

**PATENT EXPLOSION AND PATENT WARS:
HOLDUP, ROYALTIES
AND MISUNDERSTANDINGS OVER “MARKET VALUE”**

Cristina Caffarra (CRA), Pierre Régibeau (CRA. Imperial College), Robert Stillman (CRA)¹

February 2012, Revised January 2014

1. MOTIVATION AND OVERVIEW

“Patent wars” Have now been raging for a few years and signs of appeasement are few. Following Microsoft’s assertion of its “ActiveSync” patents, Apple’s litigation based partly on its “design” patents and design rights, and the Rockstar consortium acquisition of Nortel patents Google acquired MMI to get hold of MMI’s mobile communication patent portfolio and continue ongoing litigation against Apple and Microsoft. This has led to a flurry of suits and countersuits and injunctions have reached mainstream news – notably with the granting of an injunctions against Samsung’s Galaxy products in the US, decisions by Judge Posner on the appropriateness of granting injunctive relieve to SEP owners and judge Robart’s proposed methodology for the computation of FRAND royalties. By now patent wars have engulfed a number of other players, including Samsung, Huawei and Nokia and have spread to other jurisdictions such as South Korea and China.

The recent patent wars are noteworthy because they are clearly part of a broader battle for supremacy between platforms in a converged mobile environment, with large asymmetries persisting between the parties involved: device manufacturers with a stash of patents deemed “standard essential” because they have been involved early in the communication protocols; Microsoft with its traditional IP strength in software; Google, coming from the open source movement and therefore with virtually no patent portfolio until recently; Apple, very good at design and integration but less so at purely “technological” innovation, and therefore with a very different patent portfolio. The convergence of the industry means we are really facing a battle for who is getting the rents in combined devices, and this is what is making the problems so thorny – as there is a concern that parties may be using their IPRs as tools to exercise market power and raise rivals’ costs (or slow down their progress) in the market for smartphones and tablets. The hope that all will eventually reach patent settlements on their own and that “everything will be alright on the night” appear to some overly optimistic. And consumers might be hurt before the industry settles into some new, more peaceful, equilibrium.

The potentially harmful consequences of patent wars are magnified for industries that are heavily reliant on standards. Standards can be successful only if two conditions are met: firms that control important related technologies have to be induced to take part in the design of the standard; and in addition, access

¹ While CRA represents a number of parties involved in this debate, these are the personal views of the authors and do not in any way reflect the position of CRA clients.

must be given to the standard on terms that allow for a competitive market for devices while insuring that the original innovations are properly rewarded. An all-out patent war makes it increasingly unlikely that those conditions can be fulfilled. On the one hand, firms that are typically very active in standards setting may have an incentive to use their “standard essential” patents as weapons in the broader IP conflict, denying proper access to the standards to some of their rivals. On the other, firms with a lesser stake in SEPs have an incentive to claim that their rivals are indeed abusing the market power that the standard setting process bestowed upon them, in the hope that regulators will then limit their rivals’ ability to rely on their SEPs to settle broader disputes.

Because of the potential consumer harm, competition authorities have waded into this – albeit in a rather patchy manner. The main focus so far has been the potentially abusive role of SEPs enforcement, with DG Comp recently opening investigations (e.g. of MMI’s and Samsung’s licensing practices), spurred by a number of complaints, “to assess whether [they] abusively, and in contravention of a commitment they gave to the European Telecommunications Standards Institute (ETSI), used certain of their standard essential patent rights to distort competition in European mobile device markets, in breach of EU antitrust rules”. There is a number of complaints against other parties, and investigations underway at the FTC and the DOJ.

As they focus on SEPs, these investigations naturally run into the problem of meaningfully defining what the “commitments” entered into by SEP holders actually are. In practice, this means putting some content behind the usual requirement that SEPs be licensed at Fair, Reasonable and Non-Discriminatory (FRAND) terms. But there is little consensus about what FRAND terms are. For example, while there is broad agreement that the royalties charged to a licensee should not reflect the additional market power that the patent’s inclusion into the standard confers, there is no accepted methodology to translate this into actual royalty levels. Because of this, both academics and competition authorities have been turning their attention to the process through which SEP holders and potential licensees might settle: if one does not know what the “right” FRAND royalty should be, then it is important that the process through which actual royalties are set is as transparent and balanced as possible. This explains the current focus on the use of injunctions in the investigations conducted on both sides of the Atlantic.

As part of these investigations some parties have also raised concerns about the fairly common practice of setting a royalty based on a percentage of the final market value of the whole product in which the relevant standard has been incorporated. In their view, such a practice imposes an “unjustified tax” on unrelated innovation developed by others. As such SEP holders’ insistence on applying their “traditional” royalty rates to producers of converged devices is seen as “not FRAND”. The purpose of this note is to explore the economic foundations of this concern.

We start by briefly sketching in Section 2 the reasons which have driven regulators to have a look at SEP-related issues on both sides of the Atlantic. We then discuss in Section 3 some economic insights on the question of sharing rents from complementary innovations. Based on these insights, we conclude that claims that “the market value of a device is not the appropriate base on which to calculate a FRAND royalty” are misplaced, as are corresponding concerns about the choice of royalty base in agreements between SEP holders and licensees. Economic analysis is clear that original innovators need to appropriate a share of the value of follow-on innovation, in order for incentives to innovate to be preserved. Properly understood, the issue of whether a given rate is “FRAND”, or whether one should carve out functionalities before applying a given rate, are all about the share of value that original innovators can legitimately appropriate of follow-on innovations. The practical question is how that value

can be pragmatically arrived at, and whether there may be feasible, reasonable ways of limiting rent extraction in a way that does not distort incentives for innovation on all sides. We discuss some of the ideas being discussed (and their limitations) in Section 4.

2. HOLDING UP COMPLEMENTARY INNOVATION THROUGH IPR ENFORCEMENT?

2.1. THE HOLD-UP PROBLEM

Progress in the ICT sector typically takes the form of incremental innovations which build on the existing stock of products and features. Where complementarities are pervasive, and each component is protected by patents, then the “next” inventor making decisions on its own investment in patents or R&D may be discouraged if it faces someone with a large existing patent portfolio, from whom it may be difficult to licence complementary technology. Without ex ante commitment to licensing, or the ability to negotiate ex post effectively with the patent holder, new investments can thus be *held up*.

