresholds are used in order to manage the load or the number of users on the frequency carriers:

– $T_{admission}$: new call admission threshold;

– $T_{HO\ accept}$: hand-over requests admission threshold;

– $T_{HO\ depart}$: hand-over departure threshold;

– T_{drop}: drop threshold, to start disconnecting some mobile users to ensure that certain quality of service is maintained (drop calls anyway since outage conditions).

Difference between T_{drop} and T_{HO_depart} should be defined by outage risks to warrant the best trade-off between outage probability and overall resources utilization.

4.1.1.3 Description of the proposed method

In case of the shared frequencies, two sets of these four thresholds are defined:

– *Default threshold set* refers to values of thresholds that are used to prevent the use of the shared frequency carriers.

– *Target threshold set* refers to values of thresholds when shared frequency carriers are used.

Each operator has the freedom to vary the four thresholds used on the shared frequency, *provided that they remain within the default and the target threshold sets.* The target and default threshold sets can be pre-defined or changed dynamically during the operations.

Basically, each operator uses a target set of thresholds that is different from the other operators. Priority rules are generated automatically by the relative difference of target thresholds between the operators. For example, if operator A uses a T_{drop} of 14 dB while operator B uses a T_{drop} of 20 dB, operator A begins to drop its mobile users before operator B starts to do so. *Dynamic setting of the priority rules is enabled in this way*. Moreover, the cost of using these frequency carriers can be easily derived by comparing the respective threshold values used.

4.1.1.3.1 Load control on the shared and proprietary frequency

In the following, we will illustrate the method using noise rise thresholds. However it should be noted that other measurement criteria are available, such as total power transmitted by BS, system outage, etc.

There is a set of thresholds for each operator on the shared frequencies, that are dynamically adjusted, depending on the loading conditions, as explained in the two cases below.

4.1.1.3.1.1 Proprietary frequency carriers are not fully loaded

The mobile users on the shared frequency carriers are to be handed-over to the proprietary frequencies. Admission of mobile users on the shared frequency carriers is stopped and the hand-over of the users from the shared to the proprietary carriers has to be favoured.

In this case, $T_{admission}$ and $T_{HO\ accept}$ decreases, and $T_{HO\ depart}$ as well to make the users leave the shared frequency. The T_{drop} can remain at the same level in order to control the outage on the shared frequency.

The adjustment of the thresholds should not be blind in order not:

– to overload the proprietary frequency carriers;

– to overload the system by performing too many inter-frequency handovers.

The final result is that all users will leave the shared frequency, either by hand-over and if not, at the end of their call.

4.1.1.3.1.2 Proprietary frequency carriers are fully loaded

In this case the objective is to increase the load on the shared frequencies.

New call requests are placed on the shared frequency carriers. Here, $T_{admission}$, as well as $T_{HO\ depart}$ and T_{HO_accept} are increased.

$T_{admission\_proprietary}$ and $T_{Hoaccept\_proprietary}$ are the thresholds used on the proprietary frequency.

Step 1: The noise rise is checked on the proprietary frequency in order to evaluate its loading.

Step 2: The noise rise is compared to the admission threshold.

Step 3: If noise rise is less than the admission threshold, the proprietary frequency is not fully loaded and is able to accept resource requests. As a consequence, the thresholds on the shared frequency have to be decreased within the limits given by the default and target sets to prevent utilization of shared spectrum as long as proprietary spectrum is underused.

Step 4: If noise rise is higher than the admission threshold, it is compared to the handover threshold.

Step 5: If noise rise is smaller than the handover accept threshold, the proprietary frequency is not able to accept new users but can handle users coming through handover from other cells. Therefore, the admission threshold has to be increased on the shared frequency in order for new users to be handled directly by the shared frequency.

Step 6: If the noise rise is higher than the handover threshold, the proprietary frequency will not accept any new user, it is fully loaded. In this case all the thresholds on the shared frequency have to be increased.

4.1.1.3.2 Increased flexibility: dynamic adjustment of the default and target sets

The default and target thresholds can be determined at the I of the operations. However, more flexibility is provided when the default and target thresholds can be dynamically adjusted. This is possible by using a central controller (CC) to communicate with a certain number of operators in a given geographical area, (e.g. to notify the use of the shared frequency carriers by some operators, evaluate the values of thresholds for each of the shared frequency carriers assigned to the operators).

4.1.1.3.2.1 Information exchanged

The amount of information to be exchanged is limited. Moreover, this kind of information will be exchanged only in case of major event, such as major overloading, following slow variations of traffic (busy hours or days, etc.).

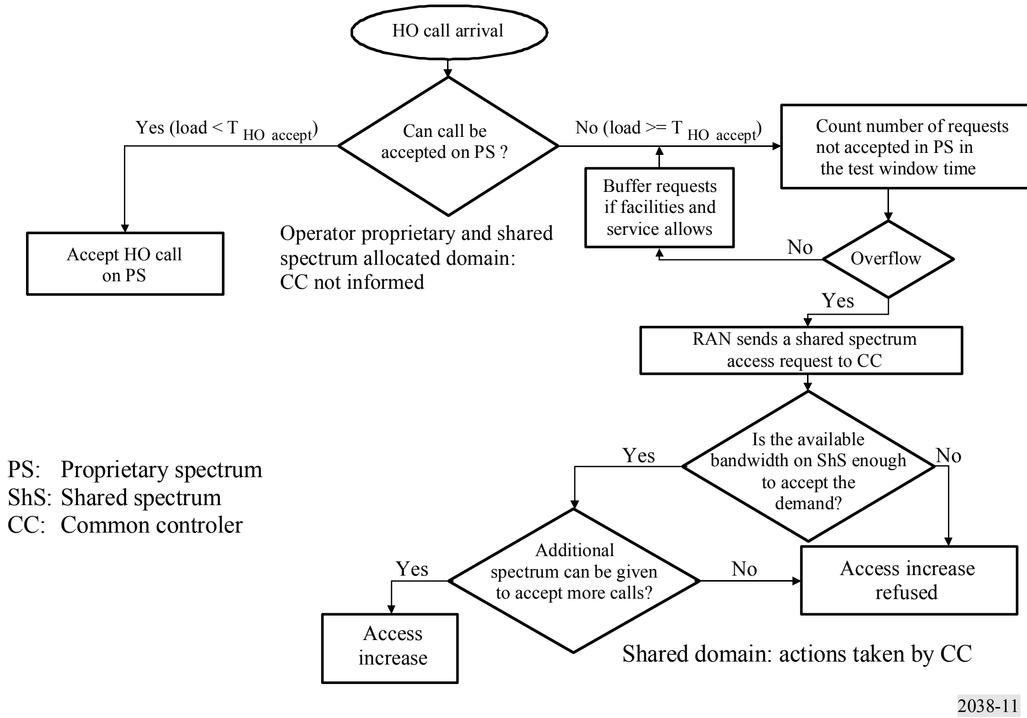
After analysing the access request, based on the current utilization of the shared spectrum and other concurrent access requests, the CC will send only an ACCESS INCREASE message if some bandwidth can be offered to the operator B on the shared spectrum. Note that the message will contain the set of target thresholds to be applied by the operator. This is enough to inform the operator B it can access the shared spectrum.

4.1.1.3.2.2 Shared spectrum management with the CC

Figure 11 displays the call admission procedure for this method. The example is given for HO calls but is equally applied to new calls.

FIGURE 11

Call admission in dynamic inter-operator spectrum sharing



Shared spectrum is accessed only when the proprietary spectrum is saturated. Before sending a SHARED SPECTRUM ACCESS REQUEST message to the CC, the overflow situation is first determined by the local controller. This is done by counting the number of unsatisfied requests over the observation time. During this observation time the access requests may be queued if such facilities exist and services might tolerate some waiting delay. Otherwise they are simply dropped.

After receiving the SHARED SPECTRUM ACCESS REQUEST, calls which have been buffered waiting for resources are accepted on the shared spectrum if enough bandwidth available. Otherwise, calls which have been eventually queued have to be dropped. This can be done either by sending a specific message to the local controller or simpler by timeout operations.

4.1.2 Conclusion and perspectives

Two sets of thresholds (default and target) are used on the shared frequencies as limits within which the actual thresholds can be varied. This enables to distribute the load between proprietary and shared frequency and to prioritize the use of the proprietary frequencies.

Priority rules and distribution of the cost between the operators are automatically derived from the specific values given to the default and target sets.

More flexibility in the adjustment of resource available to each operator can be provided by the notification of the availability of a frequency for sharing and the dynamic adjustment of the thresholds sets.

As mentioned earlier, four principles have to be respected to enable sharing of frequency between operators:

– This method consists of an algorithm that enables dynamic adjustment of resources available to each operator.

– By using the algorithm, these shared frequency carriers are to be used only when all proprietary frequency carriers are fully loaded. Regulatory controls and agreements can further enforce this.

– Priority between operators sharing these frequency carriers is determined dynamically by the values of the thresholds, enabling optimal load management policies and control of interference. The cost of the use of the shared frequency can be shared between the operators, the proportion for each operator being derived from the values of target threshold set.

– Exchange of load or confidential or sensitive information concerning each operator is not needed.

In short, the proposed approach makes the sharing of frequency carriers feasible by dynamically controlling a set of adjustable thresholds for call admission and termination, resulting in better utilization of the scarce spectrum.

4.2 Flexible spectrum sharing and cooperative mode identification and monitoring

A substantial problem raised by dynamic spectrum sharing is that information on alternative carriers and/or modes varies significantly with respect to time.

What is needed, then, are more dynamic methods of collecting and collating information on alternative modes and carriers.

A "brute force" approach, of just using the (scarce) gaps in transmit/receive patterns in existing standards, is not only difficult to schedule but is also very resource intensive; such an approach would reduce terminal battery-life to a level not seen for over a decade.

A static database orientated system would involve significant overheads due to frequent updates. A measurement of a given carrier in a given mode would, at first glance, appear to be unaffected by which operator was using that carrier. However, as soon as different operators start using that carrier, they would be using their own base-stations, which may well be in different locations to the previous operator's base-stations, hence affecting the propagation characteristics of that carrier.

Cooperative mode monitoring (CoMM) is the idea that, rather than a roaming multi-mode terminal having to (temporarily) reconfigure in order to monitor alternative modes, it is simpler to obtain an idea of coverage in these alternative modes by "asking" terminals that are already operating in that mode for their measurements, or by extrapolating from other measurements taken by terminals in that location. Whether measurements are requested on demand, or are collated centrally over a period of time, will depend on the resultant network loading, and how quickly measurement values "expire".

4.3 Conclusions and perspectives

This Report presented new solutions to two fundamental issues in flexible spectrum sharing: how to enable sharing of frequency carriers between different operators without exchanging load information and how to detect and monitor RATs in blind or cooperative ways. The proposal presented above extends the framework required for flexible spectrum sharing application. Thus, studies should concentrate on evaluating service negotiation strategies combined with the efficiency of spectrum sharing. The suitability of such schemes has to be assessed as they relate to network performance and user satisfaction levels.

4.4 References

ARIYAVISITAKUL, S. *et al.*, [May 1996] Performance of simulcast wireless techniques for personal communication systems. *IEEE J. Select. Areas Comm.*, Vol. 14, **4**, p. 632-643.

SPILLING, A. G. *et al.*, [May 1999] Adaptive networks for UMTS – An investigation of bunched base stations: *Proc. IEEE VTC*, p. 556-560.

4.5 Bibliography

SPILLING, A. G. *et al.*, [June 2000] Self-organization in future mobile communications. *Electron. & Comm. Engin. J.*, p. 133-147.

Annex 2

Technology solutions to support traffic asymmetry

1 Technical aspects

In the following, details of the different concepts in § 3.1.2 are summarized. Further details are presented in Recommendation ITU-R M.1036.

1.1 Frequency band asymmetry

With FDD a different amount of uplink and downlink spectrum, separated in the frequency domain, may be allocated for the respective direction.

Three cases can be distinguished (typically multiple users share each carrier):

– In the first case, the uplink and downlink carrier are of equal bandwidth. However, more carriers are allocated for the downlink compared to the uplink (or vice versa) in order to provide more overall downlink capacity compared to the uplink capacity (or vice versa).

– In the second case, the downlink carrier bandwidth is larger than the uplink carrier bandwidth (or vice versa).

– In the third case, the same technology is used, in both directions, but well-known multiple access techniques are used to share the uplink carrier while continuing with a dedicated downlink carrier.

The above cases require a reasonable *a priori* estimate of the expected asymmetry for a spectrum efficient frequency allocation and spectrum usage.

In order to support flexible asymmetric capacity by means of frequency band asymmetry using the first two methods above, the equipment needs to have the capability for variable duplex distance.

With the third method the same carrier bandwidth is used in both directions, but the multiple downlink channels have a common relationship to their own block edge. The lower downlink blocks can provide conventional symmetric operation and some asymmetric operation, while the upper additional downlink blocks with fixed carrier separation to the lower downlink blocks provide support for asymmetric operation (for details see Recommendation ITU-R M.1036). Thus the same equipment can be common to all operators. The duplex spacing for equipment using the same blocks is common and instead of variable duplex capability several fixed duplex spacings are required, thus simplifying terminal complexity.

Opening up the upper blocks can be timed as additional spectrum and enhanced band equipment becomes available. Note that there is no requirement for the upper blocks to be adjacent to each other. This could be overlaid on “existing” 2G or IMT-2000 technologies, where symmetric bands are already deployed.

1.2 Time slot allocation asymmetry

TDD is a duplex technique where both uplink and downlink traffic takes place on the same carrier. Downlink and uplink channels are separated in the time domain by dividing the time-frame into slots. Each time slot can be allocated to either uplink or downlink traffic. By allocating different numbers of slots for uplink and downlink, asymmetric capacity can be obtained.

A frame comprises N time slots, where n time slots are used for the downlink and N-n time slots are used for the uplink. Another option is to use variable-size time slots rather than multiple time slots.

For a TDD system to work properly and achieve good performance a cell should generally avoid having a downlink/uplink configuration different from co-channel or adjacent-channel cells, especially in case of a frequency reuse equal to one to avoid severe interference. However, this can be solved to a large extent if all operators both synchronize their networks and agree on the downlink/uplink configuration in all cells.

In the single-operator case and with a frequency reuse larger than one different downlink/uplink configurations in different cells are feasible due to the additional de-coupling by an increased geographical distance between same carrier frequencies. Some degree of frequency planning and coordination would be required.

1.3 FDD or TDD duplex arrangement combined with other technologies

Modulation asymmetry is a general technique, applicable to systems operating with either FDD or TDD, and where different modulation schemes may be used on uplink and downlink channels, respectively, in order to obtain different data rates and can provide some capability for asymmetric traffic. The maximum ratio and direction of asymmetry is basically limited by the equipment design and the modulation formats implemented in practical systems.

However, higher order modulation schemes or coding with reduced overhead require higher *S*/*I* than the scheme in the other link. Therefore, such concepts can only be applied as a trade-off between link capacity and coverage for packet services in a link adaptation mode. Using different modulation techniques supports increased peak data rates for users having good radio conditions.

This method could be used to enhance both FDD and TDD systems in future extensions of the IMT-2000 standards and systems beyond IMT-2000. This means that the important regulatory question whether new spectrum should be reserved for TDD or FDD systems does not need to include the consideration of modulation asymmetry.

An additional method to improve asymmetric capacity is the use of adaptive antennas or more advanced detection schemes to increase the link capacity for one link. This would provide some additional possible asymmetry for a given frequency allocation. These techniques will add capabilities either alone or in combination to both FDD and TDD based systems. Similar radio propagation properties for TDD uplink and downlink show some advantages in the application of adaptive antennas for low mobile speed.

Advanced detection schemes to mitigate the impact of co-channel interference can be applied to both FDD and TDD.

2 Comparison of the various methods to provide asymmetric traffic capability

In general, the duplex scheme is one of many factors that determine the overall spectrum efficiency of a system. In terms of efficient support for asymmetric traffic, both FDD and TDD schemes have inherent advantages and disadvantages.

2.1 FDD scheme with symmetrical spectrum allocation

The maximum available user data rate per link is fixed.

This scheme has the following advantages:

– It allows for continuous (non-bursty) transmission in the uplink and downlink. This also allows for faster signalling of feedback information for, e.g. power control, link adaptation, and fast channel-dependent scheduling.

– For wide area coverage the range is primarily limited only by the system margin.

– No additional particular requirements for adjacent channel isolation or co-planning of systems in adjacent channels compared to TDD.

– Multi-operator co-location of BSs is possible depending on system design and frequency reuse independent of overall spectrum asymmetry.

– No inherent relation between the range of available maximum service data rates and the degree of asymmetric capacity, which is the case for TDD.

– Flexibility, to a certain extent, to traffic asymmetry.

Potential disadvantages to be considered are:

– Symmetric paired spectrum with a minimum duplex distance is required.

– The spectrum efficiency of the arrangement depends on the relation between the symmetric spectrum and the actual network traffic asymmetry.

2.2 FDD scheme with asymmetrical spectrum allocation

The maximum available user data rate per link is fixed.

This duplex scheme has the following advantages in addition to § 2.1:

– Flexible pairing of uplink and downlink carriers is possible that allows for asymmetric capacity. The spectrum usage is most efficient if the selected bandwidth ratio for both bands corresponds to the traffic asymmetry. The asymmetric spectrum can be used either as having more carriers in one direction, or as having wider carriers in one direction or a combination thereof.

– The multibandwidth alternative allows for higher peak rates in the direction with the wider band.

– Availability of additional unpaired spectrum is sufficient.

Potential disadvantages to be considered are:

– Asymmetrical paired spectrum requirement.

– Estimation of future spectrum demand per direction is required and may be difficult in advance. An immediate adaptation may be difficult to implement but indications are that more spectrum in both uplink and downlink are needed and later adaptation is then possible.

– The requirement for flexible duplex spacing, for the methods that require it, slightly increases the terminal implementation complexity.

– The spectrum efficiency of the arrangement depends on the relation between the degree of actual network traffic asymmetry and the degree of spectrum asymmetry.

– Multirate, multiple bandwidth capability for channels of different width are required if the uplink and downlink carriers are designed to have different bandwidths.

2.3 TDD scheme

The maximum available service data rate per link depends on the ratio of asymmetry.

This duplex scheme has the following advantages:

– Availability of unpaired spectrum is sufficient. The identification of single blocks of spectrum may be easier than for paired spectrum.

– Flexibility is available with respect to the degree of traffic asymmetry, depending on the co-channel and adjacent channel interference conditions. The spectrum usage is independent of the location of the switching point between uplink and downlink transmission.

– The spectrum efficiency of the arrangement is less dependent on the actual network traffic asymmetry since TDD can vary the degree of asymmetry within a specified range.

– If the neighbouring cells/systems agree on the same slot configuration, depending on system design and frequency reuse, the range of asymmetry is given by the number of time slots.

– Capacity increase by using adaptive antennas can be further improved by using the reciprocity of the radio channel for low mobile speeds.

Potential disadvantages to be considered are:

– Services in adjacent bands must be able to cope with both uplink and downlink interference.

– Synchronization and coordination of uplink/downlink of neighbouring cells is required with a small frequency reuse; in the case of a sufficiently large reuse cluster size no coordination is necessary within an operator's frequency allocation but it is still needed between operators having frequency bands adjacent to each other.

– Multi operator co-location of BSs depends on the system design, the frequency reuse and the frequency separation of co-located operators.

– Isolation between adjacent channels is required.

2.4 Comparison of the alternative technologies to provide asymmetric traffic capability

Asymmetric modulation has the following advantages:

– No additional frequency allocations or time slot allocations are required.

– A better throughput for the same channel, however, only for lower range or cases with higher S/N ratio (link adaptation necessary).

– This method could be used to enhance both TDD and FDD systems in areas with favourable radio conditions.

Potential disadvantages to be considered are:

– Influence of direction on system margin.

– Only limited possible ratio of asymmetry.

– Trade-off between coverage and maximum available data rate.

– Link adaptation is required.

– Mainly applicable for packet services.

– More complicated planning and implementation.

– If capacity on one link can be improved, the same method does also increase capacity on the other link; therefore, no inherent advantage with regard to improving asymmetric service provision.

Application of additional techniques provides the following advantages:

– Improvements can be applied to both, FDD and TDD.

– Improvement of capacity by adaptive antennas in the downlink only by transmit antenna for simple terminals.

– Adaptive transmit antennas are more effective for TDD than for FDD.

Potential disadvantages to be considered are:

– More capacity will be needed in the downlink, which would increase terminal receiver complexity compared to the base station for advanced detection schemes.

– These methods would put higher complexity in the terminal than in the BS.

The selection of a suitable duplex arrangement depends on the application (short range, wide area, flexibility, etc.). The duplex scheme is only one of several aspects to be considered (other aspects are, e.g. interference sensitivity, switching point flexibility, etc.).

2.5 Comparison of normalized spectrum usage for FDD and TDD

The spectrum usage can be characterized by the total average throughput in downlink and uplink normalized to the allocated total bandwidth for the downlink and uplink (normalized throughput) for a given spectrum allocation and traffic asymmetry. The spectrum allocation is fixed, where the traffic asymmetry may vary. For FDD a symmetric spectrum allocation is assumed here.

In FDD the throughput in both links can be adjusted independently between $0 < a_{\uparrow} < 100\%$ and $0 < a_{\downarrow} < 100\%$. The spectrum usage for traffic asymmetry is maximized if one of the links is completely used with $a_{\uparrow}$ or $a_{\downarrow}$ equal 100% and the other parameter may vary with respect to the ratio of asymmetry with less than 100%. In the case if both $a_{\uparrow}$ and $a_{\downarrow}$ are less than 100%, the spectrum is used less efficiently.

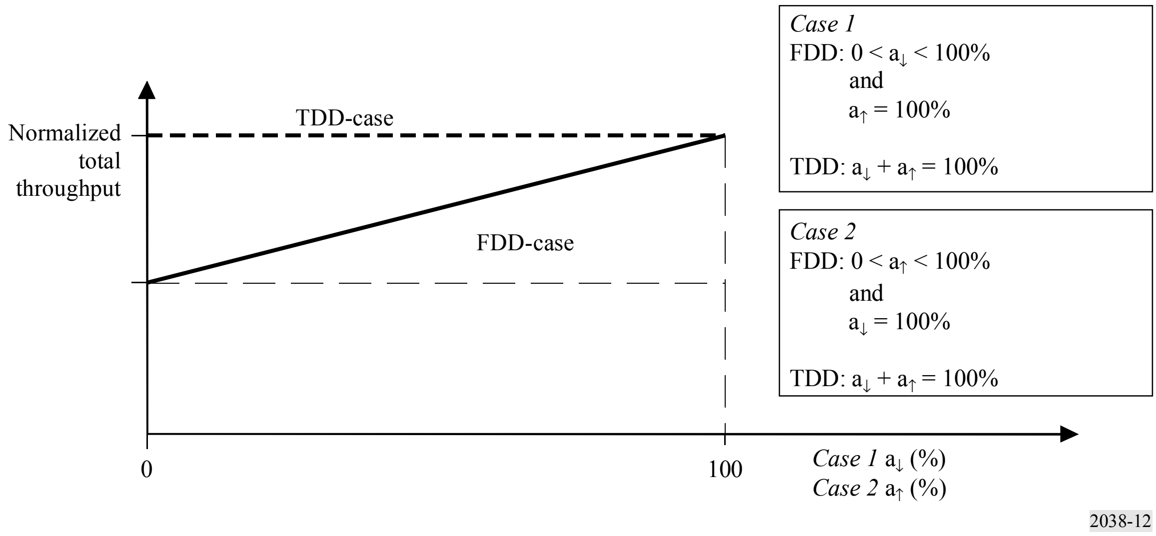
In the case of TDD $a_{\uparrow}$ and $a_{\downarrow}$ cannot be adjusted independently, because under optimum conditions the sum of both equals 100%. If the time slots can be adjusted according to the traffic asymmetry and all time slots are used, then the spectrum usage is optimal.

Under these optimum assumptions for FDD and TDD Fig. 12 shows a comparison of the spectrum usage for both duplex schemes versus $a_{\uparrow}$ and $a_{\downarrow}$. The adjustment of the location of the switching point between uplink and downlink transmission in TDD may be constrained (e.g., to maintain synchronization with other cells, traffic asymmetry from other users in the same cell). In that case, the throughput of TDD would approach that of FDD.

The selection of a suitable duplex arrangement depends on the application (short range, wide area, flexibility, spectrum availability of potential new paired or unpaired spectrum etc.). The duplex scheme is only one of several aspects to be considered (other aspects are, e.g. interference sensitivity, switching point flexibility, etc.). This Annex provides information about the trade-off between efficient spectrum usage and maximum throughput per direction including other advantages and disadvantages, which have different impacts depending on requirements. The spectrum availability will not be known before WRC-07. Therefore, no general guideline can be derived yet.

FIGURE 12

Comparison of normalized spectrum usage in the case of FDD and TDD



Annex 3

Advanced system innovation using TDD

Members of the TDD Coalition, for example, have developed important technical innovations through the implementation of TDD.

System advantages can be obtained from the use of reciprocal channels – a unique feature of TDD systems. Channel reciprocity for single carrier frequency shared by uplink and downlink allows an easier access to channel-state information for advanced signal processing techniques. For instance, channel reciprocity ensures that the fading on the uplink and downlink are highly correlated. Since the channel characteristics are same in both directions, any signal processing resources for doing space/time/equalization/frequency processing can be shared between the transmitter and receiver. Hence, TDD is a uniquely suited technology for advanced signal processing in the areas of open-loop power control, novel multipath and antenna combining, and time-space processing techniques, with a lower cost adder.

For example, adaptive antenna arrays can be added by implementing advanced signal processing at the BS and sharing the channel weighting information with the subscriber units. This allows the spectral efficiency of the system to be increased by an order of magnitude without increasing the CPE cost. Another example is mesh networks. These innovative systems, using an architecture hitherto more common for military HF systems, are better facilitated by TDD implementation. The network and frequency planning in a substantial FDD mesh (MP-MP) system is significantly more difficult than for conventional multi-way P-P or P-MP (area coordinated) systems.

Another area of innovation is in MAC. The TDD operation allows for highly dynamic and various configurations of physical layer time frame. TDD systems have a much higher flexibility to handle the dynamic up/down traffic, since the boundary between uplink and downlink duty cycle could be adaptively adjusted to accommodate the service requirements. Dynamic TDD systems are far more bandwidth efficient than the traditional FDD systems for the future data-centric multimedia traffic. By implementing intelligent (channel-aware) MAC protocols and use of the superior architecture provided by TDD, throughput multiplication, statistical multiplexing gain, and reduction of packet delays can be achieved.

1 Introduction

TDD offers solution for asymmetric high data rate services and provides flexibility in deployment of networks in a variety of environments including busy urban, hotspot and busy indoor environments as well as wide area applications at low cost. TDD supports all voice and data applications, providing efficient use of spectrum for the most data-intensive services. It is the most effective air interface for asymmetric, "bursty" data applications in "always on" mode. This capability is crucial as the number of wireless Internet applications and multi-media services for consumers and corporate/business users increase over the next few years. The TTD technology thus provides operators with an opportunity to be able to deploy sufficient capacity and capability in order to increase their average revenue per user (ARPU) through offering bandwidth-hungry and asymmetric data services.

TDD is based on the concept of transmit and receive on the same frequency which means that both uplink and downlink channels experience more or less the same radio channel conditions. This reciprocity in the uplink and downlink can be used to the best advantage to introduce new and innovative techniques where most of the intensive signal processing can be carried out at the base station or the user terminal and the BS will be able to utilize information from the channel to the best advantage of the system. These techniques (described in the following sections) can be used to improve both coverage and capacity.

TDD is also cost-efficient for network deployment as it can leverage the infrastructure of an FDD-only roll-out to offer scalable capacity for "hotspots" where combined voice and data traffic will be supported through a multi-tier architecture of macro-, micro- and picocells.

Overall, TDD offers a platform for systems beyond IMT-2000. This is further described in the following section.

2 Enhanced TDD – A key platform for systems beyond IMT-2000

Several of the key inherent features of TDD as well as the developments currently undertaken by standards working parties and study groups make TDD an ideal platform for systems beyond IMT-2000. Examples of key features include the following.

2.1 High data rate support

TDD technology is enhancing towards higher and higher data rate support. This is in line with the increasingly data centric usage projected. Higher order modulations combined with fast link adaptation will provide adaptive modulation and coding techniques that reduce the *S*/*N* requirements and allow a more efficient data communications, in effect increasing system data transport capacity. In addition, improved diversity techniques and smart antenna techniques will help support high rate data traffic. Furthermore, techniques such as H-ARQ and its variations will make data transmission more efficient from the air interface point of view. These features and others are being introduced to allow TDD support ever-increasing data rates.

TDD capability to provide both time as well as code multiplexing, make it more attractive for high speed access schemes, because it allows more flexible and efficient use of physical channel resources, which consent to better integration of different service types (i.e. voice, data, etc.). This in turn provides higher capacity as well as higher spectral efficiency.

In addition, the inherent reciprocity in uplink and downlink in TDD would make it more feasible to deploy techniques such as smart antennas and diversity techniques to be able to support higher data rate traffic.

2.2 Higher spectral efficiency

In addition to the signal processing techniques and modulation schemes mentioned earlier, very efficient resource allocation algorithms make TDD spectrally efficient, especially when supporting asymmetrical data such as for Internet related services. The so called slow and fast dynamic channel allocation algorithms make sure that resource units are allocated optimally for uplink and downlink transmission.

