Electricity tariffs for a distributed future

Draft version 1.0

Deliverable for the project

Projeto Cooperado de P&D sobre
Modernização das Tarifas de Distribuição de Energia Elétrica

Developed for

Paolo Mastropietro, Pablo Rodilla, Carlos Batlle

October 2018

Contents

1. The case for a change in electricity tariffs

Since their creation, at the end of the nineteenth century, power sectors evolved quite linearly. Generation technologies improved, both in terms of efficiency and emissions, but the fundamentals of the functioning of electricity systems did not vary significantly. The generation segment has been, in the majority of cases, organised around very large generation facilities, which exploit economies of scale and are commonly located close to energy and water resources and far away from load centres. Electricity then flows at high voltages through the transmission grid, before being distributed at lower voltages through the distribution network and finally supplied to consumers. This configuration leads to a mono-directional use of the power grid, from generation centres to consumption ones, as depicted in Figure 1. Demand, especially the domestic one, has had a very limited role; it has been commonly considered as totally inelastic and its participation in the market has been limited.

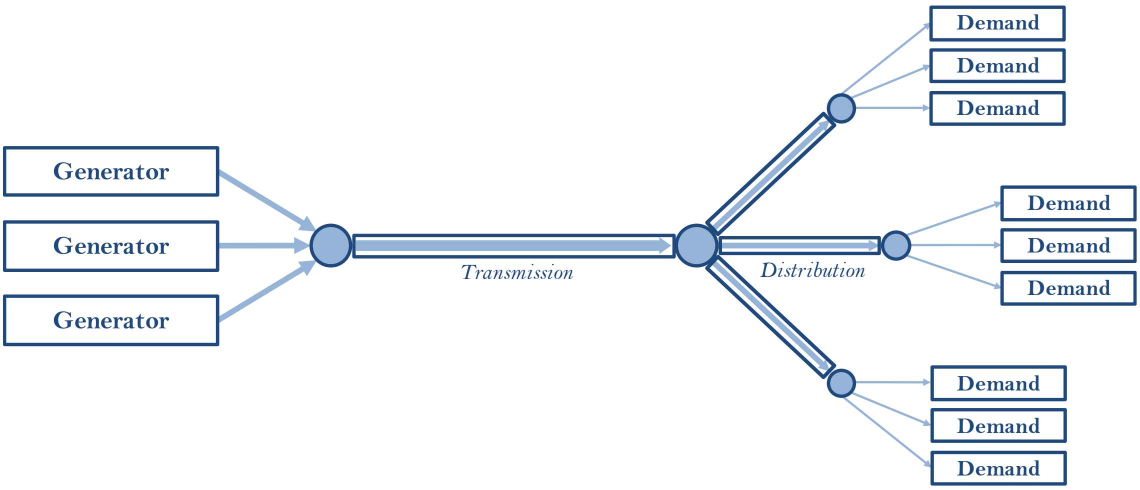

Figure 1. Conventional functioning of a power system

However, electric power sectors, and consequently their regulation, are nowadays at a crossroads. In the last decade, several factors of change have been observed, which envisage a transition to a different model. The most relevant driver of this transition is doubtless climate change. Global warming requires the decarbonisation of the energy industry and much of this effort lies on power sectors. Different types of support mechanisms, implemented all over the world, allowed a fast-paced development of renewable technologies and large-scale renewable power plants are now competitive in many electricity markets.

The most important change of paradigm, however, is not expected to happen in the wholesale market, but at the end of the electricity supply chain. Recent technological developments made available a broad variety of Distributed Energy Resources (DER) and Services (DES), whose deployment may revolutionise the functioning of power systems. A list of these disrupting technologies includes (but is not limited to):

- Distributed generation, either from renewable or from non-renewable energy sources, connected to the distribution network and installed close to demand.

- Electricity storage, whose performance is improving rapidly and whose cost is decreasing.
- Plug-in electric vehicle, which are supposed to be a pivotal element in the decarbonisation of the transport sector.
- Smart meters, whose rollout is close to be completed in many jurisdictions, which allow a more active role of electricity consumers.
- Domotics, which provide residential consumers with a new generation of appliances and management solutions that can drive a more efficient usage of electricity.