Standards also complicated matters. They are desirable because they ensure compatibility and interoperability, but the hold-up problem can be particularly severe because standards involve perfect complementarity. FRAND commitments were designed to solve the hold-up problem, as a commitment against exploitation ex post, however what FRAND exactly means is not clear – so the concept does not solve the problem of limiting extraction (and therefore the potential hold up).

But why is this not simply an ex ante contracting problem? If standard setting organisations write the right contracts and those contracts are then properly enforced by the Courts, in principle we should be able to get around the hold-up problem. Why then have these problems not been resolved effectively, and why are regulators now thinking of using competition law to intervene? The reality is that the contracting problem is very difficult when there are many parties with technologies that are complements when it comes to setting a standard but products, such as smart phones or tablets, that are actually competing hard in final markets. As different firms are likely to have different relative strengths in the technology markets and in the product markets, SSO-determined rules for the licensing of SEPs will always be compromises that might not fully address the SSO-related hold up issue. The SSO rules themselves should therefore be seen as the result of a battle between parties trying to appropriate the rents created by the SSO’s activities. In other words, it should be no surprise if SSO rules are as much about *who gets the rents from whom* as about preventing SEP holders from exploiting the market power that they derive from the inclusion of their patents in a given standard.

One should add that, in practice, SSOs also face a trade-off between setting and enforcing rules that prevent the ex post exploitation of the additional market power granted by the standard-setting process and the need to get all firms with important patents to take part in this process. Given that SSOs have no power to regulate the behaviour of firms that decide not to participate, they would not in general be able to design contracts that enforce the first best outcome.

Faced with such shortcomings, regulators appear to have also developed a concern about whether the Court system’s own ability to deal effectively with standard-related hold-up. Because of the SSOs’ own limitations, the task of the Courts is not simply to ensure that some well-defined contract is enforced. Rather, they often find themselves in a position where they have to determine what the obligations of SEP holders were or should have been, where they must be vigilant that their own procedures (such as injunctions) do not allow one of the parties to gain the upper hand in negotiations towards an eventual

settlement and where they might eventually have to rule on what the elusive FRAND conditions for licensing actually are. At the very least, there is a sense that the national Courts do need some form of guidance in taking out such a difficult set of tasks.

2.2. THE FRAND/BASE PROBLEM

In addition to injunctions (which are not discussed further here), one aspect which is attracting attention is the question of the “base” on which the percentage royalty should be calculated - i.e. the “value” to which the percentage should apply. This is *not* a new problem: there is a long recognition that rate and base need to be “commensurate” – i.e. the higher the base the lower the percentage and vice-versa – after all, what matters is the overall royalty payment which is made. We also know that there is great heterogeneity in commercial negotiations between parties about the *structure* of the royalty payment (often starting from a percentage royalty, and then progressing to lump sums, combinations of ongoing royalties and lump sums, compensatory payments, reductions in cash royalties based on reciprocal patent licences or licences for other products, and co-operation/technical support for a whole portfolio).

Still, the issue which has attracted concern – with the decline of feature phones and the explosion of combination devices and of tablets – is that the *value* of these devices has increased significantly as a result of the inclusion of features which are technically unrelated to technologies which read on the SEPs in question. Is then a royalty calculated on the overall market value of the product a way of extracting undue rent from any other innovation that is unrelated to the standard itself? This concern is said to remain (in some form) even when the percentage opening position in discussions with potential licensees in practice is superseded in negotiations. Even though the final terms are the result of negotiation, the royalties per unit realised tend to be positively related to the net price of licensee products that make use of the patented technology. A royalty structure that provides the patentee with higher royalties per unit on higher-priced products – the concern goes – effectively allows him to realise a share of the added value created by the introduction of features that have nothing to do with the patented technology.

Experts for the patent holders in litigation have responded that this argument ignores that the *value* of additional product features can be affected by the *foundational technology*, even if the technology that enables the additional feature does not directly depend on the foundational technology. These experts have explained that *synergies* can exist between the original patents and the gross profits that a licensee realizes from “non-infringing features”. These features can themselves be divided into two subgroups. On the one hand we have what we will call “dependent features” which are those that, even though they do not infringe the SEPs could not be enabled without the technology covered by the SEPs. For example, a feature that relies on the availability of large quantities of data downloaded from the web would be of no interest without access to the patents that covers such data download.. On the other hand we have what we call “symbiotic features”, which are features that could be used without access to SEPs but which become more attractive if access to the technologies covered by SEPs is secured. For example, viewing videos on a smart-phone is much more enjoyable if one can use the latest communication standard than if one has to rely on an earlier, less powerful, specification. One of the experts provided the following example relating to the synergies between the cellular and Wi-Fi technologies covered by certain mobile patents and the various features of Apple’s iPhone:

“...cellular and Wi-Fi capabilities are central to consumer demand for the iPhone. Without those capabilities, the iPhone would essentially be a music player/camera device. Apple could not charge the same price for such a device as it charges for the iPhone. In addition, it is unlikely that such a device could have achieved the same level of sales as the iPhone (even at a lower price).”

Thus while this problem is not new in the context of patent litigation, it is one of the aspects of the issues which are now considered by regulators. In the rest of this brief paper we consider this issue in particular.

3. ON THE QUESTION OF REASONABLE ROYALTIES AND ROYALTY BASE

3.1. OVERVIEW

Expressed in economic terms, the concern is that imposing a variable royalty on the whole value of devices which incorporate multiple functionalities amounts in effect to “taxing” value-creating components of such devices that are not directly related to the patentee’s IP. The broader policy concern is that a royalty scheme that defines payments on such a broad base decreases incentives to invest in future improvements of the devices covered by the licensing agreement. Our discussion below explains that:

- First, it is important to distinguish clearly between the total (expected) *level of royalty payments* made by the licensee, and the royalty *base* that merely determines how this total payment is made. Because the choice of royalty base is independent of the total amount received by the licensor, it is incorrect to conclude that using a broad royalty base gives rise *in itself* to an illegitimate “tax” on the licensee’s income.
- Secondly, the appropriate measure of a technology’s contribution to a complex device – such as a smartphone – is likely to be rather broad. Stated most generally, the contribution equals the difference between the licensee’s actual profits, and those it would realise if its products did not have access to the patented technology. On an *ex post* basis, this means considering the reduction in profits that the licensee would suffer if it were no longer permitted access to the patented technology. On an *ex ante* basis, this means considering the difference between the licensee’s actual profits and the profits it would have realised if the relevant standards had been based on the next-best alternative to the patented technology. Either way, a technology’s contribution would have to include not only the “standalone” value of product features that depend directly on (and could not operate without) the patented technology, but also the *additional value of features that are enhanced by the patented technology*. Moreover, the economic contribution of the technology would also have to include *additional value obtained from features that are independent of the technology, to the extent that total sales of the device are increased by the presence of the technology*. Indeed even complementary products that are not physically connected to the device that embeds the infringed technology *should also enter into the proper valuation of the technology’s contribution*, to the extent that larger sales of the device promoted by the presence of the technology also generate extra sales of complementary goods and therefore additional profits for the licensee.