2.3 Improved cell planning and coverage

TDD provides a second "dimension" in cell planning when it is deployed together with FDD systems. When addressing hotspots, splitting cells, and serving data centric usage areas, TDD and FDD systems could be laid out in a way that takes advantage of the strengths of each. When doing the cell planning, FDD and TDD coverage areas can be treated almost independent and provide the operator with a new plane of coverage maps, given the coexistence practices are followed.

2.4 Flexible and IP data centric deployment

As data rates demanded become higher and higher, cell sizes shrink and evolve towards micro and picocells. TDD technology can also be offered in picocell format and address high data rate users in support of IP-centric applications requiring high bandwidth. With TDD, the architecture of BSs and user terminals are planned to take full advantage of multi-user detection algorithms, diversity and antenna processing techniques required for such environments.

2.5 Integrated multi-mode offerings

Dynamics of the standards bodies and their activities so far indicate that more than one standard is expected to emerge as systems beyond IMT-2000 with multiple evolution paths. TDD technology will be offered as part of the multiple-solution products that support more than one standard. Inherent features of the enhanced TDD offer the opportunity for cost-effective common architecture alongside other systems beyond IMT-2000.

3 TDD current and emerging system innovation

TDD offers an excellent opportunity for future systems innovation through its inherent features. One of its key features is the ability to transmit and receive on the same frequency thus allowing innovative techniques to take advantage of the reciprocity of the uplink and downlink channels. Key examples of innovative techniques that are being considered are as follows.

3.1 Channel sensing and reciprocity

Characteristics of a wireless channel vary in time and frequency. The uplink and downlink channels of a wireless communication system are said to be reciprocal if the channel impulse response does not vary significantly between the transmissions in uplink and downlink. For FDD systems, duplex gap requirements on the separation of uplink and downlink frequencies eliminate channel reciprocity. System advantages, however, can be obtained from the use of reciprocal channels – a unique feature of TDD systems. The uplink and downlink channel responses of a TDD system are reciprocal if the dwell time is reasonably small. Channel reciprocity for single carrier frequency shared by uplink and downlink allows an easier access to channel-state information for advanced signal processing techniques. For instance, channel reciprocity ensures that the fading on the uplink and downlink are highly correlated. Since the channel characteristics are same in both directions, any signal processing resources for doing space/time/equalization/frequency processing can be shared between the transmitter and the receiver. Hence TDD is a uniquely suited technology for advanced signal processing in the areas of open-loop power control, novel multipath and antenna combining, and time-space processing techniques, with a low additional cost.

As an example of the benefits of channel sensing, a base station equipped with adaptive beam forming arrays can sense the environment in the uplink but must extrapolate the channel conditions to the downlink unless TDD is being used. While beam forming techniques introduce improvements to both TDD and FDD systems, using the antenna array to improve downlink performance of an FDD system is usually a more difficult problem than the uplink, due to the lack of direct measurement of downlink channel responses. Traditional methods of FDD downlink beam-forming such as direction of arrival (DoA)-based approaches use the uplink signals to construct the downlink channel response. Such techniques require very complicated computations and do not perform well in the presence of severe multipath. Also, applying blind downlink beam-forming that utilizes the uplink spatial channel characteristics yields to sub-optimum performance due to the direction of arrival-direction of departure (DoA-DoD) angular offset caused by multipath channel. On the other hand, the channel de-correlation in an FDD system causes blind downlink optimum combining schemes to perform sub-optimally when the duplex gap is greater than only a few MHz. Channel reciprocity in TDD, on the other hand, acts as an inherent feedback and allows the adaptive antennas to perform at their best for both uplink and downlink.

3.2 Adaptive antennas

Adaptive antenna arrays can be added by implementing advanced signal processing at the base station and sharing the channel weighting information with the user terminals. In TDD systems, this allows the spectral efficiency of the system to be increased by an order of magnitude without increasing the cost of the user terminal.

3.3 Multi-user detection

In CDMA systems, users are simultaneously active on the same channel, differentiated by their specific orthogonal codes. The orthogonality of these codes protects users against multiple access interference. This orthogonality is, however, lost to some degree in the presence of frequency-selective fading. Multi-user detection techniques can be used to combat the effect of multiple access interference. All these techniques require knowledge of the channel impulse response. Estimation of the channel, especially in the downlink, can be carried out in a much simpler, more efficient way with TDD, as discussed earlier in § 3.1.

Annex 4

Adaptive antenna concepts and key technical characteristics

1 Introduction

This Annex identifies the key adaptive antenna concepts and briefly describes their technical characteristics. The traditional approach to the analysis and design of wireless systems has generally been to address antenna systems separately from other key systems aspects, such as:

– propagation issues;

– interference mitigation techniques;

– system organization (access techniques, power control, etc.);

– modulation.

Adaptive antenna technologies are best implemented with an overall system approach, where all the system components, including the antenna system, are integrated in an optimal way, leading to substantial coverage improvements.

This Annex reviews the various concepts of adaptive antennas, including the concept of "spatial channels" provides a theoretical analysis of the potential of the technology and identifies the key characteristics.

2 Antennas and adaptive antenna concepts

2.1 Antenna and coverage

Adequate for simple RF environments where no specific knowledge of the user's location is available, the omnidirectional approach scatters signals, reaching target users with only a tiny fraction of the overall energy radiated into the environment (or, conversely, for emissions from the users towards the BS).

Given this limitation, omnidirectional strategies attempt to overcome propagation challenges by simply boosting the power level of the signals. In settings where numerous users (hence, interferers) are relatively close to each other, this makes a bad situation worse in that the vast majority of the RF signal energy becomes a source of potential interference for other users in the same or adjacent cells, rather than increasing the amount of information conveyed by the link. In uplink applications (user to base station), omnidirectional antennas offer no gain advantage for the signals of served users, limiting the range of the systems. Also, this single element approach has no multipath mitigation capabilities. Therefore omnidirectional strategies directly and adversely impact spectral efficiency, limiting frequency reuse.

Sectorized antenna systems take a traditional cell area and subdivide it into sectors that are covered using multiple directional antennas looking sited at the BS location. Operationally, each sector is treated as a different cell. Sectorized cells can improve channel reuse by confining the interference presented by the BS and its users to the rest of the network, and are widely used for this purpose. As many as six sectors per cell have been used in commercial service.

2.2 Antenna and multipath

In a step towards "smarter" antennas, space diversity antenna systems incorporate two (or more) antenna elements whose physical separation is used to combat the negative effects of multipath.

Diversity offers an improvement in the effective strength of the received signal by using one of two methods:

– *Switched diversity (SWD):* Assuming that at least one antenna will be in a favourable location at a given moment, this system continually switches between antennas (connecting each of the receiving channels to the most favourably located antenna) to select the antenna with the maximum signal energy. While reducing signal fading, SWD does not increase gain since a single antenna is used at any time, and it does not provide interference mitigation.

– *Diversity combining:* This approach coherently combines the signals from each antenna to produce gain. Maximal ratio combining systems combine the outputs of all the antennas to maximize the ratio of combined received signal energy to noise.

In contrast to SWD systems, diversity combining uses all antenna elements at all times for each user, creating an effective antenna pattern that dynamically adjusts to the propagation environment. Diversity combining is not guaranteed to maximize the gain for any particular user, however. As the algorithms that determine the combining strategy attempt to maximize total signal energy, rather than that of a particular user, the effective antenna pattern may in fact provide peak gain to radiators other than the desired user (e.g. co-channel users in other cells). This is especially true in the high interference environments that are typical of a heavily loaded cellular system.

2.3 Antenna systems and interference

More sophisticated antenna systems can mitigate the other major limiting factor in cellular wireless systems, co-channel interference. For transmission purposes, the objective is to concentrate RF power toward each user of a radio channel only when required, therefore limiting the interference to other users in adjacent cells. For reception, the aim is to provide peak gain in the direction of the desired user while simultaneously limiting sensitivity in the direction of other co-channel users. This assumes an antenna system with instant beam steering capabilities: This can be achieved with phased array technology, in particular with digital beam-forming techniques.

In addition, using a larger number of simple antenna elements gives a new dimension to the treatment of diversity.

2.4 Adaptive antenna systems

The advent of powerful and low-cost digital signal processors, general-purpose processors and ASICs, as well as the development of software-based signal-processing techniques, have together made advanced adaptive antenna systems a practical reality for cellular communications systems. Arrays of multiple antennas, combined with digital beam forming techniques and advanced low cost baseband signal processing, open a new and promising area for enhancing wireless communication systems.

Terms commonly used today that embrace various aspects of "smart" antenna system technology include intelligent antennas, phased arrays, spatial processing, digital beam forming, adaptive antenna systems, etc. Adaptive antenna systems are customarily categorized as either "switched beam" or "adaptive array" systems. However, they both share many hardware characteristics and are distinguished primarily by their adaptive intelligence.

At the heart of an adaptive antenna system is an array of antenna elements (typically 4 to 12), the outputs of which are combined to adaptively control signal transmission and reception. Antenna elements can be arranged in linear, circular, planar, or random configurations and are most often installed at the BS site, although they may also be implemented in the mobile terminal. When an adaptive antenna directs its main lobe with enhanced gain to serve a user in a particular direction, the antenna system side lobes and nulls (or directions of minimal gain) are directed in varying directions from the centre of the main lobe. Different switched beam and adaptive smart antenna systems control the lobes and the nulls with varying degrees of accuracy and flexibility.

2.4.1 Switched-beam antenna

Switched-beam antenna systems form multiple fixed beams with heightened sensitivity in particular directions. These antenna systems detect signal strength, choosing from one of several predetermined, fixed beams, based on weighted combinations of antenna outputs with the greatest output power in the remote user's channel, and switching from one beam to another as the mobile moves through the sector. These choices are driven by RF or baseband digital signal processing techniques. Switched beam systems can be thought of as a "micro-sectorization" strategy.

2.4.2 Adaptive array antenna

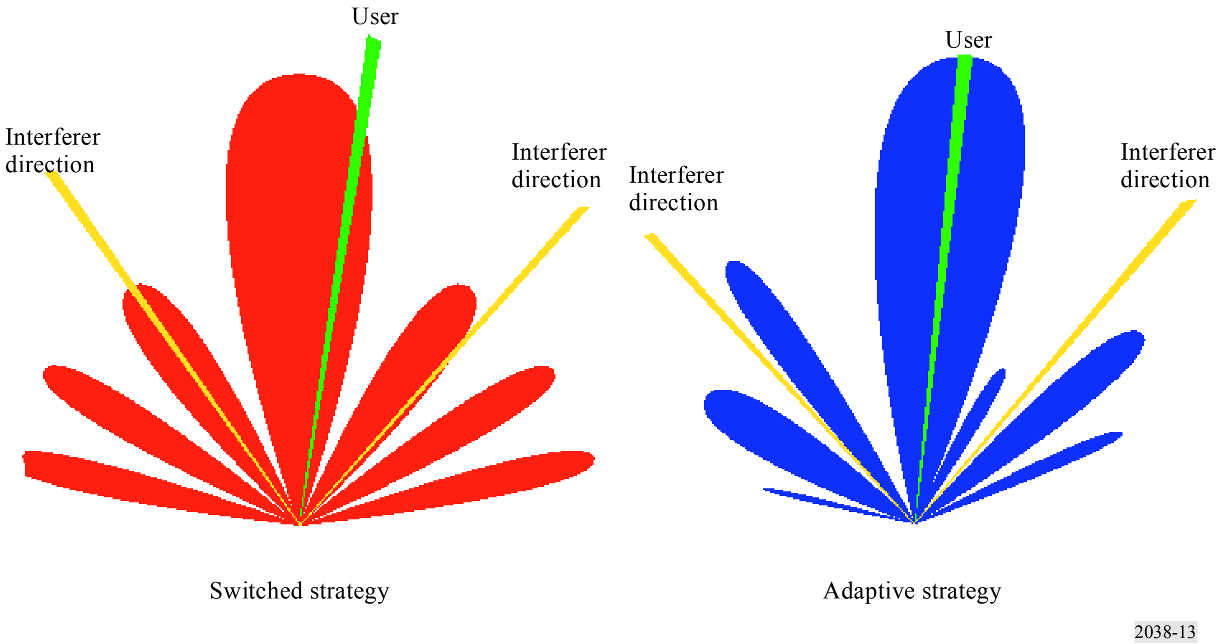
Adaptive antenna technology represents the most advanced approach to date. Using a variety of signal-processing algorithms, an adaptive system effectively aims to identify and track all the relevant signals and interferers present in order to dynamically minimize interference and maximize reception of the signals of interest. In the same manner as a switched-beam system, an adaptive system will attempt to increase gain based on the user's signal as received at the various elements in the array. However, only the adaptive system provides optimal gain while simultaneously mitigating interference. Diversity combining also continuously adapts the antenna pattern in response to the environment. The difference between it and the adaptive antenna method is fundamentally in the richness of the models on which the two systems' processing strategies are based. In a diversity system, the model is simply that there is a single user in the cell on the radio channel of interest. In the adaptive system, the model is extended to include the presence of interferers and, often, temporal history regarding the user's propagation characteristics. With this second model, it is possible to discriminate users from interferers, even at low SINRs, and provide reliable gain and interference mitigation simultaneously.

The adaptive antenna systems approach to communication between a user and the base-station in effect takes advantage of the spatial dimension, adapting to the RF environment – including the full constellation of users and other emitters – as it changes, according to predefined strategies. This approach continuously updates the BS system's radiation and reception patterns, based on changes in both the desired and interfering signals' relative configuration. In particular, the ability to efficiently track users through antenna main lobes and interferers through nulls ensures that the link budget is constantly maximized. By implementing the smart antenna strategies digitally, it is possible for the BS to support a separate, tailored, strategy for each active channel in the system via a single array and set of radio electronics.

The difference between the two approaches – adaptive and switched beam – is illustrated simplistically in Fig. 13, which shows how the adaptive algorithms behave with respect to interferers and the desired signal.

FIGURE 13

Difference between switched beam and adaptive beam



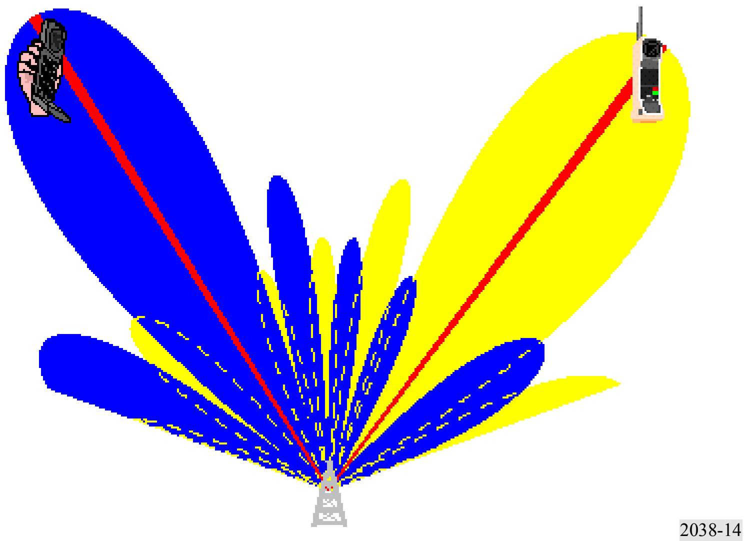
2.4.3 Spatial processing: the fully adaptive approach

Utilizing sophisticated algorithms and powerful processing hardware and microprocessors, "spatial processing" takes the frequency reuse advantage resulting from interference suppression to a new level. In essence, spatial processing dynamically creates a different beam for each user and assigns frequency/channels on an ongoing basis in real time. Spatial processing maximizes the use of multiple antennas to usefully combine signals in space, through methods that transcend the "one user-one beam" methodology.

Depending on the details of the air interface and the service definition, so-called "spatial channels" can be robustly created via spatial processing whereby each conventional temporal channel (e.g. frequency and time slot or code combination) may be reused within the cell, achieving reuse factors less than one. Figure 14 depicts such a situation for two users. Spatial channels, or intra-cell reuse, are used operationally today in commercial cellular systems. While the concept of intra-cell reuse may seem unfamiliar, it is readily supported so long as adequate spatial selectivity is available in the distribution and collection of radio energy from the cell. Depending on the air interface, as little as 10 dB of spatial selectivity or isolation for different locations in the cell may be adequate.

FIGURE 14

A two-user spatial channels strategy, reuse = 0.5



Overall the spectral efficiency as defined as bits/s/Hz/cell may be increased by some 20-40 times or more in practical deployed systems, through use of adaptive antennas (PHS, GSM for example). Globally, over 140 000 adaptive antenna-based systems have already been deployed for various microwave cellular systems, mobile and fixed (FWA).

3 Effects of adaptive antennas on TDD/FDD coexistence

The direct effect of coexistence is due to the fact that the RF energy radiated by transmitters is focused in specific areas of the cell and is not constant over time. This characteristic plays a major role in determining the likelihood of interference in coexistence scenarios, especially in the context of TDD/FDD mixed system deployments. While an absolute worst case may look prohibitive, the statistical factor introduced by the use of adaptive antennas determines the percentage of time that the worst case happens. If this percentage is satisfactorily small, any established coexistence rules may for example be relaxed, thus helping the economics of deployment. The Monte Carlo approach to statistical analysis can be used to study the improvements that can be obtained from adaptive antennas here. In realistic IMT-2000 system simulations, very significant improvements were determined in terms of safe coexistence distances (much reduced), or reduced RF isolation requirement on BSs. In some cases any additional isolation is now sufficiently modest that it should be readily achieved simply by adopting coexistence-friendly site engineering practices, and in other situations coexistence cases otherwise impossible become practical.

4 Some radio interface design considerations

4.1 Uplink

Timing alignment, or timing discrepancy, between signal bits can change because of phase variations and the arrived frequency can be offset from the transmit frequency because of differences in the local oscillators and multipath signals cause fluctuations in the received signal levels. The net result is high bit error rates or an unrecoverable signal, if not corrected. Attempts to correct these anomalies can be made at the antenna by using switched antenna diversity, or optimal ratio combining. However, analysing the signal after it leaves the demodulator and using signal feedback mechanisms to continually correct the signal is where the greatest enhancements can be achieved.

The degrees of freedom to correct these anomalies increase if different samples of the same signal can be analysed from multiple antenna/receiver chains as the same signal impinging on different antenna elements have slightly different characteristics, i.e. phase difference and amplitude. These together with other temporal parameters can be analysed separately and together to optimize the ultimate signal that is forwarded to the switching network. Another figure of merit that can be improved significantly when analysing signals in this manner is carrier-to-interference rejection (C/I), or in other words, the interference from signals that are not of interest can be isolated and rejected to clean-up the desired signal, leading to better cell coverage and call quality.

To achieve these performance improvements, the cost of adding power amplifiers in a multiple antenna BS, has the potential to be lower than the cost of the power amplifier in a conventional single antenna base station. Amplifier costs increase disproportionately with the output power and the cost of multiple lower power amplifiers, which may be integrated at the antenna, cost less today than one higher power amplifier having equivalent effective radiated power.

4.2 Downlink

Using reciprocity, it is possible to achieve the same performance gains in the downlink direction. This is straightforward with TDD but becomes more difficult when used in an FDD environment, although not insurmountable. Broadcast signalling channels also present special challenges, but can be accommodated with innovative solutions.

4.3 Subscriber unit

For certain applications it may be economically feasible to incorporate adaptive antenna processing at the subscriber unit thereby increasing uplink gain and reducing system-wide uplink interference. Significant benefits are also possible through simpler uplink strategies when adaptive antennas are present at the BS, only. Implementation of uplink power control will improve the overall performance of the network since less interference equates to more capacity. Tightly coupling subscriber power control into an overall performance enhancing strategy can significantly enhance network performance.

4.4 Optimal new radio interfaces

Over the last decade, there have been many efforts to increase the efficiency of wireless networks. This effort has largely focused on modulation types, channel coding and access methods. Below are listed some areas in which careful design with consideration to adaptive antennas can more closely approach the global optimum.

4.5 Duplexing methods

As mentioned above, the benefits of TDD over FDD are evident due to the significantly reduced de-correlation between uplink and downlink channels. Also one of the initial drawbacks to TDD technology were the problems associated with the rapid switching between transmit and receive, limiting the e.i.r.p. of a TDD BS. However, with the combined signal from various independent transmitting elements the overall power level that can now effectively be rapidly switched on and off is significantly increased. As for terrestrial wide-area, full high-speed mobility systems the use of paired bands and FDD transmission might be advantageous, while shorter-range, slower mobility systems and TDD transmission in unpaired bands can be advantageous to handle asymmetric traffic.

4.6 Carrier bandwidth

The decision to incorporate adaptive antennas will affect the choice of carrier bandwidth. Adaptive antenna systems can most effectively control the RF environment when the number of significant (as seen at a given BS) co-channel in-cell and out-of-cell users is modest. It may be better to partition the users into slices of spectrum rather than have large number of users sharing the "whole" spectrum. One may also wish to consider the signal correlation variation across the bandwidth, which tends to decrease as the channel bandwidth increases making the adaptive antenna processing more complex.

4.7 Modulation methods

In the design of an air interface that uses smart antennas the modulation methods should be selected to maximize the system throughput when network level interference and the capability of the smart antennas to reject that interference is considered. This design may lead to a variable rate modulation structure that operates efficiently over a range of C/I operating points.

4.8 Signalling, control and broadcast methods

One point that is often discussed is how smart antennas can handle broadcast channels that are common in many existing air interfaces. This is a prime instance where by designing the air interface with adaptive antennas in mind, the broadcast structures can be designed so that they work well within the a multiple-antenna structure. Certainly, the air interface may contain broadcast information, but careful design of that structure must be done or many of the adaptive antenna benefits in terms of base station costs may not be fully realized.

The same applies for "blind" channels such as paging where the information is directed to a single user (as opposed to broadcast information), but where the location of that user is not known.

4.9 Burst structures

The use of adaptive antennas will also impact the design of burst structures. The structures of the burst may include "training" data to help the adaptive antenna processing. A balance between the spectral efficiency benefits and overhead of that data has to be considered to create optimal throughput.

4.10 Frame structures

If adaptive antennas are not considered, the major consideration on frame structure is latency and resource sharing. Considering adaptive antennas in the design, one may wish to consider the rate of which the spatial information gets updated. This can impact the spectral efficiency as mobility of the user increases. In general the more rapid the updates to the spatial information the better the spectral efficiency as mobility increases. This tends to lead to frames with shorter durations, typically below 10 ms.

4.11 Media access control

The use of adaptive antennas can have broad implications on the entire protocol chain. This includes elements of the protocol that are traditionally thought to be design-independent of the RF system. One such area is media access, where it has become increasingly important for air interfaces to support random access packet switched data, as well as guaranteed QoS, for which acceptable latencies are on the orders of tens of ms.

The addition of adaptive antenna technology into the system allows for spatial collision resolution in the MAC, reducing access contention and improving performance over that achievable with traditional non-adaptive antenna systems.

5 Conclusions

Integrating adaptive antenna systems into the design of future IMT-2000 systems and systems beyond IMT-2000, will significantly improve the spectral efficiency of these new radio systems. Spectral efficiency gains from adaptive antenna systems can be used not only to reduce the number of BSs (cells) needed to deploy an IMT-2000 network, but also to obtain significantly increased data rates within a limited amount of spectrum.

6 Bibliography

COMPTON [1988] *Adaptive antennas: Concepts and performance*. Prentice Hall.

ETSI [September 2001] ETSI TR 125 942 V4.0.0 (2001-09) Technical Report. Universal Mobile Telecommunications Systems (UMTS); RF System Scenarios. European Telecommunications Standards Institute.

FARINA [1991] *Antenna based signal processing techniques for radar systems*. Artech House, Boston, United States of America.

HAYKIN, S. (Ed.) [1991] *Advances in spectrum analysis and array processing*. Vol. II. Prentice Hall.

HAYKIN, S. [1991] *Advances in array processing*. Prentice Hall.

HAYKIN, S. [1991 and 1996] *Adaptive filter theory*. Prentice Hall.

HUDSON [1981] *Adaptive arrays principles*. Peter Peregrinus.

KIM, K. *et al*., [2000] *Handbook of CDMA System Design, Engineering and Optimization*. Prentice Hall.

LAIHO J. *et al*., [2002] *Radio Network Planning and Optimisation for UMTS*. Wiley.

LIBERTI, J. C. and RAPPAPORT, T. S. [1999] Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications. *Prentice Hall Communications Engineering and Emerging Technologies*.

LITVA, J. [1996] *Digital Beam forming in Wireless Communications*. Artech House.

MOZINGO and MILLER [1980] *Introduction to adaptive arrays*. Wiley Interscience.

RAPPAPORT, T. S. Smart Antennas: adaptive arrays, algorithms and wireless position location: Selected Readings IEEE.

RAPPAPORT, T. S. [1996] *Wireless communications*. Wiley.

WIDROW, B. and STEARUS, S. D. [1985] *Adaptive signal processing*. Prentice Hall.

Annex 5

Multiple-input multiple-output techniques

1 Introduction

MIMO techniques utilize multi-element antennas at both ends of the link with signal processing algorithms which make positive use of the multipath propagation channels found in terrestrial mobile communications. For propagation within a typical urban environment this has been shown to increase the capacity of the link over that available through conventional beam forming techniques.

With the wide range of possible antenna configurations that can be considered it is usual in the literature and standardization bodies to categorize the various multi antenna schemes by two indices, the number of transmit antennas used and the number of receive antennas used. Thus an antenna combination for a link is described as type [$M_{Transmit}$, $N_{Receive}$] or more simply a type [M,N] scheme.

Current 2G systems have terminals with one antenna and therefore typically use:

[1,2] – e.g. receive diversity using maximal ratio combining in the uplink

[2,1] – e.g. transmit diversity in the downlink

In addition, for some IMT-2000 3G systems, the following are being defined:

[1,4] – additional receive diversity in the uplink

[4,1] – transmit diversity in the downlink based on adaptive feedback from the mobile terminal

The classical active beam forming systems are generally of the form:

[*N*,1] – downlink

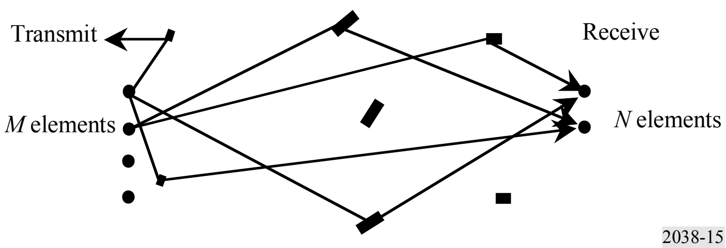
[1,*N*] – uplink.

2 MIMO antenna schemes

In a typical propagation environment in the bands used for mobile systems there are of course many paths for the RF energy to pass along between the transmitter and receiver antenna elements. In a MIMO link there will therefore be many separate paths between the different pairings of antenna elements and Fig. 15 shows a diagram of some of these paths which may exist between the transmit and receive antenna element arrays.

FIGURE 15

Illustration of multipath transmission between transmit and receive antenna arrays

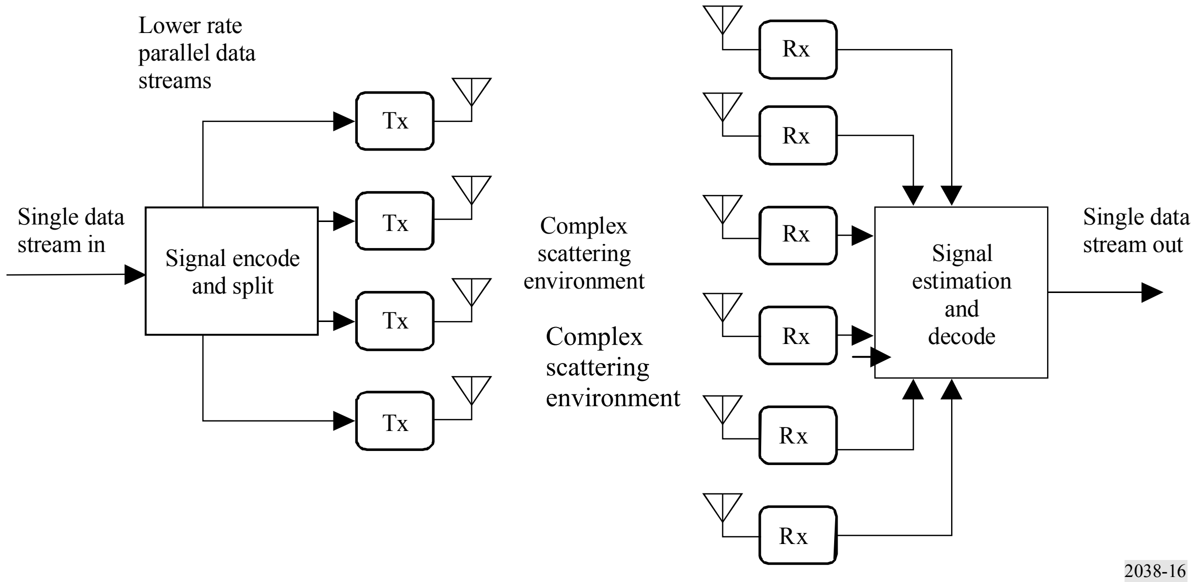


For adaptive beam forming techniques the signals to and from the antenna elements are adjusted in phase and amplitude to form a beam to select the best single path between the transmit array and the receive array, while minimizing the antenna gain in the direction of the unwanted interfering signals. With multiple antenna elements at both ends of the link it is also possible to transmit different data streams on some, or all, of the transmit elements and separate the different transmissions by several receivers which each maximize one signal while minimizing the others. This process can give significant capacity advantages over the simple beam forming approach for complex scattering environments and the bounds to the capacity achievable with this type of arrangement in typical mobile propagation environment are derived in [Foschini and Gans, 1998].

