Future management of spectrum

S Olafsson, B Glover and M Nekovee

As the heterogeneity of wireless access technologies increases, dynamic allocation and utilisation of spectrum become ever more important. The traditional rigid allocation of spectrum for technology-specific usage is not suitable for the increasingly dynamic demand driven by the continuous emergence of technologies providing new services with different quality of service requirements. New spectrum management techniques and increasingly flexible spectrum usage rights are therefore called for. We discuss the limitations of present spectrum management techniques and explore some new alternatives including spectrum trading and opportunistic spectrum access.

1. Introduction

In the last thirty years or so wireless technologies have grown at an immensely fast rate. Progress in solid state technologies and radio-frequency hardware supporting rapid miniaturisation in handsets and other wireless components have led to exponential growth in radio systems. More recently we have seen the emergence of new communications systems offering personalised services to users on the move [1]. This trend, which is likely to continue, has resulted in rapidly increasing demand for spectrum. However, the lengthy and bureaucratic nature of the spectrum allocation process has proved unable to reallocate spectrum on a sufficiently dynamic basis to accommodate the changing needs of new and emerging technologies. Governments and regulators have recognised this trend, and the need to seek new and innovative methods to increase efficiency in the sharing of spectrum among all users.

Traditionally radio spectrum usage has been controlled by governments who have viewed spectrum as a scarce resource that needs to be controlled and allocated to users on a strict technology-defined basis. Today there is a general and broad agreement among various market players and user groups that a new multidisciplinary approach to spectrum management is required. This view is shared by regulatory bodies such as the Federal Communications Commission (FCC) [2] in the USA and the Office of Communications (Ofcom) [3, 4] in the UK. Both bodies have recently published a number of consultative documents on the ways in which spectrum is allocated and managed. The multidisciplinary approach will require inputs from technology, economics and regulation. Recognising the importance of rethinking the spectrum situation, operators, regulators and vendors alike have become active in reviewing new innovative methods for spectrum access and spectrum sharing.

The strict management of spectrum by regulatory bodies had its justification in the initial years of mobile telephony. Harmonisation and co-ordinated allocation of spectrum in Europe provided the basis for the roaming capabilities of homogeneous technologies like GSM. However, recent developments in wireless access technologies have cast serious doubt on the suitability of rigid and technology-specific allocation of spectrum.

Whereas harmonisation has clear benefits for immature and homogeneous network technologies, it is likely to act as a deterrent for the usage and take-up of new wireless technologies. Intelligent and adaptive devices that can sense the environment and the interference levels in the neighbouring spectrum would clearly benefit from the freedom to access the parts of the spectrum that are presently underutilised. These devices are currently precluded from usage of large parts of the available spectrum that had been allocated on a technology-specific basis to different operators or service providers. More flexible allocation of spectrum would both support investments in the wireless space and bring new and innovative products and services to the market more quickly, as technology-specific spectrum allocations would not be required.

The main disadvantage of a ‘once and for all’ allocation of spectrum is its inevitable rigidity as it cannot accommodate technological and demand uncertainty. Predicting the emergence of new wireless technologies, their spectral demands or user take-up is a notoriously difficult and unreliable task. Rigid and technology-specific allocations of spectrum will therefore inevitably lead to suboptimal usage. In fact, even though the GSM technology provides very efficient usage of spectrum under heavy load conditions, measurements conducted indicate that large segments of the GSM spectrum are severely underutilised over long periods of time [5]. There are also examples of spectrum having been allocated to technologies that were never implemented, such as ERMES [6] and TFTS [7].

Generally, allocating spectrum on the basis of technological expectations is very likely to result in failure. That approach amounts simply to applying centralised long-term planning procedures to the resource management of very dynamic and rapidly evolving technology. The auction of the 3G spectrum resulted in the allocation of very expensive spectrum to technology which has not followed the path it was then expected to take. The 3G spectrum is presently largely underutilised and presents perhaps the best example in recent times of the limitations of rigid resource planning.

The spectrum debate is currently very active and many different ways to implement more adaptive access to spectrum are being discussed. These include opportunistic device access to licensed spectrum slots that are presently not being used. Also, spectrum trading would enable permanent or temporary transfer of spectrum usage rights without the involvement of the regulator. Both approaches, individually or together, would allow for more flexible usage of spectrum avoiding artificial spectrum shortage caused by rigid allocation.

Some parts of the spectrum are designated licence-exempt, such as the ISM bands at 902 — 928 MHz, 2.4—2.483 GHz and parts of the range 5.725—5.78 GHz used by the IEEE802.11a standard (for a more detailed discussion of the UK Frequency Allocation Table 2004, see http://www.ofcom.org.uk/radiocomms/isu/ukfat/fat2004.pdf). Recent technologies such as Bluetooth and Wireless LANs operate in these bands. Despite the considerable interference in these bands, WLANs present the most popular and successful wireless access to the Internet today. Large-scale roll-outs of wireless access points based on the IEEE802.11 protocol are planned and have in some cases already been undertaken. The user experiences are generally good. The tremendous success of WLANs, in spite of intensively used spectrum and considerable interference, provides strong support for the case of making more unregulated spectrum available for wireless access. However, as the number of access points increases, interference may become a serious problem and reduce the quality of the user experience.

Developing more spectrally efficient technologies for usage in the licence-exempt domain has therefore become one of the most challenging research objectives in wireless communications [8]. This task requires considerable understanding of how heterogeneous devices, with diverse time and QoS requirements, co-operate in a distributed environment in optimally utilising the available spectrum.

One extreme vision of spectrum allocation is to make all available spectrum licence-exempt. Here the spectrum can be accessed by any cognitive device which satisfies certain etiquettes. The overall utilisation of the spectrum depends on the efficiency of the access protocols and other aspects of intelligent or adaptive behaviour like power control [9] and adaptive channel allocation [10].

Recent years have seen the emergence of a variety of new access technologies, in addition to the above-mentioned Bluetooth and WLAN. The coexistence of these technologies has led to heterogeneous network structure where devices require more flexible access to spectrum. This is certainly the case when one device can be equipped with many different access interfaces. As new access technologies continue to be developed and tested their successful deployment can only be secured within a framework of flexible and adaptive management of spectrum resources. Indeed, the devices should have the choice as to which parts of the spectrum to access for any given application and QoS requirements. This may be an opportunistic usage of regulated, but presently unused spectrum, or a short-term purchase of spectrum from a licence holder or a broker who manages an active portfolio of available spectrum. That portfolio would need to be updated and rebalanced in real time.