We further argue that such an approach to the economic contribution of the infringed technology, which implies that the licensee can have a right to participate even in the licensee profits generated by features of the device that do not depend directly on the patented technology, is in the same spirit as the conclusions from the extensive economic literature on sequential innovations. The overwhelming conclusion of this literature is indeed that the first innovator should be given a share of the surplus created by follow-on innovators.

- Thirdly, because the economic contribution of a technology is likely to vary with product revenues, there are good economic reasons to use product revenues as the royalty *base* for determining *how* royalties should be calculated. Royalties calculated on this basis favour efficient risk-sharing

and can help reduce asymmetric information. In that sense, a broad economic contribution directly implies payments metered on a broad royalty base.

- Finally, we turn to the effects of a royalty payment scheme on *incentives to innovate*. It is inevitably the case that *any* payment scheme that expresses royalties as a percentage of downstream sales and/or profits will have some negative effect on incentives to innovate by downstream firms – though these negative effects will be worse when royalties are based on profit than when they are based on sales.² Importantly, the literature on sequential innovation suggests that, taking into account both the value of initial (upstream) innovation and the value of subsequent (downstream) innovation, innovation incentives for the economy as a whole are optimised when the initial (upstream) innovator is provided with a share of the incremental value associated with the subsequent (downstream) innovation.

We also show that in the case of a device that incorporates both features that are enhanced by the technology and other features that are not, the overall negative effects of variable royalties on downstream innovation are not necessarily smaller if the payment is computed on sales/profits attributable only to the enhanced feature rather than on sales/profits for the whole device. *There cannot thus be a presumption that broad-based royalties are more harmful to downstream innovation than more narrowly focussed schemes.*

We finally discuss how regulators may be focusing on the issue of the “royalty base” as a springboard for some broader thinking about the design of patent rights.

3.2. OPTIMAL COMPENSATION OF UPSTREAM PATENT HOLDERS

In a patent system that encourages innovation by providing successful innovators with rewards, it is desirable for the reward to be related to the *economic value* of the innovation. This raises two questions: how should the economic value of an innovation be measured; and what is the optimal share of this economic value that should go to the patent holder? These questions are addressed below. As part of this analysis, we discuss the issues raised when a downstream licensee increases the value of its products by adding “non-infringing features” – i.e. features based on technology where there is no direct connection between this technology and the technology of the patent holder. As discussed before, such features might be either enabled or simply improved by the technology covered by SEPs.

3.2.1. Economic value of the infringed innovation

From an economic point of view, the point of departure for any analysis of reasonable royalty payments must be the contribution of the licensed innovation to the licensee’s profits. The basic principle is straightforward: this contribution is equal to the licensee’s profits with access to the innovation, minus the licensee’s profits without access to the innovation. The difficulty is of course to determine what might have occurred without access, i.e. to get the correct counterfactual.

In principle, the analysis of the economic value of an innovation should capture every possible channel through which the use of the technology might have affected the sales and profits of the infringer. While implementing this approach is never easy, this “overall profit” approach is extremely useful as a

² This is because a royalty based on profits also captures some of the return from cost-reducing efforts on the part of the licensee.

framework in that it helps us identify the various types of effects that should be included in a meaningful assessment of the economic contribution of an innovation.

This economic contribution of an innovation clearly involves any increase in the revenues obtained from selling the products that embody the innovation. Such extra revenue comes both from a price-increasing effect and a quantity-expanding effect: each individual might be willing to pay a higher price for the products cum innovation; and, for a given price, more sales of the products will be made. Methodologies that just look at the price difference between a product with the infringed innovation and one without are therefore incomplete, since they do not take into account that the infringer has likely exploited some of the advantage conferred by the innovation to expand its sales rather than increase the price of its devices.

3.2.2. The economic value of an innovation when the licensee adds “unpatented features”

Let us now consider a situation where the infringing product consists of multiple components, sold as a whole, but where the patent in question reads only on a subset of the product components. As shorthand, let us refer to the other product components as “non-infringing features” that are the result of downstream innovation.

The introduction of non-infringing features and downstream innovation to the analysis has no effect on the general principle for how the economic value of an innovation should be measured. The same principle applies: the economic value of the upstream patented innovation continues to equal the difference between (a) the licensee’s profits when it has access to the patented technology and includes the unpatented features and (b) the licensee’s profits when it does not have access to the patented technology and pursues its next best alternative.

To explain further, let $V(U,D)$ equal the profits of the downstream infringing firm with access to the upstream and downstream innovations. Let $V(U',D)$ equal the downstream profits with access to the downstream innovation but without access to the upstream patented innovation; U' is the best non-infringing alternative to the patented upstream technology. Let $V(U,0)$ equal the downstream profits with access to the upstream technology but without the downstream innovation, and let $V(U',0)$ equal the downstream profits using the best non-infringing alternative to the patented upstream technology (but without access to the downstream innovation).

Using this notation:

- $V(U,0) - V(U',0)$ can be described as the standalone value of the patented upstream technology. It represents the increase in the value of the downstream infringing product created by access to the patented upstream technology, in the absence of the downstream innovation.
- $V(U,D) - V(U',D)$ can be described as the incremental value of the patented upstream technology (assuming access to the downstream innovation). It represents the reduction in value if the downstream infringing product had access to the downstream innovation but had to remove the upstream patented technology.

