Visions of 4G

by B. G. Evans and K. Baughan

As the Virtual Centre of Excellence in Mobile and Personal Communications (Mobile VCE) moves into its second core research programme it has been decided to set up a fourth generation (4G) Visions Group aimed at harmonising the research work across the work areas and amongst the numerous researchers working on the programme. This paper outlines the initial work of the group and provides a start to what will become an evolving vision of 4G. A short history of previous generations of mobile communications systems and a discussion of the limitations of third generation (3G) systems are followed by a vision of 4G for 2010 based around five elements: fully converged services, ubiquitous mobile access, diverse user devices, autonomous networks and software dependency. This vision is developed in more detail from a technology viewpoint into the key areas of networks and services, software systems and wireless access. It has been based upon a set of user scenarios that have been developed elsewhere in the Mobile VCE but which are summarised in the paper.

1 Introduction

Pick up any newspaper today and it is a safe bet that you will find an article somewhere relating to mobile communications. If it is not in the technology section it will almost certainly be in the business section and relate to the increasing share prices of operators or equipment manufacturers, or acquisitions and take-overs thereof. Such is the pervasiveness of mobile communications that it is affecting virtually everyone's life and has become a major political topic and a significant contributor to national gross domestic product (GDP).

The major driver to change in the mobile area in the last ten years has been the massive enabling implications of digital technology, both in digital signal processing and in service provision. The equivalent driver now, and in the next five years, will be the all pervasiveness of software in both networks and terminals. The digital revolution is well underway and we stand at the doorway to the software revolution. Accompanying these changes are societal developments involving the extensions in the use of mobiles. Starting out from speech-dominated services we are now experiencing massive growth in applications involving SMS (Short Message Service) together with the

Abbreviations

ANSI	= American National Standards Institute
API	= Application Programming Interface
BRAN	= Broadband Radio Access Network
CAC	= Connection Admission Control
CDMA	= Code Division Multiple Access
DAB	= Digital Audio Broadcasting (a standard)
DVB	= Digital Video Broadcasting (a standard)
EDGE	= Enhanced Data rates for GSM Evolution
ETSI	= European Telecommunications Standards Institute
FDD	= Frequency Division Duplex
GPRS	= General Packet Radio Services
GSM	= Global System for Mobile Communications
HAN	= Home Area Network
HAPS	= High Altitude Platform Stations
HIPERLAN	= High Performance Local Area Network (an ETSI standard)
HTML	= HyperText Markup Language
ITU	= International Telecommunications Union
IMT-2000	= International Mobile Telecommunications 2000 (the 3G ITU standard)
LAC	= Link Access Control
MAC	= Medium Access Control
MAP	= Mobile Applications Protocol
OFDM	= Orthogonal Frequency Division Multiplex
PAN	= Personal Access Network
QoS	= Quality of Service
SMS	= Short Message Service
TDD	= Time Division Duplex
UMTS	= Universal Mobile Telecommunications System (the 3G ETSI standard)
UTRA	= UMTS Terrestrial Radio Access
VAN	= Vehicle Area Network
VHE	= Virtual Home Environment
WAP	= Wireless Area Protocol
W-CDMA	= Wideband Code Division Multiple Access
WLAN	=Wireless Local Area Network
WML	= Wireless Markup Language
XDSL	= X (various) Digital Subscriber Line

start of Internet applications using WAP (Wireless Application Protocol) and i-mode. The mobile phone has not only followed the watch, the calculator and the organiser as an essential personal accessory but has subsumed all of them. With the new Internet extensions it will also lead to a convergence of the PC, hi-fi and television and provide mobility to facilities previously only available on one network.

The development from first generation analogue systems (1985) to second generation (2G) digital GSM (1992) was the heart of the digital revolution. But much more than this it was a huge success for standardisation emanating from Europe and gradually spreading globally. (The meaning of the name 'GSM' evolved from 'Groupe Spécial Mobile' to 'Global System for Mobile Communications'.)

However, world-wide roaming still presents some problems with pockets of US standards IS-95 (a code division multiple access [CDMA] rather than a time division muliple access [TDMA] digital system) and IS-136 (a TDMA variant) still entrenched in some countries. Extensions to GSM (2G) via GPRS (General Packet Radio Service) and EDGE (Enhanced Data rates for GSM Evolution) (E-GPRS) as well as WAP and i-mode (so called 2·5G) will allow the transmission of higher data rates as well as speech prior to the introduction of 3G.

Mobile systems comprise a radio access together with a supporting core network. In GSM the latter is characterised by MAP (Mobile Applications Protocol), which provides the mobility management features of the system.

GSM was designed for digital speech services or for low bit rate data that could fit into a speech channel (e.g. 9·6 kbit/s). It is a circuit rather than a packet oriented network and hence is inefficient for data communications. To address the rapid popularity increase of Internet services, GPRS is being added to GSM to allow packet (Internet Protocol [IP]) communications at up to about 100 kbit/s.

Third generation (3G) systems were standardised in 1999. These include IMT-2000 (International Mobile Telecommunications 2000), which was standardised within ITU-R and includes the UMTS (Universal Mobile Telecommunications System) European standard from ETSI (European Telecommunciations Standards Institute), the US derived CDMA 2000 and the Japanese NTT DoCoMo W-CDMA (Wideband Code Division Multiple Access) system. Such systems extend services to (multirate) high-quality multimedia and to convergent networks of fixed, cellular and satellite components. The radio air interface standards are based upon W-CDMA (UTRA FDD and UTRA TDD in UMTS, multicarrier CDMA 2000 and single carrier UWC-136 on derived US standards—see the list of abbreviations). The core network has not been standardised, but a group of three—evolved GSM (MAP), evolved ANSI-41 (from the American National Standards Institute) and IP-based—are all candidates. 3G is also about a diversity of terminal types, including many non-voice terminals, such as those embedded in all sorts of consumer products. Bluetooth (another standard not within the 3G orbit, but likely to be associated with it) is a short-range system that addresses such applications. Thus services from a few bits per second up to 2 Mbit/s can be envisioned.

For broadband indoor wireless communications, standards such as HIPERLAN 2 (High Performance Local Area Network—ETSI's broadband radio access network [BRAN]) and IEEE 802.11a have emerged to support IP-based services and provide some QoS (quality of service) support. Such systems are based on orthogonal frequency division multiplexing (OFDM) rather than CDMA and are planned to operate in the 5 GHz band.

Whereas 2G operates in 900 and 1800/1900 MHz frequency bands, 3G is intended to operate in wider bandwidth allocations at 2 GHz. These new frequency bands will provide wider bandwidths for some multimedia services and the first allocations have been made in some countries via spectrum auctions (e.g. in the UK, Holland and Germany) or beauty contests (in France and Italy). The opportunity has also been taken to increase competition by allowing new operators into the bands as well as extending existing operator licences. These new systems will comprise microcells as well as macrocells in order to deliver the higher capacity services efficiently. 3G and 2G will continue to coexist for some time with optimisation of service provision between them. Various modes of delivery will be used to improve coverage in urban, suburban and rural areas, with satellite (and possibly HAPS—high altitude platform stations) playing a role.