The emergence of these services is already causing the appearance of new business models in the electricity market. DES affect all the segments of the power sector, from generation to retail, but they have a large impact especially on networks, which are required to adapt to these changes in order to be able to assimilate these new technologies. The typical power flows through distribution networks are being affected, as presented in Figure 2. The mono-directional network that distributes electricity from the transmission substation to final customers is being substituted by a grid which can be seen as the connection point for several agents, either withdrawing or injecting electricity, and in which the direction of the power flow depends on the activity of these different agents.

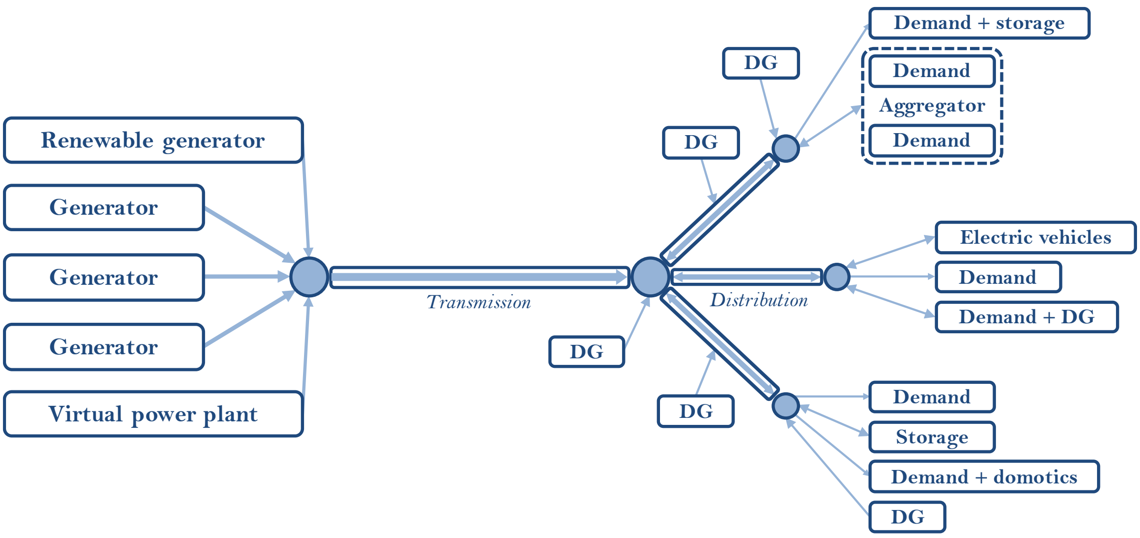

Figure 2. Typical flows in a power system after the introduction of DES

This paramount change of paradigm is already taking place. For the transition to happen efficiently, many elements of power sector regulation must be adapted to the new conditions. A central element of this adaptation is represented by electricity tariffs. Tariff design still reflects, in many jurisdictions, the old “top-down” scheme of functioning. Only minor adjustments have been carried out in the last decade and many consumers still pay an additive volumetric charge ($/kWh) which prioritises simplicity over efficiency in the cost allocation process. These tariffs are not able to “guide” an efficient development of distributed energy services. Centralised and distributed resources are sited in different locations and have different sizes and temporal patterns. The only way for these two categories of resources to jointly and efficiently operate and compete is to establish a comprehensive system of economic signals with adequate granularity to capture important variations in the value of a specific service across time and location. This system of signals is supposed to drive not only the operation but also the planning of new resources and it will probably define the equilibrium between centralised and distributed services in the future. The rest of this document is therefore dedicated to present and analyse electricity pricing methods for an efficient assimilation of DES in future power systems.

2. Guidelines for an efficient tariff design

2.1 Basic principles

The importance of designing economic signals that reflect power system costs and serve as efficient drivers for decision-makers is a well-known topic in the electricity sector. Even in the absence of DERs, there are clear benefits in using efficient economic signals, since they result in more effective response of demand connected at all voltage levels and also in enhanced efficiency of operations and investment in the bulk power system. The growing integration of DERs simply exacerbates the importance of well-designed economic signals and the ramifications of poorly designed signals.