The signal handling in these MIMO systems will generally be as shown in Fig. 16. The data streams may be coded independently, which makes for simpler decoding at the receiver [Foschini, 1996], or the full potential of the method can be achieved by including all of the data streams in a two dimensional coding scheme [Foschini, *et al*., 1999]. The ultimate capacity is achieved when a feedback from the receiver is used to adjust the power split, coding and modulation at the transmitter.

FIGURE 16

Schematic of MIMO transceiver functions



These MIMO techniques are currently the subject of a great deal of research activity and the following sections are included as general guides to the gains that can be achieved, and the type of propagation environments where these techniques have the greatest potential to increase the capacity of a network. They are not intended as definitive results and more detailed analyses are available in the references given.

3 Spectral efficiency of MIMO systems in isolated links

The capacity (bit/s/Hz) of an [M,N] MIMO link is given by:

$$C = B \log_2 (\mathbf{I}_N + r\mathbf{H}\mathrm{H}^H/M)$$

where:

B: bandwidth,

$\mathbf{I}_N$: N by N identity matrix,

r: average S/N

$\mathbf{H}$: M by N matrix whose (m,n)-th element is the complex amplitude between the m-th transmitter and n-th receiver.

We have assumed perfect channel state information at the receiver (in other words, the entries of the matrix H are known exactly. In a rich scattering environment, the entries in H are independent and identically distributed complex Gaussian random variables. Under this condition and the additional condition that r is much larger than one, as the number of antennas gets large and $M = N$, the capacity approaches:

$$C = BM \log_2(r/e)$$

On the other hand, using beam forming the array becomes more directive as M increases, resulting in a linear increase in S/N. Hence the capacity is given by:

$$C = B\log_2\left[1 + Mr\sum_{n=1}^{N}|h_n|^2\right]$$

where h_n is the complex amplitude at the n-th receiver. Hence in a sufficiently complex scattering environment the capacity is directly proportional to the number of antennas at each end of the link for the [N,N] MIMO system operating, but only proportional to the log of the number of antennas for phased array beam forming techniques.

(See footnote for a grossly simplified explanation of the basis of this effect.)[4]

---

[4] The spectral advantages of the MIMO approach are more easily shown by a further grossly over-simplified model:

The capacity limit of a single channel is taken as the classical Shannon expression:

$$C = B\log_2(1 + SNR_{Omni})$$

where:

SNR_{Omni}: S/N at the receiver for a single channel between two conventional omnidirectional antennas.

If the transmit power is spread equally between M transmit elements and M receive elements are used with conventional phased array beam forming techniques at both ends of the link, then the capacity would approach:

$$C_{Beams} = B\log_2(1 + M^2 SNR_{Omni}) \Rightarrow B\{2\log_2(M \cdot SNR_{Omni}\}$$

for reasonable S/N.

However, if under similar ideal conditions the power is split over M separate independent channels with the same path loss, then the capacity can approach M times that of each [1,M] link:

$$C_{MIMO} = MB\log_2(1 + SNR_{Omni}) \Rightarrow MB\log_2(SNR_{Omni})$$

for reasonable S/N.

This is of course, a huge over-simplification to illustrate the effect and it assumes that the separated signals from the scattered channels are independent and unaffected by each other.

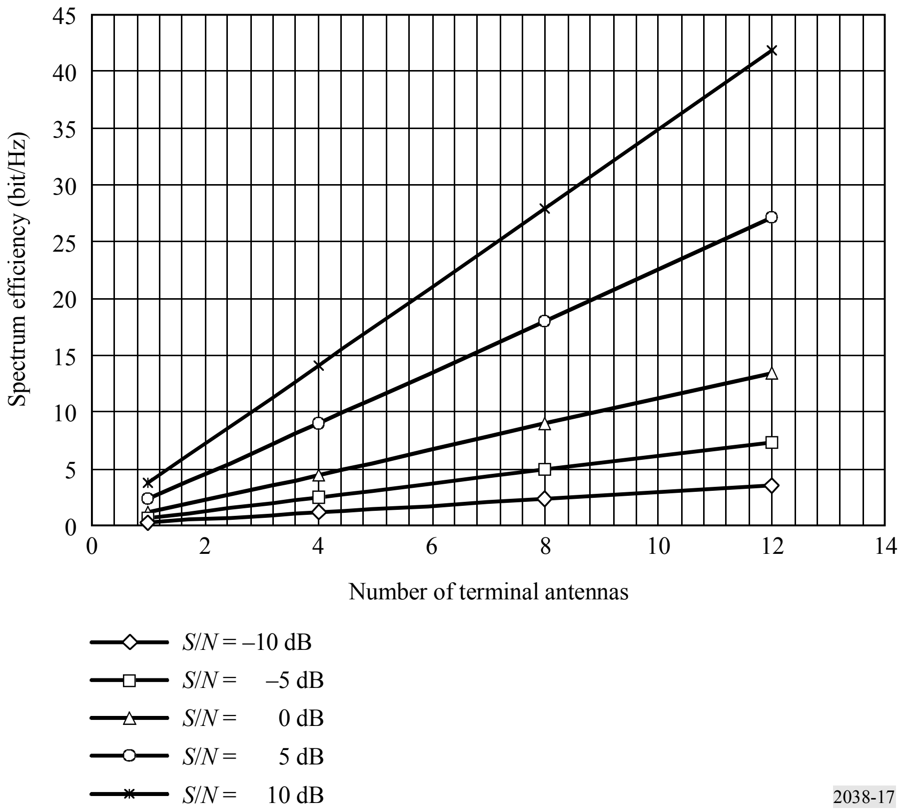
4 Spectral efficiency gains in an interference limited cellular system

Figure 17 shows the spectrum efficiency that can be achieved based on Shannon capacity [Telatar, 1999]. The number of antennas at the base station and each terminal are equal.

Similar related results for an intereference limited mobile network environment can be found in [Catreux *et al.*, 2001].

FIGURE 17

Simulation results for potential spectral efficiency gain using MIMO techniques



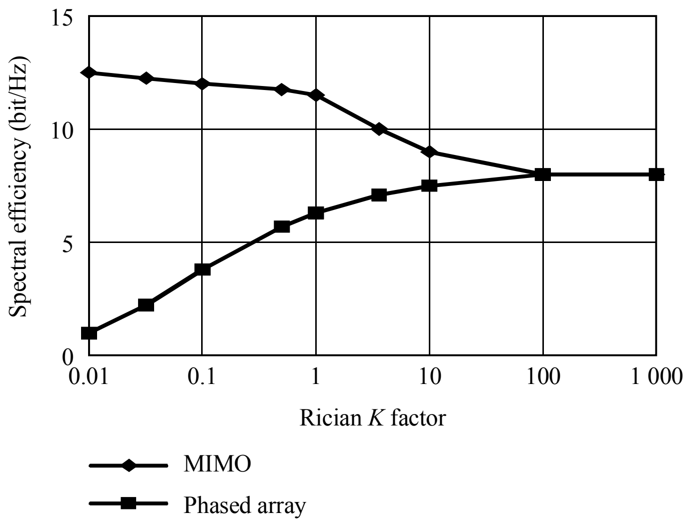
5 Variation in spectrum efficiency gains using MIMO techniques for different scattering environments

In order to achieve the potential gains of MIMO systems it is necessary to have a complex scattering environment and so it is essential to quantify the gains for different scattering environments. A simple demonstration of the gains possible is shown in Fig. 18 where the results are shown of some simulations of the spectral efficiency achievable for a single link of a mobile system operating in a range of scattering environments. The simulated environments are modelled as Rician channels with a range of K factors to cover the cases from fairly clear line of sight ($K > 10$) down to rich scattering environments ($K < 0.1$).

These results were obtained under idealized conditions of perfect channel state information at both the transmitter and receiver at 10 dB S/N.

FIGURE 18

Spectral efficiency of a link using an antenna* scheme with MIMO and phased array beam forming for different scattering channel conditions



* [Raleigh and Cioffi, 1998; Foschini *et al.*, 1999]

2038-18

6 Variation of MIMO link capacity within a cellular network

In order to be able to plan the networks which employ MIMO techniques it is necessary to have models for the distribution of capacity for the links to the individual terminals to enable engineering estimates of system capacity, total throughput, and capacity outage probability.

This topic is the focus for a substantial amount of research work, e.g. in [Smith and Shati, 2001 and 2002b] which shows that the distribution of MIMO capacity for classical Rayleigh channels tends towards a Gaussian distribution as the number of transmit and receive antennas increases. Simulations in these papers show that the Gaussian approximation to the capacity is reasonably accurate even for systems with three or more antennas at each end of the link and exact expressions for the mean and variance of the Gaussian capacity for any given numbers of transmit and receive antennas are derived in [Smith and Shati, 2001].

It is shown [Shati and Smith, 2002] that the distribution of MIMO capacity for Rician fading conditions can also be expressed as a Gaussian random variable with a similar dependence on the number of transmit and receive antennas.

These topics are further developed in [Smith and Shati, 2002a] which examines the MIMO capacity under the influence of various power allocation algorithms at the transmitter and it is shown that the Gaussian approximation still holds good for the different power allocation strategies studied. These included equal power for each antenna, the classical “water filling” technique with perfect knowledge of the channel conditions and a new strategy giving better capacity than the first two for less than ideal estimates of the channel conditions. All of these power allocation methods are shown to give link capacity that can still be approximated by a single Gaussian random variable for both Rayleigh and Rician fading channels.

7 Implementation issues

The introduction of MIMO techniques into wireless communication systems introduces a number of implementation challenges. At the BS the greatest impact is likely to be in the increased RF and cabling requirements due to the increased number of transmit/receive antenna elements until more integrated antenna, receiver, transmitter structures are developed. The greatest challenges, however, lie within the terminal where the size, power and cost constraints must be overcome.

Research initiatives must address the viability of terminals employing MIMO or diversity techniques, with particular emphasis being placed on maximizing the performance of the terminal antenna system in realistic macrocellular deployment scenarios and within the restricted form factors of future terminals such as laptops, PDAs and handsets. Key challenges will include the design of antennas with low correlation within such a confined space and good performance in realistic indoor and outdoor nomadic and mobile propagation environments and the difficulties of minimizing interactions between different functions within the terminal (EMC).

Diverse terminal antennas have already been investigated for various form factors, including picocell base units [Smith *et al.*, 1997 and 1999] and fixed wireless access terminals [Kitchener and Smith, 1998]. In the first two cases the antennas are purely internal to a standard housing, while in the third the terminal form factor was adapted to enable best antenna performance. Both of these approaches are options for future implementations of diverse/MIMO terminal antennas. In the context of a free-standing terminal, e.g. a laptop, an important aspect is whether the terminal antenna elements are flat on a table (e.g. in the base unit) or vertically oriented (in the display), and making the design robust against several deployments is part of the challenge [Smith et al., 1997 and 1999].

A key element in the development of MIMO antenna systems is optimization of the design to work in the MIMO propagation channels of the target deployment scenarios. The MIMO propagation channel is a current study area in both 3GPP and 3GPP2 standards organizations, with a series of submissions from various companies. Also, the COST 259 [5] research project team has used various propagation measurement results to arrive at an outdoors-to-outdoors channel model, and COST 273 [6] aims to extend this work, with subgroup activities including MIMO systems, handset antennas, channel measurements and channel modelling.

It is expected that ongoing research by the industry will yield results in the following areas:

– MIMO propagation channel models for macro-, micro-, and picocellular deployment scenarios which are sufficiently detailed to allow theoretical evaluation of MIMO/diversity terminal antenna configurations.

– Generic antenna system designs for laptops, PDAs, stand-alone units, and handsets.

---

[5] http://www.lx.it.pt/cost259/

[6] http://www.lx.it.pt/cost273/

– Understanding of the interaction between multi-element antenna design and the complex localized propagation environment.

– Designing efficient signal processing algorithms and associated coding schemes which achieve much of the capacity gains allowed by the theory, but which can be implemented with much lower levels of processing power [Ariyavisitakul, 2001].

There are clearly substantial implementation issues to be solved before MIMO techniques can used to increase the capacity of mobile communication networks. However, it is useful to consider the complexity of achieving high data rates within the same channel bandwidth using conventional high order modulation schemes. For a MIMO link using [4,4] antennas and 4 state modulation the equivalent single channel link would need to use a 256-point modulation constellation for the same symbol rate. With a not unreasonable 16-point modulation on the [4,4] MIMO system it would be necessary to implement a somewhat unrealistic 4 096-point modulation on the single high-speed channel.

8 Conclusions

MIMO antenna schemes have the potential to greatly increase the capacity of mobile systems. In particular, they have great potential to supply both high speed links and increased capacity in the densest urban environments where the demand for capacity is at its highest. As such they are a complimentary technique to the conventional beam forming and diversity techniques used with similar antenna arrays. Therefore it is expected that advanced networks will adapt the processing of the signals to and from the antennas to operate in all modes simultaneously at a single base station site. This will optimize both the total cell capacity and the individual data rates to the terminals with different antenna arrangements as they move between different radio environments.

9 References

ARIYAVISITAKUL, S. L. [August 2001] Turbo Space-Time Processing to improve Wireless Channel Capacity. *IEEE Trans. Comm.*, Vol. 49, **8**, p. 1347.

CATREUX, S., DRIESSEN, P. F. and GREENSTEIN, L. J. [August 2001] Attainable Throughput of an Interference-Limited Multiple-Input Multiple-Output (MIMO) Cellular System. *IEEE Trans. Comm.*, Vol. 49, **8**, p. 1307-1311.

FOSCHINI, G. J. [Autumn 1996] Layered Space-Time Architecture for Wireless Communications in a Fading Environment When Using Multi-Element Antennas. *Bell Labs Tech. J.* p. 41-59.

FOSCHINI, G. J. and GANS, M. J. [1998] On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas. *Wireless Personal Comm.*, **6**, p. 315-335.

FOSCHINI, G. J., GOLDEN, G. D., VALENZUELA, R. A. and WOLNIANSKY, P. W. [November 1999] Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE J. Selected Areas in Comm.*, Vol. 17, p. 1841-1851.

KITCHENER, D. and SMITH, M. S. [24 February 1998] Low cost antennas for mobile communications. Proc. IEE Colloquium 1998/206 on «Low cost antenna technology».

RALEIGH, G. G. and CIOFFI, J. M. [March 1998] Spatio-Temporal Coding for Wireless Communications. *IEEE Trans. Comm*., Vol. 46, **3**, p. 357-366.

SHAFI, M. and SMITH, P. J. [January 2002] MIMO Capacity in Rician Fading Channels and its Relationship to the K-Factor. *IEEE Trans. Lett. Comm*.

SMITH, M. S., BUSH, A. K., GWYNN, P. G. and AMOS, S. V. [September 1999] Microcell and Picocell base station internal antennas. Proc. WCNC 99, p. 708, New Orleans, United States of America.

SMITH, M. S., KITCHENER, D. K., DALLEY, J. E. J. and THOMAS, R. R. [April 1997] Low cost diversity antennas for low power wireless base stations. Proc. ICAP 97, p. 1.445, Edinburgh, United Kingdom.

SMITH, P. J. and SHAFI, M. [August 2001] On a Gaussian Approximation to the Capacity of Wireless MIMO Systems. *IEEE Trans. Comm*.

SMITH, P. J. and SHAFI, M. [2002a] Water Filling Methods for MIMO Systems. Proc. of 3rd AusCTW, (4-5 February) Canberra, Australia.

SMITH, P. J. and SHAFI, M. [2002b] On a Gaussian Approximation to the Capacity of Wireless MIMO Systems. Conference Proc., ICC 2002, (April 2002). New York, United States of America.

TELATAR, E. I. [November 1999] Capacity of Multi-Antenna Gaussian Channels. *Euro Trans. Telecom*. Vol. 10, p. 585-595.

10 Bibliography

NAGUIB, A. F., TAROHK, V., SESHADRI, N. and CALDERBANK, A. R. [October 1998] A Space-Time Coding Modem for High-Data-Rate Wireless Communications. *IEEE JSA*C, Vol. 16, **8**, p. 1459-1477.

DRIESSEN, P. F. and FOSCHINI, G. F. [February 1999] On the capacity Formula for Multiple-Input Multiple-Output Wireless Channels: A Geometric Interpretation. *IEEE Trans. Comm.*, Vol. 47, **2**, p. 173-176.

JEFFRIES, A. and SMITH, M. [March 2002] Multiple Antenna Terminals Incorporating MIMO and Diversity Technologies, WWRF5 Research Submission, Tempe, Arizona, United States of America.

Annex 6

Software defined radios

1 Introduction

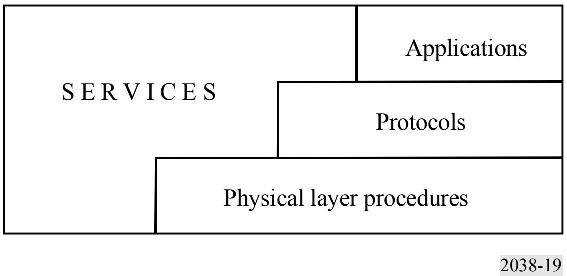
This Annex gives more detailed information on the SDR system architecture concept. It is a complement to the text of § 3 (Overview of major new technologies), which contains a definition of SDR and describes the requirements for SDR.

This Annex describes functionalities and a system architecture to fulfil the requirements stated in § 3.

SDR concerns therefore basically all communication layers (from the physical layer to the application layer) of the radio interface (see Fig. 19) and impacts both the mobile terminal and the network side.

FIGURE 19

Communication layers subject to SDR



As key objectives, SDR shall provide means for:

– adaptation of the radio interface to varying deployment environments/radio interface standards;

– provision of possibly new applications and services;

– software updates;

– enabling full exploitation of flexible heterogeneous radio networks services.

The logical SDR-architecture has to support the following functions:

– management of terminal, user and service profiles in the network entities and the terminal;

– efficient download control and reconfiguration management for terminals and network entities;

– negotiation and adaptation functionalities for services and RATs (e.g. vertical handover);

– assurance of standard compliance.

These functions are logical functions, i.e. they can be implemented in different places in the network. Moreover they can be distributed within the network and between network and terminal.

2 Terminal system architecture

In the past user terminals for mobile communication were typically interacting with intelligent network server located in distributed network architecture. Communication services were mostly deployed at network side which meant that most of the work, for example establishing a connection has be to done by network server. Meantime capabilities of mobile terminals increased more and more and nowadays terminal capabilities are sophisticated enough to offer an acceptable counterpart for deployment of high-speed data exchange and application services at any time in any place. But the requirements necessary for the use of different services, based on different terminal resources, are flexible interfaces within a terminal architecture that provides terminal reconfiguration. In particular network and terminal APIs have to provide mechanisms to add, replace and remove software modules or components for application and protocol layers.

2.1 Terminal components and API architecture for reconfiguration

2.1.1 Definition and requirements of API

An API can be defined in the following terms:

– a set of definition of ways in which one piece of software communicates with another;

– a method of achieving abstraction, usually (but not necessarily) between lower level and higher level software;

– often consists of sets of functions, procedures, variables, structures, etc.

An API can then be seen as an abstract interface definition, which is a description of the relationships among related software and/or hardware modules, such as bidirectional flow of data and control information. It describes the relationship of modules, not the implementation of those relationships. The interfaces should be independent of the implementation.

2.1.2 Terminal components

Based on the architecture shown in Fig. 20 there are four main APIs and layers identified necessary for an efficient and flexible reconfiguration process:

APIs:

– *Terminal API:* Interface to the "outside world", including any interaction to services of provider or 3rd parties, connection to operator network, local area connection and interaction with users.

– *Core API:* Interface which supports and enables the reconfiguration process between the radio configuration layer and the core software layer.

– *Hardware abstraction layer API:* Interface which allows a computer operating system (OS) to interact with a hardware device at a general or abstract level.

– *(RT)OS API:* Interface, controlling and monitoring any kind of running applications as well as device status and terminal capabilities.

Layers:

– Radio configuration layer.

– Core software layer.

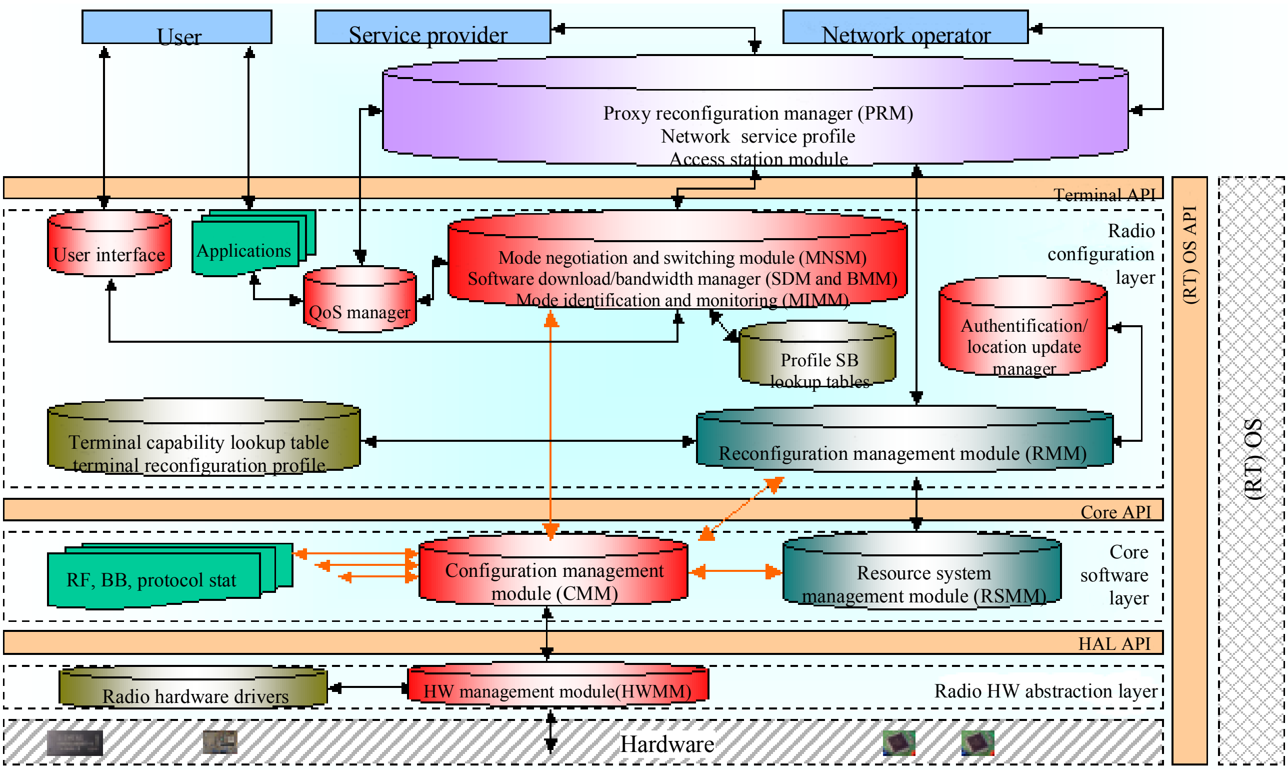
– Radio hardware abstraction layer.

– (RT)OS layer.

The radio configuration layer includes all software components, which are responsible for any interaction with the "outside world". So mode negotiation and switching module (MNSM), software download module (SDM), bandwidth management module (BMM) and mode identification and monitoring module (MIMM) are used to communicate to network server of service provider or network operator, also to the lower core software layer supported by lookup tables, profiles and reconfiguration management module (RMM). So the terminal API enables interfaces to the outside actors like user, network, provider or local area connection (e.g. Bluetooth, RLAN, etc.). A proxy configuration manager (PRM) should reduce bandwidth, terminal resources and connection time by supporting autonomous services and negotiations in the manner of WebService or information broker. A personalized user interface helps the user to interact with complete terminal functionality, he is allowed to access. This includes handling of user applications (e.g. PIM, Office, …) as well as editing of preferences and configurations for network communication and services.

FIGURE 20

API and layer architecture of reconfigurable terminals



The SDR core framework layer (SDR-CF) contains the CMM responsible for instantiating, monitoring and controlling of core radio software components. The core API supports and enables the reconfiguration process between radio configuration software layer and the core software layer.

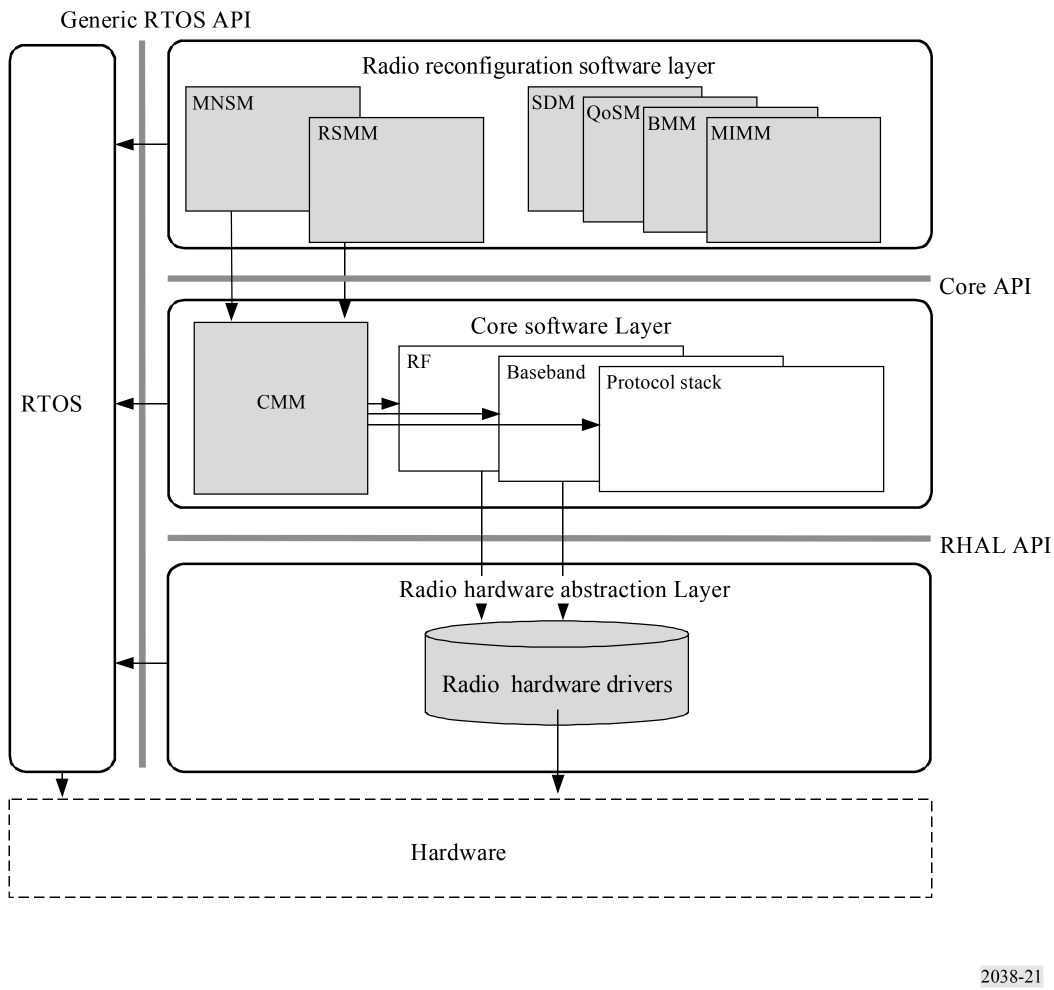
The OS layer controls current/future terminal configurations, supports domain managers and terminal agents, provides access between SDR-CF and lower layers as well as data and resource management (lookup tables, capabilities, profiles, ...).

For simple visualization the following two Figures reflect the above-mentioned API architecture in a compressed form. Only the terminal API was unattended because this API contains a lot of standardized interfaces, which do not leave much scope for new development elements. Both Figures, Fig. 21 using a radio hardware abstraction layer whereas Fig. 22 does not, are taken as a basis to explain the process and interactions between APIs, interfaces and participated components (software and hardware) in an abstract way.

Starting at the bottom of the Figure and working up, the first item is the terminal hardware itself. Examples include processing hardware such as DSPs and microcontrollers, function specific hardware such as ASICs, mixers and oscillators, reconfigurable hardware and input/output devices. Everything above the hardware is some form of software. There are four software packages; the three shown horizontally are *software layers*, whilst the fourth is shown vertically and may be thought of as a *software slice*. This slice is the operating system. In an embedded real-time system, it will be a real-time operating system (RTOS). The RTOS provides standard services (e.g. task scheduling, memory management, interrupt handling, …) to all of the software layers. Implementation of these services varies from one RTOS to another whether it be commercially sourced or developed in-house. To preserve a common interface to these standard services, for any given RTOS. This API is shown on Fig. 21 as the vertical bar sandwiched between the software layers and the RTOS.