The story of the evolution of mobile radio generations is summed up in Fig. 1.

Already, as we move from 2G to 3G the convergence of communications and computing is central to the realisation of the new generation of services and applications. Digital technology enables dynamic adaptation of systems, and intercommunicating software embedded in networks and terminals allows efficient control of the new networks. This is accentuated as we move from 3G to 4G, extending the range and bit rate of services and bringing about the convergence of fixed, mobile and broadcast networks, service provision and terminal types.

In this paper we introduce the basic ideas and thinking behind the second phase research programme (1999–2003) of the UK's Virtual Centre of Excellence in Mobile and Personal Communications (Mobile VCE)[1] in the form of 'visions for 4G'. A Visions Group has been set up to produce and maintain an evolving picture of 4G and to communicate these ideas down to the work areas and researchers. The aim is to provide an umbrella vision to harmonise the research work in the various areas.

In the next section we explain the limitations of 3G systems and derive the drivers for 4G. In subsequent sections we then present 'the 4G vision' and some of the research challenges that this presents. The approach that is taken here is one of developing a technical vision. However, it is based upon likely user scenarios that have been developed within the Mobile VCE[2].

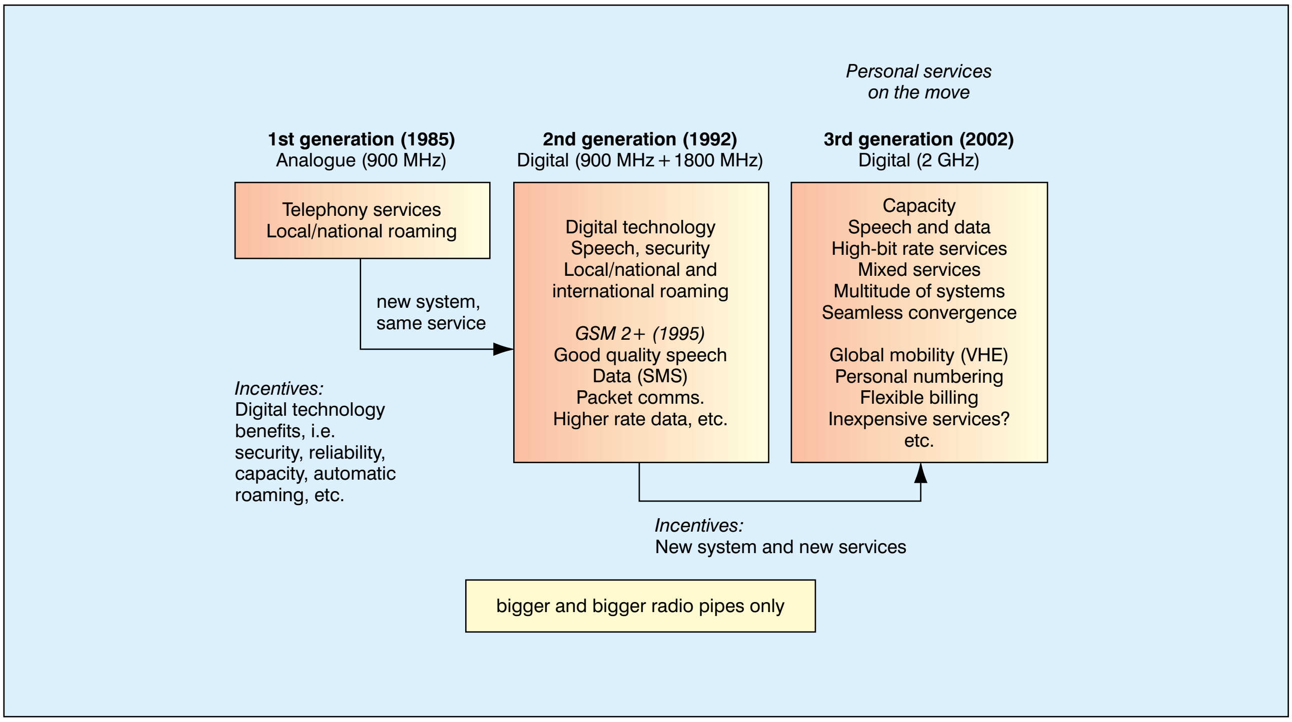


Fig. 1 Evolution of mobile radio generations

2 Limitations of 3G and drivers for 4G

From its basic conception to the time of roll-out took around ten years for 2G; a similar period will apply to 3G, which will commence service in 2001/2 and reach full deployment by 2005. Thus by 2010 it will be time to deploy 4G networks and, working backwards with the ten year cycle, it is clear that the year 2000 is appropriate to start with visions for 4G and a research programme aimed at the key issues. The Mobile VCE's second phase research programme has been constructed to meet this aim.

The starting point was to look at current trends. Here we see a phenomenal growth in mobiles with an estimated global user base that will exceed one billion by 2003. Already mobile communications exceed fixed communications in several countries and it is foreseen that mobile communications will subsume fixed by 2010 (fixed–mobile convergence will be complete). Currently short messaging is booming, especially among the younger generation, with averages of upwards of 100 messages per month dominating monthly bills. Business take-up of SMS via information services is also increasing and providing a start for mobile e-commerce, but this is currently very much limited by the bit rates available. This will be improved with the introduction of GPRS. At the time of writing we are at the beginning of the next stage on from messaging, which is to provide an efficient and convenient user interface to enable Internet content to be accessed on mobile devices—the 'internet to the pocket era'.

In Europe the WAP system (using Wireless Markup Language—WML) has been slow to gain market ground; in contrast, in Japan NTT DoCoMo's 'i-mode' system had over 10 million subscribers by summer 2000 and is picking up 50 000 new customers per day. Customers are already browsing the Internet, exchanging e-mail, conducting banking and stock transactions, making flight reservations and checking news and weather via HTML-based (HyperText Markup Language) text information on their phones. Java is expected to be available on i-mode phones soon, allowing the download of agents, games etc. and the introduction of location-based services. In Japan, the number of net phones has now passed the number of wired Internet customers and is setting the trend that others will surely follow when 3G opens up more bandwidth and improved quality.

Thus 3G will provide a significant step in the evolution of mobile personal communications. Mobility appears to be one of the fundamental elements in the evolution of the information society. As service provision based on 'network centric' architectures gradually gives way to the 'edge-centric' architectures, access is needed from more and more places at all times. But can 3G deliver?

It is true that 3G can support multimedia Internet-type services at improved speeds and quality compared to 2G. The W-CDMA based air-interface has been designed to provide improved high-capacity coverage for medium bit rates (384 kbit/s) and limited coverage at up to 2 Mbit/s (in indoor environments). Statistical multiplexing on the air also improves the efficiency of packet mode transmission. However, there are limitations with 3G as follows:

- Extension to higher data rates is difficult with CDMA due to excessive interference between services.
- It is difficult to provide a full range of multirate services, all with different QoS and performance requirements due to the constraints imposed on the core network by the air interface standard. For example, it is not a fully integrated system.