Many authors defined the basic principles that electricity tariff design should follow. Reneses et al. (2013) listed the following principles:

- <u>Cost recovery</u> or <u>economic sustainability</u>. This principle is the essential point of departure for tariff design. Any company that conducts a regulated business must be able to finance its businesses as well as any new investment required to be able to continue to operate in the future.
- <u>Economic efficiency</u>. Efficiency can be achieved by establishing signals that prompt each consumer to use the amount of electricity that is most efficient for the system as a whole.
- <u>Equity</u> in cost allocation. According to this principle, the rates applied should not provide a given competitor (in this case, customers) any advantage over any other within the electricity system. Recently, this principle has been used also to highlight the importance that changes in tariff design take into account the impact they may have on low-income consumers.
- <u>Transparency</u>. The ratemaking methodology should be transparent and the results of its application to each activity within the power sector should be made public. Furthermore, the publication of tariffs and a clear and understandable description of the method used to establish them is the sole instrument available to verify whether or not the other regulatory principles are being honoured.
- <u>Additivity</u>. End-user rates should be the result of the sum of all the cost items applicable to each group of consumers and they should be calculated from the bottom up. Hence, the sum paid by all consumers for each item should be equal to the total recognised cost of that item.
- <u>Simplicity</u>. The aim of this principle is to facilitate, as far as possible, comprehension and acceptance of the tariffs, while attempting not to forfeit other more important principles.

- Stability. The methodology used must be stable, so that power sector agents are subject to the lowest regulatory uncertainty possible.

- Consistency with liberalisation and the regulatory framework in place in each country at any given time.

Unfortunately, not all of these principles can be totally fulfilled at the same time and an equilibrium point has to be met. A useful simile to understand this problem is that of the too-short blanket: in order to completely fulfil one principle, another one will be left "uncovered". This situation is depicted graphically in Figure 3

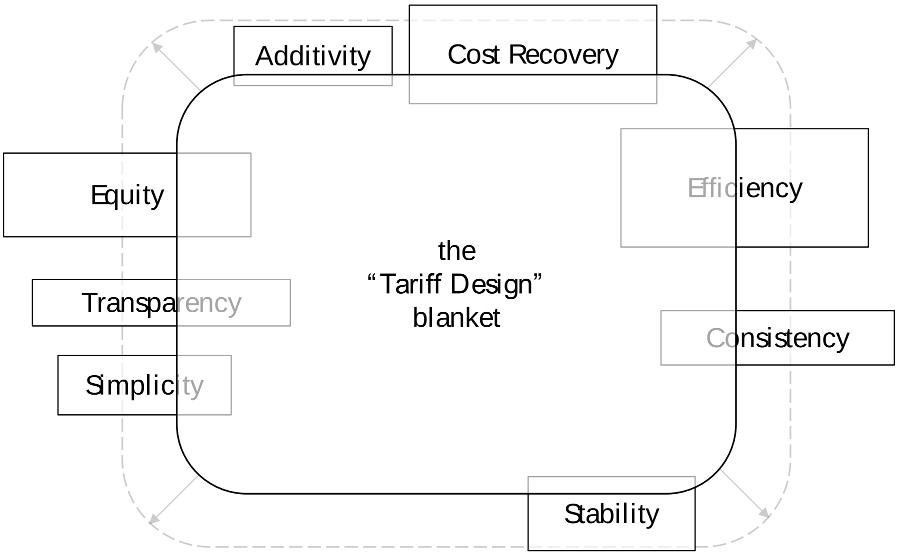

Figure 3. Theoretical principles for tariff design

Other authors directly focus on the principles that should guide tariff design with high penetration of distributed energy services. Among other documents, ENA (2017) lists the following principles (which focus on network charges, but that can be applied to the entire tariff):

- Efficient. Pricing structures that maximise economic efficiency have to be cost-reflective and achieve the following objectives: a) to signal future network capacity investment costs; b) to enable recovery of residual fixed and sunk costs in the least distortionary manner; c) to be equitable through ensuring that price fairly reflects the cost incurred as a result of individual consumer actions; and d) to signal the costs of providing network services relative to the cost of other energy technologies.

- Actionable and simple. To be effective in practice, pricing structures must provide price signals that consumers can understand and choose to respond to.

- Durable and flexible. Pricing structures have to be independent of market, technology, and policy changes. Ideally, pricing structures will have the flexibility to respond to changing external circumstances (e.g., technology, consumer profiles).

- Stable and predictable. To be a workable pricing solution, the pricing structure should enable accurate financial planning by consumers and provide stability over time, avoiding volatility.

