The pricing of radio spectrum: using incentives mechanisms to achieve efficiency[1]

Chris Doyle

Centre for Management under Regulation
Warwick Business School
January 2007

[1] This paper has been prepared for the ITU Workshop "Market mechanisms for spectrum management", Geneva, 22 and 23 January 2007. I am grateful for comments received from Martin Cave and William Webb. All errors are my responsibility. Contact details — Email: chris.doyle@cdoyle.com or chris.doyle@wbs.ac.uk; Cell: +44 7970 458809; Fax: +44 1926 328673; Office: +44 2476 574296.

1. Introduction

The regulation of radio spectrum is a costly activity which in practice is typically recovered through licence fees paid by radio spectrum users. This approach results in spectrum having a price. For example, in the United States the FCC applies two types of fees – application fees and regulatory fees which cover the administrative costs. As with all administratively determined prices, if they are set too high it can result in under-utilization of the spectrum, while if set too low hoarding and congestion may arise. Finding the right balance, which is achieved through choosing the right prices, is critical to ensure that economic efficiency is achieved.

However determining spectrum prices based upon the recovery of administrative costs, while widely practiced, fails to make use of one of the most powerful incentive mechanisms available to encourage more efficient use of radio spectrum. Rather than base spectrum charges only on administration costs, a spectrum manager can do better by setting incentive based prices that reflect economic value. The application of incentive based prices for radio spectrum licenses has been termed Administrative Incentive Pricing (AIP).[2]

In this paper I describe how the setting of incentive based radio spectrum licence fees or AIP can be undertaken to promote efficient use of spectrum in ways that go beyond the recovery of administration costs. A few countries have adopted variants of AIP. In the UK the regulator Ofcom applies AIP based on opportunity cost principles. In a review of spectrum management policy in 2002 commissioned by the UK government it was stated:

> "The fundamental mechanism by which the spectrum management regime could contribute to economic growth is through ensuring that users face continuing incentives towards more productive use of this resource. The review considers that these incentives should be financial and based on the *opportunity cost* of spectrum use. In this way, spectrum would be costed as any other input into the production process. Price signals about the cost of using spectrum would be disseminated throughout the economy. This information should enable dispersed economic agents to make their own judgments about their use of spectrum and the alternatives open to them to meet their organisational goals." Opportunity cost is defined as: "the value of an asset or resource in the next best alternative that is foregone by virtue of its actual use."[3]

Economists have shown that where firms buy inputs on competitive markets this tends to promote efficient use of these inputs and ensures that outputs are produced at the lowest

---

[2] For an introduction to Administrative Incentive Pricing see the InfoDev and ITU ICT Regulation Toolkit: http://www.ictregulationtoolkit.org/en/Section.1280.html. See also Cave, Martin, Doyle, Chris and William Webb (2007) *Essentials of Modern Spectrum Management*, Cambridge University Press, forthcoming.
[3] Cave, Martin (2002) Review of Radio Spectrum Management, for Department of Trade and Industry and H M Treasury, March.

possible cost. In other words, when firms face the right incentives they choose inputs carefully to ensure that their operational costs are kept as low as possible. When firms use inputs like radio spectrum efficiently, economists describe this as a situation of productive efficiency.

The paper is structured as follows. Section 2 looks at economic efficiency and radio spectrum, and Section 3 focuses on the productive element of this efficiency. Section 4 examines how pricing radio spectrum can be done to promote efficiency. In Section 5 the Smith-NERA method of calculating spectrum prices is discussed, which is an algorithm based on the application of the economic opportunity cost principle. Section 6 outlines how the Smith-NERA method might work in practice. Section 7 looks at the calculation of AIP in practice and focuses on a case study of fixed links in the UK. Section 8 discusses in brief incentive based spectrum charges in some other countries. Section 9 concludes the paper.

2. Economic efficiency and radio spectrum

Many spectrum management authorities are required to promote efficient spectrum assignment policy. In the United States one of the core principles of effective spectrum management governing the operations of the FCC and the NTIA (National telecommunications and Information Administration) is to maximize the efficient use of radio spectrum. In the UK, for example, the Communications Act 2003 states that spectrum assignment policy should ensure:

> "the efficient use in the United Kingdom of the electro-magnetic spectrum for wireless telegraphy"[4]

Efficiency in this context is usually understood to mean *economic efficiency*: the attainment of outcomes consistent with the application of what economists call the Pareto criterion. The Pareto criterion asserts that economic activity is efficient when it is not possible to find an alternative way of undertaking the activity to improve the well-being of one individual without harming the well-being of at least one other individual.

Economic efficiency has three dimensions relating to production, consumption and the use of resources over time.

- Productive Efficiency – production of goods and services ought to be undertaken at the lowest possible cost (cost is measured in terms of inputs). Users of radio spectrum should choose inputs, capital, labour and spectrum such that production of services is at the lowest cost.

- Allocative Efficiency – the mix of goods and services produced must be optimal in the sense that no other mix can increase the well-being of one economic agent

[4] Communications Act 2003, para. 152 section 5.

without harming the well-being of another economic agent. Spectrum should be allocated across different uses in a way that results in the Pareto criterion being satisfied.

- Dynamic Efficiency – resources are deployed in a way that encourages the most desirable level of research and development and innovation. The use of radio spectrum should allow for the right amount of innovation.

Economic analysis has identified the structural and behavioural characteristics that are needed in an economy if efficiency is to be achieved. When all the conditions are met to ensure efficiency, this is called ‘perfect competition’. In a perfectly competitive economy demand and supply for different goods and services, including radio spectrum, are brought into equality by the workings of the price mechanism and prices reflect opportunity costs.