In addition, the bandwidth available in the 2 GHz bands allocated for 3G will soon become saturated and there are constraints on the combination of frequency and time division duplex modes imposed by regulators to serve different environments efficiently.

By the year 2010, one of the key enabling technology developments will be embedded radio—the widespread availability and use of the $1 radio chip, which will evolve from short-range wireless developments such as Bluetooth. Embedded radio will eventually become as common as embedded microprocessors are today, with perhaps 50 such devices in the typical home, the user being mostly unaware of their presence. As they interact, in response to the user arriving home for example, they will form a home area network (HAN). Similarly, such devices will be present in large numbers in vehicles (the vehicular area network, or VAN), in personal belongings (the personal area network, or PAN), in the public environment, etc. Such chips will serve as a means of short-range communication between objects and devices, offering capabilities for monitoring and control, in most cases without the knowledge or intervention of the user.

As a person moves between these environments such short-range links will allow their personal profiles and preferences to move with them, with the hotel room automatically configuring itself to their personal preferred temperatures, TV channels/interests, lighting etc. However, the integration of such links with wide-area mobile access will enable far more powerful service concepts, as mobile agents access this pervasive network of sensors and access information on the user's behalf to perform and even pre-empt their needs and wishes.

In the 1G to 2G transition, as well as a transition from analogue to digital we saw a mono-service to multi-service transition. From 2G to 3G, as well as a mono-media to multimedia transition we are also seeing a transition from person-to-person to person-to-machine interactions, with users accessing video, Internet/intranet and database feeds. The 3G to 4G transition, supported by such technologies, will see a transition towards a predominance of automated and autonomously initiated machine-to-machine interactions.

Such developments will of course be accompanied by ongoing evolution of already anticipated 3G services, such as:

- send/receive e-mail
- Internet browsing (information)
- on-line transactions (e-business)
- location-dependent information
- company database access
- large-file transfer.

These services in themselves represent an increase in requirements for accessing information, for business and commercial transactions, as well as for a raft of new location-dependent information services, all including significantly higher bit-rate requirements. There is a requirement for a mixture of unicast, multicast and broadcast service delivery with dynamic variation between application services both spatially and temporally. Above all, there is a demand for ease of user access and manipulation, with minimal user involvement—complexity hidden from the user—and intelligence to learn and adapt with use.

From the above it will be seen that 4G will need to be highly dynamic in terms of support for:

- the users' traffic
- air interfaces and terminal types
- radio environments
- quality-of-service types
- mobility patterns.

4G, then, must itself be dynamic and adaptable in all aspects, with built-in intelligence. Thus a 'software system' rather than a hard-and-fixed physical system is indicated. Integration, needed to reflect the convergence issues already mentioned, is also a key to 4G, in particular integration of the radio access and the core network elements, which must be designed as a whole rather than segmented as in the past. Key drivers to 4G will be:

- a multitude of diverse devices (distributed, embedded, wearable, pervasive)
- predominance of machine-to-machine communications
- location-dependent and e-business applications
- the extension of IP protocols to mobility and range of QoS
- privacy and security
- dynamic networking and air-interfaces
- improved coverage mechanisms
- improved and dynamic spectrum usage.

3 4G visions mapping to research topics

The Mobile VCE vision for 2010 is embodied in the five key elements shown in Fig. 2 and detailed as follows:

- *Fully converged services:* Personal communications, information systems, broadcast and entertainment will have merged into a seamless pool of content available according to the user's requirement. The user will have access to a wider range of services and applications, available conveniently, securely and in a manner reflecting the user's personal preferences.
- *Ubiquitous mobile access:* The dominant mode of access to this pool of content will be mobile, accounting for all voice communications, the majority of high-speed information services, and a significant proportion of broadcast and entertainment services. Mobile access to commercial and retail services will be the norm, replacing current practices in most cases.
- *Diverse user devices:* The user will be served by a wide variety of low-cost mobile devices to access content conveniently and seamlessly. These devices will commonly be wearable—in some cases disposable—and will normally be powered independently of the mains. Devices will interact with users in a multi-sensory manner, encompassing not only speech,

hearing and sight but also the other human senses, and biological and environmental data pertinent to the application. Special devices tailored for people with disabilities will be common place
- *Autonomous networks:* Underlying these systems will be highly autonomous adaptive networks capable of self-management of their structure to meet the changing and evolving demands of users for both services and capacity. Efficient and cost-effective use of the radio spectrum will be an essential element of their operation, and here, too, autonomy and self-management will be the norm.
- *Software dependency:* Intelligent mobile agents will exist throughout the networks and in user devices, and will act continually to simplify tasks and ensure transparency to the user. These mobile agents will act at all levels, from managing an individual user's content preferences to organising and reconfiguring major elements of networks.

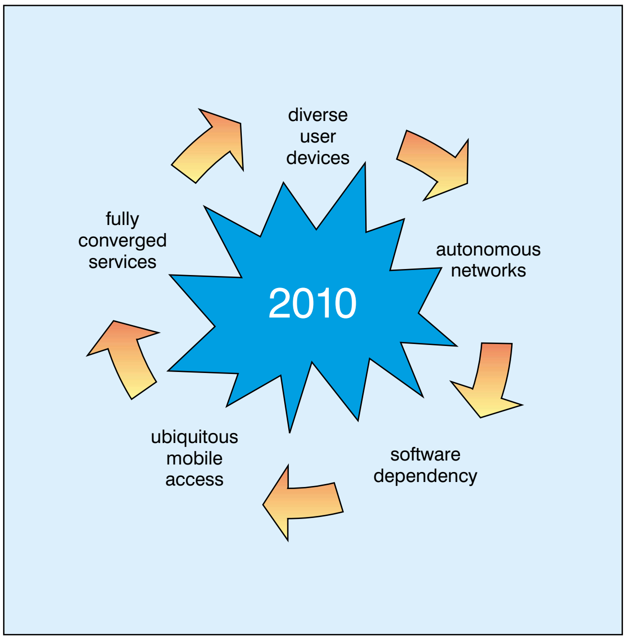


Fig. 2 Key elements of 4G vision

Analysis of the underlying technical challenges raised by the above vision and its five elements has produced three research areas, each having a number of distinct research needs. These form the basis of the Mobile VCE Phase 2 research programme. They are by no means exclusive, and have to some extent reflected the research skills available in our academic research teams. These areas are as follows:

(*a*) *Networks and services*
- evolved IP protocols for mobile systems
- *ad hoc* networking
- resource management for multimedia mobile systems
- mobility management in all-IP networks
- network dimensioning for multimedia, multirate systems.

(*b*) *Software based systems*
- mobile middleware to support resource management, reconfiguration and service demands
- software networks and radio
- software agents
- distributed mobile management systems
- integrity and security mechanisms.