- Support retail competition: It is important that the price structure can be applied across different distributors consistently, with limited distortion to retail competition.

Finally, MITEI (2016) identifies two "dominant" principles in tariff design and that should be given priority over other principles:

- Allocative efficiency. Efficient economic signals should try to capture and reflect the marginal or incremental costs of the production and utilization of electricity services. Such signals serve as the key tools with which to coordinate all the planning and operational decisions made by the diverse range of power sector agents to achieve efficient outcomes. For services provided competitively, the corresponding markets generally provide the required prices. For other services, regulated charges must be designed to send efficient signals reflecting each user's marginal or incremental contribution to regulated costs (such as network capacity).

- Sufficiency to recover the regulated costs. Prices and charges should enable the economic sustainability of regulated services via recovery of regulated costs (such as distribution network costs and policy costs). While prices and charges that provide economic signals by reflecting marginal or incremental costs contribute to recovery of regulated network costs, such prices and charges alone are unlikely to be sufficient for full cost recovery. Regulated costs not recovered via cost-reflective prices and charges, the so-called "residual costs", should be recovered in a minimally distortive manner.

The same report remarks how only the second principle is routinely met in practice, while much effort must be spent to enhance the efficiency of future electricity tariffs. MITEI (2016) also identifies two tariff principles that specifically apply to distributed energy services. According to the authors, prices and charges for electricity services should be non-discriminatory and technology-neutral. Any cost-reflective component of prices and regulated charges should be based exclusively on the individual injections and withdrawals at the network connection point, regardless of the specific technology producing those injections or withdrawals. In fact, for the power system, it does not make any difference whether a change in the power withdrawn or injected at a specific time and place has been caused by reducing demand, discharging a battery (or reducing the battery charging), or injecting power from a distributed energy source. The impact on the system is not dependent on the technology involved, thus prices and charges should not depend on technology either.

Another principle that should guide the development of DES is that cost-reflective prices and charges should be symmetrical. A marginal injection at a specific place and time should be compensated at the same rate that is charged for a marginal withdrawal at the same place and time. Non-symmetrical prices and charges would incentivise strategic decisions regarding the location of distributed energy resource behind or in front of the meter.

2.2 The elements of electricity price

An electricity tariff is composed by a combination of prices and charges that have to recover the different cost elements incurred in the power supply chain. Such cost elements can be divided among (MITEI, 2016):

- Electric energy
- Energy-related services, as operating reserves or firm capacity
- Network-related services
- Policy costs, as taxes or costs related to renewable and energy efficiency support

As already mentioned, for cost elements related to services provided competitively, the efficient signal should be conveyed by a price defined in the corresponding market; on the other hand, cost elements related to regulated activities should rely on allocation methodologies based on cost-causality (beneficiary-pays, causer-pays principles). In both cases, efficient economic signals should reflect, whenever possible, the marginal or incremental cost of electricity services.

Each of the cost items listed above will therefore have a different efficient allocation methodology (or a combination of these). Nonetheless, not all costs can be allocated efficiently, or at least not entirely. For some cost elements (as, for example, network-related costs), prices and charges that reflect the marginal or incremental cost of a service are not sufficient to achieve full cost recovery. For other cost elements (as taxes or institutional costs), there may be no obvious application of the cost-causality principle. All these expenses are commonly grouped in the broad category of residual costs. The latter, which, as mentioned, cannot be assigned efficiently, should be recovered in the least distortive manner. Figure 4 depicts graphically the different cost elements and the identification of residual costs.