Adding and subtracting the standalone value to this expression for the incremental value, the incremental value of the patented upstream technology can be described further as:

$$[V(U, 0) - V(U', 0)] + [V(U, D) - V(U, 0)] - [V(U', D) - V(U', 0)]$$

This re-expression highlights that if the downstream innovation adds more value when the upstream patented technology is present – i.e. if $[V(U, D) - V(U, 0)] > [V(U', D) - V(U', 0)]$ – then the incremental

value of the patented upstream technology is greater than the standalone value of the technology and includes some portion of the added value created by the downstream innovation. Put differently, *if there are synergies between the downstream innovation and the upstream patented technology, these synergies can be regarded as part of the incremental value of the upstream patented technology.*³

Note that in determining the portion of the value created by the downstream innovation that should be included in the incremental value of the patented upstream technology, it does not matter whether the downstream innovation is technically related in some way to the patented upstream technology. The only relevant consideration is whether the downstream innovation adds more value when the patented upstream technology is present. If the answer is “yes”, then the incremental value of the patented upstream technology is greater than its standalone value.

A special case arises when, absent the infringed technology, the infringer would simply not have entered the market for the infringing device. In this case, the “extra sales” imputable to the infringed technology are simply *all* sales, so that the economic value of the infringing technology is properly seen as the total profits obtained from the sales of the infringing devices.⁴

Note that this logic can also be applied in the case where the device that embeds the infringed innovation is usually sold with a number of other peripheral products. One can think of computer screens or printers that work with a PC, various external add-ons to smartphones (e.g. headset) or even the applications or content that can be displayed on infringing devices. Given that these peripheral products are clearly complementary to the device itself, the essential economic situation is really the same as in the case of multiple components embedded into a single device. At the very least, the infringed technology is the proper source of the additional profits that the infringer – if it sells those goods – makes from selling a larger number of these peripheral products because he/she sells a larger number of the infringing devices that trigger a demand for these products. Moreover, the increase in sales and/or prices of these devices due to a possible enhancement of their performance because of the infringed technology should also be part of a proper economic evaluation of the contribution of the infringed technology.

It is also worth mentioning that this type of indirect contributions will be larger if the infringer sells its own proprietary system, as the lack of (perfect) compatibility with peripheral products made by others ensures that the infringer will capture the bulk of this correlated increase in peripheral product sales.

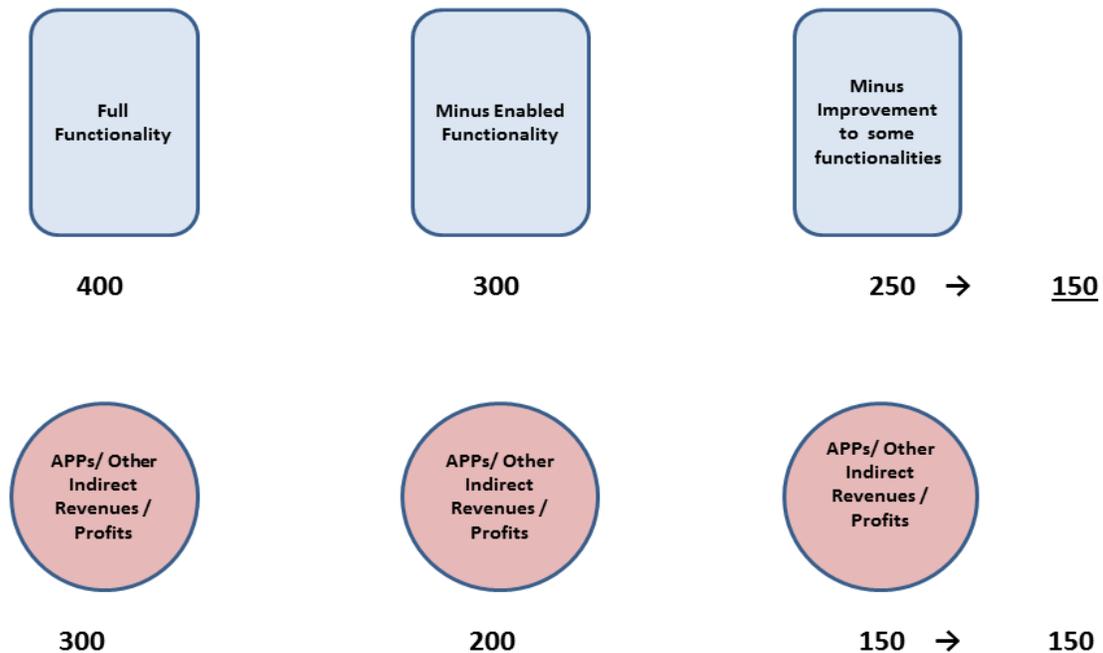
The principles that we have just laid down in general terms are illustrated in the following example presented in figures 1. With full functionality, the smartphone sells for 400 and generates linked sales of 300. If one were to eliminate the “dependent” functionalities, i.e. those that crucially depend on the SEPs, the smart phone would retail for 300 and generate only 200 worth of linked sales. Finally, if one also considers that the performance of some other “symbiotic” features would be lessened without access to the SEPs, the value of the phone drops a further 50, as does the value of linked sales. Overall then, the total revenues from the sale of the phone and the linked sales would decrease by $150 + 150 = 300$ if access to the SEPs could not be secured. This number represents the proper economic contribution of the SEPs. It is therefore that amount that needs to be divided up between licensor and licensee through

³ Downstream innovation also requires compensation, and therefore the introduction in the analysis of no-infringing features may have implications for how the economic value of an upstream innovation is *shared*. This subject is discussed further below.

⁴ In terms of the notation above, where the incremental value of the patented upstream technology equals $V(U,D) - V(U',D)$, if $V(U',D) = 0$ then the incremental value of the patented upstream technology simply equals $V(U,D)$.

negotiations. In this precise sense then, this number is the proper *base* for the determination of royalty payments: the higher this base, the higher the royalty payments to SEP-holders should be.

Figure 1: Value of the smartphone with different levels of functionality and total contribution of the patented technology



Entire Market Value rule

Overall, then, if there are synergies between the downstream innovation and the patented upstream technology, the economic contribution of the infringed technology will always include some elements (such as the increase in the total sales of the device) that relate to the value and sales of the whole device. In this sense, there is in fact a very robust economic logic for using what has become known in the debate before US courts as the “Entire Market Value” (“EMV”) rule to determine the economic value of an infringed innovation.

The EMV rule essentially adopts as the relevant base for royalty calculations the market value of the infringing product in all cases where there is a reasonable presumption that the infringed innovation is “at the core” of the infringing product, and is therefore a major determinant of that product’s overall market value. In this sense the EMV rule is closely linked to the discussion above of the economic contribution of an innovation.