(*c*) *Wireless access*
- flexible and reconfigurable radio architectures
- self planning and dynamically reconfigurable access
- intelligent and co-operative capacity enhancement
- dynamic spectrum allocation
- resource metric estimation
- HAPS and satellite integration.

4 Research challenges

In the previous section we have evolved a top-level vision, which we will now attempt to expand in order to draw out more details of the research challenges that such a vision produces. In doing this we hope at the same time to amplify the visions. It would be unreasonable to expect this vision to remain static—it should evolve and grow with time and research experience. What is given here is a glimpse of the footpath that will lead to the highway of 2010.

Networks and services

The aim of 3G is 'to provide multimedia multirate mobile communications anytime and anywhere', though this aim can only be partially met. It will be uneconomic to meet this requirement with cellular mobile radio only. 4G will extend the scenario to an all-IP network (access + core) that integrates broadcast, cellular, cordless, WLAN (wireless local area network), short-range systems and fixed wire. The vision is of integration across these network–air interfaces and of a variety of radio environments on a common, flexible and expandable platform — a 'network of networks' with distinctive radio access connected to a seamless IP-based core network (Fig. 3).

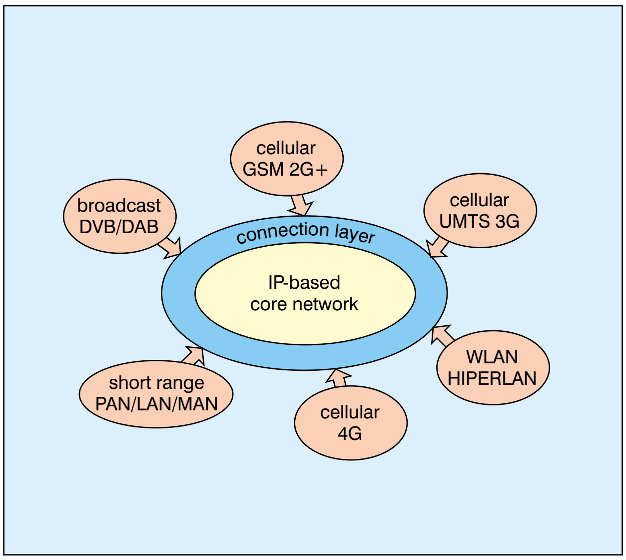
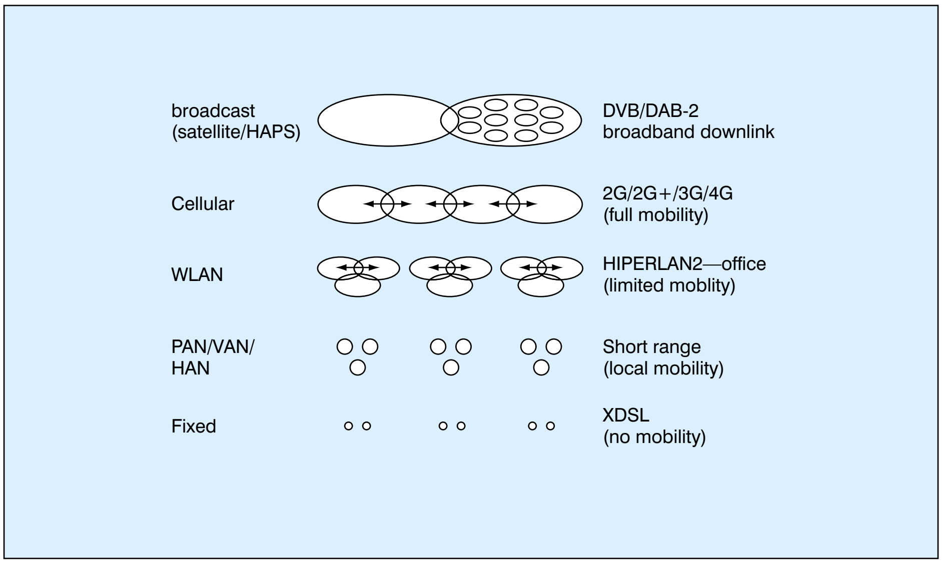


Fig. 3 Seamless connection of networks

Fig. 4 Vertical hierarchical networks (see the contribution of W. Mohr to Reference 3). It is assumed that DVB (Digital Video Broadcasting) and DAB (Digital Audio Broadcasting) will by 2010 have moved on to a new merged standard, DVB/DAB-2.



The functions contained in this vision will be:

- a connection layer between the radio access and the IP core including mobility management
- internetworking between access schemes — inter and intra system, handover, QoS negotiations, security and mobility
- ability to interface with a range of new and existing radio interfaces

A vertical view of this 4G vision (Fig. 4) shows the layered structure of hierarchical cells that facilitates optimisation for different applications and in different radio environments. In this depiction we need to provide global roaming across all layers.

Both vertical and horizontal handover between different access schemes will be available to provide seamless service and quality of service.

Network reconfigurability is a means of achieving the above scenario. This encompasses terminal reconfigurability, which enables the terminal to roam across the different air interfaces by exchanging configuration software (derived from the software radio concept). It also provides dynamic service flexibility and trading of access across the different networks by dynamically optimising the network nodes in the end-to-end connection. This involves reconfiguration of protocol stacks, programmability of network nodes and reconfigurability of base stations and terminals.

The requirement is for a distributed reconfiguration control. Fig. 5 demonstrates both internal node and external network reconfigurability.

For internal reconfiguration the functionality of the network nodes must be controlled before, during and after reconfiguration and compliance to transmission standards and regulations must be facilitated.

External reconfiguration management is required to monitor traffic, to ensure that the means for transport between terminals and network gateways (or other end-points) are synchronised (e.g. by conforming to standards) and to ensure that the databases/content servers needed for downloadable reconfiguration software are provided.

The research challenges are to provide mechanisms to implement internal and external configuration, to define and identify application programming interfaces (APIs) and to design mechanisms to ensure that reconfigured network nodes comply with regulatory standards.

An example of evolved system architectures is a combination of *ad hoc* and cellular topologies. A 'mobile *ad hoc* network' (MANET) is an autonomous system of mobile routers (and connected hosts) connected by wireless links. The routing and hosts are free to move randomly and organise themselves arbitrarily; thus the network wireless topology can change rapidly. Such a network can exist in a stand-alone form or be connected to a larger internet (as shown in Fig. 6).

In the current cellular systems, which are based on a star-topology, if the base stations are also considered to be mobile nodes the result becomes a 'network of mobile nodes' in which a base station acts as a gateway providing a bridge between two remote *ad hoc* networks or as a gateway to the fixed network. This architecture of hybrid star and *ad hoc* networks has many benefits; for example it allows self-reconfiguration and adaptability to highly variable mobile characteristics (e.g. channel conditions, traffic distribution variations, load-balancing) and it helps to minimise inaccuracies in estimating the location of mobiles.