In the equilibrium of a perfectly competitive economy the price mechanism establishes relative prices such that the cost to society of producing good X (say cellular telephony services) in terms of good Y (say broadband services) reflects consumers’ willingness to pay for such a transformation (the opportunity cost). This result is known as the First Fundamental Theorem of Welfare Economics, and it is often used to lend support to the claim that competitive markets are desirable.

The insight that equilibrium prices are consistent with economic efficiency in a perfectly competitive market economy is a useful guide for spectrum pricing policy. In the context of spectrum management, it suggests that choosing prices for spectrum which equate supply with demand is likely to be consistent with efficiency.

However, it is also well known that in a perfectly competitive economy in which externalities feature (for example where the actions of some firms may impact on other firms) efficiency may be sacrificed – particularly where property rights may be difficult to define. As unregulated use of radio spectrum would give rise to too many interference externalities, we cannot rely solely upon the market as a mechanism for achieving efficiency. Nevertheless, a spectrum manager can use market incentives, such as AIP and spectrum trading, to achieve superior outcomes than alternatives such as command and control.

3. Productive efficiency and radio spectrum

In this section I use a simple example to illustrate the relationship between the use of spectrum and productive efficiency. This is used to identify the conditions that need to be satisfied for productive efficiency in spectrum use, which in turn is useful when assessing spectrum prices.

Assume the available spectrum lies on a line between zero and one (the unit interval $[0,1]$) and that it can be used in two sectors 1 and 2 in the economy.[5] The sectors could represent broadcasting and telephony. To produce the final outputs in each sector, firms choose a mix of labour *l* and spectrum *s* – that is labour and spectrum are substitute goods.[6] We suppose that firms within a sector are identical. The total amount of labour is fixed and equal to $L = l_1 + l_2$ where l_1 is the labour used in sector 1 and l_2 is the amount of labour used in sector 2, and the price of labour, the wage rate w, is assumed to be determined on a competitive market. We assume also that the prices of all final outputs produced in the economy are determined in competitive markets. Spectrum, however, is allocated to each sector via an administrative process rather than via a market, and the price is assumed to be zero, with $\bar{s}$ be the amount of spectrum allocated to sector 1; and $1-\bar{s}$ is allocated to sector 2.[7]

Each firm using radio spectrum in each sector seeks to maximize profits and chooses an output level (and hence inputs labour and spectrum) to achieve this objective. Note that the firms face a spectrum constraint. At the allocation $\bar{s}$ total output produced in sector 1 is denoted $Q_1(\bar{s},l_1)$ and output in sector 2 is $Q_2(1-\bar{s},l_2)$.

Given an administrative allocation $\bar{s}$, there are three possible scenarios with regard to spectrum:

1. Demand for spectrum in each sector is equal to spectrum supply in each sector;
2. Demand for spectrum is no greater than spectrum supply in each sector; and
3. Demand for spectrum in one or both sectors is greater than spectrum supply in one or both sectors.

From a policy perspective the interesting scenario is 3, where demand for (free) spectrum exceeds the fixed spectrum supply in one or both sectors. As the quantity of spectrum is fixed and finite, excess demand in one or both sectors raises the issue of whether a re-allocation of spectrum could bring about a gain in efficiency. Alternatively, could a reallocation of spectrum free-up labour resources (the other input) without necessitating a reduction in the quantity of output produced in each sector? If the latter were possible, then the released labour resources imply that additional output could be produced and this indicates the initial allocation is inefficient

To examine whether a re-allocation of spectrum could deliver efficiency gains (or whether the current use of spectrum is consistent with productive efficiency), consider the effect of a hypothetical small change in spectrum allocation (a re-allocation). Suppose that for a small increase in spectrum $\Delta\bar{s}$ allocated to sector 1, the initial output level in the sector $Q_1(\bar{s},l_1)$ can be produced using Δl_1 units less labour. There are implications for

[5] Alternatively consider the assignment of frequencies within a band used by two users.
[6] Assume that many other sectors exist in the economy, but these do not use spectrum as an input. However, the other sectors make use of labour and other inputs such as capital.
[7] For simplicity assume spectrum management costs are recovered by general taxation.

sector 2 of this reallocation; there is less spectrum available but on the other hand more labour is available. If it is possible for output in sector 2 to increase as a result of this reallocation of spectrum (and labour) it would indicate that the initial administrative allocation is inefficient.

Although spectrum does not command a price in this example, it is possible to calculate its value by looking at the prices of other goods traded on markets. By doing this we can determine what might be termed the implicit price of radio spectrum and this could guide us in the setting of administrative prices. In other words, we are seeking to value of spectrum in terms of the value of other goods used in the economy.

This can be illustrated by noting that output in sector 1 can be kept constant after the radio spectrum reallocation by lowering the amount of labour used in the sector by the amount Δl_1.Because labour is traded on a competitive market at a wage rate *w*, this implies that the value of the marginal spectrum can be measured by the wage rate times the change in labour necessary to maintain output constant: that is $w\Delta l_1$. The same reasoning can be applied to sector 2, where the value of $\Delta\bar{s}$ is $w\Delta l_2$, when assessed at $Q_2(1-\bar{s}, l_2)$. Efficiency occurs when the opportunity costs of spectrum are equal across the two sectors.

The values $w\Delta l_1$ and $w\Delta l_2$ are estimates of the opportunity cost of spectrum. Where $\Delta\bar{s}$ is allocated to sector 1, the economy foregoes $w\Delta l_2$, the value of the input resources that would be saved by allocating $\Delta\bar{s}$ to sector 2 to maintain production at output level $Q_2(1-\bar{s}, l_2)$. If the marginal unit of spectrum $\Delta\bar{s}$ were allocated to sector 2, by analogous reasoning the economy foregoes a saving worth $w\Delta l_1$ in sector 1. The values $w\Delta l_1$ and $w\Delta l_2$ allow AIP to be calculated correctly but in practice require an understanding of the relationship between radio spectrum and other close substitute inputs.