FIGURE 21

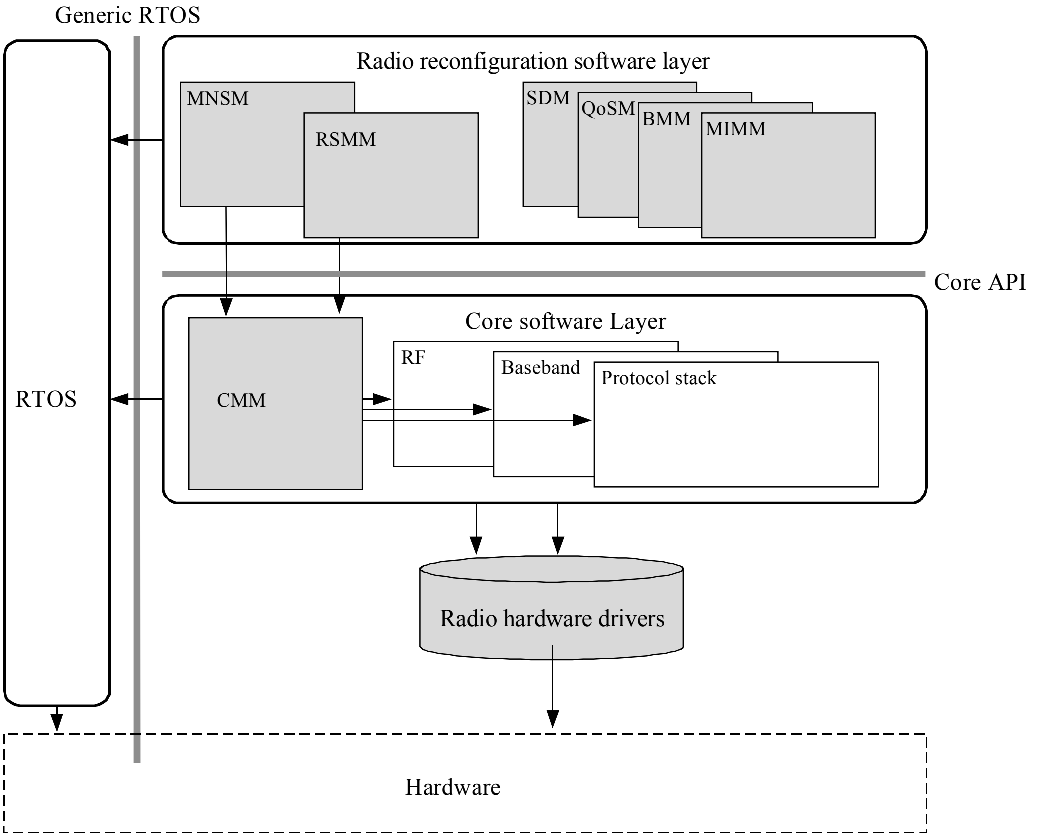
Low-level architecture (including RHAL)



An RTOS may also support other services; in particular it may support device I/O using *device drivers* for standard hardware devices. In a reconfigurable terminal many of the hardware devices will be associated with radio functionality; e.g. mixers and oscillators in the RF domain and ASIC accelerators in the baseband domain. In addition reconfigurable-hardware devices will enable flexible radio hardware accelerators. Since the number and type of these devices will vary from one class of terminal to another it is unlikely that generic operating system support for these will be possible. Direct radio hardware access will be supported through a dedicated *radio hardware abstraction layer*, shown in Fig. 21 as the first software layer above the hardware.

FIGURE 22

Alternative low-level architecture



This layer contains a library of *radio hardware drivers*, each of which publishes a software interface to the services of the radio hardware for which it is supplied. Software applications in higher layers access these services through the *radio hardware abstraction layer API* (RHAL API). This mechanism for direct access to radio hardware, without an intermediate operating system, is also essential for optimum speed and power performance.

It can be seen that the RHAL API will be dependent on the class of terminal because of variations in supported hardware. One might therefore argue that it is not an API; that it is not general at the conceptual level of a reconfigurable terminal. Taken to the extreme we obtain an alternative architecture shown in Fig. 22. This architecture does not contain a radio hardware abstraction layer and API; the radio hardware drivers are accessed directly from a library contained within the terminal. The former architecture is preferred however, because of the possibility that some hardware components may be standardized at the level of a generic reconfigurable terminal and also because the actual software mechanism for accessing hardware drivers may also need to be standardized.

Continuing to move up the architecture of Fig. 22 we come to the *core software layer*. This layer contains the core radio software components for the current radio configuration. Typically this would include RF, baseband and protocol stack components. In addition the CMM is responsible for instantiating, monitoring and controlling the core radio software. An API, the *core API*, is provided so that next layer up, the *radio reconfiguration software layer*, may access the services provided by the core layer.

The radio reconfiguration layer is concerned with all aspects of reconfiguration and contains the following components:

– reconfiguration management module (RMM);
– software download module (SDM);
– QoS manager;
– mode negotiation and switching module (MNSM);
– mode identification and monitoring module (MIMM);
– bandwidth management (BMM).

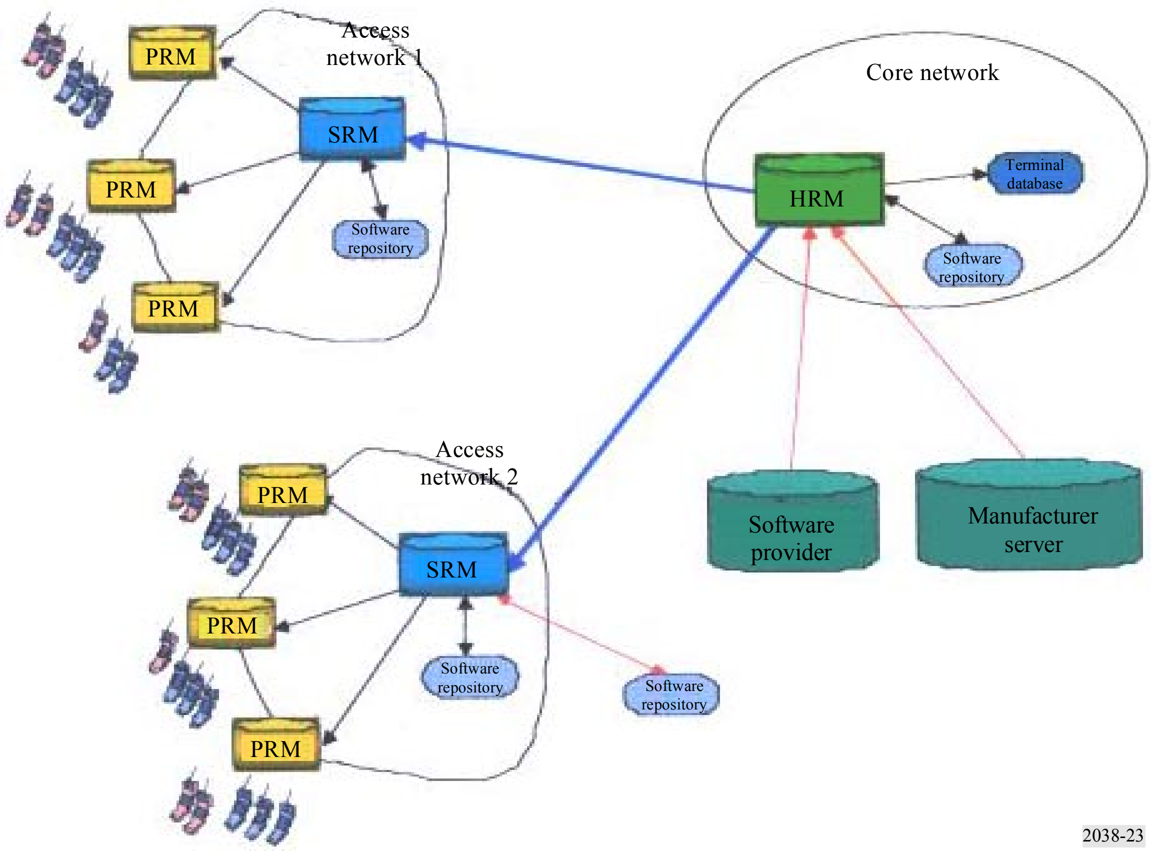
Both the MNSM and the RMM would invoke services of the CMM through the core API; the former to modify the core software configuration, the latter to obtain resource usage figures. In an open-layered architecture it would also be possible for one layer to access the services of any layer below if it is necessary. For example it might be desirable for the RMM to obtain resource usage figures directly using the RHAL API.

3 Network architecture supporting reconfigurable terminals

An hierarchical model for the network architecture supporting reconfigurable terminals is based on a network-centric approach (Fig. 23) involving the association of home reconfiguration manager (HRM), serving reconfiguration manager (SRM) and PRM. This architecture extends the classical cellular radio access networks.

FIGURE 23

Reconfiguration managers enlarging classical cellular networks



In the case the necessary mode software for another radio access technology is not already stored on the terminal, the required software modules must be delivered by the PRM. Because of the large number of different radio access schemes and different terminals, the PRM does not store every possible software module. Therefore this reconfiguration architecture includes a hierarchically reconfiguration management architecture. To speed up the reconfiguration process, every PRM caches the most frequent used modules within his access network.

One difficulty in the speed up reconfiguration process occurs, if the necessary software modules are neither available at the current PRM nor at the new PRM. For this the PRM contacts his SRM and informs him about the appropriate software. Now the SRM is responsible for the provision of the software and forwards it to the requesting PRM.

Interactions between terminal and network are crucial as the available bandwidth on the wireless link is a limited resource that should be used for services rather than negotiations. Furthermore, resources on the terminal itself are usually also limited. In order to relieve the terminal from the burden of frequent interactions with network entities, information from the network could be generally obtained via the PRM, which is located in the radio access network. It serves as a proxy instance for negotiations with other network entities, in particular the SRM and the HRM.

The core entities in the reconfiguration process are the PRMs located in every radio access network. The PRMs are the contact points for every terminal attached to the radio access network concerning reconfiguration.

The PRM is in charge of negotiating and obtaining all kind of information from the network in order to minimize interactions on the wireless link and also to avoid wasting terminal resources in these negotiation and information obtainment processes. The PRM acts on behalf of managed reconfigurable terminals; here is a list of some functionality of PRM:

– Information broker for the terminal.

– Autonomous service discovery and mode negotiation (required during mode negotiation).

– Download management (inter-working with other RRM functions).

– Terminal classmark awareness.

– Records terminal classmark and capability information.

– Caches measurements of terminals operating in specific mode (required during mode monitoring).

– Caches negotiations of terminals requesting same bearer services (required during mode negotiation).

In case of a terminal initiated software download, the terminal signalizes the need of reconfiguration to the current PRM in the radio access network and the PRM is afterwards responsible for the delivery of the appropriate software module. For the mode switching support, the PRM performs additionally different measurements and informs the terminal and the neighbouring PRMs as well.

With reference to the software download the PRM stores necessary software modules in its local repository. But the overall capacity of the storage space in the PRM is not so high. The intention is to have a fast access on the most frequently used modules. For less frequent requests of required software there exists an interface between the PRM and an intermediate server database, the SRM. Thus the request is forwarded and processed by this SRM.

Another possible reconfiguration supporting functionality is the inter-PRM-interface. Therefore neighbouring PRMs are connected to each other and enabled to exchange some information about the current status of the accompanying radio access or about an ongoing mode change of a terminal.

The reconfiguration of a terminal must not only be initiated by the terminal, but can also be triggered by an external entity. In the case of a new hardware driver version it is inefficient to inform each terminal separately.

The use of multicast would help to optimize the content delivery. For the mass upgrade of terminals the multicast mechanism could be used to avoid network overload. The idea is that every terminal manufacturer, application developer, etc. has its own server. If a terminal is now registered with its profile at the PRM, the PRM knows for which components of the terminal a mass upgrade could happen. After the registration of the terminal the PRM joins a multicast session for every possible component. If there is a mass upgrade going on, which a certain server initiated, the software packets are only delivered to these PRMs which joined the multicast group.

3.1 SRM and HRM role and location

The main idea is to have a hierarchical distributed architecture, which minimizes the network load and speeds up the software download.

The HRM is located in the home network of the terminal and is informed by providers about new software upgrades. In this case the HRM notifies the availability of new software to the SRM in the radio access networks and forwards the software to them in case of a mass upgrade. If a request for a software download arrives at the HRM, it is also responsible for the authorization of the terminal in case of a request to download a licensed software. Another point considers accounting of software download. For this the HRM uses a charging repository, which is updated if appropriate software is downloaded.

The SRMs are located between the PRMs and the HRM. One SRM is connected to several PRMs and is responsible for the provision of reconfiguration software to the attached PRMs. Thus the SRM manages on the one hand a large database of software modules for the reconfiguration process and is on the other hand able to get non-available software e.g. from external servers or HRM.

As mentioned above, the SRMs are informed from the HRM about new software and distribute it to the attached PRMs. Because of the possible heterogeneity of the radio access technologies in an IP-based mobile network, the size of the individual radio access networks may vary. If the software must be transported to a large number of PRMs, the serving reconfiguration managers have the advantage to minimize the load. In addition not every available software is needed in every access network or on every access point. Therefore the PRMs are trying to reduce the delay and needed memory and are caching only a small amount of files and the SRMs on the other hand have access to large software repositories and store much more files.

Furthermore the SRM could be involved in the control of the mobility, allocation of resources and security of moving terminals. This includes procedures required for vertical handovers; location update and inter-working between different radio access technologies in order to provide the desired QoS.

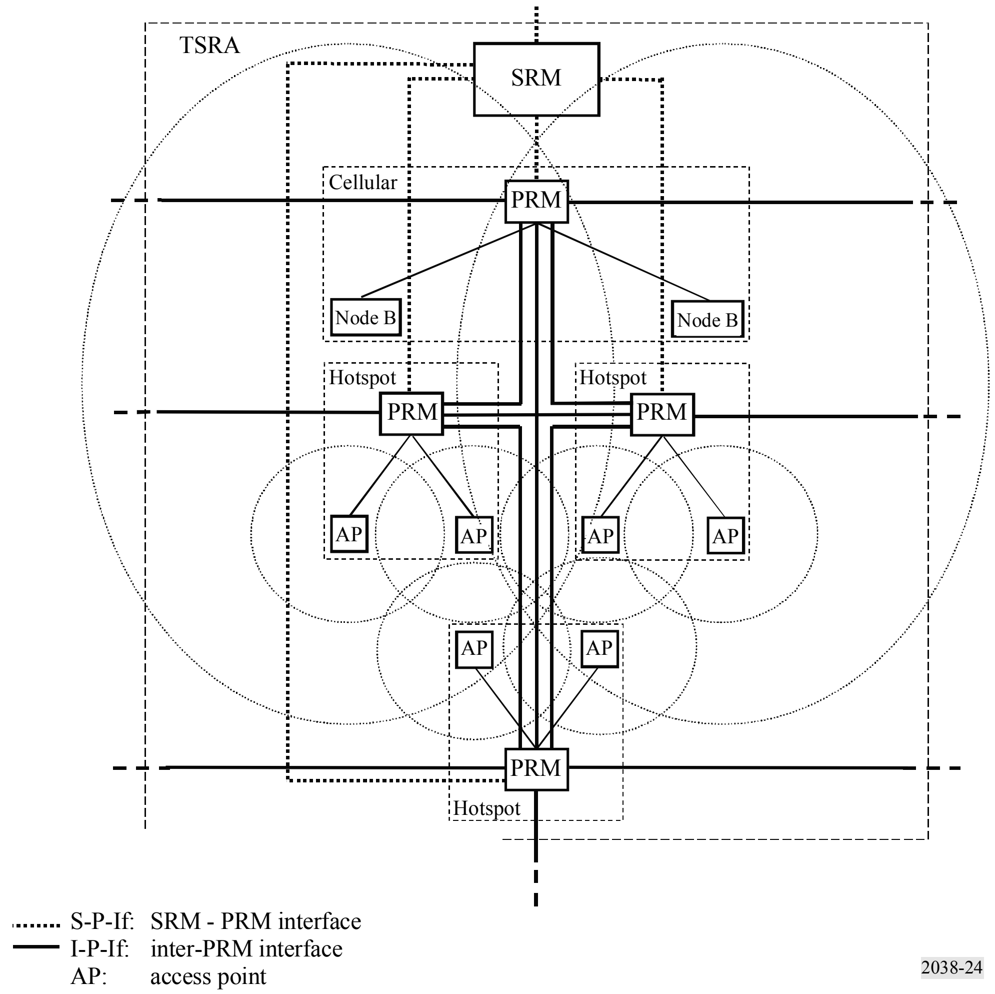
3.2 Terminal reconfiguration serving area (TRSA)

We call the area of the served PRMs by an SRM terminal reconfiguration serving area (TRSA). The TRSA cover may differ from the area of a single radio access network.

Figure 24 shows an example of a terminal reconfiguration serving area. In this area different radio access technologies are located. Three hotspot access points (e.g. IEEE 802.11 or Hiplerlan2) with lower range but high maximum available bandwidth and one cellular access point belonging to the same TRSA. The neighbouring PRMs with overlapping cell coverage are coupled with each other by the inter-PRM-interface and every PRM has a connection to the local SRM.

FIGURE 24

Example for entities and interfaces of a TRSA with different radio access technologies



3.3 Inter-PRM interface

The previous chapters have already shown that the overall functionality of the PRM is beyond the usual proxy functionality. One extension of the PRM could be a connection to neighbouring PRMs and the exchange of additional information via an inter-PRM-interface. First of all this could contain long-term information as the general supported QoS in the other radio access network (e.g. the priorities of different traffic classes, maximum bit rate, maximum delay, etc). Thus the

PRM could decide in advance which neighbouring mode is useful and for which mode the terminal should scan. After the detection of an alternative mode, which general QoS is promising, the PRM could request short-term measurements at the neighbouring PRM. The final decision should be made after the consideration of all the measurements.

After a mode switch is initiated, the old PRM could transfer useful terminal information to the new PRM over the inter-PRM-interface. In this way the new PRM can prepare the information concerning the attached neighbouring radio access networks for the new terminal and provide it to the terminal at an early stage.

As far as the measurement and resource information are concerned, every QoS supporting radio access network must provide a resource reservation mechanism. If there is a resource manager available, the PRM can request the long-term as well as the short-term information from this manager via the inter-PRM-interface and the neighbouring PRM and inform back the manager in advance if a mode switch is initiated. If there is no resource manager or other resource information entity in a network, the PRM must measure and store general results on its own.

3.4 Proxies in IP-based networks

To provide requirements for IP-based RAN architecture we start from the following all-IP scenario supporting reconfigurable terminals. The PRM is functionally split into a user and control plane servers following the methodology of the IP-based RAN. The IP-based RAN architecture is based on the following paradigms:

De-layering

It means that the logical and physical separation of the network elements on three functional planes:

– Control plane

– User plane

– Transport plane.

The *user-plane* handles the user data exchanged between the terminal and the CN. Its main task is to transform the user data into radio frames to be transmitted by the radio BS and vice versa. Due to the real time radio processing requirements, the user plane functions will be implemented by a highly-specialized hardware platform: The *user plane server (UPS).*

The *control-plane* manages the radio resources and controls the radio processing in the user plane. The control plane functions are grouped within the *radio control server (RCS)* that is typically a standard all-purpose platform. One implication is that the control functions (in the c-plane) have the capability to manage and control the radio resource in the u-plane even if they are related to different radio technologies.

IP transport protocol

All traffic transport in the whole network is based on IP datagrams. Thus, the network provides end-to-end IP-based connectivity and supports virtually all IP-based services. Radio BSs (BS, NodeB) are directly connected to the IP transport network. It means that the radio BSs are connected with the rest of the network via IP transport protocol.

Hierarchical function distribution

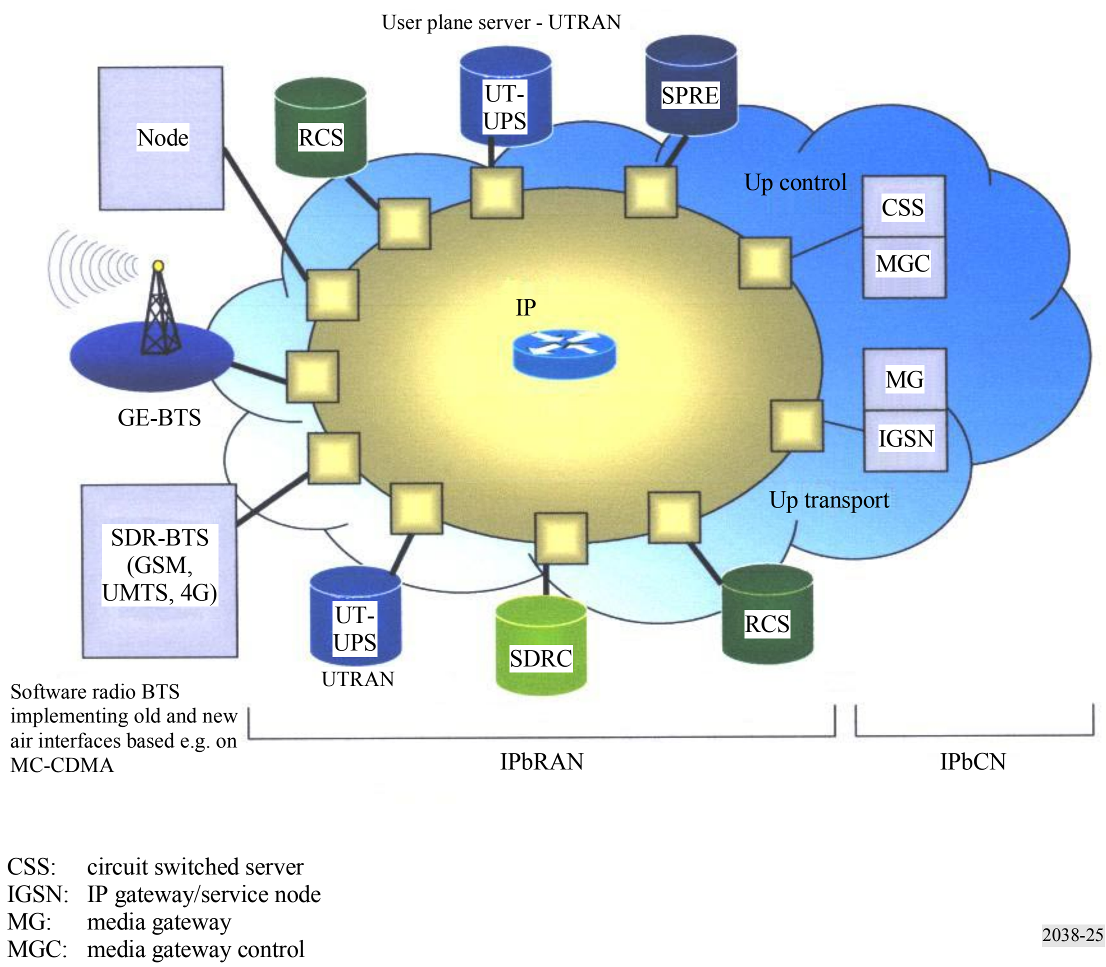
Hierarchical distribution of the functions means that network functional entities that execute functions related to a particular access technology are grouped into functional domains called *"access networks"* (IPbRAN). Functions common to all access technologies are provided by the subnetwork called *"core network" (*IPbCN).

The function split in c-plane and u-plane domain introduces scalability and load balancing for the proxies and their processing capability. Following the same approach the PRM is split into *software download and reconfiguration controller* (SDRC) that collects c-plane functions, and *software download and profile repository* (SPRE) that collects u-plane functions. Their functions, with the aim to address software download and reconfiguration functions, are explained in the following sub-chapters.

For example, a new evolution of BTS could be used to interface with SDR terminal. This kind of BTS should have an IP interface and contains some of the functionality previously held in the core network with the aim to provide multimode capabilities.

FIGURE 25

IP-based network and proxies supporting reconfigurable terminals



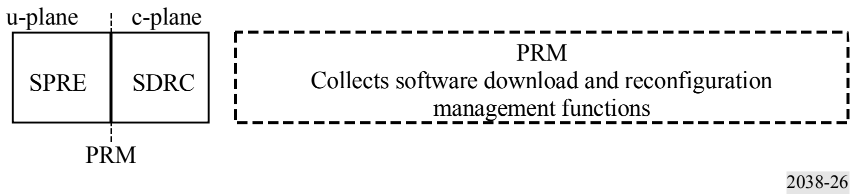
Function split in control and user plane for PRM

It follows from proxy functions, that the PRM is mainly involved in mode monitoring, mode negotiation, mode switching and software download, where information about user, terminal and services and control capability are required. Following the *de-layering* approach PRM functions can be split in u-plane and c-plane functionality. The u-plane functions collect all the profiles and the information needed for the performing of the software download. The c-plane functions collect all the functions needed to perform the control operations. We call:

– *SPRE:* Software download and profile repository (u-plane) the logical entity that collects:
  – terminal, user, service profile
  – software modules.
– *SDRC:* Software download and reconfiguration controller (c-plane) the logical entity that hosts the PRM controlling functions, in particular:
  – software download control functions
  – reconfiguration control functions.

FIGURE 26

PRM function split



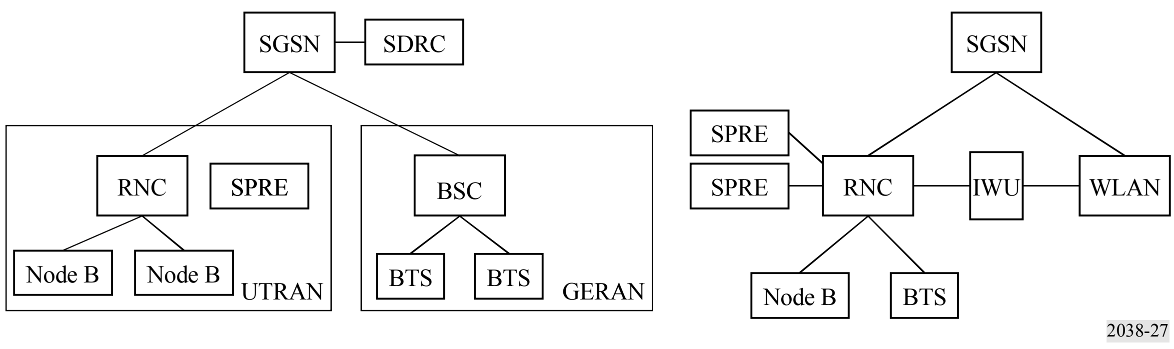
The u-plane PRM (SPRE) should be located in the RAN, because the following information has to be available and updated as fast as possible:

– Channel and system load conditions
– User service profiles
– Application QoS
– Terminal capabilities
– Network address and protocol.

Depending from the type of coupling between different RATs and the time constraints of the service and the necessity or not to perform a seamless handover, the c-plane PRM (SDRC) could be located both in the RAN (near or included in the RNC) or in the core network (near or included in the SGSN). When the SPRE and SDRC are located both in the access network, the interaction with the RRM is more direct and the micro mobility management could be managed from a dedicated SDRC function. Figure shows the location of the SPRE and SDRC in both scenarios.

FIGURE 27

Example of proxy control functions in the core network (left) and in the RAN (right)



Annex 7

High data rate packet nodes (HDRPN)

1 Cellular telephone emulates the traditional wired telephone capability

The intent of the cellular radio concept has, in the past, been to emulate as closely as possible the service that is provided to the user by a standard wired telephone. This was necessary to make it so the telephone user could use the cell phone just like a normal telephone and the telephone network would be usable to the providers of the cell phone service for long-distance interconnection. This made the cell phone much easier to sell, particularly in the initial stages. Using the telephone network to provide the long-distance element of the cellular service meant the primary task was to emulate the local access part of the telephone network. Technically this was not such a difficult task, but to provide low cost spectrally efficient solutions became a challenge. A solution that reassigned the expensive BS equipment and the spectrum on a call-by-call basis proved to be successful for telephone type voice traffic.

2 Traditional telephone network becomes part of a data network

In the meantime, the uses for the telephone itself were expanding rapidly. With the advent of the computer it became evident that the computer would be more useful if it could access remote databases and communicate with other computers. Again the telephone network was the most interconnected medium so it was decided to use it along with its local access lines to establish this interconnectivity on a worldwide basis. The result became known as the Internet. The local access lines were also a challenge for the Internet engineers. The lines could not pass data at the data rates that would make the users want to download large files, so they did everything possible to increase the data carrying capacity of the local lines, including high speed line modems and still it required the lines be connected for long periods of time; much longer than a voice call. The users have become accustomed to expect this kind of Internet service from their voice telephone. Therefore, they also expect the cellular phones to provide the same Internet service. Wired phones are available for connection to the network all the time even when a call is not in progress, therefore the local wireline access connection costs do not cost the provider more if the connection times are longer. For the cellular provider, his costs go down proportionately if he can service many users with the same equipment. Sharing the same spectrum by many users is also important for radio operators, but wireline users have no spectrum sharing problems for the local loop.