Together with the benefits there are also some new challenges, which mainly reside in the unpredictability of the network topology due to mobility of the nodes; this unpredictability, coupled with the local-broadcast capability, provides new challenges in designing a communication system on top of an *ad hoc* wireless network. The following will be required:

- distributed MAC (medium access control) and dynamic routing support
- wireless service location protocols
- wireless dynamic host configuration protocols
- distributed LAC and QoS-based routing schemes.

In mobile IP networks we cannot provide absolute quality-of-service guarantees, but various levels of quality can be 'guaranteed' at a cost to other resources. As the complexity of the networks and the range of the services increase there is a trade-off between resource management costs and quality of service that needs to be optimised. The whole issue of resource management in a mobile IP network is a complex trade-off of signalling, scalability, delay and offered QoS.

As already mentioned, in 4G we will encounter a whole range of new multirate services, whose traffic models in isolation and in mixed mode need to be further examined. It is likely that aggregate models will not be sufficient for the design and dynamic control of such networks. The effects of traffic scheduling, MAC and CAC (connection admission control) and mobility will be required to devise the dimensioning tools needed to design 4G networks.

Software systems

We have already seen in the previous subsection that to effect terminal and network node reconfigurability we need a middleware layer. This consists of network intelligence in the form of object-oriented distributed processing and supporting environments that offer the openness necessary to break down traditional boundaries to interoperability and uniform service provision. The mobile software agent approach is an especially important building block as it offers the ability to cope with the complexities of distributed systems. Such building blocks may reside at one time in the terminal and then in the network; or they may be composed of other objects that themselves are mobile. Within the mobile system there exists a range of objects whose naming, addressing and location are key new issues. A further step in this development is the application of the Web-service-model rather than the client/server principle; recent industry tendencies show a shift towards this paradigm and XML (eXtensible Markup Language) is seen as the technology of the future for Web-based distributed services. However, this technology has yet to prove its scalability and suitability for future application in mobile networks.

In addition to the network utilities there will be a range of applications and services within 4G that also have associated with them objects, interfaces (APIs) and protocols. It is the entirety of different technologies that underlies the middleware for the new 4G software system.

The 'killer application' for 4G is likely to be the personal mobile assistant (PMA)—in effect the software complement to the personal area network—that will organise, share and enhance all of our daily routines and life situations. It will provide a range of functions including:

- ability to learn from experiences and to build on personal experiences, i.e. to have intelligence
- decision capability to organise routine functions with other PMAs and network data bases, e.g. diary, travel arrangements, holidays, prompts (shopping, haircut, theatre, birthdays, etc.)
- a range of communication modes: voice, image (with image superimposition via head-up displays such as glasses or retinal overlays), multiparty meetings (including live action video of us and our current environment), etc.
- provision of navigation and positioning information and