A central issue in the spectrum discussion focuses on the balance between licensed and licence-exempt spectrum. It is likely that in the short term this balance will mainly be determined by the regulator. In the long term, however, we believe that this ratio will be determined by a combination of regulation, technological progress and spectrum trading. Clearly, from an individual operator's point of view, owning spectrum provides security in terms of being able to guarantee wireless services to its customer base. However, it might also create unnecessary overhead costs and lead to large underutilisation of the spectrum.

How the radio spectrum will be managed in the future is rather unclear at the moment. Many ideas are making the rounds and to develop (some of) them successfully requires an interdisciplinary approach with the involvement of research areas such as radio technology, regulation and economics. Tools from economics will help with the pricing issues as well as providing measures for the economic efficiency of the achieved resource utilisation.

We also believe that the application of game theory will prove very useful in encouraging users to express their expected benefits from the usage of common resources. Viewing spectrum as a commodity that may be traded directly between its users, or by the intermediation of spectrum brokers, may also improve its utilisation. Finally, the economics perspective is required to develop some fairness measures for the allocation of spectrum to its users.

If new modulation, transmission or receiver technologies emerge that hugely improve the utilisation of spectrum it is likely that the economic considerations will be less important. Also, abundance of spectrum will have an impact on the way in which spectrum is acquired and whether it will be traded. Oversupply of spectrum would certainly reduce its value as a commodity. In that situation the spectrum commons [11] approach is the most likely scenario where access is permitted to all devices provided they use the accepted efficient access technologies.

The structure of this paper is as follows. In the next section we discuss the traditional command and control approach to spectrum allocation. Section 3 then discusses the need for more dynamic spectrum allocation procedures based on the increasing variety and complexity of coexisting access technologies. Section 4 considers the nature of spectrum inefficiency and some possible ways to break it. In section 5 we give a brief discussion of spectrum usage and property rights and ways in which these can be implemented and monitored. Following that we introduce the concept of cognitive radios and discuss their future role in enhancing spectrum efficiency. In section 7 we discuss the role of spectrum trading in future spectrum management, followed, in section 8, by the discussion of the digital dividend. In section 9 we discuss some future research challenges. We end with a summary and conclusions.

2. Traditional allocation of spectrum

Traditionally the spectrum has been controlled and allocated by governments on a licensing basis. Initially, allocations were on a strict technology basis, i.e. defined parts of the spectrum could only be used by specific radio technologies. Licence holders were not allowed to reallocate the spectrum to different technologies or to other users who might have better use for it. Furthermore, the allocation procedures were lengthy and bureaucratic opening up the possibility that the decision making process could be influenced by non-relevant factors.

In this section we discuss very briefly two different ways that have been used to assign spectrum to different users. Both methods have been used on different occasions in different countries, generally with results that have failed to please everybody.

2.1 Beauty contests

Submissions for spectrum are generally written ones, which may be supplemented by hearings and interviews where each contender for spectrum can present his or her case to an elected select committee. Like any other process that requires a judgement to be made on the basis of a written and/or oral presentation, beauty contests are potentially influenced by non-relevant factors and can therefore be prone to fraud and waste. They can also be subjective and discriminatory as they lack transparency. However, beauty contests are well suited in that they place a ceiling on the spectrum fees imposed [12].

2.2 Auctions

The assumption behind auctions is that potential bidders should have a clear notion of what value the spectrum has to them. Consequently, the spectrum should go to those who value it most. Accurate valuation of the spectrum, however, is notoriously difficult as revenue predictions are based on assumptions on spectrum utilisation, market take-up and the cost of future technologies. The auctions of the 3G spectrum led to hugely distorted prices [13] and had subsequent severely negative financial and operational impact on some of the major bidders. With hindsight it appears that the value of the 3G spectrum was grossly overestimated given that, at the time of the auction, the technology was not ready for imple-mentation and that, due to licence restrictions, operators were unable to make use of subsequent alternatives.

The main shortcoming of the auction process was the fact that it was technology specific. In the meantime, other technologies have emerged which are not able to operate in the presently underutilised 3G spectrum. It would therefore be more efficient to auction spectrum on a technology neutral basis.

Perhaps the issue is not so much what method is used for the primary allocation of spectrum but more importantly the need for dynamic and adaptive secondary reallocation of spectrum to other technologies and users. This would support the aim of allocating spectrum to those who appreciate it most, and therefore understand its real value. This mechanism can only be provided by the market where spectrum usage rights could be transferred from one user/ technology to another.

3. The need for dynamic spectrum allocations

It is becoming clear that strict command-and-control management of the spectrum is not suitable for the increasingly dynamic nature of spectrum usage. The regulators simply do not have sufficient understanding of the evolving spectrum requirements of new and emerging technologies. Consequently they will struggle to implement efficient spectrum allocation strategies. Regulators have realised the need to involve market makers in the wireless technology space, including service providers, equipment manufacturers, network operators, entrepreneurs, researchers and of course the users. Their involvement is vital for the development of more efficient spectrum allocation strategies, providing support for innovation and entrepreneurship as well as to the whole economic growth.

Clearly, technology-specific spectrum allocation and spectrum licensing have been very important and perhaps crucial for the development and the global take-up of certain wireless technologies. A good example is the European standard for GSM, which supported the fast take-up of mobile phones and also enabled the easy roaming across countries. Licensing agreements and strict technology harmonisation are essential at some stage of technical development. This is the case for ‘simple’ and homogeneous technologies like GSM. However, with the emergence of more flexible or heterogeneous access technologies the need for harmonisation is substantially reduced. For the new technologies an essential requirement is the ability to access spectrum on a dynamic basis and to be free from the technology-specific restrictions imposed by traditional spectrum allocation (see Fig 1).

Fig 1 As wireless access technologies develop from simple homogeneous technologies to more complex heterogeneous ones, the need for harmonisation and cross-technology standardisation is substantially reduced. In a world of complex heterogeneous devices spectrum liberalisation is required for improved spectrum utilisation.