3 Traditional telephone access network needs upgrading to support advanced Internet

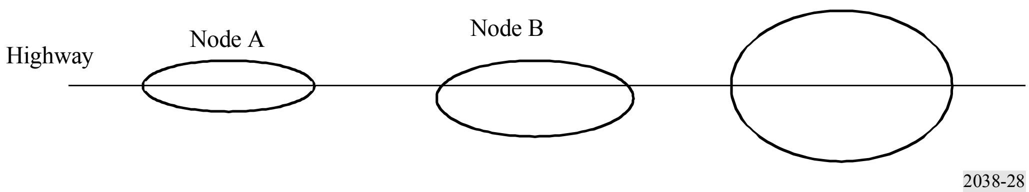
Based on the previous discussion, the range of data rates that it is necessary for a future cellular system to support in a ubiquitous manner is up to 56 kbit/s, because that is the data rate the wireline can normally support which is provided by the local telephone service. Actually, a large portion of the world cannot expect to achieve even 28 kbit/s. Higher data rates are available from wireline providers, but special measures are needed to obtain them and they are not available everywhere anytime. In many areas it can take weeks to obtain even DSL, T1 or ISDN service from the local wireline providers. Therefore, a wireless service that provides higher data rates at specific locations under the right conditions is consistent with the way high data rate services are delivered by the wireline providers. The cellular provider may decide to make 144 kbit/s or higher available over the entire region, but that becomes a cost and spectrum availability determination by the provider.

4 HDRPNs provide asymmetric wideband data rates

A concept for a future system that provides up to 56 kbit/s over the entire service area with high rate packet nodes that deliver hundreds of megabits of data over specific regions can be an attractive cost efficient solution. The HDRPNs are designed to transmit low power even when transporting high data rates, thus reducing the interference into adjacent bands and to remote co-frequency sites. This allows the spectrum to be reused many times using inexpensive nodes.

First consider the case where the high rate packet nodes are placed along a major highway. These nodes would create small areas of coverage as shown in Fig. 28. The size of the coverage areas need not be the same size as also indicated in Fig. 28. These small super cells can be selected to cover toll booths, major intersections, persistent congestion points or spaced periodically to regulate message delay characteristics. Using this type of structure for the packet transfers the average throughput for the data can be shown to increase dramatically, as much as nine to one[7]. The delay is also an important consideration and analysis shows the best results are achieved when the cells are small so the node can transmit higher packet rates[7]. When the user is a pedestrian or a vehicle that is expected to traverse a more random path the route may not come close to a HDRPN, as defined above, for some time. In this case it is important to have a large number of economical nodes. Placing nodes at intersections, plazas, train stations and allowing businesses that have a major feeder trunk entering their facility to mount external antennas for public use will provide significant coverage. Experience will determine the location for additional nodes if they are necessary. Analysis of this two dimensional case also shows the possible advantages[7] of HDRPNs.

FIGURE 28



[7] FRENKIEL, R., BADRINATH, B. R., BORAS J. and YATES, R. [April 2000] The Infostation Challenge: Balancing Cost and Ubiquity in Delivering Wireless Data. IEEE Personal Comm., Vol. 7, **2**, p. 66. Rutgers University United States of America.

5 Location services and specific message control provide practical solution

Location services are a significant adjunct to effectively utilizing the high rate packet node concept. Location services allow the user, with aid from the system, to predict the time he will enter the coverage of a high rate packet node. Using location services the system can notify a user he has a message waiting, with information about the message, when it can be delivered and what the delivery charge will be. Given this information the user can choose if he wants to receive the message or wait till he gets home to open it. This call placement information can appear on a screen or be announced audibly. The location of the user can be determined in several ways depending on the capability offered by the system. The user can also utilize the information from a GPS by having the ability to forward his GPS coordinates to the cellular system[8]. It is also possible for the system to aid the user terminal in storing the last calculated position of the user terminal in the user terminal memory file where it can be used like a set of GPS coordinates[8]. The user can also initiate the process for sending large data files or while browsing the Internet. While browsing he may request a large file, again, if the size of the file exceeds a reasonable size the system sends the same message as when a large message was waiting which includes elements that give the size of the requested file, the time when it can be delivered and the delivery charge. Again, if he does not accept the request the data file is not delivered. The operator will determine reasonable file size based on the required delivery data transmission rate. Rates that do not exceed normal Internet line modem rates should be considered reasonable. The location services and the transmission of the information to set up the high speed packet transfer are conducted on the supporting ubiquitous low speed cellular network. The system will make the determination as to how the high rate packet transmission is to be accomplished. If there is sufficient capacity remaining in the cellular system one of its cellular nodes can act as the HDRPN when the user terminal gets close to the cellular BS[9,10]. If this is not feasible the system can identify a specific high rate packet node for the high speed data transfer. Similarly a low data rate satellite system can utilize a high rate packet node concept to transfer large data files. In remote areas the delay may be greater due to the distance between HDRPNs. However, since the number of potential routes in remote areas is much less the number of high rate packet nodes required to cover expected routes may be a reasonable number.

---

[8] RIDGELY BOLGIANO, D. y LaVEAN, G. [25 March 1997 – Filed September 6, 1994] Wireless Telephone Distribution System With Time and Space Diversity Transmission for Determining Receiver Location. United States Patent No 5,614,914. InterDigital Technology Corporation, United States of America.

[9] FURUSKAR, A., MAZUR, S., MULLER, F. and OLOFSSON; H. [June 1999] Edge: Enhanced Rates for GSM and TDMA/136 Evolution: *IEEE Personal Comm. Mag*., Vol. 6, **3**.

[10] CHUANG, J. QUI, X. and WHITEHEAD, J. [September 1999] Data Throughput Enhancement in Wireless Packet Systems by Improved Link Adaptation with Application to the EDGE System. Proc. IEEE VTC '99-FALL, Amsterdam, The Netherlands.

6 HDRPNs enabling technology

As the packet transfer rate becomes very high the radio channel transmission characteristics may be different than what would be optimum for a ubiquitous cellular system. These high transmission rate channels, that are probably asymmetric, would be well suited for a TDD implementation, see Annex 3, where the channel distortion of the received signal can be used to precondition the high rate transmission signals. This high rate packet capability may require a modified air interface tailored for this application, see Annex 9. The TDD structure is also very flexible and can provide several different data rates at the same time or rapidly reconfigure the entire capacity into a single channel. This flexibility is important since new users enter and leave the coverage areas very quickly. The TDD structure also allows the mobile uplink to access a reconfigured channel to transmit large files to the network. This reconfiguration can assign a large portion of the time slots to the uplink transmission, but when the transmission is completed the time slots can be instantly assigned to other users, probably downlink users. Since the uplink is not expected to be used as frequently as the downlink it is spectrally efficient to have the flexibility to reassign the same spectrum between uplink and downlink. With a FDD structure the uplink spectrum would have to be reserved to handle large data throughput even if it is only used occasionally. Reservations also help these reconfigurations and the location services capability can provide accurate reservation information. Antennas also play a major role in the ability to transmit high data rates. The energy from the node needs to be concentrated into a small area to increase the signal strength for the target receiver and to reduce the amount of interference to the rest of the region. Antenna systems such as adaptive antennas (see Annex 4 and Annex 5) distributed antennas[11], (see also Annex 1) and intelligent cells[12] as possible solutions for concentrating the transmitted energy into small areas.

One possible solution for control of the process would require a network algorithm that searches for the addressee of the packet file and determines the approximate location and quarries that region looking for the addressee, see Annex 9. The response to this particular quarry will include a geographical location and cause a second answer to be sent approximately 10 s later. Both transmissions are needed to determine direction and velocity of travel. Using the position and velocity vector for a targeted mobile the system calculates the arrival time for the targeted mobile to reach the next high rate packet node. The system sends, utilizing the ubiquitous cellular network, a message telling the addressee such things as how long it will take before he receives the message, who the message is from, the length of the message and how much he will be charged for deliver of the message. The addressee is asked if he wants to receive the message. If the answer is yes a reservation[13] is made and the packet is delivered by the next high rate node and the addressee is

---

[11] RIDGELY BOLGIANO, D. y LaVEAN, G. [12 de enero de 1999 – Registrado el 29 de agosto de 1997] Wireless Telephone Distribution System With Time and Space Diversity Transmission. United States Patent No 5,859,879. InterDigital Technology Corporation, Estados Unidos de América.

[12] LEE, W. C. Y. [1995] *Mobile Cellular Telecommunications – Analog and Digital Systems*. Air Touch Communications Inc., Second Edition, Publisher McGraw Hill Inc., p. 563.

[13] GANGULY, S., NATH, B. and GOYAL, N. [February 2002] Optimal Reservation Schedule In Multimedia Cellular Networks. Tech. Report. Rutgers University, United States of America.

billed for delivery. The system needs to also implement a special algorithm to rapidly synchronize and transfer packets. Probably a special error correcting technique is needed for this relatively unique data transmission channel[14]

The mobile terminal will have to be an adaptive terminal to provide this premium service and interact with the HDRPN, but still have capability to function effectively in the cellular network. Fortunately there is much progress in the field of software defined radios, see Annex 6 and Annex 11. These radios will also benefit from the work in RF MEMS, see Annex 12, as an enabling technology for the future terminal. This is not expected to have much impact on the cost of the terminal over a normal high end IMT-2000 terminal.

Annex 8

Internet technologies and support of IP applications over mobile systems

1 Introduction

Current IMT-2000 systems include IP-based conversational services carried over IP-based networks connected to related radio access technologies. The basis of IP-based conversational service is to provide the IMT-2000 terminal with IP support including IP addressing. This IP layer operates above the IP-based bearer services in the IMT-2000 networks (core network and radio access networks).

When discussing IP requirements the above has to be kept in mind, so that support for all-IP conversational service and above applications are not mixed with network internal requirements related to the usage of IP for its internal usage. It should then be remembered that the networks are defined for optimized usage with the IMT-2000 radio technologies.

In the ongoing enhancement of IMT-2000, an all-IP conversational architecture has been selected as the base for the all-IP conversational service definition for several of the radio access technologies. This means that necessary adaptation of Internet Engineering Task Force (IETF) protocols to effectively support all-IP applications run in a mobile system radio environment is being done. Example of such work is the header compression protocol progressing in IETF ROHC Working Group (WG).

All-IP multimedia services are not the evolution of the circuit switched services but represent a new category of services, mobile terminals, services capabilities, and user expectations. Any new multimedia service, which may have a similar name or functionality to a comparable standardized service, does not necessarily have to have the same look and feel from the user's perspective of the standardized service. Voice communications (IP telephony) is one example of real-time service that would be provided as an IP multimedia application.

[14] MAO WU, H., EVANS, J. y CAGGIANO, M. [April 1999] Report WINLAB-TR-178. Improvement of IP Packet Throughput with an Adaptive Radio Link Protocol for Infostations' Technical.

2 Technologies to support more efficient IP applications over mobile systems

Several of the IMT-2000 technologies already support end-to-end IP applications by means of the radio access bearers defined. Moreover, more advanced IP header compression has been added in order to support IP applications even more efficiently according to ongoing work in IETF ROHC WG. Currently, there is work ongoing in order to further enhance the IP application SIP signalling support by means of optimizing the transport of IP application signalling, possibly with a dedicated RAB.

2.1 Robust header compression

IP header compression is a technique to significantly reduce the size of the IP header while maintaining the header information contents 100%. Several compression schemes exist in today's Internet. However, for the mobile environment these schemes are not enough since they rely on very low header loss rate. In the mobile environment with fairly high loss rate, more robust header compression is needed. This requirement has been acknowledged by IETF in the ROHC WG that has developed a robust header compression scheme an RFC.

2.2 Signalling compression

The signalling needed for all-IP is envisaged to consist of SIP/SDP signalling which are not optimized for mobile environment in the sense that it is ASCII based. This leads to excessive signalling delays and loss of performance. Signalling compression is therefore essential in order to successfully support all-IP services.

3 Reuse of Internet technologies in the radio access networks

Although the radio access networks of the existing IMT-2000 technologies are to a large extent using other means of transport between their internal nodes than IP transport, work is ongoing in order to be able to use IP the transport technology. It is essential that the introduction of IP as transport technology in the radio access network does not degrade operability, performance and end-to-end QoS. The necessary mechanisms for IP transport between nodes in a radio access network are being developed for the ongoing enhancements of IMT-2000. Especially IPv6 is considered.

4 IP-based RAN

4.1 Architecture and transport network

4.1.1 Wireless access technology independence

– IP-based RAN shall at least support the current 2G and 3G radio technologies such as W-CDMA and cdma2000. This requirement shall be based on the support of the complete backward compatibility and inter-operability. This requirement also makes sure that

IP-based RAN architecture supporting UTRA can accommodate IP-based RAN architecture supporting cdma2000 just by replacing the radio-specific function from UTRA to IP-based cdma2000.

– IP-based RAN architecture may be flexible to support another non-cellular wireless technology, e.g. Wireless LAN and Bluetooth. This requirement also is based on the flexible architecture to support the radio-specific function with plug and play operation.

– IP-based RAN shall support inter-working and inter-operation to support the handoff among different wireless access technologies.

4.1.2 Backward compatibility

– IP-based RAN architecture and associate protocol set shall support the backward compatibility with the current RAN architecture from the ground up.

– From the service perspective, no service and performance degradation are expected in IP-based RAN comparing to the current RAN architecture. This requirement shall need some criteria to verify the backward compatibility.

– IP-based RAN architecture and associate protocol set shall provide at least the equivalent functionality to the current RAN architectures. The main changes in IP-based RAN architecture comparing to the current RAN architecture shall be easily identified.

4.1.3 Interoperability with legacy (2G/3G) networks and mobile terminals

– IP-based RAN architecture shall support interoperability between current 2G/3G core network (including GSM MAP, ANSI-41 Core Network) and IP-based RAN.

– IP-based RAN architecture shall support interoperability between current 2G/3G radio access networks and IP-based RAN.

– IP-based RAN architecture shall also provide support for legacy (2G/3G) mobile terminals.

4.1.4 Forward compatibility

– IP-based RAN architecture and associate protocol set shall allow the forward compatibility with the architecture accommodating the new radio technologies.

– IP-based RAN architecture shall easily accommodate any change to be expected by introducing new services (e.g. IP broadcast/multicast services).

4.1.5 Interoperability with all IP network and IP-based mobile terminals

– IP-based RAN architecture shall support the interoperability between the All IP core network and IP-based RAN.

– IP-based RAN architecture shall support the interoperability between the next generation following 2G/3G radio access networks and IP-based RAN.

– IP-based RAN architecture shall also provide support for IP-enabled mobile terminals.

4.1.6 Layered independent architecture

– IP-based RAN architecture shall support the layered architecture.

– IP-based RAN architecture shall support the separation among the user plane function, the control plane function, and the transport plane function.

4.1.7 Open interface support

– IP-based RAN architecture shall support open interfaces between any network entities in IP-based RAN that may be implemented by operators/ISPs and manufacturers as separate systems, subsystems, or network entities.

– IETF protocols shall be considered and adopted in these open interfaces wherever possible. For example, mobile IP mechanism can be another alternative for IP mobility in the RAN.

4.1.8 QoS support

– IP-based RAN architecture shall support the means to enable end-to-end QoS at least within RAN scope.

– IP-based RAN architecture shall satisfy the policy-based QoS architecture. This requirement can raise the issue, where the policy enforcement point and the policy decision point are.

– The resolution of QoS in IP-based RAN shall be consistent with the edge-to-edge (ETE) QoS on the core network level.

– IP-based RAN architecture should be capable of simultaneously supporting multiple levels of static QoS (negotiation of parameters before the session setup) as well as dynamic QoS (negotiation of parameters while the session is in progress) including in handoff scenarios.

– IP-based RAN architecture shall support the QoS enabled routing/handoff procedure. That is to say, the best selection of the routing path/handoff path satisfying the QoS required by the user shall be possible.

– IP-based RAN architecture shall support the configuration with load balancing for supporting the different level QoS requirement per user.

– IP-based RAN architecture shall support IPv6 enabled QoS resolutions.

– IP-based RAN architecture may support IPv4 enabled QoS resolutions.

4.1.9 IP transport

– IP-based RAN architecture shall transport bearer and control/signalling traffic based on IP technology.

– IP transport in the RAN shall be independent of the L1, L2 technology.

– IP-based RAN architecture shall support IPv6/IPv4 addressing mechanism.

4.1.10 Distributed dynamic configuration

– IP-based RAN architecture shall support the multiple dynamic configuration between functional entities. For example, the Node B functionality can select an appropriate RNC functionality for satisfying the QoS and the robustness (non-drop) during the call.

– IP-based RAN architecture shall support the distributed radio control and bearer control functions.

– IP-based RAN architecture shall support the distribution of cells dependent radio and bearer control functions towards the radio access points.

4.1.11 Radio resource management

– IP-based RAN architecture shall support the efficient radio resource management (allocation, maintenance, and release) in order to satisfy the QoS required by the user and the Policy required by the operator.

– IP-based RAN architecture shall support the function to optimize and negotiate the radio resource among the different wireless access technologies that are supported.

4.1.12 Performance

– The performances in IP-based RAN architecture regarding to link utilization, QoS, call drop rate, easy handoff, and so on shall be equal to or greater than those in the current RAN architecture.

– IP-based RAN architecture shall support increases in capacity without architectural impact.

4.1.13 Scaleable architecture

– IP-based RAN architecture shall provide network operators the ability to expand specific RAN function entities independently of other entities.

– IP-based RAN architecture shall allow network operators to gradually deploy network entities and expand their networks.

4.1.14 Security

– IP-based RAN architecture shall provide functions to protect its network resources and traffic from unauthorized access.

– IP-based RAN architecture shall handle multiple radio link authentication protocols (e.g., CAVE for IS-95, A5/1 for GSM).

– IP-based RAN architecture shall allow AAA to be present in the RAN or in the core network for access authentication, authorization.

4.2 Radio network

– IP-based RAN architecture shall optimize the use of the bandwidth for end-to-end IP transport for certain class of real time applications.

  – IP protocols have a large amount of overhead that will reduce spectral efficiency, in particular when used for voice applications. It may also suffer intolerable delays because bandwidths for wireless mobile applications are still limited and sometimes expensive to obtain. In case of predominant IP/UDP/RTP protocol stack, the size of the combined headers is at least 40 bytes for IPv4 and at least 60 bytes for IPv6, while the voice data is typically shorter than the IP/UDP/RTP header; various header adaptation or multiplexing techniques can be therefore applied. (e.g. header compression or PPPmux).

– IP-based RAN shall support the very flexible allocation of resources among different cells and also dynamically uplink and downlink based on the unpredictable change of IP traffic over the overall access network.

– IP-based RAN shall support inter-working/interoperability of the QoS mechanism developed for the radio access network and the QoS mechanism used in the IP core network (e.g. MPLS, DiffServ).

  – New parameters may be, if necessary, introduced into the QoS mechanism of the radio access network or the IP core network.

– IP-based RAN shall maximize spectrum efficiency over the air.

  – Improved statistical multiplexing should be provided to support mixed services (e.g. real time variable bit rate and non-real time bursty data stream).

  – Optimal source and channel coding should be enhanced for various IP multimedia applications.

– IP-based RAN architecture shall provide protocol stacks supporting a range of services with different QoS requirements in the access network.

– IP-based RAN shall enhance medium access control and radio link control (RLC) for different IP multimedia applications.

  – Improved radio access and resource allocation scheme should be applied for different IP applications in medium access control. Fast uplink access procedures can be, for example, provided on the uplink and downlink for a certain type of service.

  – Radio link control functions should be differentiated for each flow with more delicate flow classification (e.g. mapping of IP traffic onto appropriate radio bearer in the access network.)

– IP-based RAN shall optimize physical layer mechanisms to guarantee the quality of some applications such as voice over IP.

  – Requirements for real time IP applications can be for instance applied to the optimization of physical layer mechanisms for guaranteeing QoS of the services. In this case, a couple of parameters for VoIP service (e.g. type of codec, echo control, voice packet size and de-jittering delay) can be used for the setting of coding rate, interleaving span in the physical layer; whereby the quality of voice over IP can be further improved.

– IP-based RAN shall consider interactions with functionalities of layered protocols for optimal IP packet transmission. Interactions between IP protocols (e.g. TCP) and radio protocols (e.g. radio link control) shall be in particular investigated for this purpose.

  – IP protocols in conventional wired networks may be modified to be qualified for wireless access networks. For instance, typical TCP protocol interprets packet loss as congestion and erroneously reduce throughput; some solutions can be therefore applied to IP-based RAN.

  – There are various control loops that will operate simultaneously by higher and lower layer protocols (e.g. TCP flow control and RLC error controls). These functions in the same protocol stack should be, hence, optimized for efficient packet data transmission.

  - RLC mechanisms may consider interaction with error detection and recovery function of other lower-layered radio protocols.

  - Optimal implementation of bandwidth adaptation methods in a scalable audio and video codec should be considered for spectral efficiency of the access network.

- IP-based RAN shall offer radio technologies (e.g. radio protocols and physical mechanisms) to support a variety of broadcast and multicast services (e.g. the multimedia message service and the Internet radio broadcast service).

- IP-based RAN shall support bearer differentiation capability at the access network for multiplexing different types of IP traffic over the air to achieve maximum spectrum utilization.

  - Radio technologies should be optimized for different bearer at the access network. For instance, different coding and access schemes can be applied to various radio channels.

- IP-based RAN shall optimize connection admission for efficient radio resource usage in both uplink and downlink for a mixture of data flows with different QoS per IP address.

  - The characterization of different IP packet data streams may be applied to connection admission function in terms of service requirements such as bandwidth and delay.

  - Effects of various deployment scenarios (e.g. spectrum availability) and traffic mix such as voice and data on spectrum efficiency should be taken into account.

- IP-based RAN shall offer diverse protocol states and radio channels, and besides, fast and dynamic transition among them to support a wide range of services for different IP multimedia applications. (e.g. signalling, real time, non-real time, connection oriented and connectionless services, and combinations of these services).

  - The type of the radio channels can be dynamically changed to accommodate different types of IP packet streams for the same connection.

- IP-based RAN shall support the fast resource assignment and release procedure on the uplink and downlink for some of the IP applications that are characterized as an on-off traffic pattern. (e.g. browsing and FTP).

  - With this feature, link utilization can increase due to non-continuous bandwidth management.

- IP-based RAN shall provide handover procedures minimizing packet loss and delay for robust and seamless IP packet transmission support.

  - Lossless data transfer mechanisms should be applied to achieve the robust IP packet transmission during handover procedures of IP packet streams.

– IP-based RAN shall support advanced radio technologies that are expected to emerge in future, such as adaptive antennas, link adaptation, OFDM, software defined radio and multi-user detection.

5 Wireless Internet for heterogeneous networks

5.1 Requirements

A major challenge for the future-generation wireless Internet is that the architecture will have to be very flexible and open, capable of supporting various types of networks, terminals, and applications. The fundamental goal is to make the heterogeneous networks transparent to users. Another goal is to design a system architecture that is independent of the wireless access technology. These considerations lead to a set of requirements that are specifically relevant to heterogeneous networks. The major elements are:

a) multiservice user terminal (multi-module/SDR-based) for accessing different RANs,

b) wireless system discovery,

c) wireless system selection,

d) unified location update and paging,

e) cross-system handover,

f) simple, efficient, scalable, low-cost,

g) energy-efficient

h) secure,

i) QoS support,

j) personal mobility/universal ID.

5.2 Overview of the concept

5.2.1 Basic entities

To meet the above-mentioned requirements, here we present an architectural concept of heterogeneous networks for future wireless systems [Wu *et al.,* 2002; Inoue *et al*., 2002]. It consists of three major entities.

(A) Common core network (CCN): This can be a managed IPv6 network providing a common platform through which all multi-service user terminal (MUT, described below) will communicate with correspondent nodes in the Internet. In principal all access points of RANs are connected to this network. The network provides QoS-guaranteed routing and seamless handover among RANs. This enables natural integration of various heterogeneous networks.

(B) Basic access network (BAN): It provides a common control/signalling channel to enable all MUTs to access the common platform (CCN). The network is basically used to provide location update and paging and support wireless system discovery and vertical (cross-system) handover for all other wireless systems. Consisting of base stations and basic access components, the BAN will have a broad coverage area, preferably larger than that of the RANs it supports.

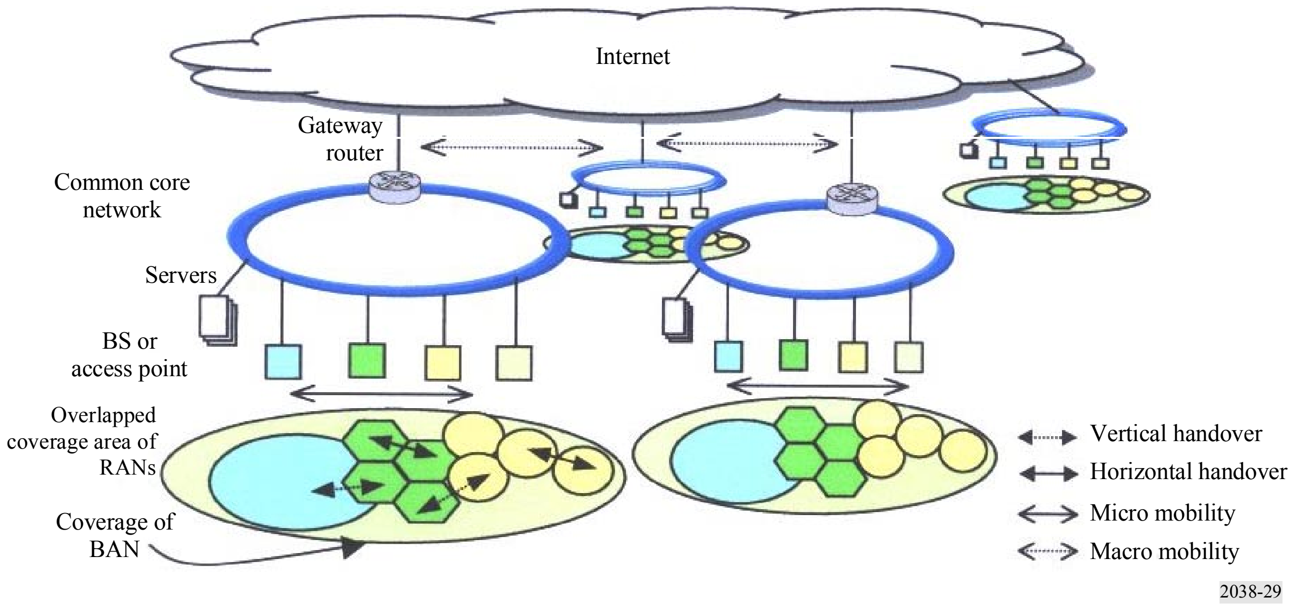
(C) Multi-service user terminal (MUT): The MUT is equipped with a multi-radio system. All terminals have a basic access component (BAC) to communicate with the BAN. Apart from this radio system, an MUT is equipped with one or more radio subsystems (preferably SDR-based) to access the CCN.

5.2.2 Network model

Figure 29 illustrates the network configuration. CCNs are connected to the Internet via gateway routers. A CCN provides services for several RANs. In general, the RANs will overlap, and a mobile host can have access to several RANs in one location.

FIGURE 29

Heterogeneous network concept



Mobile IP can be used for connecting CCNs and providing global (macro) mobility management. In a CCN-managed area, fast handover between BSs often belonging to different RANs with high-speed wireless access requires local (micro) mobility management. BSs are connected to (or integrated with) a regular IP forwarding engine. These engines are connected through some network topology that allows packets to be transmitted between the BSs and the gateway. Different RANs handle only those tasks that are specifically related to a certain radio access technology. In general, wireless access radio incorporates the physical and the data link layers only. Communication between RANs belonging to the same CCN is based on lower network layers (link or network layer).

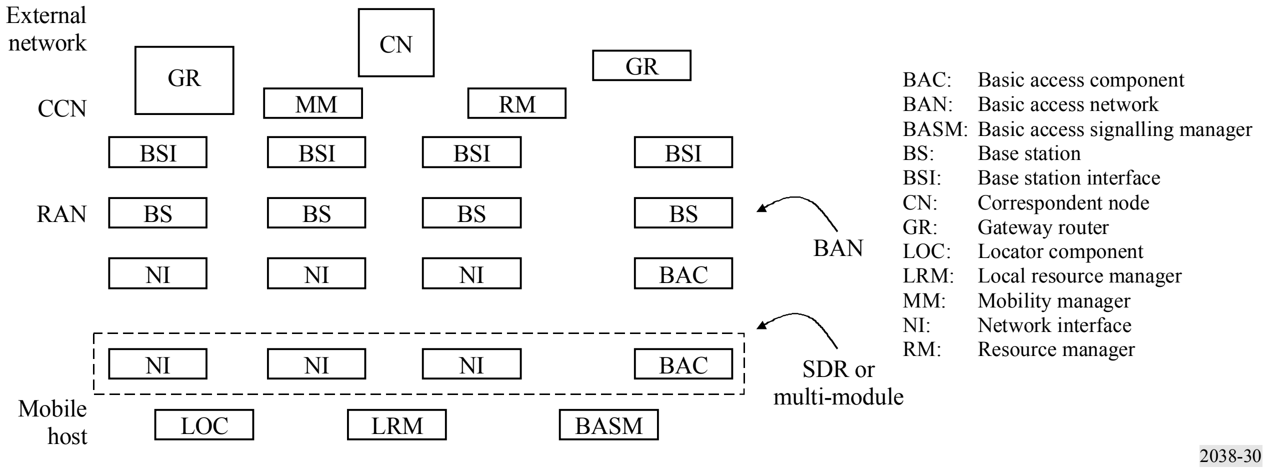
In general, the mobility related terms used in this section have the same meaning as in the definitions of IETF [Manner and Kojo, 2003]. However, there is one important difference when it comes to a heterogeneous networks environment. Micro-mobility is not confined to only a single access network. Instead it is confined only to an IP domain and access points belonging to different radio access networks may be connected to the domain. Inter-access network handover will not necessarily invoke macro-mobility.