Where some applications of EMV appear to go wrong, and where the EMV rule understandably has attracted criticism, is in failing to draw a distinction between whether an innovation is “at the core” of the infringing product and the incremental value of the innovation. Even if an innovation is “at the core” of an

infringing product, the proper measure of the economic contribution of the innovation remains the incremental effect of the innovation on the firm's EMV – i.e. the difference between the firm's value with access to the technology in question, and its value if it does not use the relevant technology and pursues its next-best alternative. *This principle of measuring economic contribution based on incremental value can be described as the Entire Marginal Market Value ("EMMV") rule.*

3.3. THE APPROPRIATE LEVEL OF COMPENSATION FOR AN IPR OWNER

EEMV provides the correct basis for measuring the economic value of an innovation. The more difficult subject, both conceptually and practically, concerns going from an appropriate valuation of the infringed innovation to an appropriate *level* of compensation for the patent owner. One issue that arises in the context of SEPs (in a situation where the patents are truly essential) is that, because value is zero without access to any of the SEPs, the ex post incremental value of each patent equals the total value of the downstream infringing product. This highlights a conundrum when there are complementary upstream patents – paying a significant share of incremental value to each upstream patent holder may exceed the total value of the infringing downstream product. In this situation, some rule for sharing the value created by the set of SEPs as a whole will be required.

Allocating value across complementary upstream patents is a difficult problem that is beyond the scope of this paper. The focus of this paper is on the situation where a downstream infringing product requires access to a (single) upstream patented technology and demand for the downstream product is also enhanced by downstream innovation. We have already discussed (above) how the incremental value of the upstream technology should be calculated in this situation: the incremental value of the upstream technology equals the difference between the value of the downstream product with access to both the upstream and downstream technologies, and the value of the downstream product without access to the upstream technology.

By the same logic, the incremental value of the downstream technology equals the difference between the value of the downstream product with access to both the upstream and downstream technologies, and the value of the downstream technology without access to the downstream innovation. Given these incremental values, and given the need to provide incentives for innovation at both the upstream and downstream levels, what can we say about the compensation that should be provided to upstream and downstream innovators?

In the absence of any downstream innovative activity on the part of the infringer, or any upstream complementary patents, incentives to innovate would be maximised by allocating the whole EMMV to the upstream patent holder. In practice, however, the downstream licensee may be able to capture through bargaining some of the added value made possible by the patented innovation.

But if the downstream device incorporates other innovations that are not covered by the infringed patent, one must ensure that these innovations are also properly rewarded. Building on the economic literature on sequential innovation, it is clear that to encourage socially optimal innovation at the upstream level, the upstream patent holder should benefit from the increase in demand for the infringing product made possible by the downstream innovation – even if the downstream innovation is unrelated to the upstream patent in a technical sense.

The problem of providing adequate incentives for upstream innovation in foundational technology and downstream innovation in product features is analogous to the problem analysed in the economic literature on sequential innovation where later firms innovate “on top of” earlier innovations. The robust

conclusion of this literature is that social welfare is maximised by allowing the enabling innovator to capture the full standalone value of its innovation (i.e. the value that the initial innovation would create in the absence of the follow-on innovation) plus a fraction of the further value from the follow-on innovation that is enabled by the initial innovation.⁵ While we do not provide here a review of this economic literature, a major insight is that it is socially efficient to increase the reward of the first innovator by reducing (by at least some amount) the private reward of second innovators. The intuition for this result can be seen most clearly in the case where the first innovation has no commercial value by itself but makes the development of commercially valuable follow-on innovations possible. If the initial innovator did not get part of the private reward generated by the follow-on innovation, there would never be any innovation.

3.4. THE FORM OF PAYMENT AND THE ROYALTY BASE

The previous section showed that, if there are synergies between downstream innovation and upstream patented technology, the incremental value of the upstream patented technology includes these synergies. A separate question is *how these synergies should be divided between upstream and downstream innovators*. Building on the literature on sequential innovations, the previous section also showed that *the upstream patent holder should realise some portion of any synergies between the upstream and downstream innovations* in order to ensure optimal innovation incentives at both the upstream and downstream levels.

We now assume that the total *amount* of compensation to the upstream patent holder has been determined, and that T is the total (expected) reward that the infringed patent holder ought to receive. We then consider whether the *form* of this payment matters, in particular with respect to innovation incentives at the upstream and downstream level.

While the royalty payment can of course be made according to a large variety of formulae, we will restrict ourselves to simple and commonly used schemes: lump sum royalty payments, per unit royalties and royalties expressed as a percentage of some royalty base (e.g. value of downstream sales).

In considering this question, note first that one would expect upstream patent holders and downstream licensees to settle on the form of payment that maximises their joint surplus since, in a bargaining framework, this would ensure that both parties end up better off than if they settled on any other form of payment. Thus, one would expect both parties – the upstream patent holder and the downstream licensee – to treat *all aspects* of the form of payment as subject to negotiation. This includes the size of lump sum payment, the royalty rate (if any) and the base to which this rate is applied. It would be a mistake to think of some aspects of the payment form, such as the royalty base, as being “imposed” by the licensor onto an unwilling licensee. If the joint surplus could be increased by changing the form of the payment, it would be in the mutual interests of the parties to do so. Applied to this case, this “Coase

5

See Scotchmer and Green (1995), “On the Division of Profits in Sequential Innovation”, *Rand Journal of Economics*, 26:1, pp. 20 – 33, Matutes, Regibeau and Rockett (1996), “Optimal patent design and the Diffusion of Innovations”, *Rand Journal of Economics*, 27:1, pp. 60-83, O’ Donoghue (1998), “A Patentability Requirement for Sequential Innovation”, *Rand Journal of Economics*, 29:4, pp. 654 – 679, De Nicolo (2000), “Two-Stage Patent races and Patent Policy”, *Rand Journal of Economics*, 33:3, pp. 488 – 501, or Hunt (2004), “Patentability, Industry Structure and innovation”, *Journal of Industrial Economics*, 52:3, pp. 401 – 425, as well as K.E. Rockett, 2011, “property Rights and Inventions”, in B.H. Hall and N. Rosenberg, *The Handbook of the Economics of Innovation*, North-Holland, Chapter 7.

Theorem” observation implies that if royalty payments tend to be calculated by multiplying a royalty rate times the value of downstream sales, there should be a presumption that this is economically efficient.