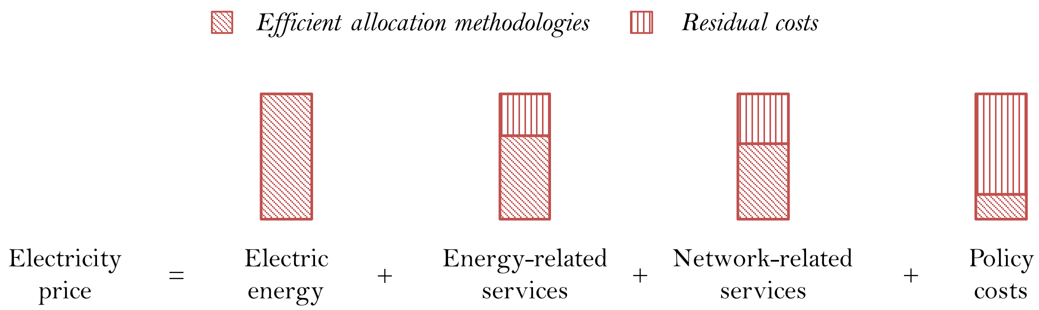

Figure 4. Cost elements of electricity supply and allocation methodologies

Efficient allocation methodologies for each cost element will be analysed in section 2.3, while the allocation of residual costs, and its impact on grid defection, is studied in section 2.4.

Beyond this initial classification, it is important to remark that each cost element should be associated to the relevant cost driver (energy demand, power demand, time of demand, location of demand, connection point, etc.) and charged accordingly in the proper format ($/kWh, $\$/kW_{contracted}$, $\$/kW_{peak}$, $/year, etc.). If, through an efficient allocation methodology, a power supply cost is efficiently assigned to each consumer according to her responsibility in the occurrence of the cost, but then it is charged in the wrong tariff format, then it will still convey an inefficient signal. EURELECTRIC (2016) tried to depict graphically the gap between cost drivers and charge formats in current tariffs in Spain, as shown in Figure 5.

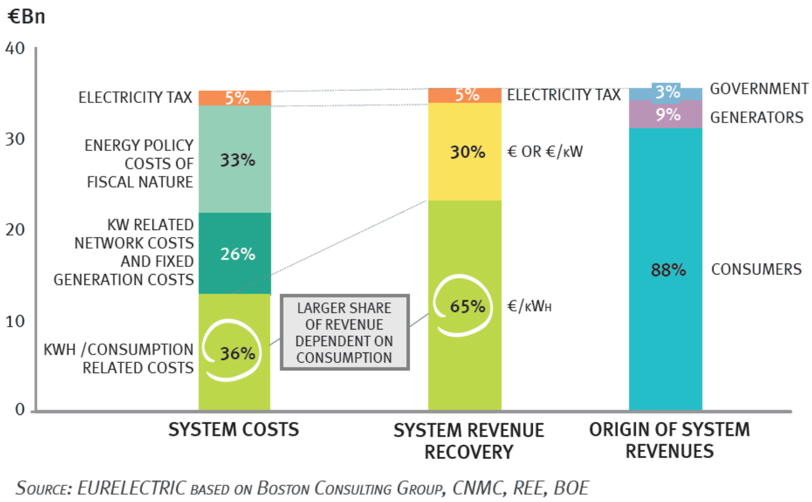

Figure 5. System costs and revenue recovery structure for Spain in 2015; chart from EURELECTRIC (2016)

It must be remarked that the choice of the proper charging format is important not only for those cost elements that can be assigned through an efficient methodology, but also for the allocation of residual costs.

2.3 Efficient allocation methodologies

As already mentioned in the previous sections, the most efficient cost-allocation methodology may be different for different cost elements and this is why electricity tariff should be additive. Following this line of thinking, this section analyses efficient methodologies for each cost item.

2.3.1 The price of electric energy

Marginalism is known to be the most efficient way to price electric energy. The cost of supplying a marginal increment in electricity demand represents an efficient signal for both operation and expansion of the system. Marginal prices, however, can be calculated with very different granularities, both in space and time, and the efficiency of the economic signal will be highly affected by this granularity.

Time granularity

The marginal cost of electricity varies depending on the time when it is consumed, due to load patterns and generation costs, and this variation could be significant. Changes in the cost of energy over the course of a day are incompletely reflected in tariffs based on flat volumetric rates, or even in time-of-use (TOU) rates. The mismatch between a flat or TOU rate and the true marginal cost of energy is particularly apparent at times of scarcity, when large cost savings could be achieved through relatively small reductions in demand.

Moreover, it must be remarked that the marginal cost cannot be accurately estimated in advance. Hourly prices calculated in the day-ahead market are usually computed 12 to 36 hours in advance and they may fail to reflect the actual operating conditions of the power system. Theoretically speaking, therefore, the price of electricity should be calculated for time intervals as short as possible and as close as possible to the real time.