The core feature of heterogeneous access technologies is that free choice of available spectrum supports the always-connected philosophy. Users simply opt for the access technology that best suits their temporary requirements and spectrum availability. In that process they try to maximise their utility function which presents their personalised requirements. This situation is schematically presented in Fig 2.

The user utility function depends on the application. For example, in the case of time-critical data transmission, a short delay or a minimum data jitter are important criteria. On other occasions low bit-error-rate may be far more important than time criticality. These facts will be reflected in the users’ utility functions, which they seek to optimise. In an environment of heterogeneous access technologies with heterogeneous requirements, we anticipate that devices will evolve strategies which improve their collective usage of the spectrum. This scenario requires that devices develop the ability to accurately model the radio environment and act according to that knowledge and their requirements (see Fig 3).

The realisation of devices with the capabilities displayed in Fig 3 are still outside the reach of present technology. However, ongoing research in industry and academia has already delivered some very promising results to build on. Before we discuss some of those in section 6, we will comment on some of the steps required to break the spectrum inflexibility so deeply rooted in the ways in which spectrum is presently managed.

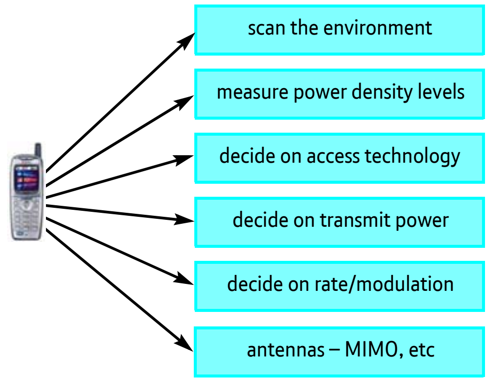

Fig 3 Devices need to scan the radio environment for spectrum occupancy. Following that they tune their actions by, for example, selecting spectrum for transmission, transmit power, rate control or even the usage of transmit antennas.

4. Breaking the spectrum inflexibility

At the core of the spectrum inflexibility is the close coupling between the three elements of ‘spectrum’, ‘ownership’ and ‘applications’. This tight relationship is underpinned by the present regulatory framework formally presented in Fig 4.

In order to implement more dynamic management of the spectrum, the interdependence of these three elements needs to be broken. There are various ways to achieve this, but only the following enablers will be mentioned here [14]:

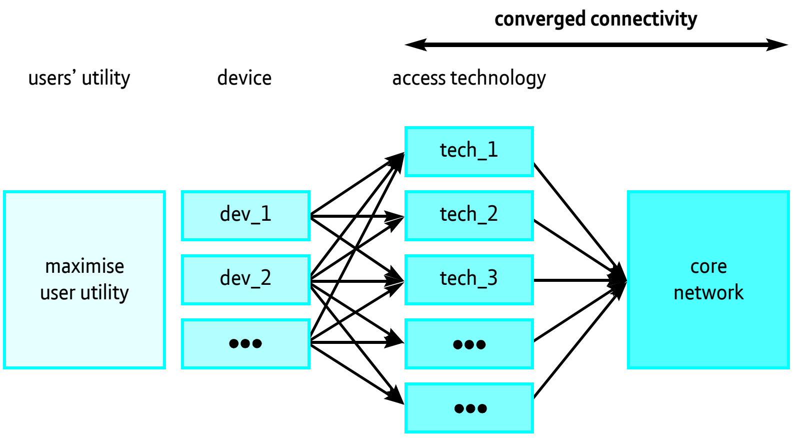

Fig 2 Different user devices can use different access technologies to suit their demands and adapt to present spectrum load distribution.

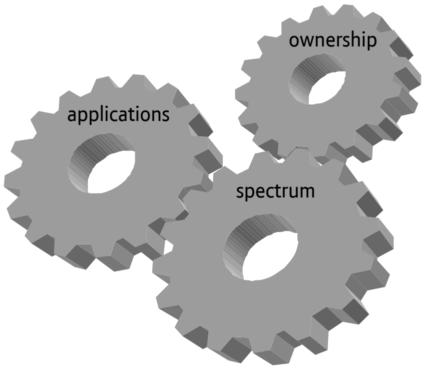

Fig 4 The tight regulatory knit between applications, ownership and spectrum is one of the core reasons for the inflexibility in spectrum usage as we know it today.

- a regulatory outlook that has confidence in the market, that does not try to predetermine technology winners and positively encourages innovation,
- regulatory, technical and market tools (and licences) that facilitate spectrum trading and change of use within time-scales compatible with the pace of innovation,
- effective ways of defining spectrum property rights and obligations that strike the right balance between the avoidance of interference and the flexibility to respond to the increasingly demanding wireless market,
- standards that are not specific to particular technologies and that support the right of the market to choose preferred solutions,
- attractive, low-cost, scalable and easy-to-deploy systems.

In spite of overwhelming support for increasing flexibility in spectrum allocation and spectrum usage [15], there are still some disagreements on how to strike the right balance between remaining restrictions and increased liberalisation. It is recognised that there will be trade-offs between the benefits of flexibility, including the ability to trade spectrum, and the disadvantages of possible fragmentation of the spectrum.

In the following section we comment briefly on recent thinking on spectrum usage (and property) rights (SUR). The focus will be on attempts to find the right balance between liberalised usage rights and the need to minimise interference.

5. Definition of spectrum usage and property rights

The original aim of issuing licences was to manage interference by placing transmission restrictions on each spectrum licence. Neighbouring spectrum users were able to use these limitations to determine the likely levels of interference to expect. In this way, system designers could plan deployments to limit the impact of interference from neighbouring systems. The nature of spectrum licensing has largely remained the same since its introduction. It has undoubtedly been a success but recent developments have started to challenge the fundamental idea behind spectrum licensing, i.e. the best way to manage interference is by limiting the actions of the transmitter. Therefore more innovative approaches to spectrum management have been sought.