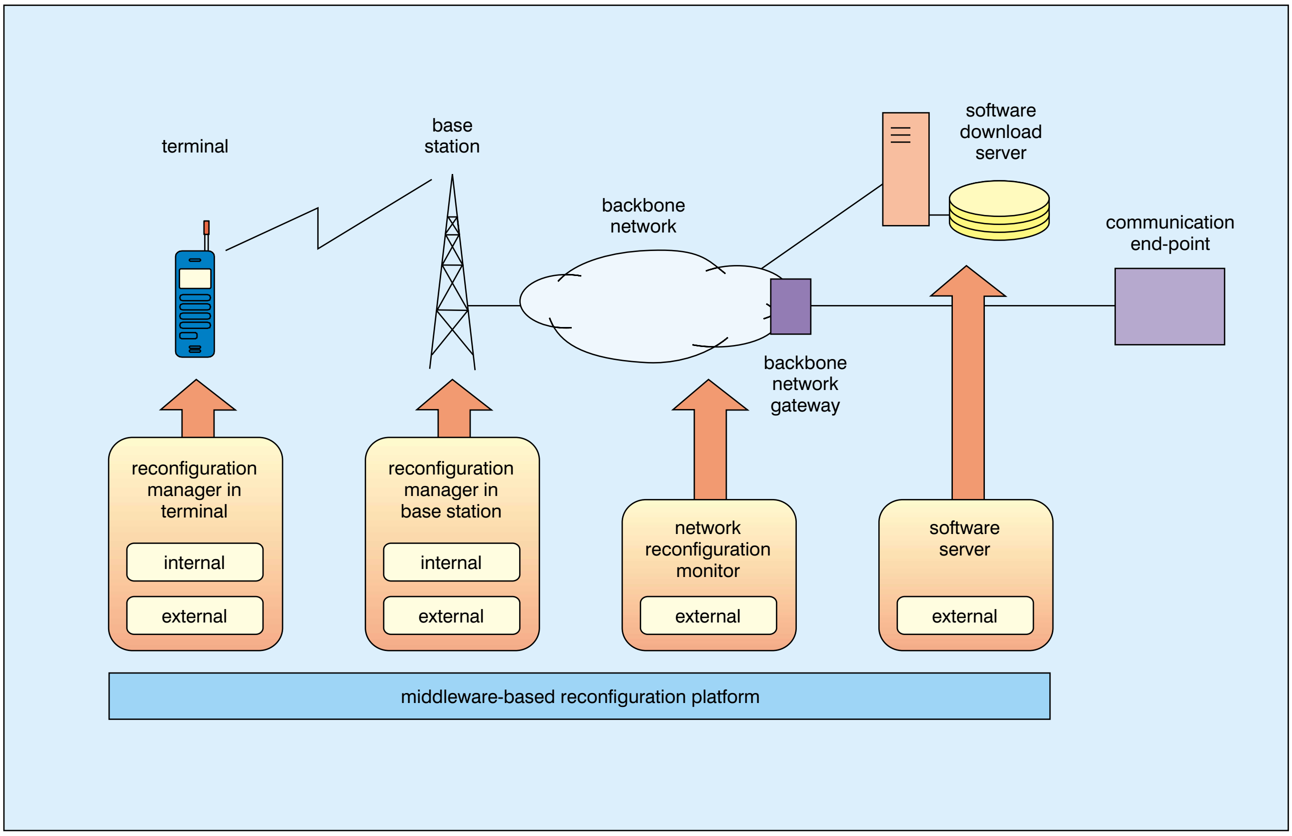


Fig. 5 Reconfiguration of mobile systems

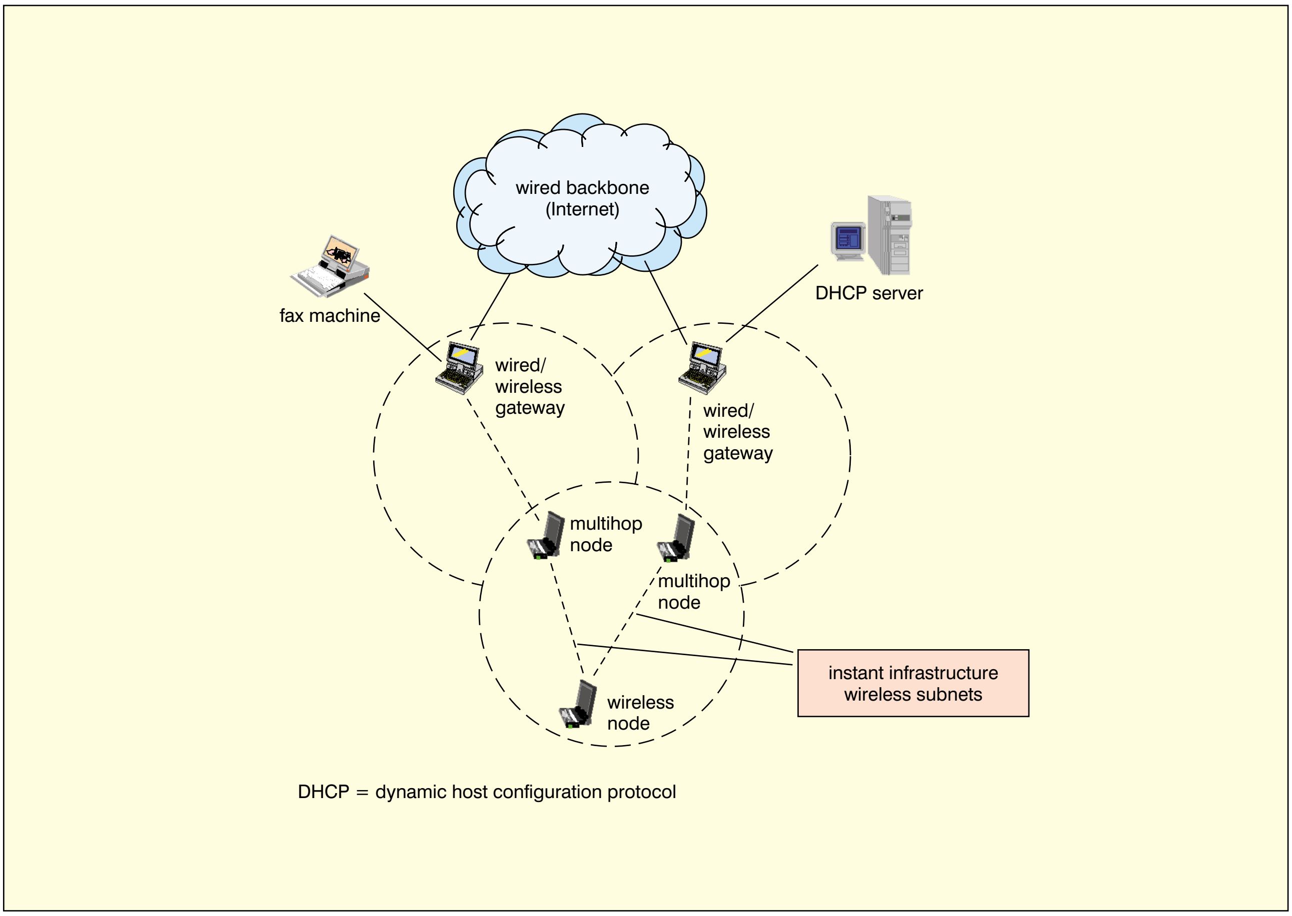


Fig. 6 An integrated *ad hoc* wireless system

thus of location-dependent services:

- detecting and reporting the location of children, pets and objects of any sort
- vehicle positioning and route planning, auto pilot and pedestrian warnings
- automatic reporting of accidents (to insurance companies, rescue services and car dealers)

- knowledge provision via intelligent browsing of the Internet
- e-business facilities for purchasing and payment
- health monitoring and provision of warnings
- infotainment: music, video and, maybe, virtual reality.

Of course the key to all this is 'mobility'—we need to have the 'PMA' whenever and wherever we are, and this places additional complexity on network and service objects and the agents that process them.

Specifically we need to consider what the metrics are that determine which objects follow the user. Some objects can move anywhere; others can move in some directions or within a constrained area. If they can move, how will the existing service determine if resources are available to support them in their new (temporary) home? Will they still be able to function? What kind of computing architecture and middleware platforms will be capable of supporting thousands, perhaps millions, of such objects?

Aspects of security pervade the whole of this area. Rules of authentication, confidentiality, scalability and availability must now be applied to objects that are continuously mobile. A whole set of conditions that are valid at one time and place may be invalid if transferred to another. Integrity and correctness issues must be considered when mechanisms that support applications are used in practice in the presence of other, distributed algorithms. For issues such as liveness, safety and boundedness—consistency, isolation and durability—execution semantics need to be evidenced for extension to the mobile environment.

Distributed management tools, in a complementary way, will allow a certain level of monitoring (including collection of data for analysis), control and troubleshooting. The management tools currently available do not encompass mobility efficiently and hence this is another important area of research.

The aim of the research in this area is to develop tools that can be used in 4G software systems. The following specific scenarios are being addressed in order to focus the issues:

- e-commerce, including microtransactions, share trading and internal business transactions
- home services, ranging from terminal enhancements (e.g. enhancing the display capabilities by using the TV screen as a display unit for the terminal) to security systems and housekeeping tasks
- transportation systems: Itinerary support, ticketing and location services are to be targeted in this area.
- infotainment on the move: This will demonstrate the

need for software and terminal reconfiguration and media-adaptation.
- telemedicine and assistance services: Emergency team support, remote/virtual operations and surveillance of heart patients are possible stages for this scenario.

This list of scenarios can be expanded arbitrarily and also into non-consumer areas (i.e. military and emergency services), however the preconditions for service delivery and demands on the network infrastructure remain the same: they will have to be adaptable to meet the user-requirements current in 2010. Support for these scenarios may be given by intelligent agents, which may represent the terminal within the network to manage the adaptations or customisations of the communication path. On an application or service layer they may additionally be used to complete business transactions for the user (e.g. booking a theatre ticket or a flight) or to support other services. Furthermore, distributed software entities (including the variety of models from objects, via agents, to the Web-service model) will encompass management and support for applications and services as well as for user and terminal mobility.

Wireless access

In the previous two sections we have looked at the type of network and the software platforms needed to reconfigure, adapt, manage and control a diversity of multimedia, multirate services and network connections. We have seen that there will be a range of radio access air interfaces optimised to the environments and the service sets that they support. The reconfigurability and the middleware flow through to the wireless access network. The radio part of the 4G system will be driven by the different radio environments, the spectrum constraints and the requirement to operate at varying and much higher bit rates and in a packet mode. Thus the drivers are:

- adaptive reconfigurability—algorithms
- spectral efficiency—air interface design and allocation of bandwidth
- environment coverage—all pervasive
- software—for the radio and the network access
- technology—embedded/wearable/low-power/high communication time/displays.

It has been decided within Mobile VCE not to become involved in technology issues or in the design of terminals. This is a large area, which is much closer to products and better suited to industry. The remaining drivers are all considered within the research programme.