The values $w\Delta l_1$ and $w\Delta l_2$ are referred to by economists as the <u>marginal benefits</u> of spectrum. Marginal benefits typically decline, reflecting the presence of decreasing returns to scale for firms. In other words, as firms produce more output the effect of a unit of an input on production falls. In Figure 1 the marginal benefits of spectrum are shown as declining in each sector. Sector 1's marginal benefit function is on the left and sector 2's marginal benefit is on the right hand side in the figure. As the amount of spectrum allocated to sector 1 increases, the marginal benefit of spectrum in sector 1 declines, whereas the marginal benefit in sector 2 increases.

In Figure 1 it can be seen that at the initial administrative allocation $\bar{s}$ the marginal benefit of spectrum in sector 1 is greater than that in sector 2. As stated above, the allocation shown at $\bar{s}$ is inefficient because the marginal benefit values, which are estimates of the opportunity costs of spectrum, are not equal. Efficiency is satisfied at the allocation s^* in Figure 1.

It is very unlikely that a radio spectrum manager will be able to calculate s^* and achieve efficient spectrum management. The informational burden of assessing the correct allocation and assignment of radio spectrum administratively is onerous. Instead radio managers should, where possible and where desirable, look towards market mechanisms to help identify efficient allocations. In the following sections I address this by looking at how a radio spectrum manager can use estimates of opportunity costs to determine spectrum prices and use these to promote efficiency.

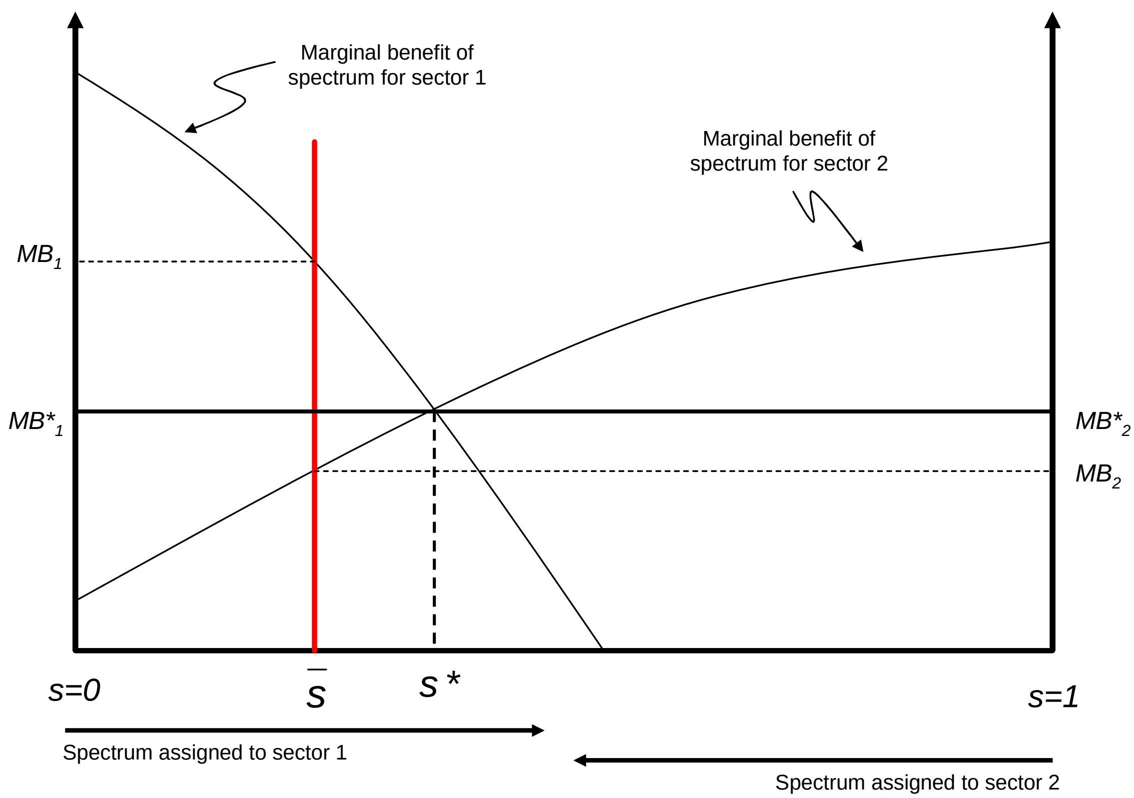

Figure 1: Marginal benefit of spectrum

3. *Pricing radio spectrum to achieve economic efficiency*

The above suggests that to implement spectrum prices with efficiency in mind requires detailed information about the relationship between radio spectrum and other close substitute inputs. Alternatively, in the context of Figure 1 it requires a spectrum manager to have a good understanding about the shape of the marginal benefit functions.

While information about the shape of the marginal benefit functions is very useful, it is very demanding to expect a spectrum manager to be able to acquire all the necessary information easily. However, as I show in the next section, it is not necessary to know in detail the entire marginal benefit functions in order to compute spectrum prices that promote efficiency.

Prices that lead to improvements in the use of radio spectrum and shift allocations and assignments in the direction of efficiency can be devised using information about estimated marginal benefits at current allocations and assignments. One such method for calculating prices based on current assignments and allocations is known as the Smith-NERA method.

4. The Smith-NERA method of calculating spectrum prices

The Smith-NERA spectrum pricing methodology is a pricing algorithm used to calculate spectrum prices based upon opportunity costs.[8] I start by outlining a simple hypothetical example to illustrate the Smith-NERA method.