5.2.3 Heterogeneous network architecture

In this section, we introduce the functional entities of the heterogeneous network architecture and the required protocols. The architecture as depicted in Fig. 30 is composed of four major building blocks: a mobile host, RANs, a CCN, and an external network. Within the external network, there are correspondent nodes (CNs). One or more gateway routers (GR) connect the external network to the CCN. Two important functional entities within the CCN are a resource manager (RM) and a mobility manager (MM). They are primarily responsible for traffic distribution and mobility-related problems.

FIGURE 30

System architecture



The CCN supports communication with the BSs, and thus with RANs. A base station interface (BSI) is primarily used to provide a uniform access mechanism for the BSs to access the CCN. The BSI can be a component of a BS. The BSs deal with wireless access problems in the normal link layer and collect status information of the wireless network they support. They use a network interface (NI) to access the network.

A Functional entities of common core network

The main goal of the architecture is to integrate different access technologies into a common architecture. To achieve this goal, the main tasks to be fulfilled by the architecture are resource management to coordinate traffic distribution in the system and mobility management to support roaming mobile hosts.

The RM is thus responsible for resource allocation and admission control to support traffic distribution in the CCN. It selects a RAN that can provide the service requested by the mobile host in the most efficient way. In essence, it combines multiple wireless access systems and exploits their specific strength to provide services in a spectrum-efficient way. Another task of the RM is to interact with IP QoS architectures (such as IntServ and DiffServ) that may be used in the external network.

The MM deals with all mobility-related issues. It keeps track of the location of mobiles, and determines which access networks are available to a mobile host at a certain location. The RM uses this information. The other main task of the MM is to provide handoffs, both local within the CCN and for the external network. To provide these handoffs, it needs to interact with the RM. If a mobile host moves within the core network, the mobility is transparent to the network layer, and the system tries to maintain IP flows and IP QoS parameters.

B Functional entities of basic access network

Here we enumerate some major functionalities or usage of BAN [Mahmood *et al.,* 2002].

BAN is mainly used to support heterogeneous paging. In a mobile environment, systems must be energy-efficient since terminals rely on batteries to operate. We expect that wireless IP communicators will be "reachable" continuously (i.e. "always on"), although not be necessarily communicating most of the time. In essence, mobile hosts will be in an idle state, but passively connected to the network infrastructure. It is then extremely inefficient to have to scan all RANs, and wait for a paging message. Moreover, since wireless networks are optimized for special services, they may not be very efficient for paging messages. A wireless network that is optimized for this kind of traffic is more efficient.

BAN can provide wireless system discovery. The BAN enables common access; every mobile host can use this BAN. The network provides the terminal with information about currently available wireless networks, so that the terminal does not have to scan all possible RANs.

BAN is used as a signalling network especially to enable vertical handoffs. Such a dedicated network can provide this service efficiently and securely.

BAN is used as a medium for most signalling and control messages. This simplifies the design of new wireless access services, since signalling is performed by another entity (BAC) or network.

C Functional entities of mobile hosts

Mobile hosts include all standard transport protocols and wireless specific control services. Control messages are transparently sent between the core network and mobile hosts' functional entities.

As shown in Fig. 30, a mobile host will contain a BAC and an SDR-based NI or multiple built-in or pluggable NIs. The BAC is used as a primary component to communicate with the BAN. Using an embedded positioning capability or the locator component (LOC), the BAC sends out location update data for paging (coarse update) when the mobile host moves across the paging boundary, and for system discovery (fine update) when the mobile host initializes a call or requires a vertical handoff.

The LRM deals with the local resources of the terminal and interacts with the resource manager in the CCN to determine what network should be used and when it will be operational. The BASM works in coordination with LRM and MM to manage the signalling over BAN.

6 IETF mobility related terminology for mobility

To facilitate understanding of this Annex, this section provides terminology, related to the support of mobility as defined by the IETF. At this time these definitions have not been completely aligned with all relevant ITU vocabulary Recommendations. The scope of the definitions supplied in this section is, therefore, limited to the text in this Annex.

According to the IETF [Manner and Kojo, 2003], multiple architectural options should be supported for mobility management. It takes the following view regarding mobility management in IP-based mobile networks:

> Different sorts of mobility management may be required of a mobile system. We can differentiate between user, personal, host and network mobility.
>
> User mobility
>
> > Refers to the ability of a user to access services from different physical hosts. This usually means the user has an account on these different hosts or that a host does not restrict users from using the host to access services.
>
> Personal mobility
>
> > Complements user mobility with the ability to track the user's location and provide the user's current location to allow sessions to be initiated by and towards the user by anyone on any other network. Personal mobility is also concerned with enabling associated security, billing and service subscription authorization made between administrative domains.
>
> Host mobility
>
> > Refers to the function of allowing a mobile host to change its point of attachment to the network, without interrupting IP packet delivery to/from that host. There may be different sub-functions depending on what the current level of service is being provided; in particular, support for host mobility usually implies active and idle modes of operation, depending on whether the host has any current sessions or not. Access network procedures are required to keep track of the current point of attachment of all the MNs or establish it at will. Accurate location and routing procedures are required in order to maintain the integrity of the communication. Host mobility is often called "terminal mobility".
>
> Network mobility
>
> > Network mobility occurs when an entire network changes its point of attachment to the Internet and, thus, its reachability in the topology, which is referred to as a mobile network. Two subcategories of mobility can be identified within either host mobility and network mobility:
> >
> > Global mobility
> >
> > > Same as macro mobility.
> >
> > Local mobility
> >
> > > Same as micro mobility.

Macro mobility

Mobility over a large area. This includes mobility support and associated address registration procedures that are needed when a mobile host moves between IP domains. Inter-AN handovers typically involve macro-mobility protocols. Mobile-IP can be seen as a means to provide macro mobility.

Micro mobility

Mobility over a small area. Usually this means mobility within an IP domain with an emphasis on support for active mode using handover, although it may include idle mode procedures also. Micro-mobility protocols exploit the locality of movement by confining movement related changes and signalling to the access network.

Local mobility management

Local mobility management (LMM) is a generic term for protocols dealing with IP mobility management confined within the access network. LMM messages are not routed outside the access network, although a handover may trigger Mobile IP messages to be sent to correspondent nodes and home agents.

7 References

INOUE, M., WU, G., MAHMUD, K., MURAKAMI, H. and HASEGAWA, M. [September 2002] Development of MIRAI System for Heterogeneous Wireless Networks. PMRC02, Lisbon, Portugal.

MAHMUD, K., WU, G., INOUE, M. and MIZUNO, M. [September 2002] Basic Access Network – the Signalling-only Network for Power-Efficient Mobile Multi-service Terminals in MIRAI Architecture. IEEE VTC Fall 2002, Vancouver, Canada.

MANNER, J. and KOJO, M. (Eds.) [April 2003] *Internet Engineering Task Force Internet-Draft, Mobility Related Terminology*.

WU, G., HAVINGA, P. and MIZUNO, M. [February 2002] MIRAI Architecture for Heterogeneous Network. *IEEE Comm. Mag.*, p. 126-134.

Annex 9

IP broadband wireless access technologies

1 Introduction

Ever-increasing demands for mobile communications require the continual evolution of systems, and development of new systems where required, for multimedia applications such as high speed data, IP-packet and video. The addition of a new IP broadband wireless access standard to the IMT-2000 family, offering unprecedented bandwidths and transparent access to all the content of the Internet and any content, public or private, based on the IP, could become a powerful element to consolidate the growth of the wireless industry.

In order to identify and to clarify this new concept Fig. 31 is used as a reference:

FIGURE 31

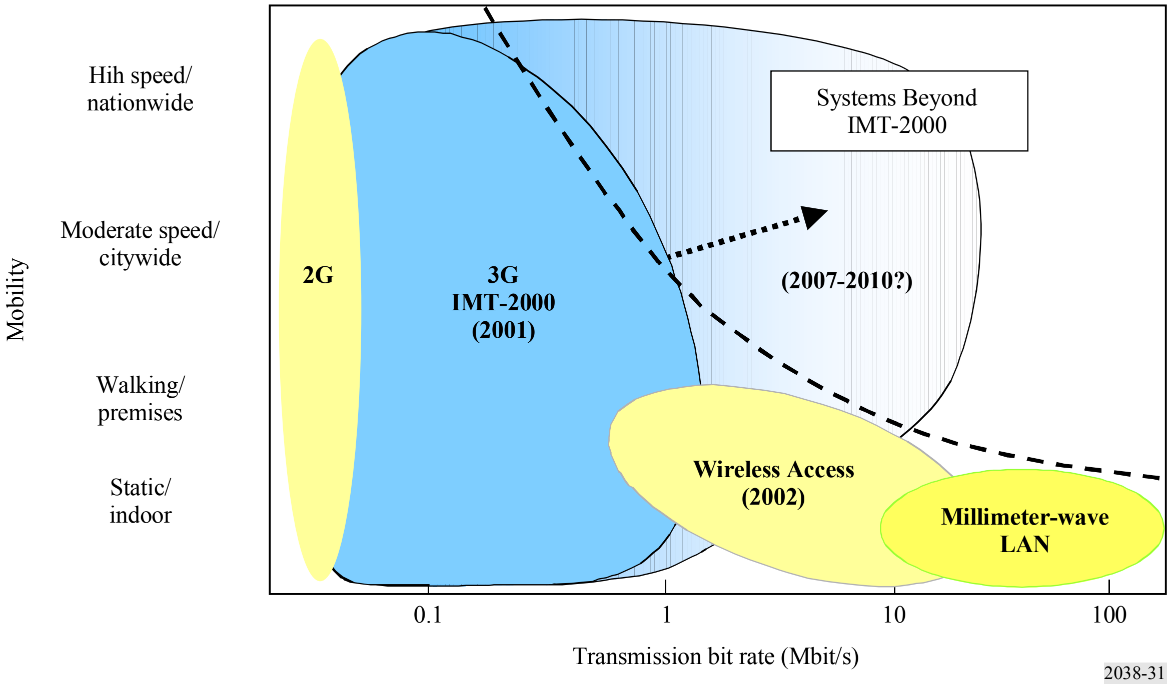


Figure 31 describes mobile systems with respect to two basic parameters, mobility and transmission rate, and proposes that there will be an expansion of the capabilities of mobile systems simultaneously towards both high mobility and high data rates. This dual evolution may be somewhat paradoxical, at least from a purely technological standpoint, as may have been recognized by the dotted line of the Figure, which may be understood as a trade-off or trend-line. In addition, a more detailed description of the possible evolution would include additional parameters such as the inclusion of more spectrally efficient technologies (such as adaptive antennas) in the present or planned mobile systems. However, even with significant development of technology, for fundamental physical reasons, it can be safely stated that:

a) high mobility,

b) high spectral efficiency,

c) high transmission rates cannot be efficiently or economically combined within a single system.

Thus, it may perhaps be a misconstruction to derive from Fig. 31 above that a single air-interface should cover the whole envelope of parameters a), b), and c) above, not to mention the added complexity of backward compatibility with various legacy switched-circuit systems.

This leads then naturally to the concept of optimized air-interface modules to serve market niches or rather markets spheres: One could then anticipate products that are designed to respond optimally to specific user profiles, defined broadly enough across substantial geographical markets, and incorporating specific subsets of these modules.

This modular approach would be in line with the current technological development of multi-standard/multimode terminals and with SDRs concepts. Nonetheless, there will continue to be implementation and integration issues.

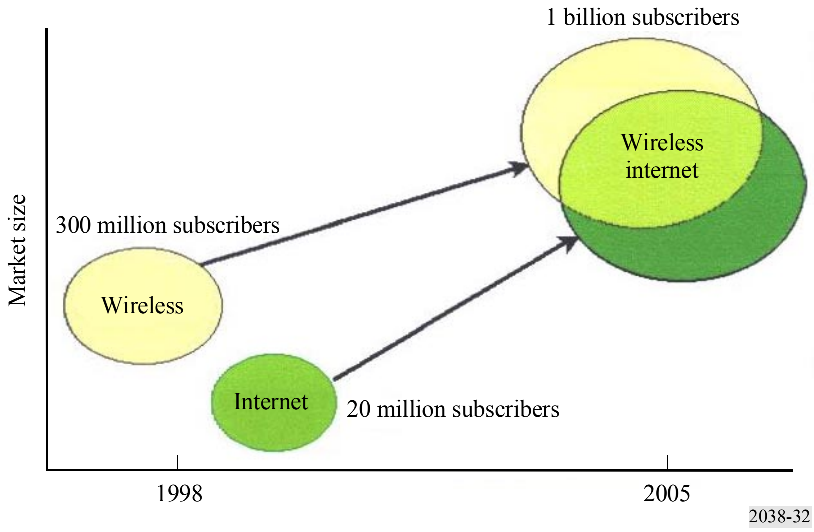
This Report proposes expanding the performance of core IMT-2000 systems simultaneously toward higher bandwidth and higher mobility is best achieved through the concept of air interface modules to serve specific market segments, which can be adopted within the IMT-2000 family of standards.

2 An example of focus area: IP broadband wireless access

Wireless communications have created a generation of users who are entirely dependent on portable devices for personal connectivity. The underlying technologies have matured to a level where portable telephony is ubiquitous, and a very large and dynamic market has formed. One promising new area for the transparent convergence of the Internet and wireless is portable broadband Internet access.

The advent of the Internet into personal and commercial communications is creating new opportunities as well as new challenges for telecommunication system planners, operators and equipment designers and manufacturers. In particular, it is agreed that wireless access to the Internet will soon represent an enormous market (see Fig. 34). Satisfying the needs of the wireless users, while continuing to meet the requirements of the myriad service providers who are offering their wares on the fixed or wired Internet creates additional challenges. Wireless providers may not be able to meet these challenges with today's offerings, and will be hard-pressed to meet even with the next generation of wireless networks that are in various stages of planning. However, new enabling technologies such as packet radio, adaptive antenna systems and Internet-derived architectures may make access to the rich content (streaming video, etc.) of the Internet as pervasive as cellular telephony today.

FIGURE 32



The entire Internet industry has grown due, in part, to the low entry barrier for a vast variety of content and services. But whereas in many markets wired residential consumers have seen bandwidth rise from 9.6 kbits/s to over 1 Mbit/s at somewhat affordable prices, the same consumers still have great difficulty to gain web-browsing access to the Internet when they leave their fixed, wired connections. Although an industry has evolved around "broadband wireless" systems as identified on Fig. 31, these are targeted at providing high-speed connections to fixed locations and buildings, not people.

Historically, despite the mobile wireless industry's repeated attempts at providing wireless data services, customers were often reluctant, though there are now some widespread consumer success stories: for example, the numerous applications and end-user devices in the Personal Handiphone and i-Mode networks in Japan, and the enormous adoption of short messaging service in GSM networks. But the adoption of packet-based networks for wireless systems has so far been gated by the need to provide circuit-switched voice services on these networks. Consequently, the adoption of end-to-end packet networks has been slowed in the past by the lack of consumer demand. In addition, most end-user devices are based on telephony concepts rather than designed as Internet appliances. Only recently have portable Internet appliances designed for wireless connectivity started to emerge.

Mobile wireless data applications can be categorized into segments that become increasingly more demanding in complexity, bandwidth and transparency to Internet content and protocols:

– Basic Internet content (e.g. weather, stocks, and news) is widely available today in most commercial mobile networks, using the Wireless Application Protocol (WAP) and other Web clipping techniques, and is delivered to mobile phones in text form.

– Over the next few years, network-enhanced applications – those that require some level of intelligence and transactional capability in the network – such as geolocation-ready applications, will emerge on these devices. Those applications have real value and are not demanding of bandwidth – only of network intelligence and preprogrammed interaction between the network and the user.

– Next, there is a variety of applications that require some level of security and reliability, such as mobile secure commerce and corporate access to intranets. Here, the issue is more complex in that the end-user has some specific and very demanding needs, such as security and encryption; those problems are being solved today.

– Undoubtedly, the IMT-2000 family of systems will allow for a considerably enhanced Internet user experience. However, from a practical standpoint it could be argued, given the ever increasing needs for voice access, that unless VoIP on a packetized network will be available, wireless access would still lag behind wireline access in terms of transparent Internet access ease, quality and affordability. Thus there is a case for specifications and standards for wireless system architectures that would enable providing business users and consumers very fast data-rate connection to the Internet, with freedom to move, and an always-on experience. Such a standard would, alongside the main components of the IMT-2000 family, focus on the arguably vast niche of data-only access mechanisms, and be wholly complementary of the other IMT-2000 family components.

The growth of the Internet is the key engine behind the need for such a standard. The greater the use of the Internet in day-to-day life and the greater the breadth of applications on the Internet, which users will experience at fixed stations, the greater will be the need for a service that allows them the same unfettered access while at a different, but un-served location. One of the goals of establishing such a standard would be to offer to all an "untethered multimedia experience". A new breed of application developers will extend broadband applications to the portable domain, as well as invent them specifically for that domain, such as broadband geolocation services and content. Such services would include telework, telehealth, tele-education, entertainment, tourism, gaming, and instructional content.

3 Breaking the wireless access bottleneck: IP broadband wireless access

The conjunction of the above trends creates an immediate need in industrialized countries, soon to be followed by emerging markets, for wireless data communication systems with the following characteristics, as seen by users and network operators:

– high data rates packet access;

– synchronous and asynchronous applications;

– high spectral efficiency;

– high efficiency in asymmetric traffic (TDD);

– "always on" connectivity;

– freedom to move (with low mobility);

– low cost;

– intuitive (transparent) interfaces for access to rich content.

Development of such new standards as stated in this contribution would rely on suitable advances in enabling new signal processing technologies that have been already developed and tested in various modern commercial communication systems.

This Report proposes that this concept deserves to be recognized as a high priority for the following reasons:

– Broadband data carrying Internet content transparently must be provided cost-effectively to many users at any location (and not only indoors).

– The combination of mass wireless connectivity (mobile systems) and mass broadband Internet connectivity (at desktops and homes through wired systems) will create a substantial "market pull" for broadband wireless systems, and standardization of the latter will benefit harmonization between systems and ultimate success.

– Developing countries' need for broadband Internet access, a potentially important booster of economic activity, may be served more cost-effectively and in a more timely manner through ubiquitous wireless connectivity.

In summary:

– The current focus of the mobility wireless industry is to emphasize value-added services with content and applications, rather than broadband IP.

– Providing ubiquitous IP broadband wireless connectivity is a new and different focus area.

4 Conclusion

This Report provides the elements of definition for a typical example that, with due market considerations to justify the development of new IMT-2000 wireless access standards. The addition of a new IP broadband wireless access standard to the IMT-2000 family, offering unprecedented bandwidths and transparent access to all the content of the Internet and any content, public or private, based on the IP, could become a powerful element to consolidate the growth of the wireless industry.

In line with the needs of developing countries as identified in Question ITU-R 77/8 this Report should be considered as the basis for the development of a preliminary draft new Recommendation for the development of new IMT-2000 wireless access standards.

Annex 10

Radio on fibre (RoF)

1 Introduction

This Annex describes a generic RoF system as defined in § 3.2.5, and identifies RoF requirements and functional specifications.

2 RoF system description

The RoF system consists of the following elements: BTS repeater(s), optical network, antenna repeater(s), and control module.

The system with its interfaces is represented in Fig. 33. A simplified schematic representation is also shown as an illustrating example in Fig. 34.

FIGURE 33

RoF system and interfaces

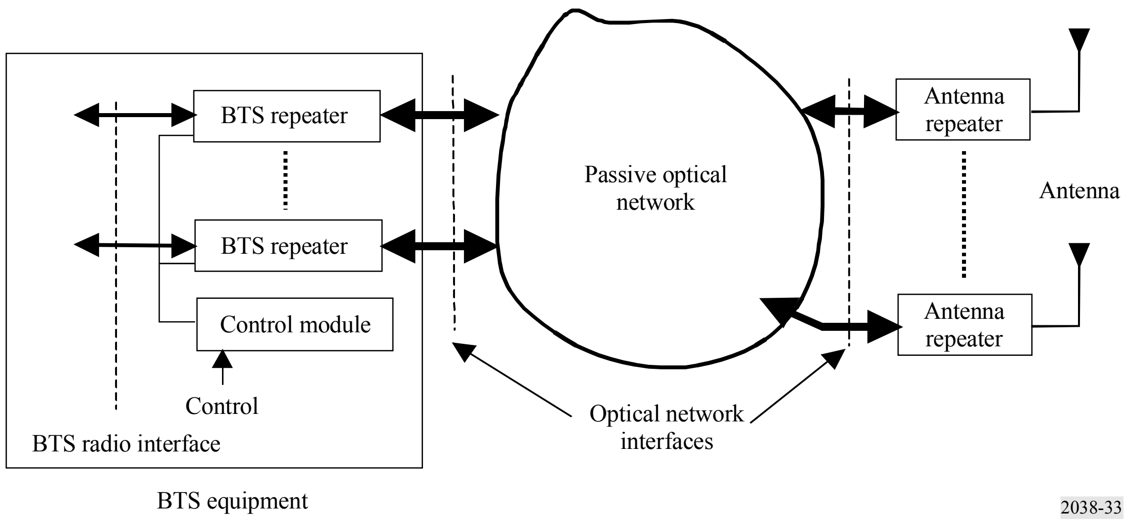
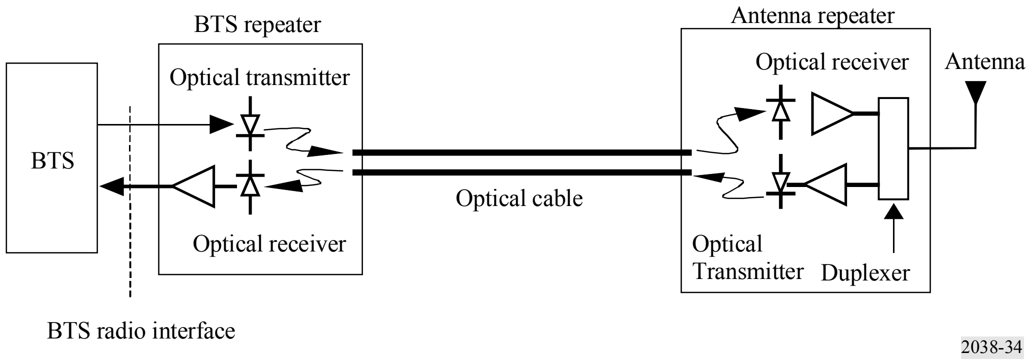


FIGURE 34

RoF simplified scheme



2.1 BTS repeater definition

The BTS repeater adapts the BTS RF interface to the optical network interface, network side. In the simple version depicted in Fig. 34, an RF carrier coming from the BTS modulates the intensity of an optical transmitter whose output is fed into an optical fibre (downlink). Correspondingly, another optical signal, intensity modulated by an uplink RF carrier, is received from another fibre and fed into an optical receiver, whose amplified output is injected into the BTS receiver port.

2.2 Optical network definition

The optical network is passive, unless explicitly stated otherwise. It connects the BTS repeaters housed in one common room with the antenna repeaters, through their corresponding optical network interfaces. Besides one or several optical fibres, it might include connectors, splitters, wavelength division multiplexers and demultiplexers and, in general, any passive optical device. In the simple version depicted in Fig. 34, it is a set of two optical fibres, for downlink and uplink transmission, respectively, connecting one BTS repeater to one antenna repeater.

2.3 Antenna repeater

The antenna repeater adapts the optical network interface, antenna side, to the transmission and receiving antennas. In the simple version depicted in Fig. 34, an optical carrier, intensity modulated by a downlink RF carrier, is fed into an optical receiver, whose output is filtered, amplified and injected into a transmission antenna. In the uplink path, the RF signals coming from the receiving antenna are filtered, amplified and fed into the intensity modulation port of an optical transmitter, whose output is in turn fed to an uplink optical fibre.

2.4 Control module

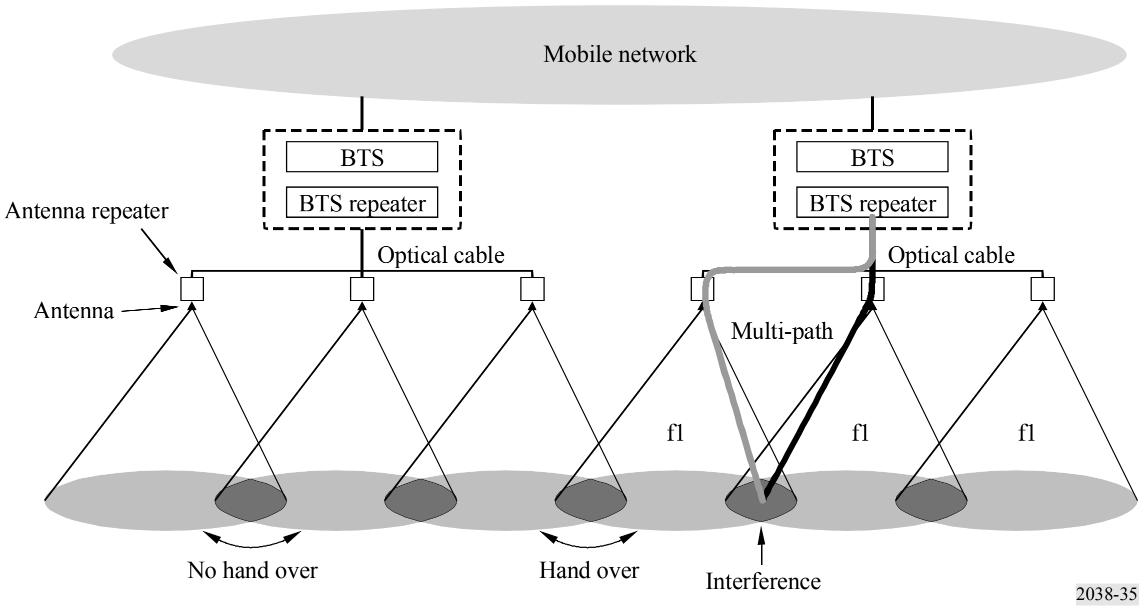
This is a logical entity, which can also be a physical module or have its functionality embedded in the BTS repeater. There is one control module per RoF system. Its mission is to provide a control interface to an external management system. Communications with the management system can be performed by means of the same wireless system, or through a wireline connection.

2.5 Group cell architecture

RoF group cell architecture shown in Fig. 35 is very effective for street-cell or hotspot cell, since handover between cells under the same BTS is not required. Therefore, high mobility of user's terminal might be achieved in the cells under the same BTS shown in [Fujise, 2001; Harada *et al*., 2001].

FIGURE 35

Group cell by RoF



3 RoF requirements and functional specifications

3.1 Transport capacity

For a given wireless system, the transport capacity of a RoF system is defined in terms of the following parameters:

– Number of bidirectional RF carriers it can transport.

– Number of antenna repeaters it can service simultaneously.

– Number of RF carriers per repeater.

– Downlink broadcast capability, or number of antenna repeaters which can radiate the same RF carrier.

3.2 Input/output performance

– Maximum output power per carrier.

– Downlink and uplink power gain.

– Degradation parameters like uplink equivalent noise figure, frequency/phase error, intermodulation characteristics, etc.

3.3 Optical plant

– Maximum optical plant insertion loss.

– Maximum optical path length.

Additionally, every RoF system sets a requirement on the maximum optical plant return losses, which depend on the optical transport technique and wireless system characteristics.

3.4 Management functions

An RoF system should support several management functional areas, like configuration (BTS and antenna repeaters) and fault reporting.

4 References

FUJISE, M. [August 2001] Radio over Fiber Transmission Technology for ITS and Mobile Communications. IEICE Trans. Fundamentals, Vol. E84, **8**, p. 1808-1814.

HARADA, H., SATO, K. and FUJISE, M. [December 2001] A Radio-on-Fiber Based Millimeter-Wave Road-Vehicle Communication System by a Code Division Multiplexing Radio Transmission Scheme. *IEEE Trans. on Intelligent Transportation Systems*, Vol. 2, **4**, p. 165-179.

Annex 11

Terminal architecture

In IMT-2000 the combination and convergence of the different worlds of information technology (IT) industry, media industry and telecommunications will integrate communication with IT. As a result, mobile communications together with IT will penetrate into the various fields of the society. The user expectations are increasing with regard to a large variety of services and applications with different degree of QoS, which is related to delay, data rate and bit error requirements. In particular users will expect from their mobile platforms not only the same breadth and depth of applications and capabilities that they currently enjoy on their PC platforms, but they will expect the same "Moore's Law" rate of advancement of new applications and capabilities.

Therefore from the user's perspective, IMT-2000 and systems beyond IMT-2000 represent a fundamental change in expectation. Rather than merely expecting a "new and improved" but "static" collection of applications and services; the user will have an expectation of a dynamic, continuing stream of new applications, capabilities and services.

Such a continuing stream will flow from a healthy ecosystem of general-purpose programmable platforms supported by a large, robust, and vibrant developer community.

1 General-purpose programmable platforms

As generally defined by literature on the subject, a general-purpose technology possesses two key features that enable it to produce profound and lasting economic benefits:

– Technological dynamism; the technology has the ability to continually advance in performance.

– Broad application; the technology can be used for a wide variety of purposes.

As the cost of the technology decreases, innovators can apply it to more and more areas of human activity; proliferating its deployment through large segments of the population, thereby increasing its impact on the whole society.