It is possible, in principle, to increase significantly the effective bit rate capacity of a given bandwidth by using adaptive signal processing at both the base station and the mobile. In 3G systems adaptive signal processing has been restricted to the base station and so the challenge is to migrate this to the terminal and, most importantly, to make the two ends co-operative. Such techniques require close co-operation between the base and mobile stations in signalling information on channel quality, whilst making decisions and allocating resources dynamically. In addition, the capabilities of both ends of the link must be known reciprocally as the channel varies in both time and space. In order to optimise a link continuously, the wireless network must acquire and process accurate knowledge of metrics that indicate the current system performance, e.g. noise, inter- and intra-system interference, location, movement variations, and channel quality prediction. Such information and its accuracy must be passed to the higher layers of the system protocol that make decisions and effect resource allocation. The emphasis on the base station in 3G systems is obvious as this has the resources, real estate and capacity to implement the spatial–temporal digital signal processing needed for antenna arrays together with advanced receiver architectures. The challenge will be to migrate this to the much smaller terminal via efficient electronics and algorithms that will still allow a range of services and good call time. The availability of individual link metrics can also be used at a network level to optimise dynamically the network radio resources and to produce a self-planning network.

Arguably the most significant driver in the wireless access is the bandwidth availability and usage and whereabouts in the spectrum it will fall. Currently 3G technology is based around bands at 2 GHz, but limited spectrum is available, even with the addition of the expansion bands. The higher bit rates envisaged for 4G networks will require more bandwidth. Where is this to be found? The scope for a world-wide bandwidth allocation is severely constrained and, even if this were feasible, the bandwidth would be very limited. The requirements are thus for much more efficient utilisation of the spectrum and, perhaps, new ideas for system co-existence. If the bandwidth is fixed we need to seek a spectrally more efficient air interface and this involves a consideration of various multiple access, modulation, coding, equalisation/interference cancellation, power control, etc. schemes. In view of our previous comments it is clear that all components of this air interface must be dynamically adaptive. As the whole network is to be IP based this will mean extremely rapid adaptation on a burst basis. In 4G systems we need to accomplish this at much higher and variable bit rates as well as in different environments (indoor, outdoor, broadcast, etc.) and in the presence of other adaptive parameters in the air interface. In time-domain systems equalisers would need to be adaptive and this raises questions of complexity. For CDMA, systems could use multicodes and adaptive interference cancellation, which again raise complexity issues. Alternatively one could move to OFDM-like systems (as in WLANs), which offer some reduction in complexity by operating in the frequency domain but raise other issues, such as synchronisation. The choice of the air interface's multiple access scheme and adaptive components will need to be based upon the ease of adaptation and reconfigurability and on the complexity. There are also significant research challenges in this area

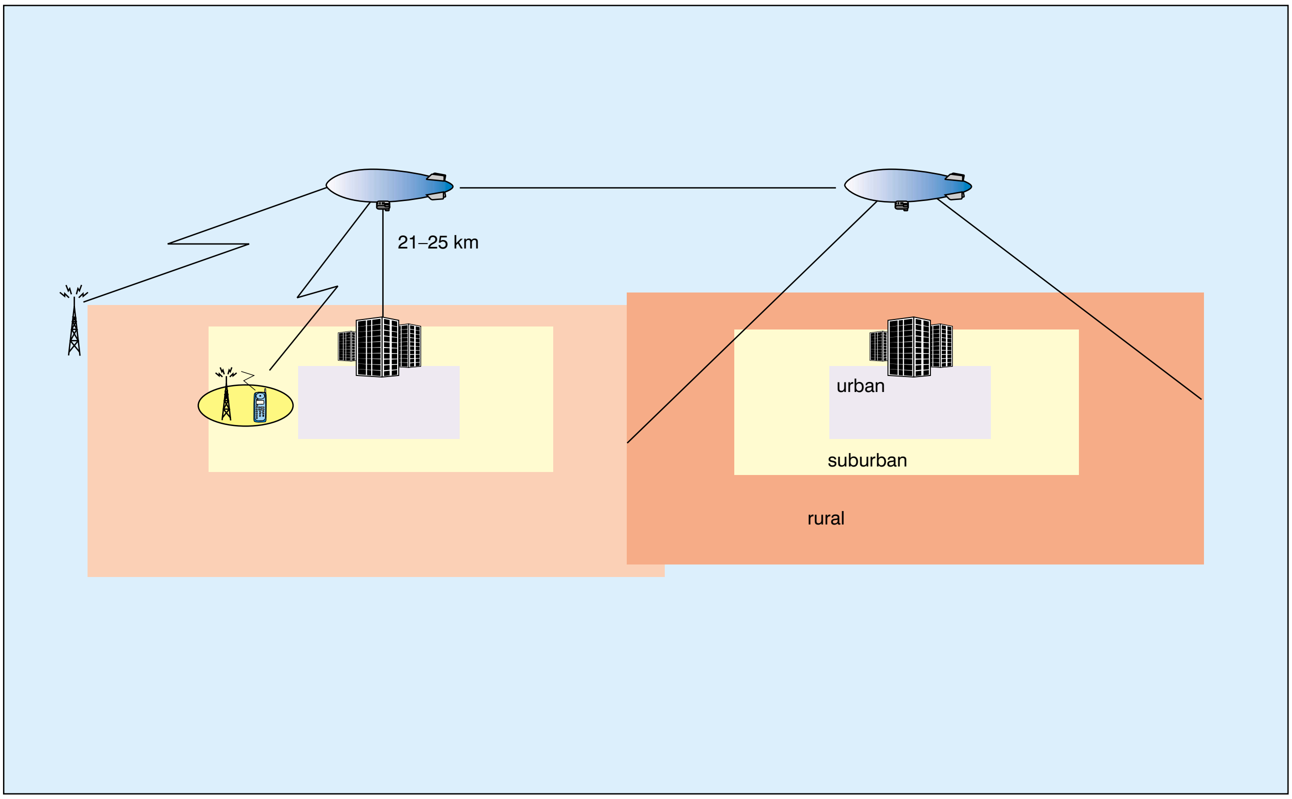


Fig. 7 HAPS providing integrated coverage

of flexible advanced terminal architectures that are not rooted solely in physical layer problems.

A further aspect of spectrum efficiency relates to the way in which regulators allocate bandwidth. The current practice of exclusive licensing of a block of spectrum is arguably not the most efficient. It would be much more efficient to allow different operators and radio standards to co-exist in the same spectrum by dynamically allocating spectrum as loading demands. Indeed, the higher bit-rate services may need to spread their requirements across several segments of spectrum. There would then be a need for a set of rules to govern the dynamic allocation of the spectrum—a self organising set of systems to maximise the use of spectrum and balance the load. Given the degree of co-operation and the processing already envisioned this should be a realistic aim.