We should also add that negotiations between the licensor and the licensee tend to occur *after* the licensor has made most of the investment that is likely to affect the total expected joint surplus from the relationship, while it seems natural that the licensee has not yet sunk as large a share of the investments that will help determine the total surplus obtained over the course of the agreement. This implies that, when agreeing on the terms of the licensing contract – and in particular when agreeing on the form of payment – the parties will pay significant attention to the potential effect of such terms on future innovation by the licensee, and will mostly disregard any potential effect on the licensor’s own incentives to innovate. The concern that the form of licensing payment agreed upon might unduly restrict further innovation by the licensee appears therefore rather odd.

Note further than even if the upstream patent holder “calls the shots” and is not willing to negotiate over the form of the license agreement, imposing a royalty base that would restrict future innovation by the licensee would not be in the unilateral interest of the licensee. Even with complete bargaining power, the licensor has an incentive to maximise the size of the surplus that it can appropriate.

Building on this backdrop, we now consider how the form of a royalty payment might affect joint surplus. We first consider the case where the features of the downstream product are fixed and there is no further downstream innovation (“static effects”). This is useful to highlight We then consider the possibility of further downstream innovation.

3.4.1. Static effects

A static framework (where all features of the infringing device have already been developed) is still useful to isolate two main mechanisms through which the form of payment might affect the total surplus shared by the two firms – so that some forms of payment might actually be privately preferred to others.

The first mechanism is the *potential effect of the royalty payment on the marginal cost of production of the infringing devices*. With lump sum payments, or with a royalty-free cross-licensing agreement, the licensing contract does not affect the marginal cost of production of the downstream firm, so the total amount that the licensee must pay has no direct effect on the price of the relevant devices. On the other hand, any scheme such that the total amount paid increases with the number of units of the devices sold increase the marginal cost of production of the licensee and leads to higher consumer prices. This is true whether we consider a per-unit royalty or a royalty expressed as a percentage of a base (such as total sales) that increases with output (over the relevant range).

The effect of an increase in marginal cost on the joint surplus available to the licensor and licensee depends on whether or not the licensor competes with the licensee in the relevant downstream markets. If they do not, then the joint surplus is maximised by making the licensee as competitive as possible, i.e. by choosing a payment scheme that does not increase its marginal cost.⁶ The parties would therefore be both better off with lump sum payments or royalty-free cross licensing. If, on the other hand, the licensor and licensee are rivals in the downstream markets, then their joint profits might be enhanced by raising

6

We are dealing with the single relationship between the licensee and the licensor. The situation where the licensor licenses a number of firms that compete with each other downstream would differ.

the costs of the licensee in order to decrease the intensity of downstream competition. If so, then we would expect the parties to agree on some form of output-related royalty base.

The second mechanism relates to *risk aversion*. Realistically, how much the infringed innovation will actually contribute to the profits of the licensee is uncertain. The parties might agree about the expected value of the EMMV, but its actual realisation cannot be known for sure when the licensing contract is signed. If the parties are risk-averse, then they would prefer not to assume the full risk associated with this uncertainty. In particular, the licensee might be reluctant to settle for a single upfront payment and be left exposed to the risk that the EMMV actually turns out to be much lower than expected. In such a context, the total surplus to be shared between licensor and licensee is maximised by allocating risk between parties so as to minimise their joint “disutility” from risk-aversion. Typically, this is achieved by agreeing a form of payment that ensures that the licensee pays more if EMMV turns out to be unexpectedly high, and pays less if it turns out to be unexpectedly low. This is best achieved by choosing a royalty base that is closely linked to EMMV, and applying a percentage royalty to that base.

A closely related mechanism arises because of *asymmetric information* between the licensor and the licensee. If the licensee has a better idea of how useful his technology is likely to be, then a licensor who is confident in his technology should be willing to “put his money where her mouth is” and agree a form of payment such that her total compensation will be greater when his technology proves effective than if it does not. Again, this involves choosing a base that is closely related to EMMV and applying a percentage royalty to that base.

This suggests there are good reasons why the parties might want to use a royalty base that is closely related to the EMMV. This begs the question of why not use the EMMV itself. From our discussion of the EMMV concept, it should be clear that actually computing the EMMV would be a rather challenging exercise. Not only does it require the specification of an adequate counterfactual, but it would require information on the effects of the licensed technology on possibly a large variety of products, including some products that do not directly incorporate the licensed technology. Pragmatically, then, the best that one can hope for is to find an easily computed and easily verified royalty base that one can reasonably expect to be linked fairly closely to the notion of EMMV.

One candidate for a royalty base related to EMMV could be a measure of the licensee’s *profits on his/her relevant product line*. The principal problems with using profits as a royalty base are measurement and verification. Profits equal revenues minus costs. While revenues might be fairly easy to measure and monitor, costs are anything but. Not only would complex issues of allocation of common costs have to be resolved, but it seems reasonable to assume that the licensee would have much better information about the cost involved and could therefore manipulate the measure to the detriment of the licensor.

3.4.2. Form of royalty payment and economic incentives with further downstream innovation

Downstream innovation and the addition of new product features are likely to keep evolving over the lifetime of a licensing agreement. In particular, new features enabling new functionalities are likely to be added and the performance of existing features might be enhanced. Both adding new features and enhancing existing ones require investment. For simplicity, let us assume that the necessary investment would be undertaken by the licensee. The relevant question then is: would some types of royalty schemes – and in particular some types of royalty bases – be more favourable to continuing investment by the licensed party?

Before we address this question directly, remember that, to the extent that access to the patented upstream technology has positive effects on the increase in downstream revenues and profits resulting from new features, these synergies should be regarded as part of the incremental value of the patented upstream technology – even if the new features do not rely in any technical sense on the upstream patent and even if the downstream innovations in question take place many years after the development of the patented upstream technology. Moreover, as the future development of the device’s functionality is likely to be very uncertain, agreeing on a percentage royalty computed on a base that includes the revenues generated by such future investments helps achieve some risk-sharing benefits.

The main potential problem with a percentage royalty expressed on a base that contains future innovation by the licensee is that such agreements will tend to have some negative effect on the licensee’s incentive to carry out such future improvements. The intuition is straightforward: the licensee is more likely to invest in a project if it can expect to capture 100% of its return, than if it has to turn over X% of the return to the licensor. If royalties were paid to the upstream patent holder on a lump-sum basis, this negative effect on incremental downstream investment incentives would disappear. In this sense, there is a trade-off between the innovation-reducing effect of running royalties and their risk-sharing benefits.