Assume that radio spectrum is in three non-overlapping frequency bands {*a*,*b*,*c*} in the interval [0,1]. Further assume there are three competing uses for this radio spectrum: I, II and III. Assume that a historical administrative allocation and assignment has occurred such that: Use I is allocated frequency band a, Use II is allocated frequency band b, and Use III is allocated frequency band c. The marginal benefits of the different frequency bands across the different uses are shown in Table 1. In addition, the marginal benefit of a non-spectrum input is also shown in the final column.

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	100	75	0	0
II	35	60	30	0
III	10	10	15	5

Table 1: Marginal benefits of spectrum

At the initial administrative assignment and allocation, spectrum in band *a* has a marginal benefit of 100 in Use I, whereas spectrum in band *b* has a lower marginal benefit of 75 in Use I and frequency band *c* has a zero marginal benefit in Use I. Thus frequency band *a* is the most valuable for Use I, and the frequency in band *c* has no value (that is, frequency band *c* cannot support the applications in Use I). The highlighted cells in Table 1 indicate the opportunity cost estimates that a radio spectrum manager could estimate at the current assignment and allocation – as data could be observed in the field to assist their computation.

[8] Smith-NERA (1996) "Study into the use of spectrum pricing", report for the Radiocommunications Agency by NERA and Smith System Engineering, April.

The discussion on productive efficiency above suggests that marginal benefits across different uses for the same frequency band should be equalized if efficiency is to be achieved. In the highlighted cells in Table 1 it is clearly the case that the marginal benefits do not equal the values elsewhere in the same column (the same band across different uses) and therefore efficiency of spectrum use is not achieved. For example, frequency band *a* has a marginal benefit 100 to Use I, whereas in Use II it has a marginal benefit of 35 and in Use III a marginal benefit of 10. As all frequency band a is currently allocated to Use I, there can be no gain from reallocating frequency band a to Use II as the spectrum is exhausted.

However, frequency band *b* has a higher value at the margin in Use I than it does in its current Use II. This suggests that by reallocating spectrum in frequency band *b* away from Use II to Use I there is the potential for a gain in efficiency. As more of frequency band *b* is allocated to Use I the marginal benefit in Use I of using frequency band *b* will fall below 75, while the marginal benefit of using frequency band *a* in Use I will fall below 100. Similarly, when frequency band *b* is taken away from Use II, the marginal benefit to Use II of frequency band *b* will increase above 60 and the marginal benefit for Use II of frequency band *c* will increase above 30.

The effect of shifting some frequency band *b* to Use I away from Use II leads to the revised marginal benefits shown in Table 2.

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	90	70	0	0
II	38	70	32	0
III	10	10	15	5

Table 2: Marginal benefits of spectrum following re-allocation of frequency band *b*

Table 2 differs from Table 1 in that the marginal benefits of frequency bands *a* and *b* in Uses I and II have changed. This reflects the fact that Use I has more spectrum and Use II less spectrum. It is also the case that the marginal benefit of frequency band *b* between Uses I and II is equal at 70. The equalization of the marginal benefits indicates that frequency band *b* is allocated efficiently across uses. With Use II having less spectrum, the marginal benefit of frequency bands *a* and *c* in Use II have increased. The effect of re-allocating spectrum between Uses I and II also has a knock on effect on the value associated with frequency band *c* in Use II.

In Table 2 above the highlighted cells indicate those which a radio administrator may more easily be able to calculate in practice. It can be seen that there is scope for a further efficiency gains by re-allocating some of frequency band *c* to Use II from Use III. By doing this, however, the marginal benefit of frequency band *b* in Use II will fall (as total spectrum in the use increases). Table 3 below presents the marginal benefits after re-allocation consistent with efficiency.

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	87	68	0	0
II	36	68	25	0
III	12	12	25	4

Table 3: Marginal benefits of spectrum following re-allocation of frequency band *c* and further re-allocation of frequency band *b*

In Table 3 efficiency occurs where there is equality in marginal benefits across uses in the two highest values. It is not possible to find further reallocations of spectrum that can yield better outcomes than shown in Table 3 and therefore this is an efficient solution. It can be seen that to arrive at the efficient outcome the radio administrator needs to know about marginal benefits of frequency bands in neighbouring uses having the highest values. Furthermore, the process of attaining efficiency is iterative – reallocations are made one at a time (or several may occur simultaneously) and after each change the new marginal benefit values are assessed. The new marginal benefit values then inform the spectrum manager about the direction of further reallocations or reassignments.

The next section discusses this iterative procedure in more detail and describes the least-cost-alternative method.

5. Setting spectrum prices to achieve efficiency using the Smith-NERA method

Tables 1-3 suggest an iterative approach is likely to be required in practice in order to achieve spectrum efficiency via administrative incentive pricing. Below I outline a framework for a radio spectrum manager that would support the implementation of an iterative procedure:

1. Identify all frequency bands and associated uses (resulting in a matrix with a much larger dimension than the illustrative ones shown above in Tables 1-3). Populate the cells with estimates of the marginal benefits MB_{ij} appearing in use *i* (row) in frequency band *j* (column) applying the least-cost-alternative. The least-cost-alternative is where a user substitutes spectrum with the least cost alternative input, such that output is unchanged following a small change in the amount of spectrum. For example, for a radio point to point fixed link the least cost alternative could be a fibre optic cable. In cellular telephony, the least cost alternative could be more or less base stations.

2. Use the estimates of marginal benefits to identify the direction of change in spectrum re-allocation. For example, Table 1 above informed us that at the margin frequency band *b* is more valuable in Use I than in Use II. Hence there ought to be a movement of frequency band *b* away from Use II to Use I. Similarly frequency band *c* is more valuable in Use II than in Use III, so some frequency band *c* should move from Use III to Use II. The radio spectrum manager looks at the value of marginal benefits in a column (across the different uses) and seeks to move spectrum away from the use with the lowest marginal benefit to the use with the highest marginal benefit.