One approach is to issue technology-neutral licences. Here, the holder of the licence can use the spectrum for any service that can be delivered by the available spectrum. However, to limit potentially harmful interference to other users (co- or adjacent-channel interference), limits need to be imposed on the power that can be transmitted inside or outside the allowed band. These limits are generally referred to as ‘EIRP’ (equivalent isotropic radiated power) limits or ‘spectrum masks’. This approach has several advantages, one of which is greater flexibility as no restrictions are put on technology or applications. Also, neighbouring devices have a clearer understanding as to what interference levels they can expect from each other. How the policing of compliance to the specified transmit power limits can be implemented is presently the subject of ongoing discussions between Ofcom and various interested parties.

5.1 Spectrum masks

Licensing has traditionally focused on restricting the power of transmitters through the use of spectrum masks (see Fig 5) to limit the power that may be transmitted at a particular frequency.

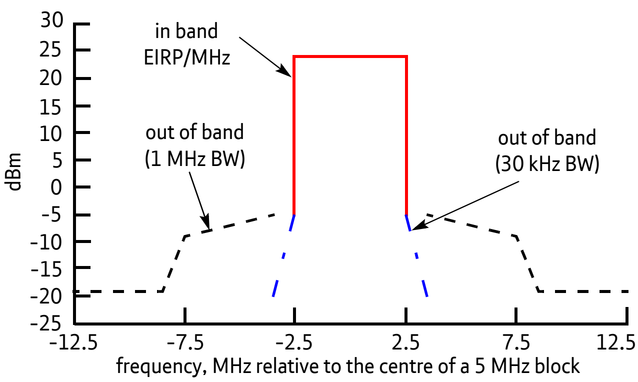

Fig 5 An example of a spectrum mask.
[Source: Ofcom — Award of available spectrum: 2500—2690 MHz, 2010—2025 MHz and 2290—2300 MHz]

One of the nice features of spectrum masks is that they are relatively easy to monitor and enforce. Device manufacturers can test all devices at the point of manufacture and, once deployed, testing devices for compliance in the event of dispute is relatively simple.

In addition to spectrum masks, licences have placed restrictions on both the technology and the services that can be offered using a particular band. These additional restrictions allow neighbouring spectrum users to make assumptions about the levels of interference they can expect.

5.2 The problem

Spectrum masks have worked as a licensing tool for many decades, but they are not without problems. The limits placed on the technology used or the services offered are widely thought to cause inefficiencies in spectrum allocation. Current operators are often unable to deploy new technologies in their existing spectrum, even if it is more profitable or more efficient. These limitations have prompted a review of how licences are defined.

One option is to maintain the use of spectrum masks, but drop the restrictions on services and technologies. This would limit the power of transmitters while allowing licence-holders to select their desired technology and services. However, there are problems with this approach.

In the following (admittedly extreme) example, we consider the case of a broadcast operator that has taken the decision to start operating a mobile service.

When the original broadcast network was established, the terms of the licence included a spectral mask for the transmitter. This spectral mask, coupled with knowledge of broadcast systems, allowed neighbouring users to make assumptions as to the nature of expected interference.

If the licensee decided to replace the single broadcast transmitter with a cellular mobile network, the structure of the network changes dramatically. Each cell contains a single base-station and multiple mobile devices. Each of these can act as a transmitter (see Fig 6). How does the original spectrum mask translate into spectrum masks for the mobile devices or the cellular base-stations? The same mask cannot be applied to all base-stations or the interference experienced by neighbours is likely to increase dramatically. As mobile devices move around, so the interference levels change. Neighbouring users are no longer able to predict expected interference.

This is not a problem for spectrum licences that are initially allocated to cellular or mobile networks as the licence accounts for these issues when the spectrum masks are determined. The problems only arise when a change of use is considered.

This is an extreme example and, in reality, situations similar to that described may not occur that often. But the use of spectrum masks is also problematic in other areas. Smart antennas and cognitive radios both promise increased spectral efficiency by dynamically exploiting the changes in the radio-frequency environment. Cognitive devices may limit their power as they approach the boundary with neighbouring users. Smart antennas may focus radiation away from the boundary. In both cases, a limited spectrum mask will prevent the licence-holder from fully exploiting their spectrum.

5.3 Proposed solutions

Given the problems identified with spectrum masks and the difficulties associated with finding an alternative, the search for a more robust definition of spectrum licences has yielded a number of possible solutions. The thrust of the proposals is to move the focus away from transmitter restrictions, to interference limits at the licence boundary. The principle behind these proposals is simple: Licensees are free to make the best use of their licence as long as their actions do not harm the experience of others. However, the practicalities of such a scheme are, as yet, unsolved.

One proposal put forward as part of Ofcom's consultation into spectrum usage rights [16] is the idea that

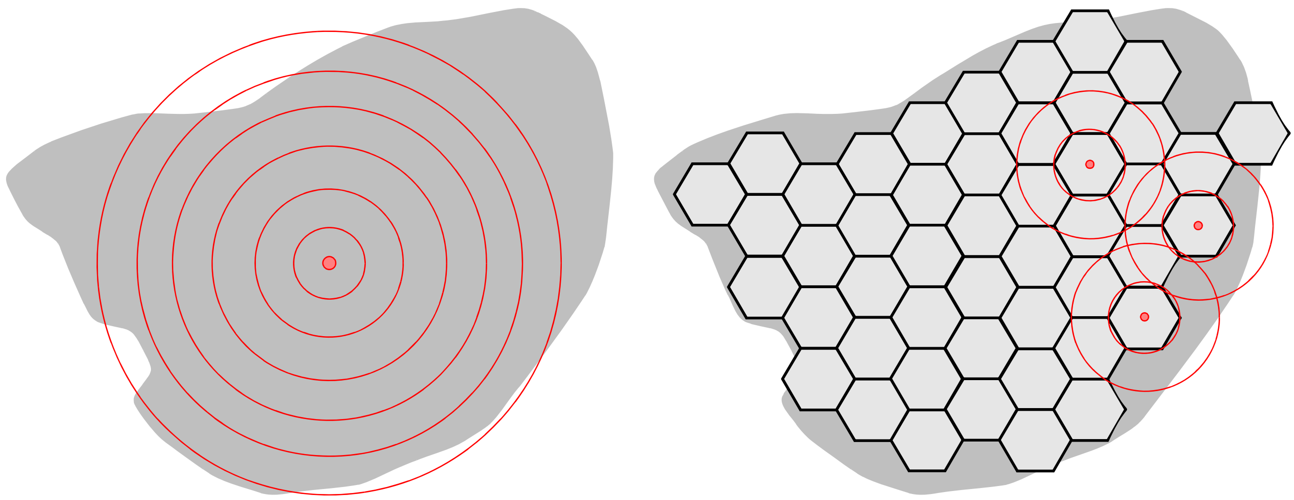

Fig 6 Diagram highlighting the issues regarding flexible use of spectrum. On the left, a single broadcast transmitter is used to reach all users within a well-defined region. The diagram on the right indicates how changing to a cellular-based system changes the nature of interference experienced at the boundary.

licences should be defined in terms of both geographical and spectral boundaries. Limits should then be set on the power flux density (PFD) that is allowed to flow through the boundaries. Such a scheme allows the licence-holder to deploy any network they choose, as long as the levels of interference experienced by neighbours do not exceed the licence limits.