Is the inclusion in the royalty base of sources of value that are not directly related to the infringed innovation (i.e. are not part of EMMV) therefore a problem? Concretely, is the use of a scheme based on the sales of a device problematic given that some of the future improvements on the device are neither enabled nor enhanced by the infringed technology? Regulators seems to be moving towards a view that innovation incentives could be sharpened if the royalty base excluded revenues related to downstream innovations that were unrelated to the upstream patented technology.

We note again that one would expect the chosen royalty payment scheme to reflect the parties’ opinion as to which of these two effects is strongest. A running royalty on a broad royalty base should therefore be seen as a sign that the parties believe that their joint surplus is likely to be higher under such a scheme even if it might have some chilling effect on further downstream innovation.

To help understand how a broad royalty base could result in *more* downstream innovation than a narrower royalty base (i.e. a royalty base that excluded downstream revenues related to features where there was no synergy between the value of these features and the upstream patented technology), let us begin with some notation. As before, let U represent the upstream patented technology. Let D1 be the downstream innovation which has synergies with the upstream patented technology; let D2 be the unrelated downstream innovation. For the purposes of this discussion, let us assume away measurement problems – although in the real world, ease of measurement can be a significant issue and a reason to adopt a broad royalty base.

The value of the downstream product assuming access to the upstream patented technology and access to both types of downstream innovation is $V(U,D1,D2)$. To say that D2 is unrelated to the patented upstream technology means that the value created by D2 does not depend on whether the downstream firm has access to the patented technology:

$$V(U, D1, D2) - V(U, D1, 0) = V(U', D1, D2) - V(U', D1, 0)$$

By contrast, the value created by D1 is affected by whether the downstream firm has access to the patented technology:

$$V(U, D1, D2) - V(U, 0, D2) > V(U', D1, D2) - V(U', 0, D2)$$

Suppose that one were able to exclude the incremental value of D2 from the royalty base, and suppose that this produced an adjusted royalty base that was 50% of the unadjusted royalty base. Under these assumptions, if the agreed royalty on a base equal to the total present and future sales of the device were X%, then the agreed royalty on a narrower royalty base that excludes the effects of “unrelated innovations” would be 2X%. So, with a broad royalty base, incentives to invest on developing both enabled and non-enabled features would be hampered by the X% share of the revenues from investment implied by the royalty. This would decrease incentive to invest moderately and uniformly across various types of features.

If the royalty base were instead calculated by subtracting the value attributable to innovations that were unrelated to the upstream patented technology, then (a) the royalty scheme would have negative effects only on downstream innovations where there were synergies between the value created by the innovations and the patented upstream technology; and (b) the negative effect of downstream innovation of this type would be greater than if a broad royalty base were used (in which case the negative effects on downstream innovation incentives would be spread across the two categories of downstream innovation). As the following examples show, it is not possible to determine which type of royalty scheme/royalty base hampers innovation least without making detailed assumptions on the sensitivity of investments to expected income.

Assume that, in the absence of any running royalty, there would be an overall improvement in the value of the device equal to ΔV , with “enabled features” (D1) accounting for $\frac{\Delta V}{2}$ and “non-enabled features” (D2) accounting for the other $\frac{\Delta V}{2}$. Now assume that a “tax” of X% is small enough that all projects are still invested in but a tax of 2X% kills off all projects. In such a case, innovation would not be affected if a broad base is chosen but it would be halved if a narrower base were chosen since the royalty rate would have to be doubled in order to still give the same appropriate reward to the patent holder. Clearly then, left to their own device, the parties would choose a broader base.

Now assume that a X% royalty reduces the proportion of projects successfully completed from 1 to $\alpha < 1$ and a 2X% royalty reduces the proportion of projects completed from 1 to $\beta < \alpha < 1$. With a broad royalty base, the rate would be X% leading to a total level of innovation equal to $\alpha\Delta V$. With a narrower base, the rate would be 0% on non-enabled features so that the corresponding level of further innovation for these features would be $\frac{\Delta V}{2}$. For the enabled features, the rate would be 2X% and the resulting level of innovation would therefore be $\frac{\beta\Delta V}{2}$. The total level of innovation is therefore equal to $\frac{1+\beta}{2}\Delta V$. So the level of innovation is greater with a narrower base if $\frac{1+\beta}{2} > \alpha$ or $\beta > 2\alpha - 1$. So if a royalty rate of x involved a 10% reduction in innovation ($\alpha = 0.1$), then opting for the narrower base would increase overall innovation if $\beta > 0.8$, i.e. if the decrease in innovation is less than proportional to the increase in the royalty rate.

Therefore, there cannot be an unqualified presumption that basing the payment (not the total value) of royalties on a base that includes “unrelated” features is more nefarious for innovation than opting for a base that includes only the features that are enabled or enhanced by the licensed technology. Overall, and especially given that the broader base is easier to define and measure, it is not at all clear that a broader royalty base would be bad for further innovation.

3.4.3. Inclusion of heterogeneous devices in the royalty base

One of the features of the regulators’ current concerns appears to be a sense that it is not appropriate (or does not make sense) to have the same royalty rate apply across a range of rather different devices.

The economic contribution of a given set of technologies will vary from one device to another. This would depend not only on the proportion of the device's features that are enabled or enhanced by the technologies, but also on the set of complementary products that might accompany the device. In principle then both the total contribution of the licensor, and his contribution "per device sold" could be argued should vary across devices. If the royalty payment scheme specifies the same rate applied to the same type of base (e.g. net sales) for all devices then, clearly, some devices make a larger contribution to the total transfer made to the licensor while others make a smaller contribution to this transfer.

But it would be very hard in practice to discriminate across products and determine precise economic valuations over a wide product range. It then makes sense to adopt a single rate for devices that belong broadly to the same family, while possibly allowing for different rates for sufficiently different devices (e.g. smartphones and game consoles). Whether the benefits of such modulation outweigh the additional transaction costs such an approach would necessarily introduce is something that the parties seem best placed to determine during their negotiations. As for the choice of royalty base, again it would seem that both parties would have the same interest: choose the payment scheme that maximises their joint surplus.

3.5. A (TENTATIVE) SUMMARY

Certain SEP owners have traditionally started negotiations by proposing to license their portfolio of patents for a given rate (2.25%, 2.4%, other) of the net sales of the devices deemed to infringe their intellectual property. Certain prospective licensees have made vocal complaints that imposing a variable royalty on the value of the whole devices amounts to "taxing" value-creating components of such devices that are not directly related to the relevant IP. Translated into policy terms, this may point to a concern that a royalty scheme that defines payments on such a broad base would decrease incentives to invest in future improvements of the devices covered by the licensing agreement.