New mobile user equipment (UE) are assuming these characteristics by:

– Containing high power general-purpose processors that follow Moore's law of dramatically increasing price/performance.

– Providing a flexible, programmable platform that can be used for an ever-increasing variety of uses.

2 High-level concerns for mobile user equipment

The convergence of wireless connectivity and a general-purpose programmable platform heightens some existing concerns and raises new ones, so that environmental factors as well as traditional technology and market drivers will influence the architecture of these devices.

2.1 Environmental factors

On the environmental side, the situation can be viewed as a three-person game of the following fundamental interests:

– Economic: the commercial, consumer, and societal benefits of a product or service.

– Security: protection of commercial, consumer and public assets.

– Privacy: protection of the sensitive data from unauthorized access.

2.1.1 Economic

From a business perspective this is of course creating products or services that customers find attractive. There can also be non-commercial societal values desired.

2.1.2 Security

Against the great benefits of programmability, however, we have the spectre of security risks. These can take the form of network damaging viruses and denial of service attacks, fraudulent use of the network, and the piracy of spectrum; access and damage to sensitive data behind corporate firewalls; digital content theft; and theft or damage of customer applications or data.

2.1.3 Privacy

Against the need for authentication to combat commercial fraud, and legitimate law enforcement requirements, we must also balance the need to maintain the privacy of individuals and corporations against unwarranted invasion such as unauthorized access to customer proprietary network information, and sensitive local user data such as stored in persistent memory or generated by context-aware technology.

2.2 Key technological and market drivers

Combining with the environmental factors we have traditional market and technology drivers:

2.2.1 User value pull

– The deployment of robust packet data network capabilities, which allow new data-intensive software applications that integrate Internet and multimedia applications, such as streaming video, multimedia, animated graphics, m-commerce, and network connectivity. And many of these applications may be personalized or "Context Aware" with sensitive user information.

– The desire by consumers and mobile professionals to access secure, data-intensive applications.

2.2.2 Security requirements pull

– Dynamic security algorithms, mechanisms, and technologies.

– Authentication technology including biometric devices.

– Digital rights management for valuable content protection.

2.2.3 Technology enablers

– The emergence of low-power, high-performance microprocessors, dense memory, and efficient baseband logic has created the opportunity for placing unprecedented capabilities in the hands of users.

– Low-cost, high-performance servers, deployed in the infrastructure, to address the interface between information sources and wireless clients, and to make end-and-end capabilities a reality.

– Distributed communications technologies enabled by service discovery software middle-ware.

3 High level architectural trends for mobile user equipment

To meet the needs of network security and integrity earlier generation application development and delivery mechanisms for wireless devices was a serial and slow process. Hardware (the silicon and device) was developed first; applications were then written for a particular hardware and air interface; finally hardware and applications were then tested together to ensure proper operation for each specific air interface and network. However such a paradigm would not be viable to meet the user expectations for IMT-2000 and systems beyond IMT-2000.

– Applications development would not be able to keep up with the pace of the Internet growth if this serial development process were to continue.

– In addition to application software development, SDR technologies are enabling more and more air interface and network functions to be performed in software.

– Likewise security algorithms and technologies need to be continually evolving and improving to stay ahead of malicious intent and breaches.

To maintain network and user space integrity, communications software will be “decoupled” and executed in parallel with user applications being written to a general-purpose processor running in a general-purpose execution environment. This partitioning maximizes the economic viability by allowing application development to evolve independent from communication standards, as well as enhancing security by providing autonomous network and user spaces.

Creating coexistent autonomy for the radio subsystem, application subsystem, and memory subsystems portions is evolving as a means to solve the triple environmental requirements of enabling economically viable products and services; while maintaining network and corporate security, and user sovereignty over application space and data privacy. Put anecdotally, “good fences make good neighbours”.

By greatly reducing the interdependencies of the three players (economics, security, privacy) experimentation for finding the equilibrium can occur much more quickly and at much lower cost.

3.1 Key elements of coexistent subsystem autonomy in UE architecture

Some common elements required for these architectures are:

– The architecture is open – The autonomous subsystems are connected by open physical and logical interfaces, and support a wide range of operating systems, execution environments, and air interfaces. Open interfaces enable and foster competition of multiple vendors of the subsystems, and allows hardware and software development to evolve more rapidly and independently at their own pace thereby ensuring a continuous stream of innovative solutions at the lowest cost.

– The architecture is flexible and adaptive – High levels of modularity, allowing each module to be independently tested and reused across many different systems. A modular design incorporates the ability to integrate new hardware and software features as industry standards and market needs evolve.

Annex 12

RF MEMS

1 Development of a multi-standard RF module using RF MEMS components

As IMT-2000 services are about to be commencing, backward compatibilities with PCS, CDMA, or etc., are becoming necessary and then multiband/multi-standard terminals will appear on the stage. With the existing technologies, multiband/multi-standard terminals will be bulky and expensive. Therefore, we can use MEMS technology to develop an RF module that could be applicable for various frequencies with compact and flexible structures. Furthermore, an intelligent RF module for beyond IMT-2000 terminals based on SDR will be needed.

This technology can yield small size, light weight, low power and high performance to replace discrete passive RF components so that we can produce a flexible and compact RF module for multiband/multi-standard terminals.

Table 5 shows the characteristics of each component when RF MEMS technology is used.

TABLE 5

Characteristics of an RF module using RF MEMS technologies

Duplexer	To be 13 mm in thickness, using tunable RF MEMS filter
Switch	Low insertion loss, low power consumption, programmable
Antenna	A large reduction in volume Switching capacity to be applicable for multiband Minimizing the risk from electromagnetic waves
Front-end module	With single chip integration of RFICs and passive MEMS, volume can be reduced to one fifth Easy to find a plan to reduce interference caused from activities of multifrequencies Cost reduction, using single package Most effective solution for beyond IMT-2000 systems or SDR

2 A new RF solution with SDR technology

For beyond IMT-2000 terminals based on SDR, we need a more complex multi-standard RF module, applicable for multi-standards systems, such as GSM, DCS1800, PCS, W-PAN based on TDD and CDMA, PCS, IMT-2000 based on FDD. RF MEMS technology would be a fruitful solution for multi-standard terminals.

Figure 36 shows a basic SDR system block diagram. If this system is only composed of existing components, several RF modules will be needed, resulting in a huge volume, heavy weight and high cost. On the other hand, taking MEMS components such as programmable LNA, tunable filter, programmable complicated switch, and so on, we can simply make a competitive product with small volume and light weight which also can be related to future terminals and will be a solution for RF systems in future wireless mobile systems.

FIGURE 36

Multiband/multi-standard system block diagram

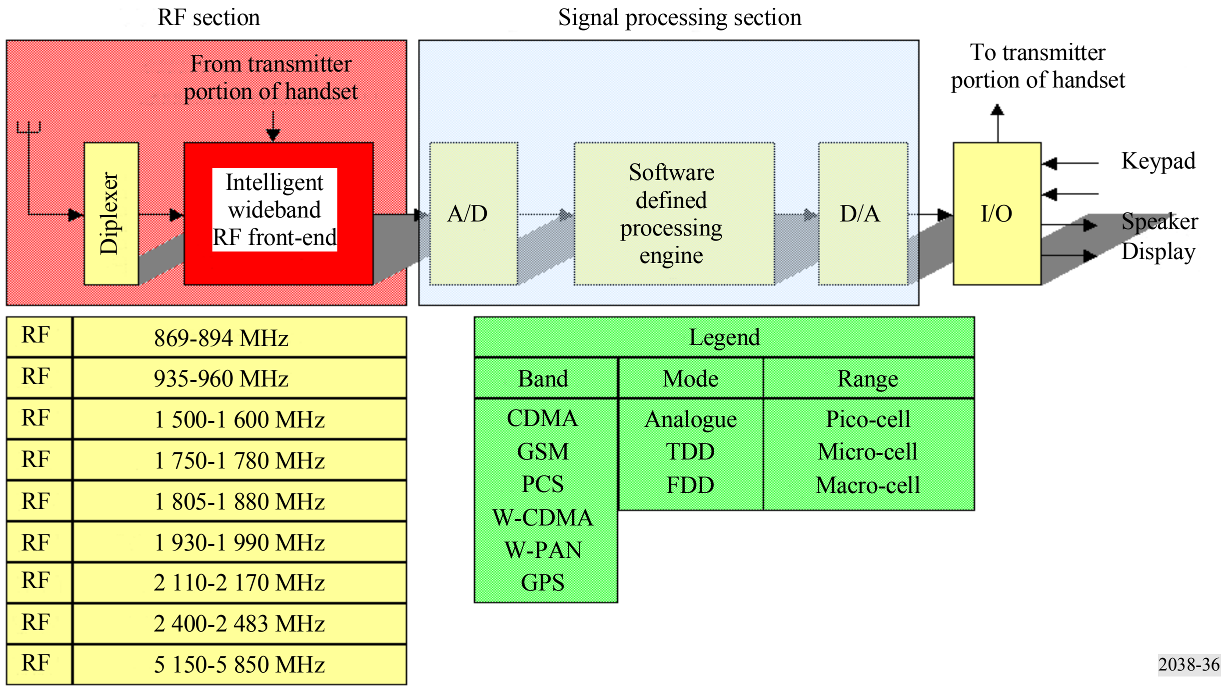
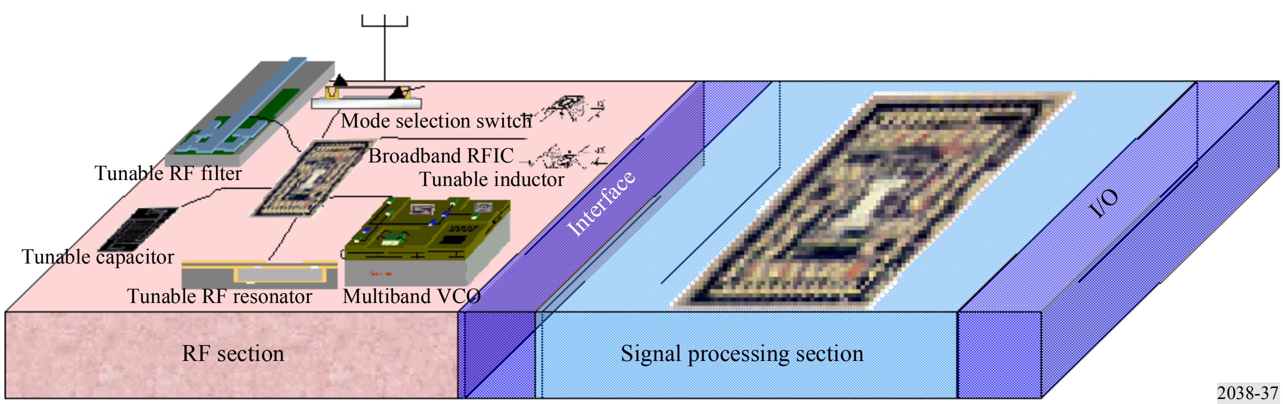


Figure 37 shows a new concept for a future terminal, which makes a system on a single chip package, combining the RF section with the signal processing section in one package. With existing technologies, this conceptual terminal cannot be implemented. However, when a multiband RF module with RF MEMS technology is accomplished, it would be possible to make an RF module with SDR technology as shown in Fig. 37, (RF section). What is more, a future terminal on a single chip package can be realized.

FIGURE 37

Future terminal concept with new RF solution



Annex 13

New innovative user interfaces for future multimedia wireless terminal devices

An example of a new innovative user interface:

GKOS – The global keyboard optimized for small wireless terminals

1 The new generation of terminal devices

How the user experiences new telecommunications technology, depends on the services offered and also on the usability, design and quality of the terminals. Wearable computing is a popular study item at universities worldwide, giving new ideas of man-machine interfaces applicable also for mobile terminals.

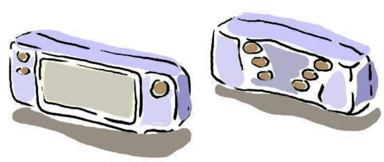
Text messaging is the killer data application of today, and multimedia messaging, including text, is expected to be the next. There must be a wide and big enough screen for displaying good quality images and videos. Combining a keyboard and a large enough display on a compact small terminal is a challenge. From the usability point of view, however, this is a must. Mobile text entry methods should also be made faster.

Most solutions offered for text input so far are not open standards but proprietary methods including IPRs. The proposed physical keyboards tend to add features and/or buttons to the conventional dialling keypad instead of decreasing the number of keys that could rather be the goal in order to minimize the space required.

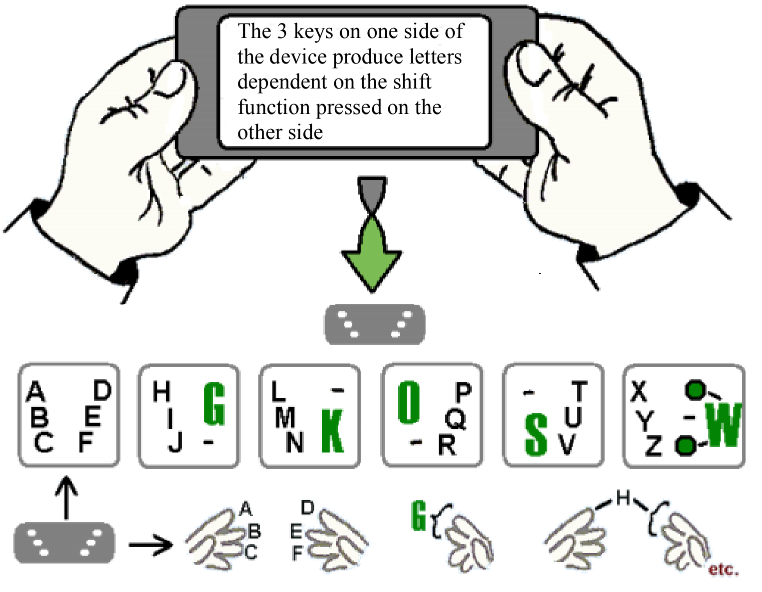
There is also a clear need for harmonization and for recommended use of common open interface standards in this area. Namely, if a user gets used to one type of keyboard and becomes a committed and skilled user of it, she or he will get frustrated if the next phone, new version or another brand, has a different or slightly different user interface solution and the learning curve must be restarted. The GKOS keyboard, a proposed open standard described below, is one solution to the problems just mentioned.

2 Keys on the back – the GKOS keyboard

The GKOS keyboard is a set of six keys on the back of the user device to keep the front panel available for the display and to enable fast text entry as well as provide all functions found on the PC QWERTY keyboard. It is intended for the tiny wireless terminals of IMT-2000 and systems beyond but can also be used in many other applications (e.g. full PC remote control). The proposed concept is an open standard and was first published on 5 October 2000. See also http://gkos.com/ for more details.



3 How to type on GKOS



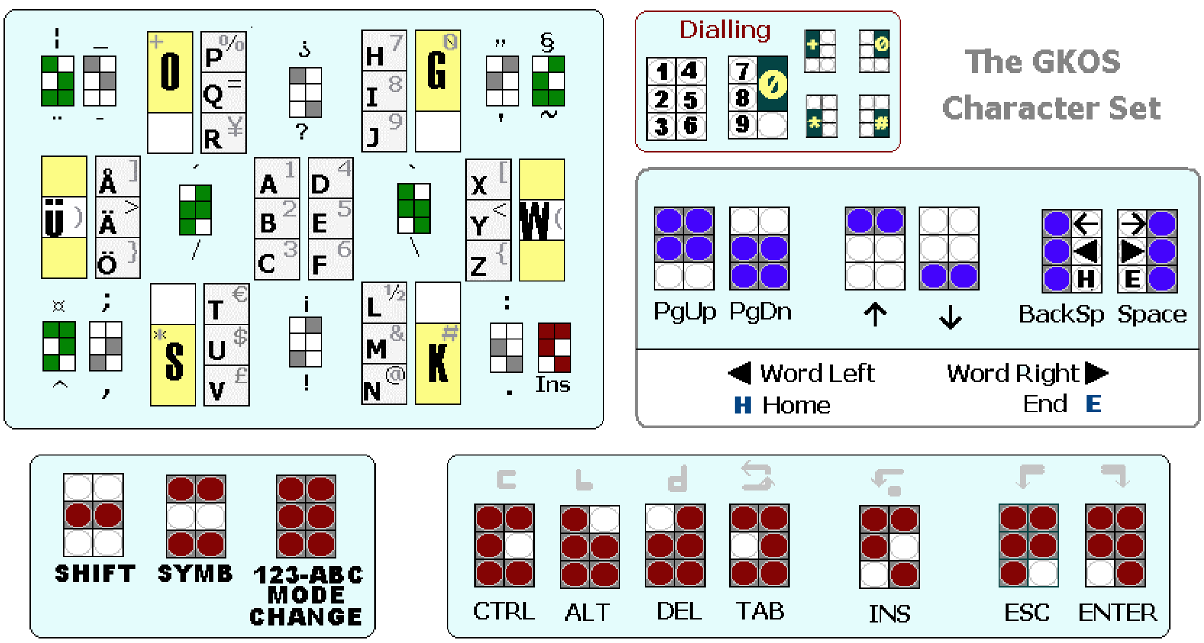
2038-37a

As described in the Figure above, for letters (and numbers), maximum two simultaneous key presses are needed per hand. Whenever one hand presses two keys (e.g. D+E keys), the other hand presses a single key (e.g. to get "H") or no key at all (result is then "G"). Letters A to F are just single keys.

Space: Press the GKOS "right bar" = all 3 keys on the right-hand side.

Backspace: Press the GKOS "left bar" = all 3 keys on the left-hand side.

Note that the instructions above are enough for entering simple text messages. Below, the full character set is shown including all PC QWERTY characters:



The yellow areas in the Figure represent two keys that must be pressed simultaneously to get the character marked on it, or to use them as a shift function to get the three other characters of the group. One group of characters is reserved for national letters.

The main principle is that, for frequently used characters, only one to three simultaneous key presses are needed but for functions more. This way typing is lighter and functions will not be activated by mistake if typing carelessly. The 123-ABC mode change toggles between the two characters sets shown above. With SYMB, single characters can be picked from the parallel set. Further, for example, semicolon is SYMB + period. Two consecutive SHIFTs set CAPS LOCK on and one SHIFT only has an effect on one character. For clarity, the numbers are shown here twice ("Dialling" above). Note the suggesting form of the control characters, and that the blue characters for navigating, scrolling and cursor movement are quite self-evident.

4 GKOS features

4.1 Advantages and disadvantages

Basically, GKOS is a chording keyboard (combination keyboard) but as the keypad is split between two hands, the number of different combinations of simultaneous key presses per hand is very low (three chords for letters and numbers, four in total, compared to 50 + chords of conventional chord keyboards), and because they are extremely simple (like just pressing two adjacent keys) no special physical skill is required.

Using the GKOS keyboard involves both hands, as is also the case e.g. with PDAs having stylus interface or QWERTY. This can be seen as one disadvantage of the GKOS. Also, there is a learning curve as this is a completely new method of typing, but the list of advantages, however, is long:

1. The keyboard takes no space on the front panel or on the display area.

2. Low cost and easy to integrate on a small terminal (only 6 keys needed).

3. Open keyboard standard, free for anyone to use, for other purposes too.

4. The same hardware suits most languages.

5. Easy enough for the beginner (on-screen virtual keyboard to start with if desired).

6. Suitable for fast typing (experienced user: 30-45 wpm, expert: 45-60 wpm).

7. Does not require full attention of the typist (no table of characters to watch continuously etc., very low eye strain).

8. Usable also in the dark or with eyes closed.

9. No need to have backlight on the keys (lower terminal power consumption).

10. In addition to text entry, includes all functions of the PC QWERTY keyboard.

11. Can be used to control all functions of a mobile phone.

12. Operates as a game controller (even as a pointer control in some applications).

13. Provides easy dialling functions.

14. Facilitates display browsing and menu selection.

15. To operate the keyboard, no desk or table is needed but the device can be used on a desk top as well.

16. Fully integrated and does not require separate tools (but easily combines e.g. with a stylus).

17. Seamlessly combines with a mouse or other pointing device.

18. Usable as a wireless personal PC keyboard if desired (by those preferring the new method).

19. Does not require special physical skills (trivial combinations/chords per hand).

20. Treats left and right-handed people equally (same hardware and software for both).

21. Does not necessarily increase the physical size of the mobile device.

22. GKOS typing skill does not interfere with QWERTY skill (different enough).

23. The GKOS keyboard and a wide screen can easily be combined on a single terminal.

24. The principle of mapping of the GKOS characters makes it possible to have a compact virtual keyboard on the display that can be used either with a stylus on the 6 keys on the back.

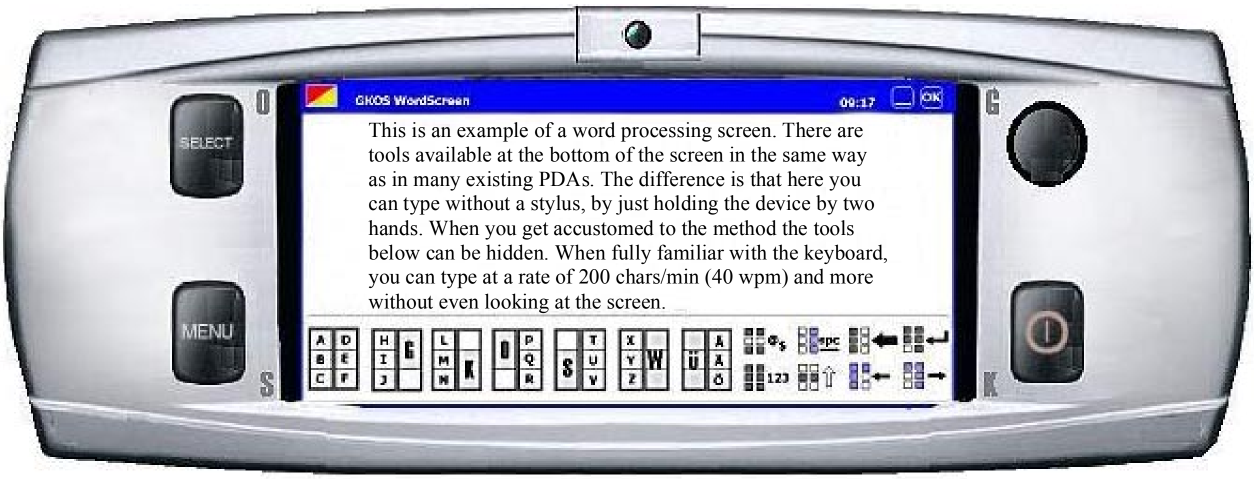
4.2 GKOS typing speed

After a small amount of practice, the typing speed exceeds that of the GSM number pad method (multi-tap). It is quite easy to reach a typing speed of 100 characters/min (20 wpm). When fully familiar with the keyboard, a typing speed of 200 characters/min (40 wpm) is obtained. Final expert typing speed can be around 300 characters/min (60 wpm). The speed naturally depends on the physical keyboard implementation and on the type of keys (light or hard to press etc.).

An experienced GKOS user can type much faster than the speed of multi-tapping or handwriting and therefore can e.g. take notes at a meeting, while being able to watch other things than just the terminal.

4.3 GKOS implementation examples

The dimensions of the GKOS keyboard depend on the form, size and weight of the terminal. In any case, there are six keys on the back of the equipment: three for the left and three for the right hand. Pinky fingers are used only for balancing the grip of the terminal and thumbs are kept on the sides of the front panel to manage the pointing device.



2038-37c

The left thumb can operate keys corresponding the left and right mouse switch (the upper “select” key and the lower “menu” key), and the right thumb then moves the pointer. This way both the mouse and keyboard functions can be fully controlled, having the fingers at the same position all the time. The virtual keyboard at the lower part of the display can be hidden after getting accustomed to the method.

The tiny mobile device below has a foldable GKOS keypad that can be used while holding the terminal or also when it is standing on a desk top (the keyboard folded out and the display facing the user).

Annex 14

Reconfigurable processors

1 Background

In recent years, new-type embedded processors have been developed. The outstanding feature of these processors is that their performance is not improved simply by increasing the clock speed or the memory bandwidth, but by providing an ability to implement a custom pipeline that allows custom instructions to be defined and executed based on what is required for the application.

There are two types of approach to realize user custom instructions. One is configurable processor and another is reconfigurable processor.

Configurable processors are promising in achieving great performance advance because they allow the creation of custom instructions and their execution units optimized for each application. However, configurable processors cannot be added new custom instructions after fabrication. In wireless communication systems, system specifications are often changed or they are slightly different from country to country. Therefore, processors that can realize variety kinds of instructions and can be added new custom instructions after fabrication are required to adapt those situations.

The latter approach fits this requirement. These processors include arrays of arithmetic-and-logic units (ALUs) and other operational units. These processors allow the connection configuration of the execution units to be changed to match the data processing flow required for the application. The amount of configuration information required is low because, unlike field programmable gate-arrays (FPGAs), the granularity of the circuit included in the processor is not small. This means that it is possible to provide multiple sets of on-chip configuration information and dynamically select one of these sets at individual clock cycles. These reconfigurable processors would be extremely fast if processing sequences could be mapped to the execution unit array.

The example shown in this Annex is an embedded processor which is based on the VLIW architecture [Suga *et al*., 2000; Okano *et al*., 2002]. This processor is designed for media applications and include an implementation of single instruction multiple data (SIMD) type media operation instructions. Therefore, they can perform fast media processing. The reconfigurable execution unit is developed so that general-purpose processor core can handle user defined custom instructions. The reconfigurable unit is not an ALU array type. By focusing to bit level data processing, it enables the processor to achieve higher performance with small area overhead.

The rest of this Annex reports the characteristics, functions, performance, and other nature of the reconfigurable unit designed for general-purpose processor as an example of reconfigurable processor.

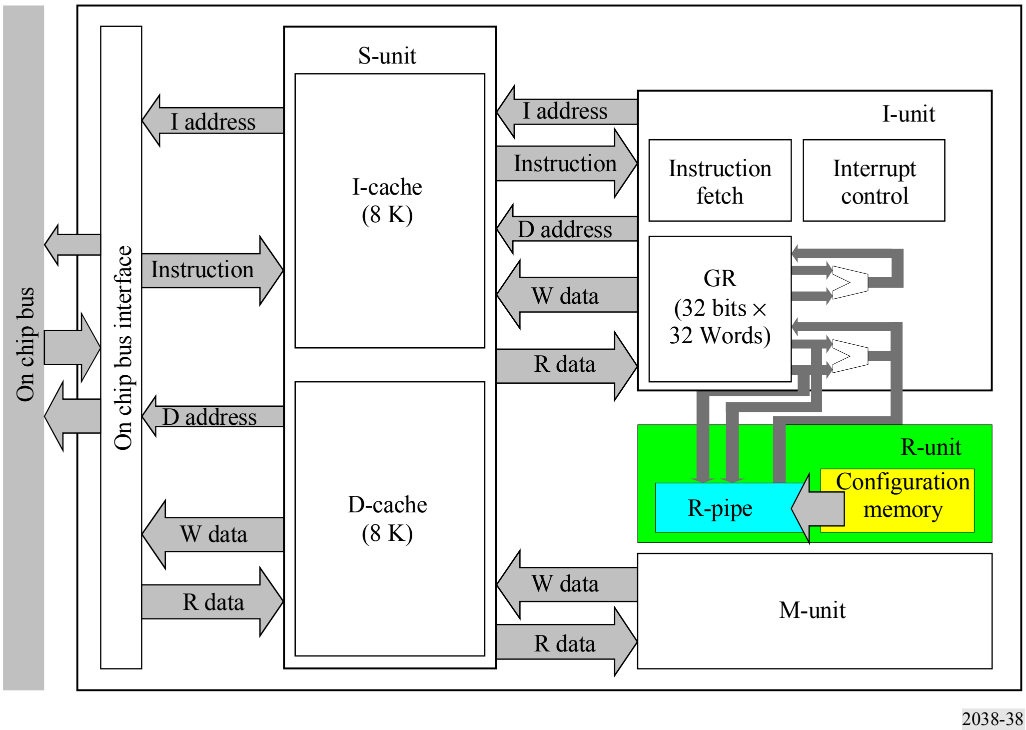
2 Processor architecture

2.1 Overall structure

Figure 38 shows an example of reconfigurable processor that includes the reconfigurable unit (R-unit). The integer unit (I-unit) consists of an instruction fetch block, various controllers, an integer register file (GR: 32 bits × 32 words), and two integer pipelines. The media unit (M-unit) consists of a media register file (FR: 32 bits × 32 words) and two media pipelines. The cache unit (S-unit) includes a two-way set-associative combination of two 8 K cache memories one for instructions and the other for data.

FIGURE 38

Block diagram of a reconfigurable processor that includes R-unit



R-unit is composed of an execution pipeline (R-pipe), which can dynamically change the internal configuration based on configuration information signals and configuration memory, which holds configuration information. One configuration within R-pipe is defined by 256 bits of information. Configuration memory is a 1 K RAM with 32 entries. It can hold configuration information represented by 32 types of custom instructions. Data in the configuration memory can be rewritten by using a configuration load instruction. This allows over 32 custom instructions to be defined and executed.

This type of general-purpose processor includes a powerful instruction set for media operation. Therefore, custom instruction processing performed within R-unit is not intended to further increase the media processing speed, but to improve the efficiency of processing at which this type of general-purpose processor is weak. More specifically, this type of general-purpose processor is efficient in word-level data processing, but is weak at bit-level data processing, such as encryption-related processing. R-unit is designed to perform bit-level data more efficiently.