While theoretically sound, limiting the PFD at the boundary of a licence results in increased complications for a licensee deploying or operating a network. It is no longer sufficient to test that the power output of each transmitter lies within the limits imposed by a spectral mask. The licensee is required to estimate the contribution of each transmitter to the flux through the boundary and ensure that the total does not exceed the licence requirements.

While it is generally acknowledged that the principle of SURs is a good one, disputes over the technicalities of implementation have cast doubt over the future of the proposals. There are also a number of issues with regard to dispute resolution and the challenges associated with identifying licence transgressions. The dominant position of incumbent networks, and their legitimate desire to protect their existing infrastructure, has resulted in significant dispute surrounding the details of practical issues such as measurement bandwidth, and the period and frequency of measurements. A recent Ofcom consultation [16] failed to resolve many of these issues and it remains to be seen how the regulator will take these proposals forward.

5.4 Negotiation of SUR parameters

As proposed, spectrum usage rights allow for the negotiation of licence parameters with neighbouring spectrum users. These negotiations may result in the net transfer of money between licensees. This directly links the concept of SURs with that of spectrum trading and offers the prospect of both increased flexibility and efficiency.

It remains unclear how negotiations would take place between commercial and non-commercial spectrum users and it may transpire that the regulator is required to negotiate on behalf of non-commercial users, or that commercial/non-commercial boundaries are fixed and non-negotiable — for more on spectrum trading, see section 7.

6. Cognitive radios

In this section we discuss briefly the concept of cognitive devices and their expected role in the future management of spectrum access.

6.1 What is a cognitive radio?

The term cognitive radio (CR) was first introduced by Joseph Mitola as '... the point in which wireless personal digital assistants (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer-to-computer communication to: (a) detect user communications needs as a function of user context and (b) to provide radio resources and wireless services most appropriate to those needs' [17]. Recently the term cognitive radio has been used in a narrower sense for radio systems that have adaptive spectrum awareness. The FCC, for example, defines it in the following way [18]: 'A cognitive radio (CR) is a radio that can change its transmitter parameters based on interaction with the environment in which it operates. The majority of cognitive radios will probably be software-defined radio (SDR) but neither having software nor being programmable are requirements of a cognitive radio'. In the rest of this section we shall focus on the latter form of cognitive radio, which is also known as spectrum-sensing cognitive radio.

Cognitive radios may operate opportunistically at frequencies that were originally licensed to an incumbent radio service or in available frequencies in unlicensed bands. An ideal cognitive radio operating in licensed bands acts as a spectrum scavenger. It first 'senses' the spectrum it wishes to use and identifies the presence, if any, of primary users. Based on that information, and regulatory policies applicable to that spectrum, the devices identify spectrum opportunities (frequency, time, space and code), and transmits in a manner that avoids the level of interference perceived by primary users. A cognitive radio may co-exist with the primary users either on a not-to-interfere basis or on an easement basis, which allows secondary transmissions as long as they are below an acceptable interference level.

6.2 Implementation issues of cognitive radios

While conceptually simple, the identification of idle spectrum in licensed bands has been shown to be a technologically difficult problem. To protect primary users against interference caused by cognitive radios, spectrum opportunities have to be identified correctly, and their usage has to be managed. Detecting the spectrum usage of primary users is by no means straightforward. First of all, different classes of primary user would require different sensitivity and rate of sensing for the detection. For example, TV broadcast signals are much easier to detect than GPS signals, since a TV receiver's sensitivity is tens of dB worse than a GPS receiver. In general, cognitive radio sensitivity should outperform primary user receivers by a large margin in order to prevent what is essentially a hidden node problem. The hidden node problem would occur, for example, when the cognitive radio is shadowed, in severe multi-path fading, or inside buildings with high penetration loss, while in a close neighbourhood there is a primary user who is at the marginal reception, due to more favourable channel conditions [19].

Cognitive radio is considered by many, including the FCC, to be an important new paradigm in wireless technology which offers the potential to make efficient and flexible use of unused spectrum, potentially allowing large amounts of spectrum to become available for future high-bandwidth applications. The attitude of the UK regulator, Ofcom, towards CR has been so far very cautious. In particular, Ofcom does not intend to allow unregulated opportunistic access to unused parts of the spectrum and cites a number of reasons why this entitlement has not been granted, and in particular the above-mentioned hidden node problem, which can result in harmful interference with primary users [20].

The development of cognitive radio is currently moving from the conceptual stage to early forms of implementation. One example of early life forms of CR is the proposed IEEE802.22 [21], an emerging radio standard for access networks, designed to operate opportunistically in TV broadcast channels. Because the IEEE802.22 group was only formed in 2004, nothing has been specified yet regarding particular functionalities of the PHYS/MAC layer. However, the IEEE, together with the FCC, is moving towards a centralised approach for resource discovery. Specifically, each access point (AP) would be armed with a GPS receiver which would allow it to measure its position. This information would be sent back to a centralised server (in the USA managed by the FCC), which would respond with the information about TV free channels in the area of the AP. However, there are proposals to allow local spectrum sensing without the requirement to be location-aware, where a cognitive AP would decide by itself which channels are available for communication. A combination of these approaches is also envisioned.