The nature of the debate is often confused. This paper has sought to clarify that:

First, one must be very careful when defining what is meant by a "tax on unrelated features". The fact that the royalties are *paid* as a function of the whole value of the device does *not* mean at all that the licensor unjustly profits from surplus created by others, including the licensee. In negotiations, the total expected value of the royalty transfers between the two parties will be naturally based on the *economic contribution* that the licensed technology makes to the devices involved. The fact that the total transfer might then be *calculated and paid* as a percentage of a base that does include some value that is unrelated to the infringed technology does *not* affect the total *level* of the transfer and cannot therefore be seen in itself as an additional "tax" of the licensee's income.

Secondly, the level of compensation for the licensor *must be linked to the economic value that its technology creates*. This economic value will typically include not only the standalone value of the features enabled by the technology, but also the increase in the value of features enhanced by the technology (synergies) and the extra profit *on the whole device* arising because of the additional sales which are induced by the increased value of the product incorporating the infringed technology. Overall then, even when the value of a number of features of the end device is completely independent from the infringed technology, the economic contribution of this technology will include some of the revenues coming from the (extra) sales of these independent features. Moreover, the economic value created by the technology also extends to the profit margin on all additional sales of complementary products if their sales are boosted by the increased sales of the device. Thus the proper measure of economic value

definitely extends to some of the value associated with “independent features”, and might also extend to the value of products that are not even physically linked to the device that embeds the infringed technology. The *measure of value* that is relevant to determine the *overall level* of compensation, reached either through bargaining or through other means, is therefore likely to be pretty broad-based indeed.

Then, because of uncertainty and asymmetric information there are good economic reasons why the base used for the *payment* of the agreed upon total (expected) transfer should be linked to the measure of economic contribution of the technology. One would therefore expect the royalty base itself to be rather broad, reflecting the breadth of the contribution to economic value.

Turning to the special case of devices that do incorporate multiple functionalities, we then ask whether a rate of royalty applied to the value of the whole product would have a greater chilling effect on future innovation than a proportionally increased royalty rate defined on the value contributed only by the features that are in fact enabled by the relevant technology. In other words, keeping the total amount of royalty transfers from licensee to licensor constant, is further innovation more limited if the royalty payment is determined on a narrow or on a broad base? We show that innovation is larger with a narrow base *if* the reduction in innovation increases less than proportionally with the rate of royalty required to keep the total transfer unchanged. There is therefore *no general presumption that choosing a broad royalty base would decrease innovation compared to the alternative of a broader base*.

Finally, both the level of the total royalty transfer and the form of its payments (including the base) are determined as part of a bargaining process between licensor and licensee. In this process, both parties have a common interest in maximising the size of the total surplus that they get to share. In particular, the licensor has no interest in unduly limiting future innovation by the licensee if this simply adds to the (expected) size of the pie to be shared. In fact, when it comes to choosing the base for royalty payments, the parties essentially have common incentives: they both want to maximise future innovation, they both benefit from alleviating asymmetric information issues and from sharing uncertainty optimally and they both gain from choosing simple schemes that are easy to implement.

Of course a licensee can still benefit from complaining about the form of payment in the hope that the clear logical and economic distinction between such form and the legitimate total level of transfer might be missed – so that an argument against the broad base would actually lead to a lowering of the overall royalty payment. This would be a logical and economic fallacy.

4. FOOD FOR THOUGHT: ROYALTY BASE AS PART AS A BROADER (RE)DESIGN OF PATENT PROTECTION?

We have so far discussed the role of the royalty base in the sole context of (FRAND) licensing. We have therefore taken the current state of the patent protection system as given. However, it is also interesting to speculate as to how regulators may be thinking of using the royalty base as part of some broader redesign of patent rights.

Academic economists have of course long complained that a “one size fits all” patent system does not make sense as the economic context in which patent rights are exercised differs significantly across various sectors of activities. It is for example widely recognised that very significant patent protection might be needed more in industries like pharmaceuticals, where a significant invention can be covered by relatively few patents, than in ITC-type industries where the development of any useful product relies on a

large number of innovations covered by an even greater number of patents. In such complex industries, extensive patent protection strengthens “patent thickets” and might be damaging to all.

More controversially, one can also argue that patent protection should be adapted to the length of the innovation cycle in the sector of application. There are indeed two good reasons why one might argue that patent life should be shorter in industries with short product/innovation cycles: First, in such fast-moving industries, the rewards from being first to market can already be quite considerable. If such “first mover advantage” is sufficient, then it might suffice to provide adequate incentives to innovate. There might be no need to provide inventors with additional profits through patent protection, or at least through long-lived patent protection. Secondly, while it is clear that early innovators should receive some of the surplus accruing to later innovators building on their initial contribution, one might feel that such a reward should decrease as the technological distance between the first innovation and its successor increases. In fast-moving industries, considerable technological distances can arise quickly.

The traditional objection to the design of “sector-specific” patent rights is that, at the time of grant, it is not possible to identify with any certainty the sectors in which the innovation covered by the patent will eventually be applied. One possibility could then be to grant patents with uniform *maximum* lifetimes of twenty years, but then specify that the effective lifetime will depend on the eventual *fields of use* of the patent. A given patent could then last ten years in one of its applications and twenty in another. The argument would be that since both IP Law and IP-related competition law already make heavy use of the “field of use” concept, such a system would not raise any new conceptual difficulties. Moreover, as fields of use would be determined *ex post*, i.e. when a product relying on the patent enters some market(s), competition authorities may be able to determine which markets the product belongs to and hence which patent protection length it should enjoy.

But we should be very careful about thinking this is where the solution lies. While the “field of use” approach might be conceptually straightforward, it still raises very significant practical difficulties. One could make a possible argument for implementing sector-specific patent length in an indirect fashion, by for instance ensuring that royalty bases become progressively “thinner” over time. Concretely, this could take the form of accepting royalties based on the price of the full devices for a period of a few years but then “freeze” the base at the value of a device with the functionalities available at the end of such a period. Again this raises practical issues as to how the continuing value of this “frozen” device would be computed in later years. What needs to be considered is also that real-life contracts achieved through commercial negotiations *already* often contain clauses which limit the royalty payment over time, for instance, or use a tapered profile. So the suggestion may be in fact redundant for all practical purposes. Still, the question seems at least worthy of debate and further thought.