Space granularity

The marginal cost of electrical energy also differs significantly by location within the network. These differences are due to the presence of losses within transmission (and distribution) lines, and the occurrence of congestions in the grid. The average difference among locational prices basically depends on the development of the network and on the distribution of generators and loads within it, but also a properly-sized network can face large differences in locational prices at specific moments in time.

Uniform or even zonal pricing mechanisms may fail in reflecting these location-related changes in the marginal cost of electricity. In theory, the best option is to calculate a price for each node of the transmission network and to apply such nodal price to both generation and consumption located in that node.

Distribution locational prices

Actually, the same reasoning can be applied to the distribution network. The marginal cost may not be the same in the entire distribution grid and, for example, it may be lower closer to the connection to the transmission network than at the end of a feeder, due to electric losses. Distribution marginal pricing is not applied in any power system, but recent studies (Caramanis et al., 2016; MITEI, 2016) demonstrated its advantages both theoretically and through case examples based on simulations[1].

The introduction of distribution locational prices may be very important for an efficient development of distributed energy services. On the one hand, by internalising energy losses in the distribution network, it permits to disclose the real value of a DER in different locations of the distribution grid. On the other hand, it may signal the limitations of the distribution network in assimilating DERs, prompting a more efficient siting of these resources within the distribution grid.

Granularity ranges

Greater granularities have clear benefits in terms of efficiency, but these come at a cost, in terms of, among others, increased computational efforts (Figure 6).

[1] MITei (2016), based on data for Spain, estimates that the consideration of marginal losses in the distribution grid may result in price differences between 11% and 17%, with spikes of up to 40% in low-voltage levels.

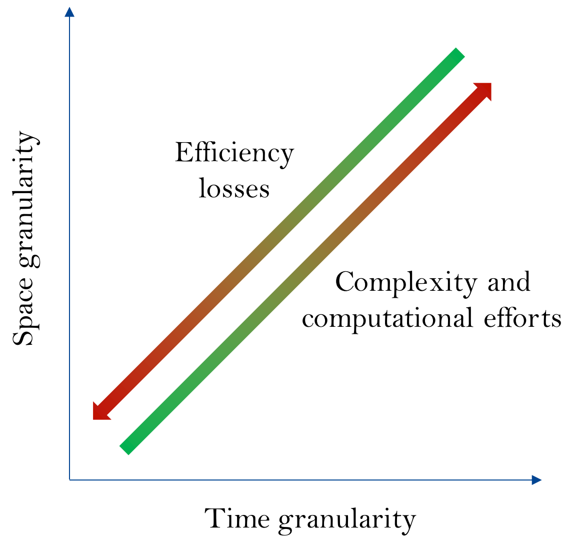

Figure 6. Trade-off between costs and benefits of increased temporal and spatial granularities in electricity prices

A trade-off between benefits and costs of increased granularities must be found and this equilibrium point will depend on the peculiarities of each power system, as its generation mix, the development of its transmission and distribution networks, and the state of deployment of distributed energy resources, among other factors.

2.3.2 The price of energy-related services

Among energy-related services, those that are most commonly found in power systems are operating reserves and firm capacity, which are analysed hereunder.

Operating reserves

The cost of operating reserves usually account for a small percentage of the final cost of electricity. However, this percentage could grow in the next decades, due to the penetration of intermittent resources. Moreover, despite the comparatively small size of the ancillary services market, this may represent a significant economic opportunity for demand-response and distributed energy resources.

An efficient pricing of operating reserves should be based on economic signals that convey the costs of reserve provision and through which the occurrence of reserves scarcity reaches all power system agents. This can be achieved either by facilitating the participation of DERs in the reserve market (eliminating unnecessary limitations and moving these markets closer to real time) or by improving the allocation methodology of the cost of reserves, establishing a system of charges that reflects the cost-causality for both the capacity reservation and its activation in real time and signals the scarcity of operating reserves.

Firm capacity

The cost related to capacity mechanisms, or, more in general, to mechanisms that pursue system adequacy and reliability, is another item of the final electricity cost that is expected to grow in the future (ISO New England, 2015). Also in this case, it is essential that future regulation allows for the participation of DERs in capacity and reliability markets and that the cost of these mechanisms is assigned efficiently, following the cost-causality principle. This efficient methodology varies depending on the characteristic of the system and its scarcity conditions. A capacity-constrained system dominated by thermal power plants has stress events related to the supply of peak demand during certain hours and its capacity mechanism is likely to remunerate the ability of resources to deliver during those hours. In this case, an efficient charge should be proportional to the capacity demand in those same hours. On the other hand, an hydro-dominated system has stress events related to dry seasons that may last for months and an efficient charge should be proportional to the energy consumption rather than to capacity.