3. Having identified the direction of re-allocation, which will depend on spectrum substitutability and marginal benefits, identify the maximum values of the MB in each column, for each column *j* call this $MB_{ij}*$. In Table 1 these are 100, 75 and 30.

4. If the maximum in step 3 occurs in a use which does not currently use the frequency band (such as Use I and frequency band *b* in Table 1), then spectrum prices should be set to lie in the interval between $MB_{ij}*$ and the current use marginal benefit. Hence, the price of frequency band *b* should lie between 75 and 60.

5. Judgment is needed with respect to the actual price(s) chosen in the interval, but any information about the characteristics of the efficient allocation could guide price setting. Thus, if the values in Table 3 were known, this could inform the selection of prices. Assuming the data in Table 3 are not known, it could be proposed that a price for frequency band *b*, for example, is set above 60 – but not significantly so.

6. If the maximum in step 3 is the value for the current use of the band then set the price at this value.

7. Having set prices for the spectrum, users will respond by changing their demands. After a period of time new marginal benefit values will emerge and the above procedure can be repeated. This may takes up to five years or so. Eventually the radio manager will converge towards an efficient allocation of assignment of radio spectrum.

5.1 Setting the spectrum price: using judgement

Step 5 in the iterative process outlined above presents a question: If an interval is observed in marginal benefit values, what point or points in the interval should form the spectrum price(s)? Clearly setting prices too high will lead to a fall in the use of a frequency band and new demand would likely be small. This is clearly inefficient as spectrum would not be used. It is better that spectrum is used and contributing to welfare, than not being used at all.

Setting a price close to, but not equal to the lower limit, would result in new demand for a frequency band, such as frequency band *b* in Use I, but existing demand by Use II would not fall by much. However, a price above the marginal benefit in Use II, but below Use I for frequency band *b* would lead to some spectrum being relinquished.

This approach in judging the right price for a frequency band serves to illustrate a more general point. Erring on the side of caution and approaching what the economists term the socially optimal price(s) (resulting in the equalization of marginal benefits) from below is better for welfare.

5.2 Comments

The above is intentionally simple for expositional reasons. In practice there are many different firms operating in a use within a frequency band. In practice some firms in a use will find the AIP price too high, and other firms will find the price lying below their marginal benefit values. For AIP to work well, the selection of the representative firm has to be undertaken carefully.

There may be several different sub-uses occupying a frequency band, reflecting different final markets (e.g. in a PMR band there may be taxi firms, utilities and couriers). This is likely to give rise to different estimates for the marginal benefit of spectrum in a given use. A single measure of the marginal benefit may be calculated by taking a weighted average, where the weights to use could be the amounts of frequency in the different sub-uses.

The above analysis also makes no distinction across geographic areas. However, this can be accommodated by looking at matrices for different regions. In some regions where excess demand occurs, opportunity costs will play a role in influencing prices, whereas in other regions this will not be an issue.

Time is not considered explicitly in the above example. Demands vary through time, and some future uses may not be known. Prices should therefore be periodically re-evaluated taking account of changes in demand and technology.

6. Applying administrative incentive prices: some issues

To apply AIP requires the application of the Smith-NERA methodology. This means a spectrum manager needs to know the input alternatives for the current radio spectrum used by an application and should understand what quantum of these alternative inputs would substitute for the current radio spectrum used. The steps facing a radio spectrum manager in the process of establishing AIP are as follows:

1. For a given frequency band identify current and other potential uses of the band.

2. Calculate the opportunity costs of spectrum for the current use of the band and other uses. This is achieved by applying the least cost alternative method described in the last chapter.
3. If there is a use with an opportunity cost higher than the current use, then set the AIP between the two values, but towards the bottom end of the range of values.
4. If there is no use with an opportunity cost higher than the current use of the band then set the AIP at the value for the current use.

In principle the calculation of AIP is straightforward, but in practice there are a number of challenges which are discussed below. Note that the application of AIP is predicated on the assumption that it is possible to reallocate spectrum administratively from the current use to other potential uses. The feasibility of achieving this over the timescale for which prices are set, say five years, needs to be considered. In bands that are shared between different uses, reallocation will be relatively straightforward. However, it may not be the case for other bands. If it is not likely to be feasible to reallocate the spectrum in these timescales[9], then the opportunity cost for the current use should be used to determine AIP.

6.1 Calculating opportunity costs

Opportunity costs are calculated using the approach discussed above. The marginal value of spectrum is the additional cost (or cost saving) to an *average* or *reasonably efficient* user as a result of being denied access to a small amount of spectrum (or being given access to an additional small amount of spectrum). The additional cost (cost saving) depends on the application and should be calculated as the estimated minimum cost of the alternative actions facing the user. These alternatives may include

- investing in more/less network infrastructure to achieve the same quantity and quality of output with less/more spectrum;
- adopting narrower bandwidth equipment;
- switching to an alternative band;
- switching to an alternative service (e.g. a public service rather than private communications) or technology (e.g. fibre or leased line rather than fixed radio link).

The value of opportunity costs calculated by a spectrum manager will differ between the cases where spectrum is taken away and where spectrum is increased. For a marginal reduction in spectrum, the calculation will overstate the ‘true’ value of spectrum, whereas for a marginal increase the calculation understates the ‘true’ value.[10] An average of the

[9] This may be because there is no equipment available for the new use in the given frequency band. Administrative processes for reallocating spectrum tend to be slow, though where administrative processes are replaced by market mechanisms, such as trading and the auctioning of overlay licences, the reallocation of spectrum may occur over a shorter time period.
[10] The difference in values arises because a profit maximising firm when faced with a change in the quantity of spectrum (having less or more) would respond by changing the output produced. However, in the least-cost-alternative approach, it is assumed that output is unchanged. While unrealistic as an assumption, it enables a proper test for efficiency.

opportunity cost values obtained from an increase and a decrease in spectrum gives a reasonable approximation to the true value of spectrum.