2.2 Custom instructions

Instruction sets used in FR-V series processors assign certain instruction codes for custom instructions. This allows users to define custom instructions. Using these instruction codes, four types of instructions, CONFIGLOAD, EXEC, LUT, and RSRMOD, are defined.

CONFIGLOAD loads configuration information to the configuration memory. When the instruction translator incorporated in I-unit detects that the next instruction from the instruction fetch block is CONFIGLOAD, it converts the instruction into four double-word load instructions (8-byte load instructions) and issues them to the execution block. The data that is transferred from the cache is not written to the register file, but to the configuration memory. The loading of configuration information to the configuration memory is thus accomplished.

EXEC executes custom instructions. To ensure that there is no restriction on the types of custom instructions that can be defined, custom instructions are not assigned operation codes in fixed combinations. The operation code area includes a field in which configuration memory. entry can be specified. EXEC executes the custom instruction that is stored in the entry specified by that field. This means that, if two custom instructions are represented by the same operation code, but their configuration memory. entries hold different information, they will be executed as different custom instructions.

The LUT instruction specifies that information from the configuration memory be not used to define R-pipe configuration, but used as table memory. The 2^n (n = 1, 2, 3) bit in the data stored in the GR specified by the operation code is replaced with the 2^n bit included in the 256 bits of configuration information.

RSRMOD operates on special-purpose registers (SPRs) RSR0 and RSR1 which have been newly provided for R-unit.

If different custom instructions are defined although they perform nearly the same processing and differ only in certain quantities, such as shift extents, some configuration memory. area would be wasted. To avoid this waste, the following arrangement is used. Parameter information, denoted by sel and pos, is obtained from special-purpose registers. As a consequence, when the same custom instruction is passed one piece of parameter information at one time and another piece of information at another time, it will be executed as different instructions at the two times. The use of the configuration memory is thus made more efficient.

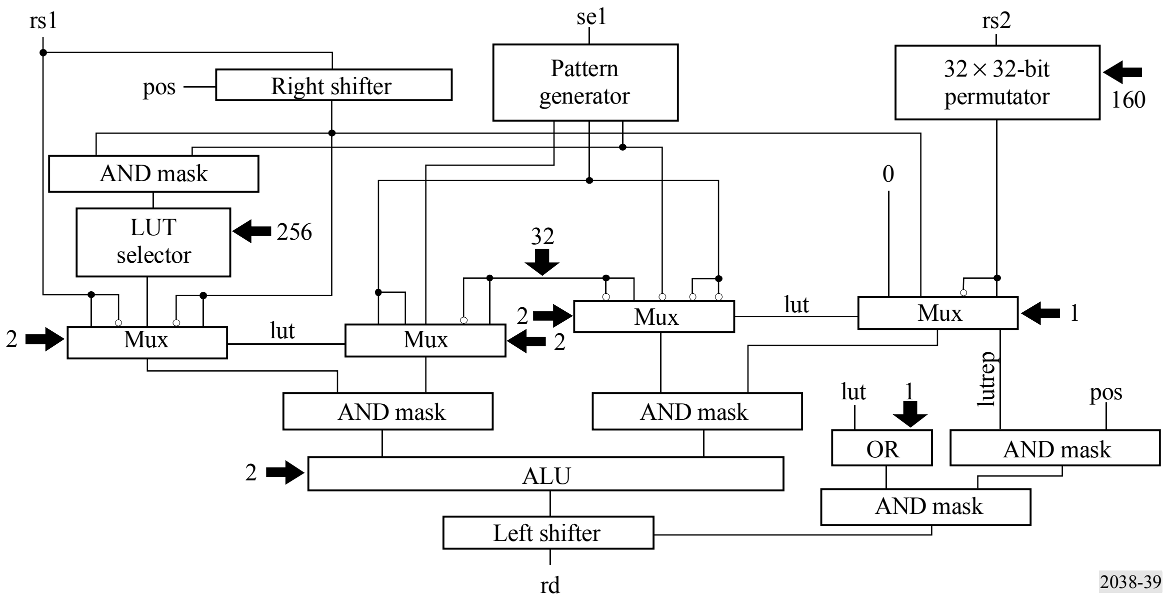
The sel and pos fields of the RSR register can be set with values by the RSRMOD instruction. However, an automatic updating function is also provided and can be used when the values of these fields need to be updated at regular intervals. This function eliminates the need for setting values each time using the RSRMOD instruction.

2.3 Structure of the reconfigurable unit

Figure 39 shows the R-pipe circuit. The bold arrows in the Figure indicate configuration information. The permutator shown at right top is a block that permutes the 32 bits of input data in a specified way. The pattern generator shown at centre top generates mask data. Actually, it can output three types of mask data based on the sel signal. The LUT selector is a circuit that extracts data when the LUT instruction uses the configuration information as a table. The other R-pipe components include shifters, multiplexers, AND masks, ALUs, among others. The internal structure of these component circuits and the connective relationship among the component circuits are defined by the configuration information. One configuration is defined by 256 bits.

FIGURE 39

Internal structure of R-pipe



Custom instruction definition examples will be provided later.

3 Performance improvement of DES application

The performance improvement was evaluated when an entire application is realized with this processor. The applications used were DES and Triple DES, which are widely used encryption algorithms.

DES is a block encryption algorithm, which encrypts 64 bits of input data and produces 64 bits of encrypted data. The encrypting procedure begins with initial permutation (IP) on the 64 bits of input data. The input bits are permuted. The next step is called the F function. It is repeated 16 times. Finally, inverse permutation IP^{-1} is performed to permute the bits again. Encryption is now complete.

Triple DES is an encryption algorithm in which DES is performed three times.

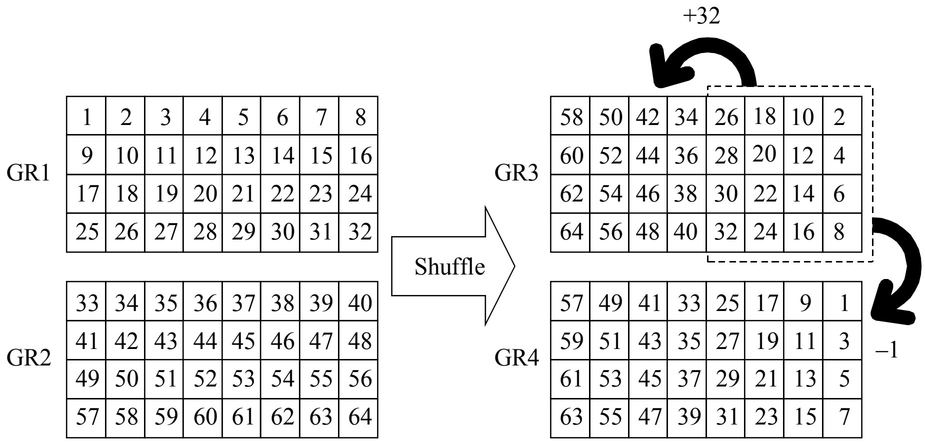
3.1 Initial permutation (IP)

This section explains how to implement the custom instruction that accomplishes initial permutation.

Initial permutation needs to achieve the bit exchange as shown in Fig. 40. The 64 bits of input data are stored in GR1 and GR2. Encrypted data is supposed to be stored in GR3 and GR4. The bit exchange appears random. Actually, however, all of the four 16-bit chunks follow the same exchange pattern. Notice the right 16-bit chunk on GR3. The set of each bit in this chunk that is 32-bit shifted is the left 16-bit chunk on GR3. Similarly, the set of each bit in the right 16-bit chunk on GR3 that is 1-bit shifted is the right 16-bit chunk on GR4. The left 16-bit chunk on GR4 is the set of each bit in the right 16-bit chunk on GR4 that is 1-bit shifted.

FIGURE 40

How bits are exchanged by the initial permutation (IP) step

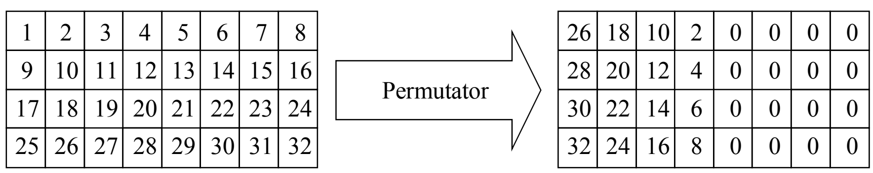


2038-40

Therefore if the permutator in R-pipe is designed to achieve the bit exchange as shown in Fig. 41, one IP custom instruction can be used for any initial permutation provided that it is combined with an appropriate shift instruction.

FIGURE 41

Permutator design for the initial permutation (IP) custom instruction

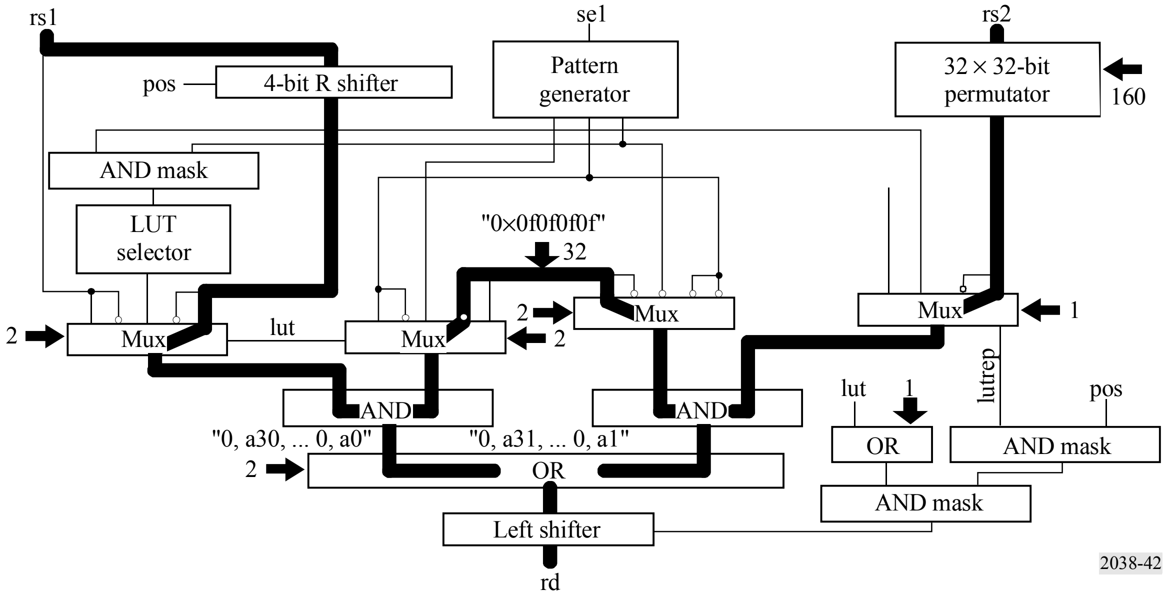


2038-41

More specifically, R-pipe is configured as shown in Fig. 42. The input from the first operand (rs1) is 4-bit right shifted by the right shifter, and then ANDed with 0xf0f0f0f0. The input from the second operand (rs2) is fed to the permutator, which performs bit permutation as shown in Fig. 41, and then ANDed with 0x0f0f0f0f. The OR of the two AND gate outputs is output as the operation result.

FIGURE 42

R-pipe configuration for the initial permutation (IP) custom instruction



The following six steps (instructions) are performed to achieve initial permutation using a custom instruction that complies with the above specification.

Step 1: Specify GR0 (zero register with all bits set to 0) for rs1 and GR1 for rs2, execute the custom instruction, and store the result in GR3.

Step 2: Specify GR3 for rs1 and GR2 for rs2, execute the custom instruction, and store the result in GR3.

The above two steps store the desired data in GR3.

Step 3: Shift GR1 to the right by 1 bit using a shift instruction, which is a general integer instruction and store the result in GR1.

Step 4: Similarly, shift GR2 to the right by 1 bit and store the result in GR2.

Step 5: Specify GR0 for rs1 and GR1 for rs2, execute the custom instruction, and store the result in GR4.

Step 6: Specify GR3 for rs1 and GR2 for rs2, execute the custom instruction, and store the result in GR4.

Steps 3 to 6 store the desired data in GR4.

As described above, six steps of executing one custom instruction can accomplish the desired initial permutation.

If an attempt was made to perform the same processing without using a custom instruction, the execution of dozens of instructions would be required because individual bits need to be manipulated for bit permutation.

3.2 Performance improvement

Other custom instructions for processing other than initial permutation, which is explained in § 3.1, are defined to ensure that DES processing can be performed at high speed. A total of 18 custom instructions were defined.

The register transfer level (RTL) simulation of triple DES processing shows the six times speed up when the R-unit was used compared with when the R-unit was not used.

With the method discussed above, custom instructions are defined to perform processing that would require a combination of large number of instructions if R-unit were not used. This means that the number of instructions required is reduced and, as a consequence, that the amount of instruction code becomes smaller. In case of triple DES processing, the program code size becomes less than one half when the R-unit was used compared with when the R-unit was not used.

4 Application to digital baseband processing

The R-unit version used this time is designed to perform bit processing at high speed. Since digital baseband processing requires extensive bit manipulation, it seems that this R-unit will also be effective for digital baseband processing.

For instance, in Bluetooth baseband processing, error correction is based on a bit repetition of code rate of 1/3 or a shortened Hamming code (15,10) of code rate of 2/3. The first method is simply to transmit the same bit three times in succession. If the permutator in R-unit is used, send data would be created efficiently. A speed increase by 20 to 30 times will be expected when compared with the case where the R-unit is not used. The second method can be made faster about four times if linear feedback shift register (LFSR) processing can be performed. A speed increase by about three times will also be possible for CRC processing, scrambling, and other processing.

It can be said that this reconfigurable processor is also applicable to 3G or systems beyond IMT-2000, because those systems require many bit level data processing.

5 Conclusion

The reconfigurable unit designed and described here is specialized for processing at which general-purpose processors are weak and is compact in that it consists of 1 kByte of memory and logic circuit with a size of about 20 K gates. When it is compared with a general-purpose processor core, it represents no more than 5% of the combination. The reconfigurable unit can achieve a great performance improvement on applications that require bit-level operation, such as encryption, although the circuit overhead involved is very small.

6 References

OKANO, H. *et al*. [February 2000] An 8-Way VLIW Embedded Multimedia Processor Built in 7-Layer Metal 0.11um CMOS Technology. IEEE International Solid-State Circuits Conference.

SUGA, A. *et al*. [February 2000] A 4-Way VLIW Embedded Multimedia Processor. IEEE International Solid-State Circuits Conference.

Annex 15

Multi-hop radio networks

1 Introduction

Within the possible evolution of IMT-2000 and systems beyond, highly increased data rates would enable a large variety of services and applications with different degree of QoS to be provided. On the other hand, high transmission rates require a wide bandwidth most probably available at high frequencies, resulting in shrinking of the cell area that one BS can cover. Instead of deploying many BSs to cover whole area, it would be better to enhance the coverage areas of the BSs by use of radio-relay techniques. This radio network architecture utilizing radio-relay functionality is called a multi-hop radio network, which provides a service area extension among various wireless access media such as the evolution of current cellular system, systems beyond IMT-2000 and non-cellular wireless media (for example, wireless LAN and Bluetooth) with high-rate data services and flexibility. From the mobile users' perspective, they would enjoy a ubiquitous communication environment and non-disconnected voice/data communication. A cellular network operator would save operating cost and wireless resources.

The relay functionality needed in multi-hop radio networks can be provided either by other users terminals – this technique is used in "ad hoc multi-hop networks" – or by fixed installed extension points – this technique is used in "structured multi-hop networks".

In order to realize the multi-hop wireless network, additional architectural elements and technologies must be considered and implemented in wireless networks, which are described in the following sections.

2 Technical aspects of multi-hop techniques

The realization of multi-hop techniques has impacts on different system aspects:

a) *Physical layer techniques:*

Because of the assumed high channel bandwidth per channel it is assumed that only single frequency repeater techniques come into question. The duplex method used has to provide a high isolation between transmission and reception path of the extension points (EPs) respective relaying terminals. TDD seems to be the best suited duplexing method to cope with the limitations caused by the bandwidth restrictions and the isolation requirements.

b) *MAC-layer techniques:*

The MAC layer controls the physical layer and establishment/release of transmission paths between mobile terminals, EPs and access points (APs). In particular it controls the duplex method used by the physical layer.

c) *Routing:*

Routing of data between mobile terminals (MTs), EPs and APs is an essential functionality in multi-hop networks. It has two main functions:

– selection of routes for the source destination pair (MT-AP via EPs/MTs); and,

– delivery of messages to their correct destination if a route becomes unavailable without QoS loss.

Different routing techniques can be used to fulfil these functions. They can be classified according to the method of control (centralized or distributed), to the dynamic behaviour (static or adaptive) and to the kind of information they base on (reactive or proactive).

d) *Radio resource management:*

The radio resource management comprises functions such as handover, power control, congestion control, packet data scheduling, etc. A typical case for terminals in multi-hop radio networks is that the terminal is out of range of a BS. Besides the signalling of control messages over several hops still under control of a BS, a distributed RRM approach seems to be feasible where RRM tasks are managed in a self-organizing manner. Alternatively, an overlay network, e.g. an existing long-range, narrow-band cellular network, might assist in an efficient manner, resulting in a hierarchical network structure. The overlay network can, e.g., determine the frequency band to be used, the amount of time a station is allowed to occupy the allocated frequency, and manage the connection and usage parameter/flow control.

3 Architectures of multi-hop wireless network

3.1 Ad hoc multi-hop wireless networks

Figure 43 depicts a possible architecture of the ad hoc multi-hop wireless network employing a cellular system and a wireless LAN system as a non-cellular network.

In a conventional network and a current network a wireless terminal is directly connected with a BS. However, in the ad hoc multi-hop network there are also wireless terminals that cannot be connected with a BS because of insufficient receiving level and channel occupation by other terminals. In that case neighbouring terminals relay data traffic of that user terminal to the BS.

3.2 Structured multi-hop wireless network

Figure 44 depicts a possible architecture of a structured multi-hop radio network. The employed technology could be a cellular system or a wireless LAN system as a non-cellular network.

Due to the limited coverage of the access points additional extension points are established in the target coverage area. The extension points are either directly or via other extension points and by means of radio links connected to an access point. the extension points possess a relay functionality that allows them to forward data/signals from/to mobile terminals, access points or other extension points. Structured multi-hop radio networks can use for EP-to-EP and EP-to-AP transmission the same radio technology as for MT-to-AP transmission (homogeneous multi-hop network). Alternatively the radio technology for MT-to-AP transmission can be different to those for EP-to-EP and EP-to-AP transmission.

FIGURE 43

Architecture example of ad hoc multi-hop wireless network

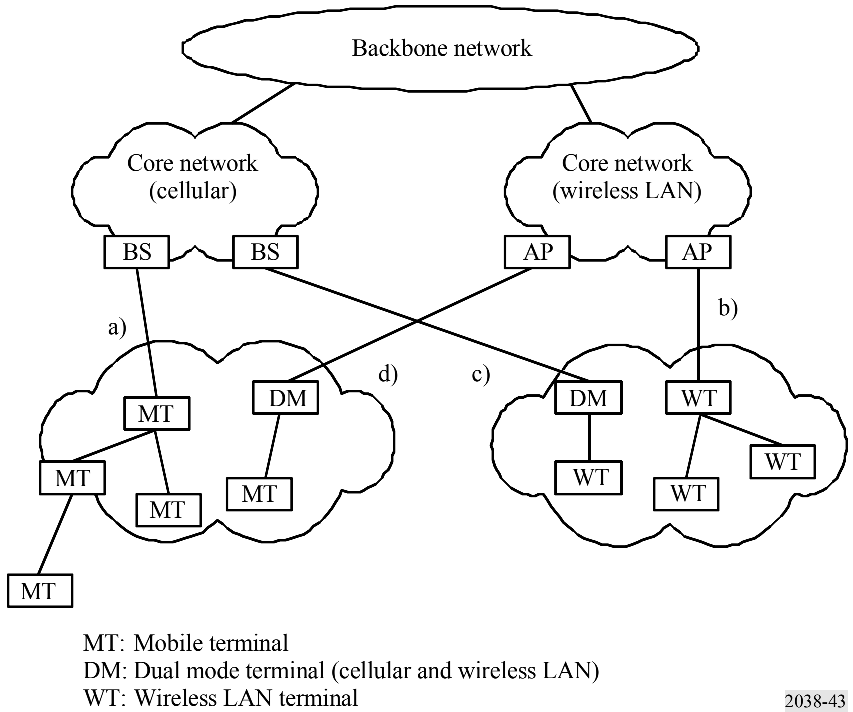
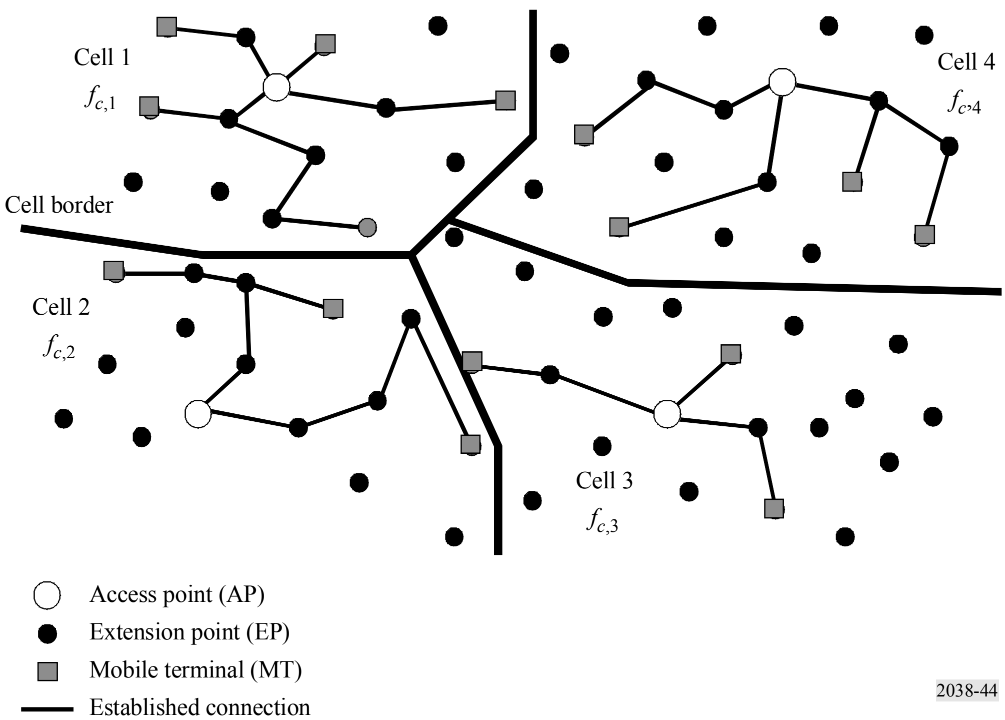


FIGURE 44

Architecture example of structured multi-hop wireless network



Since extension points only need a power supply (e.g. solar panels), their installation is easy and cheap. Since extension points are installed in well defined locations and typically with directional antennas, they increase the coverage of an access point in a reliable, predictable way.

4 Multi-hop access within cellular system

Current cellular system is based on a direct connection scheme between a cellular terminal and a base station. Once the cellular terminal moves out of the service area (e.g. inside a building or a tunnel), there is no means to communicate with the base station under a current cellular architecture.

Point a) in Fig. 43 shows one of the multi-hop access schemes for a wireless network composed in a cellular system. User terminals that have capability of relaying traffic generated from (or destined to) other terminals are used to compose a multi-hopped path between the target user terminal and the BS.

In structured multi-hop networks (see Fig. 44) the relaying of the traffic is provided by the extension points.

4.1 Media access control (MAC)

Cellular terminals (possible relay terminals) within a cell of a BS are controlled by the BS, however, cellular terminals out of the cell cannot receive control packets from the BS, resulting in a dead spot condition. In the multi-hop wireless network the relay terminal or the extension points would control the out of range terminals and deliver traffic to both directions: from the BS to the user terminal and the opposite direction.

In QoS and other performance perspective, wireless terminals including relaying terminals or extension points shall support media access method among appropriate neighbouring terminals/ extension points.

4.2 User identification

In a conventional cellular network each wireless terminal is distinguished by a unique identifier such as international mobile subscriber identity (IMSI), which is delivered to the network (for example AAA server) at the beginning of a connection setup phase and used for mobility control, accounting and other purposes.

Within the ad hoc multi-hop wireless network relay terminals must deliver such information along with user traffic to the BS and the network. The BS administers user stations along with information of the multi-hopped path and neighbouring terminal toward the user terminal. Alternatively, these kinds of functions can be realized by an existing overlay network.

Within structured multi-hop radio networks the surrounding extension points are permanently known by the access points after an initial set-up phase. The routing mechanism implemented in the access points and extension points selects the appropriate path between mobile terminal and access point.

4.3 Handover and routing

When a wireless terminal directly connected with a BS moves out of range, an appropriate handover mechanism would be initiated to maintain a continuous wireless connection. A wireless terminal must select appropriate relaying terminals or extension points and a route toward the BS. For structured multi-hop networks the terminal has just to select the extension point which provides the best link performance. The route is then given inherently by the given network structure. The handover results in a re-routing.

5 Multi-hop access within wireless LAN system

In this architecture possible user terminals could be notebook personal computers, PDA and other devices equipped with wireless LAN interfaces.

5.1 Media access control (MAC)

As a standard media access control protocol, carrier sense multiple access with collision avoidance (CSMA/CA) scheme can be adopted to the multi-hop system. Other suitable MAC schemes are based on TDMA, either slotted or un-slotted. Each wireless LAN terminal autonomously sends user data. At an access point user data are converted into a wired packet format, and are then delivered to the final destination.

Point b) in Fig. 43 shows multi-hop access within wireless LAN system. Current wireless LAN systems do not support multi-hop connections or packet relaying mechanisms, however, an ad hoc connection mode has been achieved for direct connection between two terminals.

5.2 User identification

In this architecture a wireless terminal is identified by the MAC address of the wireless LAN interface.

6 System interworking among wireless media

Points c) and d) in Fig. 43 show multi-hop connection by interworking among heterogeneous wireless systems in the case of ad hoc multi-hop networks. In this architecture dual mode terminals, for example equipped with a cellular interface and a wireless LAN interface, are used as a gateway to interconnect two systems.

In structured multi-hop networks as shown in Fig. 44 it is possible as well to have extension points that are equipped with both a cellular and a wireless LAN interface. In this case the transmissions between MT and EP and between MT and AP are performed preferably via a cellular interface, whereas the transmissions between EP and AP are preferably performed via a wireless LAN interface. Such an architecture has the advantage of:

– lower delay because simultaneous transmissions on the MT-EP and the EP-AP interfaces are possible;

– increased capacity in particular at the MT-EP/AP interface, since the EPs are operated as traffic concentrators.

7 Benefits of multi-hop radio networks

Multi-hop technology has advantages, that makes this technology important for enhanced 3G systems and systems beyond 3G. It alleviates, in particular, problems of high data rate cellular radio systems which are basically caused by the high transmission bandwidth and the expected operational frequency bands above 3 GHz.

Advantages are:

– This technology is a means to increase by orders of magnitude the coverage of APs, that is limited because of high path-loss and limited transmission power of MTs, EPs and APs.

– Since EPs can be realized as stand-alone entities (they need only a possibly solar-based power supply) with low infrastructure costs, structured multi-hop radio networks allow very economical coverage and capacity enhancements and permit an economical use of the frequency resources by operators. The usage of relaying terminals instead of EPs avoid the need of additional infrastructure at all. However, the coverage depends in this case on the availability of relaying terminals in the vicinity and on their capabilities. Therefore a combination seems to be useful, where a basic coverage is provided by a structured multi-hop network and relaying terminals may increase the performance in case of high number of wireless terminals.

– Fine and easy to achieve adaptation of the offered traffic capacity per area unit of an AP and the really needed traffic capacity per area unit is possible thereby increasing the spectrum efficiency.

– In case of structured multi-hop radio networks the radio network planning is facilitated. Radio coverage can easily be extended.

– The lower transmission power of the MTs allows longer battery times and lower electromagnetic radiation.

In the multi-hop connection scheme, there shall be two phases, multi-hopped path setup phase and user data transmission phase. In the former phase, multi-hopped path setup sequence and handover sequence that is triggered as a result of movement of user terminal or relaying terminals requires further investigation. Providing secure connection along the multi-hopped path is another issue to be solved. In a later phase, MAC protocols and relay mechanism that is to be installed on relay terminals are one of the key elements to be further investigated.

8 Issues for developing and operating multi-hop systems

Issues to be solved for developing and operating multi-hop systems include the following:

– Relay mechanism

– MAC protocol

– User/terminal authentication and accounting

– Security

– Multi-hopped path setup sequence

– Handover sequence.

Further research is needed to reduce the disadvantages of multi-hop networks as e.g. increased delay in case of transmissions via one or more EPs and potential waste of transmission capacity in case of MAC relay. Also the routing within a multi-hop network will be a key challenge.

In general, multi-hop technology can be considered as a useful complement for spectrum sharing, but could be used as well as an alternative to spectrum sharing. Currently advanced concepts – like cooperation between several extension points – are being investigated and promise higher performance.