Recently, some authors have discussed ultra-wide bandwidth (UWB) technology as a suitable transmission technique for the implementation of cognitive radios [22]. UWB technology is characterised by spectral occupancy of over 500 MHz and can encode large amounts of information over a series of impulsive base-band signals using extremely low power, generally close to the thermal noise floor [23]. UWB signals are therefore essentially non-detectable by single antenna devices more than 10 m away. Due to their spread, UWB signals overlap with narrowband radio systems making coexistence and compatibility an issue. However, due to the relatively low increase in the noise floor for the narrowband radio systems, this is unlikely to become a real problem. To address this issue, the FCC has released guidelines for UWB radio masks opening the way for the coexistence of UWB and traditional radio devices [24]. Substantial modelling and measurement work on the coexistence of UWB and narrowband systems is still required.

6.3 Future architectures for cognitive radio networks

So far much of the work on cognitive radios has focused on the operation of a single cognitive radio in a dynamic spectrum landscape. In reality, however, cognitive radios will operate in a networked architecture. Our own research at BT has explored this less-studied aspect of cognitive radio systems. In particular, we have identified a number of potential architectures for the operation of future systems involving cognitive radios and have critically examined their implementation issues [25] as well as identifying their role in a paradigm shift from static spectrum management to dynamic spectrum access and management [25, 26].

The most pragmatic, albeit less ambitious, architecture is a centralised approach in which dynamic spectrum access by cognitive radios takes place exclusively within a section of spectrum reserved by regulatory bodies. The access to this section of the spectrum is then managed by spectrum brokers and request for spectrum can be generated either by network operators on behalf of a user's cognitive devices or directly by end-user devices which participate in the spectrum-leasing process.

An alternative architecture for cognitive radio networks is a collaborative and decentralised approach in which groups of cognitive radios form a user group to co-ordinate their communications activities. Each member of this group will sense the available spectrum pool, which is divided into sub-channels. Separate control channels are then employed by each group for co-operative spectrum sensing, in order to mitigate the above-mentioned hidden node problem, and for group management. A pair of cognitive radios selects a set of sub-channels based on the estimated channel gain and the user's QoS requirement, without the need for any central control.

A final possible architecture is the fully autonomous cognitive radio network. This presents the most ambitious form of CR architecture and potentially challenges wireless network operators. The most notable example of this approach is the Defense Advanced Research Project Agency (DARPA) next generation (XG) programme. In autonomous CR networks each device operates independently, attempting to optimise its own operation by adapting its transmission and other characteristics in response to the radio activity of other cognitive radios and primary users. In order that local and decentralised actions of such cognitive radios result in optimal utilisation of vacant spectrum, while also avoiding the so-called tragedy of spectrum commons, a set of spectrum etiquette rules needs to be implemented according to which cognitive radios operate. How spectrum etiquette rules should be formulated, and whether, given a set of etiquette rules, autonomous cognitive radio networks would settle down to an optimal equilibrium state, are open research questions which also have important regulatory and techno-economic implications. Some of our future research at BT aims to address these fundamental questions using tools from non-cooperative game theory and large-scale modelling and simulations of interactive particle systems.

7. Spectrum trading

As the allocation process for spectrum (mainly through auctions) does not allow for its flexible utilisation, users may not have access to the parts of the spectrum that best suit their present requirements. These shortcomings could be rectified by the introduction of spectrum trading. Active implementation of spectrum trading would also serve as an important component in effective interference management.

Spectrum trading could take many different forms and happen on different time-scales. In its simplest form spectrum trading refers only to temporary or permanent selling of spectrum licences. A spectrum licence holder may want to provide services and use it for technology which operates better in bands for which it has no licence. Another licence holder with complementary interests may be willing to enter a trade where licences are essentially swapped with perhaps some net flow of money between the trading parties. The actual cost of the transaction will depend on the supply and demand relationship between the two spectrum bands being exchanged. An example for how spectrum trading has had an impact on the structure of the wireless space is Nextel's purchase and aggregation of regional licences to create a national mobile network in the USA.

As technologies develop and emerge on ever faster time-scales it is likely that trading will play an increasingly important role in the efficient utilisation of spectrum. This process will be accompanied by rapidly developing means to dynamically price the use of spectrum.

Spectrum trading on ever shorter time-scales will eventually create the basis for a 'real-time' and 'liquid' market in spectrum. Most importantly, this will provide the basis for transparency in pricing, which will replace the non-transparent bilateral price negotiations typical for spectrum trades so far. In these liquid markets users can acquire the spectrum that best suits their needs and pay a price determined by the market. It is likely that the resulting economic efficiency will support technical efficiency in general and effective interference management in particular.

Until very recently the main quoted criteria for allocation of spectrum have been spectrum efficiency and interference management. Modern cellular systems utilise the available spectrum very well under heavy load conditions. However, there is strong evidence that parts of the licensed spectrum are severely underutilised [5]; and this may happen when at the same time other parts of the spectrum are overloaded, causing performance deterioration to its users. We believe that a better dynamic utilisation of the available spectrum can be achieved by it being frequently traded. Of course, trading need not mean terminal sale or purchase of spectrum, but simply the sold or purchased right to access spectrum when and if the need arises. These transactions can be in the form of derivative securities, like futures, forwards, swaps or options [27].

Of course the complexities of 'near-real-time' spectrum trading are very considerable and so far no definite schemes for its implementation have been worked out. We believe that this will happen through brokers who continuously monitor the utilisation of different frequency bands and, on the basis of that knowledge, broker between those who want to buy or sell spectrum. The brokers may also have the role of a market maker and as such may be required to continuously quote the price of spectrum bands on the basis of their supply and demand situation (see Fig 7).

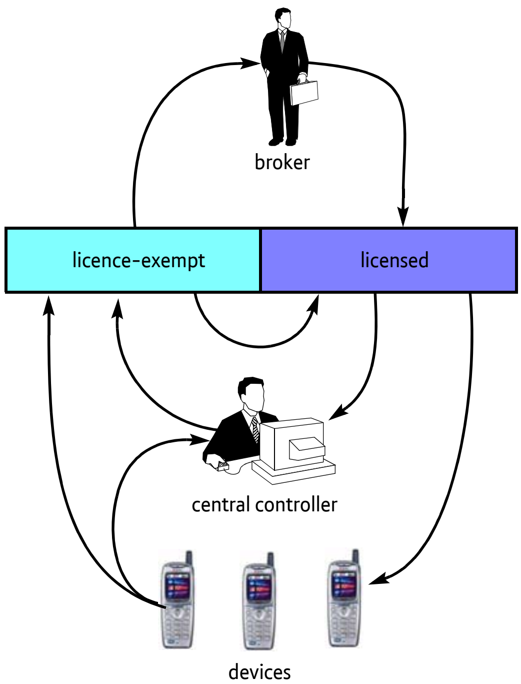

Fig 7 Devices can identify and utilise empty spectrum slots directly or through the use of a central broker. Also, the transition between licence-exempt and licensed spectrum can be mediated by a broker.