In practice, however, it is not always feasible to estimate values for both increases and decreases in spectrum and so there may be a small bias in estimates if the spectrum manager relies on one or the other.

6.2 Assumptions

To calculate opportunity costs a spectrum manager faces a number of modeling challenges. In particular, the spectrum manager will have to make assumptions in relation to a number of key questions:

- What is the appropriate size of a marginal change in radio spectrum?
- What is meant by an average or reasonably efficient user? For example, should the average user be categorized in terms of network topology and characteristics of equipment used (e.g. age, bandwidth, power), given radio communication demands (e.g. local or national, traffic levels, service quality requirements) or some other metric?
- What is the appropriate discount rate and discounting period?
- How should equipment maintenance costs be assessed?
- How does one reflect the maturity of existing networks in the calculations?

Each of these questions is considered below.

What is the appropriate size of a marginal change in radio spectrum?

The calculation of AIP should be based on an assessment of the marginal value of spectrum, where marginal is meant to be a small change in spectrum used. Therefore a marginal increase or decrease in spectrum should reflect the minimum amount that is likely to be of practical benefit to the user. For example, in the case of a cellular network this should take account of typical cellular re-use patterns.

The amount of spectrum that constitutes ‘marginal’ will also differ by service. For example, for PMR services marginal spectrum is likely to be a 2 x 12.5 kHz channel, whereas for aeronautical communications it is larger at 25 kHz and for a cellular network or PAMR services it is the number of channels required to populate a single cell “cluster”, taking account of typical planning parameters.

Thus the marginal amount of spectrum depends on the use considered.

What is meant by an average or reasonably efficient user?

Defining a reasonably efficient user for the purposes of calculating marginal values should be based on information held by the spectrum agency in its frequency tables, and from information gathered from secondary sources and from industry. In some cases (notably for fixed links), the relationship between costs and bandwidth is non-linear, as a high proportion of costs are fixed (i.e. are independent of bandwidth). Consequently, there is a wide variation in opportunity costs determined for individual link types. One approach is to identify opportunity costs for each main link type and from these determine a weighted average value reflecting the total amount of spectrum utilised by each link type. Inevitably some degree of judgment is used in deriving the assumed user profiles.

What is the appropriate discount rate and discounting period?

Current costs need to be converted into annual recurring values, as there are long-term and short-term costs associated with varying radio spectrum. The spectrum manager will need to make assumptions about discount rates, though these will be informed significantly by government assessments of discount rates and by the assessment of the cost of capital in the industry or sector concerned.

How should equipment maintenance costs be assessed?

As radio spectrum is more often than not employed by capital intensive industries, maintenance costs can be significant. Hence maintenance costs can impact materially the assessments of opportunity costs. There is no straightforward way of dealing with maintenance costs, though guidance may be found from the depreciation rates used in company accounts. In the UK a very simplistic approach has been advocated, where per annum maintenance costs are assumed to be 12% of initial capital expenditures.

How does one reflect the maturity of existing networks in the calculations?

When the amount of spectrum is varied in a use, the way a user responds by changing inputs will often depend critically on the maturity of the technology and/or network deployed. For example, in the case of GSM spectrum in Europe the networks are largely mature as they are fully developed in terms of coverage. Therefore a marginal change in spectrum will only affect the capacity of the network in areas where at peak demand the network is congested. To calculate the opportunity costs in this case the spectrum manager needs to understand how the GSM network performs at peak demand, and needs to understand the amount of infrastructure (i.e. base stations) that would substitute for spectrum.

6.3 Congestion and area sterilised

Clearly the opportunity cost of spectrum for a user or use will be related to the spectrum denied to other users, and the costs will typically be higher the greater the bandwidth used and the wider the geographic area over which use is denied i.e. the area sterilised by the service. The concept of area sterilised is appropriate for services such as mobile and broadcasting but works less well for fixed links where congestion at specific nodal sites is often the main constraint on spectrum use.

If national prices are calculated, then the opportunity costs obtained for a local frequency assignment, such as PMR or CBS (Common Base Station), can be converted to a national value by multiplying the local value by the likely amount of frequency reuse. This approach implicitly assumes that spectrum use is congested at a national level. It is important to test whether this assumption holds or not when converting marginal values into AIP. If the assumption does not hold and there is excess demand for spectrum in some but not all locations, then the national value could be calculated as relevant multiples of the congested and non-congested values where the multiples depend on the extent of congestion. An alternative approach would be to have geographic de-averaging of spectrum prices.

For some services (e.g. PMR or fixed wireless access) it may be appropriate to apply weighting metrics such as population or the number of businesses within an area where spectrum is consumed as a proxy for the degree of congestion. In other cases (e.g. fixed links), congestion may be measured in terms of the actual level of use at specific locations.

7. Calculating AIP in practice: case study of fixed links in the UK

7.1 Introduction

AIP has been in use since 1998 in the UK, following the passage of the 1998 Wireless Telegraphy Act. The application of AIP since 1998 has evolved and generally become more sophisticated. The amount of revenue that was collected in 2002 is shown in the Table 4. Initially the scope of AIP was limited to some major commercial uses, but over time it has been extended to cover spectrum used by the emergency services and the military.