A great deal of work on the characterisation of radio environments has already been performed in the 2 GHz and 5 GHz bands within the first phase of Mobile VCE's research, and spatial–temporal channel models have been produced. However, 4G systems will incorporate smart antennas at both ends of the radio link with the aim of using antenna diversity in the tasks of cancelling out interference and assisting in signal extraction. This implies that direction-of-arrival information, including all multipath components, will be an important parameter in determining the performance of array processing techniques. There is a need to augment models with such data for both the base station and the terminal station. A more open question is where to position the next frequency bands for mobile communications. An early study is needed here in advance of more detailed radio environment characterisations.

Coverage is likely to remain a problem throughout the lifetime of 3G systems. The network-of-networks structure of 4G sytems, together with the addition of multimedia, multirate services, mean that coverage will continue to present challenges. We have already seen that the likely structure will be based upon a hierarchical arrangement of macro-, micro- and picocells. Superimposed on this will be the megacell, which will provide the integration of broadcast services in a wider sense. Until now, it has been assumed that satellites would provide such an overlay, and indeed they will in some areas of the world. However, another attractive alternative could be high-altitude platform stations (HAPS), which have many benefits, particularly in aiding integration.

HAPS are not an alternative to satellite communications, rather they are a complementary element to terrestrial network architectures, mainly providing overlaid macro-/microcells for underlaid picocells supported through ground-based terrestrial mobile systems. These platforms can be made quasi-stationary at an altitude around 21–25 km in the stratospheric layer and project hundreds of cells over metropolitan areas (Fig. 7).

Due to the large coverage provided by each platform, they are highly suitable for providing local broadcasting services. A communication payload supporting 3G/4G and terrestrial DAB/DVD air interfaces and spectrum could also support broadband and very asymmetric services more efficiently than 3G/4G or DAB/DVB air-interfaces could individually. ITU-R has already recognised the use of HAPS as high base stations as an option for part of the terrestrial delivery of IMT-2000 in the bands 1885–1980 MHz, 2010–2025 MHz and

2110–2170 MHz in Regions 1 and 3, and 1885–1980 MHz and 2110–2160 MHz in Region 2 (Recommendation ITU-R M (IMT-HAPS)).

HAPS have many other advantages in reducing terrestrial real-estate problems, achieving rapid roll-out, providing improved interface management to hundreds of cells, spectrally efficient delivery of multicast/broadcast, provision of location-based services and, of course, integration. The research challenge is to integrate terrestrial and HAPS radio access so as to enhance spectral efficiency and preserve QoS for the range of services offered.

Software, algorithms and technology are the keys to the wireless access sector. Interplay between them will be the key to the eventual system selection, but the Mobile VCE's research programme will not be constrained in this way. The aim is to research new techniques which themselves will form the building blocks of 4G.

5 Conclusions

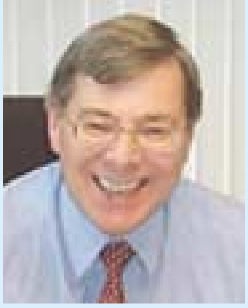
It is always dangerous to predict too far ahead in a fast-moving field such as mobile communications. Almost by definition the eventual 2010 scene will not match exactly that depicted in the 4G vision described herein. However, we feel that the key elements—fully converged services, ubiquitous mobile access, diverse user devices, autonomous networks and software dependency—will persist. Their exact technological manifestation in the eventual 4G system may depart from our initial ideas, which we outlined in Section 4, but the important thing is that we have created a scenario that points us in the right direction. Our 4G vision has been based on migration from 3G, on drivers and deficiencies of the latter system, and on perceived user trends. We have herein deliberately steered clear of 'user scenarios' that have been well researched by others[2,4]. The 4G vision is thus a technology vision, but it has been well informed by user trends and by such user visions as mentioned previously. We have also participated in an EU Visions programme (WSI)[3], which may be consulted for a wider perspective.

The 4G Vision is a living document which we intend to update and amend as time and knowledge progress. It will act as the umbrella vision to a large research programme and place in context the detailed research work that will take place in the various areas. In this respect it will help to continuously steer the research as it progresses and we hope, therefore, to make it more relevant and beneficial.

Acknowledgment

The work reported in this paper was undertaken by the Mobile VCE's 4G Vision Group, comprising:

Barry Evans (University of Surrey)
Walter Tuttlebee (Director, Mobile VCE)
Keith Baughan (Nokia Mobile Phones)
Mike Walker (Vodafone)
Peter Ramdsale (One2One)
Dave Thomas (Texas Instruments)



Professor Barry Evans received the BSc and PhD degrees in Electrical Engineering and Microwave Systems from the University of Leeds in 1965 and 1968, respectively. He was British Telecom Lecturer-Reader in Telecommunications Systems at the University of Essex from 1969 to 1983. In 1983 he was appointed to the Alec Harley Reeves Chair in Information Systems Engineering at the University of Surrey and in 1990 became the first Director of the postgraduate Centre for Satellite Engineering Research, which he built up to around 150 researchers and a spin-off company, Surrey Satellite Technology Ltd. Since 1996 he has been Director of the new Centre for Communication Systems Research at Surrey, which is now 100 researchers strong. In 1999 he was appointed Dean of Engineering Research across the three Engineering Schools at Surrey. He is a Fellow of the Royal Academy of Engineering in the UK and of the IEE. Barry Evans is editor of the *International Journal of Satellite Communications,* and author of three books and over 400 journal papers. He was a prominent member of the UK Foresight ITEC Panel, and is an adviser to OFTEL and to UK/MOD.

Address: CCSR, University of Surrey, Guildford GU2 7XH, UK.
Email: B.Evans@ee.surrey.ac.uk

Keith Baughan received BEng and PhD degrees from the University of Sheffield and went on to lecture for 3 years at the University of Aston. He then joined Plessey at the Allen Clark Research Centre, Caswell and during the following 22 years held a number of positions, including Manager of the Plessey Microwave Component business and finally General Manager of the Research Centre at Caswell. In 1990 Keith left Plessey to join Nokia Mobile Phones, where he was Vice President of Research from 1991 to 1998. During 1999 he set up a Group to identify and launch new businesses within Nokia and then, having completed this, returned to Research in January 2000. Keith has been Chairman of the Mobile VCE since its formation in 1996. He is an IEE Fellow.

Address: Nokia Mobile Phones, Summit Avenue, Farnborough, Hants GU14 0NG, UK.

References

1 TUTTLEBEE, W. H. W.: 'Mobile VCE: the convergence of industry and academia', *Electron. Commun. Eng. J.*, December 2000, **12**, (6), pp.245–248
2 IRVINE, J., *et al.*: 'Mobile VCE scenarios'. A document produced as part of the Software Based Systems work area within the Mobile VCE Core 2 research programme, September 2000. See http://www.mobilevce.com
3 'Wireless Strategic Initiative'. An EU project. See http://www.ist-wsi.org
4 Foresight ITEC Group: 'Visions'. See http://www.foresight.gov.uk

First received 2nd October and in final form 24th October 2000.