One of the most important potentials of spectrum trading is its active use for interference management. As the time-scales for trading get shorter, it can be effectively used to address short-term spectrum demands, and in such a manner support interference management which has, until now, been addressed in purely technical terms.

8. The digital dividend

Digital dividend is a term coined to describe the radio spectrum which becomes available after the transition to digital television is complete and analogue broadcasters are switched off. The spectrum required by digital television for an equivalent amount of content is estimated by some to be between a third and a half of that required by analogue television. We note, however, that other studies suggest that the amount of digital dividend that becomes available will largely depend upon whether there will be significant demand for additional commercial channels and high-definition TV over the terrestrial platform.

The potential future of this spectrum could be wide ranging, and determining how to distribute and manage this spectrum may pose problems for regulators. In a report to the EC [28] it is suggested that, due to the convergence between broadcasting and communications, broadcasters should not be treated differently, and so the spectrum dividend could be auctioned to the highest bidder, an approach which was taken in the USA and is also favoured by Ofcom [29].

9. Future research challenges

Although it is hard to predict how the management of spectrum will evolve in the future, it is evident that it will be much more dynamic than it is today. However, we expect both cognitive radio and spectrum trading to play a very important role. Also, it is to be expected that interference, which was the initial motivation for regulating spectrum usage, will be managed on a real-time basis through a combination of technology and markets rather than politics and bureaucratic procedures. The paradigm shift towards dynamic management of spectrum may promote both vertical and horizontal disintegration of the existing model for wireless services and architectures. For example, it may promote the unbundling of investment in spectrum rights, the operation of a mobile network, and the offering of mobile services. The real challenges that call for solutions are therefore of a cross-disciplinary nature, requiring contributions from technology, regulation and financial economics. At the core of the new approach is the heterogeneity of devices as well as their individualistic behaviour, based on the ability to detect relevant aspects of the radio environment and make real-time decisions on how to utilise available resources.

The realisation of devices that are aware of the radio environment as well as the activity of other devices requires the solution of some outstanding technical challenges [30]. Furthermore, on the basis of environmental conditions, including estimated spectral occupancy and signal-to-interference-noise ratio (SINR), devices need to be able to make decisions that best suit their requirements but at the same time have minimum detrimental impact on other devices. To achieve some of these goals progress is needed in the following areas:

- signal processing techniques,
- reliable and fast analysis of spectrum occupancy statistics,
- the evolution of spectrum usage strategy based on occupancy statistics — game theory,
- understanding of access etiquettes in heterogeneous multi-application environment,
- cognitive devices.

Addressing and solving some of these technical problems will require extensive collaborative efforts between academia and industry. A cross-disciplinary approach will be required with inputs from signal processing, radio and antenna technologies, regulatory thinking and economics. It is likely that some of the models and techniques required for making advances in this area are already available. To some extent the challenge is therefore to identify them and put them together into an integrated framework. That process calls for a multi-disciplinary approach.

10. Conclusions

With the increasing importance of wireless communications, an adaptive and efficient utilisation of spectrum is required. In this paper we discuss the need to depart from the 'command-and-control' system which has traditionally been used for allocating spectrum. We argue that, as the diversity of access technologies increases, the need for harmonisation and cross-technology standardisation are significantly reduced. Technology-specific spectrum allocation cannot accommodate the increasing technological and demand uncertainty that characterises wireless communications today. Predicting the emergence of new technologies and their user take-up is notoriously difficult and unreliable. Technology-specific spectrum allocations will therefore inevitably lead to suboptimal spectrum allocations.

As the complexities of wireless access technologies increase, new multidisciplinary approaches to spectrum management are required with inputs from technology, economics and regulation. Also, increased liberalisation and a technology-neutral approach to dynamic spectrum allocation is required. This process will be supported by the emergence of markets in spectrum, enabling users to acquire and sell spectrum as dictated by their needs.

However, the importance of spectrum trading will depend on the technical advances made in accessing the spectrum, such as power control, channel selection and access behaviour. If spectrum access technologies become more effective in making spectrum abundant, the need for trading it would be significantly reduced. In other words, it is the balance between supply and demand for spectrum which determines the future need for spectrum trading. This balance on the other hand is controlled by technology as well as the ratio of licensed to licence-exempt spectrum.

Future devices will have the ability and the permission to access licensed spectrum on an opportunistic basis. This requires them to be aware of the spectral occupancy in their immediate radio environment. By compiling statistics of spectrum usage, either from direct measurements or through information provided by others, devices will apply learning algorithms to evolve spectrum access strategies. This may require co-operation between devices, not only in the form of providing information on spectrum occupancy, but also relating to access etiquettes.

Whatever paths spectrum technologies take in the future, a definition of spectrum usage rights is required as a framework for user behaviour. Spectrum usage rights are the recognition of the growing trend towards considering spectrum licences as property rights that can be owned, traded or even shared. While the principle of SURs is generally accepted, the most significant hurdle to implementation is likely to be the disputes surrounding the initial definition of licence terms. Once this is overcome, SURs and the flexibility they provide will offer a solid basis for future spectrum management techniques. However, overcoming the initial hurdles is far from easy and may require a culture shift within the wireless industry.