AIP revenue raised by sector in the UK £000

	Sector	2004/2005	2005/2006
1	Aeronautical	818	931
2	Amateur and Citizen's band	1,030	883
3	Broadcasting	2,454	4,001
4	Business Radio	15,187	11,838
5	Fixed Links	18,203	20,895
6	Maritime	1,723	2,031
7	Programme Making and Special Events	1,145	1,412
8	Public Wireless Networks	63,868	63,011
9	Science and technology	112	745
10	Satellite	928	974
11	Ministry of Defence	24,314	55,398
Total		**132,168**	**164,094**

Table 4 – AIP revenue in the UK 2004-06

7.2 Setting AIP for fixed links in the UK: a case study

Fixed link services are point to point radio based services and in the UK they are used primarily for infrastructure links for mobile telecommunications networks. Each fixed link is separately licensed and there are over 40,000 links in operation. Popular frequency bands in use are 7.5GHz, 13GHz, 23GHz and 38GHz. There is also increasing interest in the 55, 58 and 65GHz bands for very short fixed infrastructure and access links.

In the Indepen study commission by Ofcom, the main alternatives to fixed links were considered to be:

1. Use of more spectrum-efficient technologies within the same frequency band, i.e. allowing less spectrum to be used to convey the same amount of data.
2. Use of a higher frequency band where there is greater capacity and less likelihood of congestion.
3. Use of a non-radio alternative (e.g. a leased line or fibre).

I illustrate the calculations of opportunity cost for fixed links in the UK by looking at 1 above, which examines different technologies using the same frequency bands.

For most fixed links applications there are competing technologies, differentiated in terms of adaptive modulation[11], which can perform the necessary data conveyance. Higher modulation schemes generally result in a lower spectrum utilisation per unit of data conveyed by the link, but cause greater interference and therefore require greater separation from other co-channel links. Thus there is a trade-off. Lower modulation schemes are spectrally inefficient but cause less co-channel interference, higher modulation schemes are more spectrally efficient but generate more co-channel interference.

The first task facing a spectrum manager in assessing AIP for fixed links is to identify the capital costs for the different technologies which can perform the necessary data conveyance. In the UK information from several sources was used to assess the costs of three different modulation schemes (QPSK, 16QAM and 128QAM) for six different data rates ranging from 2Mbps up to 155Mbps. It was assumed that a spectrally more efficient modulation scheme utilises 75% of the spectrum utilised by the next best less efficient alternative. It was also assumed that over 15 years of the lifetime of the equipment, a discount rate of 10% applied and that the maintenance costs of more efficient equipment was identical to that of less efficient equipment.

The figures from the Indepen study for a link operating at 8Mbps is shown in Table 5 below.

Link speed and less efficient scheme	More efficient option	Spectrum utilisation		Equipment costs (£)		Value per 2x1 MHz (£)	Annualised value (£ per 2x1 MHz)[12]
		Less efficient	More efficient	Less efficient	More efficient		
8Mbps/QSPK	16QAM	7 MHz	5.25 MHz	6,500	10,400	=£3,900/1.75= £2,228	266

Table 5 – Data on fixed links

In Table 5 the value of spectrum is calculated by supposing that there is a marginal decrease in spectrum. For the operator to maintain data conveyance at 8Mbps, a more efficient modulation scheme at 16QAM would be the next best alternative and this would entail slightly less bandwidth (1.75MHz). On the other hand the 16QAM technology is more costly, £10,400 versus £6,500. Hence, the value of 2x1MHz is estimated by taking the additional costs associated with the more efficient technology and dividing this by the spectrum saved. Finally this figure is adjusted to obtain an annualised sum.

[11] Adaptive modulation is used in many digital communication networks (e.g. cable modems, DSL modems, CDMA, 3G, WiFi, WiMax and point-to-point fixed links). Common techniques include quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM). These techniques can be used to increase capacity and speed in a network. Modulation is the process by which a carrier wave is able to carry the message or digital signal. There are three common methods: amplitude, frequency and phase key shifting.

[12] This is the value equivalent to the payment of a loan based on constant payments over a 15 year period for a constant interest rate (the discount rate) of 10%.

The computation illustrated in Table 5 shows that the estimated opportunity cost is sensitive to a number of variables: capital cost estimates, the perceived next best alternative, judgments about spectral efficiency gains, the discount rate (or cost of capital), and the lifetime of the equipment.

The Indepen study calculated values for a range of fixed link types and calculated a weighted average per MHz value based on the total bandwidth of each link type multiplied by the value per MHz and divided by the total bandwidth of all links in operation. The figure proposed by the consultants for the opportunity cost of fixed links was £132 per annum per 2x1 MHz per link. The latter estimate is sensitive to assessments about the bandwidth occupied by different link types and the number of links in operation across the different link types, as well as on the other factors referred to above.

The opportunity cost estimate proposed by Indepen for marginal spectrum used by a fixed link was reduced by the regulator Ofcom to £88.[13] To price an individual link the reference spectrum price estimate is adjusted by a number of factors and the formula used by Ofcom is:

> Fixed link licence fee = spectrum price × bandwidth factor × band factor × path length factor x availability factor

The spectrum price is £88 per 2x1MHz bidirectional link and is calculated in the way described above.

The bandwidth factor takes account directly of the amount of bandwidth used by a link, for example for in the 6GHz band the average bandwidth is 37.37MHz. Ofcom applies a minimum of 1 to the bandwidth factor.

The band factor reflects the balance in supply and demand on a band-by-band basis, and as such the level of congestion. The value is 1 for lower frequency bands and declines for links in higher frequency bands.

The path length factor reflects the opportunity cost of spectrum in a certain band, based on the extent to which shorter links deny spectrum to other users (of potentially longer and more efficient links) in that band. Ofcom operates a minimum path length (MPL) policy to conserve lower frequency bands for longer links which can be accommodated only in these bands. Whilst it is Ofcom’s general policy to avoid making assignments where the link path length is less than the MPL, it does so when requested. When such assignments are made, the path length factor adjusts the fee by placing a premium on the use of path lengths below MPL. This premium reflects the opportunity cost of spectrum, based on the extent to which shorter links deny spectrum to other users in that band. For a given MPL for each band and system type, the path link factor is calculated according to the following formula:

[13] See Ofcom “Spectrum pricing: A statement on proposals for setting Wireless Telegraphy Act licence fees” 23 February 2005, London.

When a link path length is at least as long as the minimum path length, the path length factor is equal to 1. When a link path length is less than the minimum path length, the path length factor equals the square root of the minimum path length divided by the path length. Ofcom caps the path length factor at 4.

Finally the availability factor determines the quality of spectrum a fixed link user receives (i.e. the probability that the fixed link user can receive a signal). A system availability requirement of 99.99% (sometimes referred to as “four nines” or “two nines”) is the normal starting point when making assignments and is the most commonly requested value. However, other availability requirements are also available to suit customer needs. In developing the algorithm, the value of unity for the availability factor has been associated with the most common availability requirement (e.g. 99.99%). Higher (or lower) availability requirements attract a higher (or lower) availability factor, reflecting the opportunity cost of the spectrum denied to other users. The availability factor applied varies from 0.7 for 99.9% availability through to 1.4 for 99.999% availability.

8. Incentive based spectrum charges in other countries

Only a few countries have deployed incentive based spectrum prices based upon opportunity cost principles. Below I describe in brief for some of those that have introduced AIP like prices or are considering doing so.

8.1 Australia

Australia has operated a system of spectrum pricing for a number of years that embodies in part the principle of opportunity costs. The fees are determined by the Australian Communications and Media Authority (ACMA), formed in 2005 from the Australian Communications Authority (ACA) and a broadcasting regulatory body.

The ACMA employs the following principles so that licence fees contribute to the efficient allocation of spectrum, and promote an equitable and consistent fee regime:

1. Charges should cover the direct administrative costs of issue, renewal and installment processing.
2. Taxes from licensees as a group should recover the indirect costs of spectrum management (such as international coordination costs).
3. Taxes should be based on the amount of spectrum denied to other users.
4. Spectrum denied should be priced at its opportunity cost (the value of the best alternative use of that spectrum).
5. If the opportunity cost is less than the indirect costs attributable to the licensee, taxes should only recover costs.

As in the UK adjustment factors are applied to meet specific conditions.

8.2 Canada and Denmark

In Canada the Government through the Minister of Industry has exclusive spectrum management responsibility of radio spectrum. Day to day spectrum management is performed by Industry Canada, a federal government department reporting to the Minister of Industry. Spectrum allocation is largely harmonized with the U.S.

The total cost of the spectrum management program run by Industry Canada is around $61m per year and Industry Canada's licence fee revenue derived from non-broadcast activities is $209m per year. The cost of managing broadcast spectrum is $13m per year and the CRTC (the broadcast regulator) raises licence fee revenue of $101m per year. The fees raised are much in excess of the administrative costs involved.

The setting of spectrum fee in Canada is based on the market value or a reasonable approximation thereof of the spectrum used. As discussed in the previous chapter, the market value of spectrum can be estimated by its opportunity cost.

Since 1996 fees in Canada have been based on the quantum of spectrum authorized in a defined geographic area, with population or households included as a variable. Industry Canada is considering the introduction of an AIP-like mechanism called Spectrum Efficiency Incentive Pricing.

Denmark's regulator Telestyrelsen is also moving towards a spectrum management regime that incorporates greater use of market mechanisms and spectrum licence fees have a factor which includes an opportunity cost element.

9. Conclusion

In this paper I have considered the setting of spectrum prices largely from a both conceptual and practical perspective. I have shown that a spectrum management agency can use prices to achieve efficiency in spectrum use, and that generally this will lead to superior outcomes than setting prices on a cost recovery basis. The economic underpinnings of efficient pricing of spectrum were illustrated and the Smith-NERA methodology was discussed.

Incentive based spectrum pricing is a tool that spectrum managers can use to encourage efficient spectrum use. Charging annual fees for the holding of spectrum is one way in which the spectrum manager can encourage current and prospective holders to make the right decisions to ensure efficient use of the spectrum.

Any use of spectrum imposes an opportunity cost on society – the value foregone of alternative use. This is because spectrum is finite and use is exclusionary – the use of spectrum for one purpose precludes its use for another. Therefore all decisions affecting current and future spectrum use should be made with a full and accurate reflection of these opportunity costs, if those decisions are to lead to the socially optimal allocation of resources in the short and long term. If the opportunity costs of spectrum use are ignored or discounted, socially sub-optimal decisions will be made. One of the best ways of ensuring that the opportunity costs of spectrum are fully and accurately reflected by decision makers is for those opportunity costs to be reflected in prices that have to be paid to hold spectrum.

This is the principle behind the use of AIP. The primary purpose in applying AIP is not, in general, to achieve any specific short-term change in the use of spectrum. Rather, the aim is to ensure that the holders of spectrum fully recognise the costs that their use imposes on society by holding spectrum (or seeking to acquire additional spectrum), when making decisions. Many holders of spectrum are not in a position to make rapid changes to their use of spectrum in response to the application of AIP, but note that in practically every case the holders of spectrum have opportunities to change their use of spectrum in the longer term.

The use of AIP is justified by the benefits that should materialise in the longer term, as better decisions are made in light of increased awareness and appreciation of the value of spectrum – better decisions that should lead to more efficient use of the spectrum. The UK regulator Ofcom cited some evidence of the success of AIP. Since 2003 significant amounts of spectrum have been returned to Ofcom for re-assignment, as a more or less direct result of AIP. 28MHz of the more valuable spectrum below 3GHz has been released by public and private sector users in response to AIP, as has 160MHz of the second-tier spectrum in the range 3-10GHz.