References

1 Analysis: 'Spectrum demand for non-government services', (September 2005) — http://www.spectrumaudit.org.uk/pdf/spectrum_demand.pdf

2 Federal Communications Commission — http://www.fcc.gov/

3 The Office of Communications — http://www.ofcom.org.uk/

4 Ofcom: 'Spectrum Framework Review', (November 2003) — http://www.ofcom.org.uk/consult/condocs/sfr/

5 The Shared Spectrum Company: 'Spectrum Occupancy Measurements', National Science Foundation (NSF) (August 2005) — http://www.sharedspectrum.com/?section=nsf_summary

6 CEPT: 'Final report from CEPT in response to the EC Mandate on the review of the frequency band 169.4—169.8 MHz', (November 1994) — http://europa.eu.int/information_society/policy/radio_spectrum/docs/current/mandates/rsc04_69_ermes.pdf

7 CEPT: 'Final CEPT report in response to EC Mandate on TFTS bands', (September 2005) — http://forum.europa.eu.int/Public/irc/infso/radiospectrum/library?l=/public_documents_2005/rsc13_october_2005/rscom05-46_tftspdf/_EN_1.0_&a=d

8 Kawade S and Hodgkinson T: 'Licence-exempt wireless communication systems', BT Technol J, 25, No 2, pp 64—75 (April 2007).

9 Olafsson S: 'Sufficient Conditions for Convergent Power Dynamics in *ad hoc* Networks', Proc Spring IEEE VTC, Melbourne (May 2006).

10 Olafsson S and Chin E: 'Integrating Adaptive Power Control and Channel Selection in *ad hoc* Networks', Proc Spring IEEE VTC, Dublin (April 2007).

11 Faulhaber G R and Farber D: 'Spectrum Management: Property Rights, Markets, and the Commons', (2003) — http://assets.wharton.upenn.edu/~faulhabe/

12 ETNO Reflection Document Commenting on Auctions and Beauty Contests: 'ETNO Reflection Document RD203', (December 2004).

13 Ofcom (Radio Communications Agency): '3G Spectrum Auction Result', (April 2000) — http://www.ofcom.org.uk/static/archive/spectrum auctions/3gindex.htm

14 Hewitt T: 'Ensuring the correct timing of transition to liberalised markets', European Spectrum Management Conference, Brussels (March 2006).

15 Responses to Ofcom Consultation on Spectrum Usage Rights (June 2006) — http://www.ofcom.org.uk/consult/condocs/sur/responses/

16 Ofcom: 'Spectrum Usage Rights', (April 2006) — http://www.ofcom.org.uk/consult/condocs/sur/spur.pdf

17 Mitola J: 'Cognitive radio for flexible mobile multimedia communications', in Sixth International Workshop on Mobile Multimedia Communications, San Diego, CA (1999).

18 Federal Communication Commission: 'Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies', ET Docket 03-108, Notice of Inquiry and Proposed Rulemaking and Order (May 2003).

19 Nekovee M: 'Dynamic spectrum access with cognitive radios: Research challenges and future architectures', Proc of the 1st International Conference on Cognitive Radio Oriented Networks (CROWNCOM 2006), Mykonos, Greece (June 2006).

20 Ofcom: 'Cognitive Radio', (December 2006) — http://www.ofcom.org.uk/research/technology/overview/emer_tech/cograd/

21 IEEE 802 LAN/MAN Standards Committee — http://www.ieee802.org/22/

22 Granelli F, Zhang H, Zhou X and Marano S: 'Research advances in cognitive ultra wide band radio and their application to sensor networks', Mobile Netw Appl, 11, pp 487—499 (2006).

23 Gu X and Taylor L: 'Ultra-Wideband and its Capabilities', in Smyth P (Ed): 'Mobile and Wireless Communications: Key Technologies and Future Applications', The Institution of Electrical Engineers (2004).

24 Federal Communication Commission: 'Revision of Part 15 of the commission's rules regarding ultra-wideband transmission systems', FCC 02-48 (February 2002).

25 Nekovee M: 'Dynamic spectrum access — concepts and future architectures', BT Technol J, 24, No 2, pp 111—116 (April 2006).

26 Glover W and Nekovee M: 'Dynamic spectrum: Going the full circle', Proc of IEEE DySpan 2007, Dublin, Ireland (April 2007).

27 Olafsson S: 'Managing Risk in Network Capacity Investments', CRG Report (January 2000).

28 Burns J, Marks P, LeBorgne F and Rudd R: 'Implications of digital switchover for spectrum management', Technical Report, EC (June 2005).

29 Ofcom: 'The Ofcom Digital Dividend Review', Press Release (November 2005).

30 Haykin S: 'Cognitive Radio: Brain-Empowered Wireless Communications', IEEE Journal on Selected Areas in Communications, 23, No 2 (February 2005).

Sverrir Olafsson is a Chief Researcher in BT's Mobility Research Centre (MRC) at Adastral Park. He leads the Complexity Research Team within the MRC and has over the years also been the manager of various BT internal and external co-operative projects. He also holds a teaching position at the University of Reykjavik. Since joining BT in 1989 he has worked in the areas of neural networks, time series analysis, adaptability, evolutionary models, game theory, complexity theory, and access control. In the last three years his main focus has been on long-term strategic mobility research, including power control, spectral efficiency, distributed resource allocation and risk management.

He is on the organising committee of various international conferences, on the editorial board of Telecommunication Systems, on the advisory board of various academic and non-academic institutions and a reviewer for various technical journals. He has published over 60 refereed papers, authored numerous reports and written more than 175 articles on popular scientific and technical matters.

He has a PhD in theoretical physics. Before joining BT he worked for ten years at various Universities where he lectured in mathematical physics and did research in elementary particle physics, and soliton theory.

Bill Glover joined BT in May 2006. He graduated with an MSc in Electrical and Electronic Engineering from Bristol University in 2005.

He then worked as a hardware/software engineer before joining BT Research at Adastral Park.

Current research includes the future of spectrum management and the possibilities afforded by new concepts including cognitive radio and spectrum trading.

Maziar Nekovee is a Senior Scientist at BT's Mobility Research Centre and a Royal Society Industry Fellow at UCL. He is the author of over 50 papers in international peer-reviewed journals and conferences, holds a number of patents, and is a guest editor of the ACM Mobile Networks and Applications Special issue on Cognitive Radio Networks. Current and past research areas include dynamic spectrum access and cognitive radio networks, theory and applications of complex networks, vehicular *ad hoc* networks, grid computing and multicast communications. He obtained his MSc (cum laude) in electrical engineering from Delft University of Technology in the Netherlands, and his PhD in theoretical and computational physics from the University of Nijmegen, also in the Netherlands. He joined BT in 2001.