2.3.3 Network charges

Network activities are customarily treated as regulated monopolies. The provision of these services must comply with some standards defined by the regulator, who also determined the remuneration for these activities.

The most efficient way to recover network cost is through the calculation of locational prices. Due to losses and congestions, locational prices generate the so-called "network rents"[2]. In the absence of economies of scale in network investment (and if other theoretical hypotheses are fulfilled), it has been proved that locational prices completely recover the network costs (Rubio-Odériz, 1999). However, in practice, network rents can cover only a small percentage of total network cost, due to, among other factors, the lumpiness of transmission investments and risk aversion to power system failures.

The remainder of the total network cost not covered through network rents can still be assigned among power system agents through an efficient allocation methodology. A method commonly applied to electricity networks is the Long-Run Marginal Cost (LRMC). In this context, the LRMC represents the increment in network costs that is caused by a marginal increment of withdrawals or injections in a certain point of the grid in the long run, thus considering the possibility of new investments in the grid. Obviously, the LRMC of the network depends on the time and location of the marginal increment; therefore, the resulting charges are supposed to consider a certain temporal and spatial granularity and to be applied to both generation and demand.

However, the application of the LRMC to network costs present many challenges, as studied in literature and recently remarked by Batlle et al. (2016). The first problem arises at the moment of setting the marginal increment. Mathematically speaking, the expression "marginal" could be interpreted as very small if compared with the actual withdrawals/injections. However, such marginal increment is likely to result in no cost at all, since it could be supplied, most of the times, through the existing network (especially considering the significant lumpiness of

[2] Network rents is a more general expression than congestion rents, since it encompasses also the effect of losses. It must be remarked that network rents result from the application of locational prices and that this applies not only to the price of energy, but also to the price of energy-related services, as operating reserves and firm capacity.

investment that characterizes network industries). No consensus can be found around the size of the increment. Some authors, as for instance Williams and Strbac (2001), proposed 500 MW; some other reports, as FSR (2005), preferred the concept of long-run average incremental cost, which is the cost of meeting large increases in demand, averaged over the size of the increment.

Once long-run marginal costs have been calculated (or approximated) for each group of grid users, they must be applied to specific cost drivers. It is evident that most network costs are driven by the demand of capacity, so the most efficient format is $/kW. However, which capacity should be used for this charge? The methodology with more support in literature is the peak-coincident network charge, through which consumers pay for grid costs according to their contribution to aggregate peak network utilization. Also in this case, several challenges arise when applying this methodology to real-world tariffs. Which is the peak demand? Is the network-wide peak demand or it is assessed at node or voltage level? Is the yearly peak demand or a set of peaks is to be defined? Are these peaks identified *ex-ante* or *ex-post*? A detailed answer to these questions is not within the scope of this position paper, but further discussion can be found in MITEI (2016).

Regardless of the design of the LRMC methodology, not all network costs will be recovered through these efficient charges. The part of network costs not covered through network rents nor through LRMC charges is referred to as residual network costs. It must be remarked that the recent evolution of power sectors may affect the suitability of the LRMC approach for network cost allocation. After several decades of fast-paced growth, many countries are now experiencing decline in electricity demand. Sometimes these decreases were expected by regulators, but sometimes they caught system planners unprepared. Beyond demand declines, the sudden entrance of distributed generation and, more significantly, of electricity storage and demand response may reduce peak power consumption, thus leaving part of the network capacity unused. For these reasons, in the near future, many networks may become oversized and present a significant surplus capacity. In such condition, long-run marginal costs would reflect such surplus (even large increments would not result in the need for new investments) and LRMC charges may decrease sharply (the same applies to network rents), reducing the quota of network costs that cannot be assigned efficiently and that must be treated as residual network costs.

2.3.4 Policy costs

Policy costs are the element of electricity tariffs that more rapidly is growing in many jurisdictions. Figure 7 shows the fast-paced growth of policy costs in the European Union and their main components.

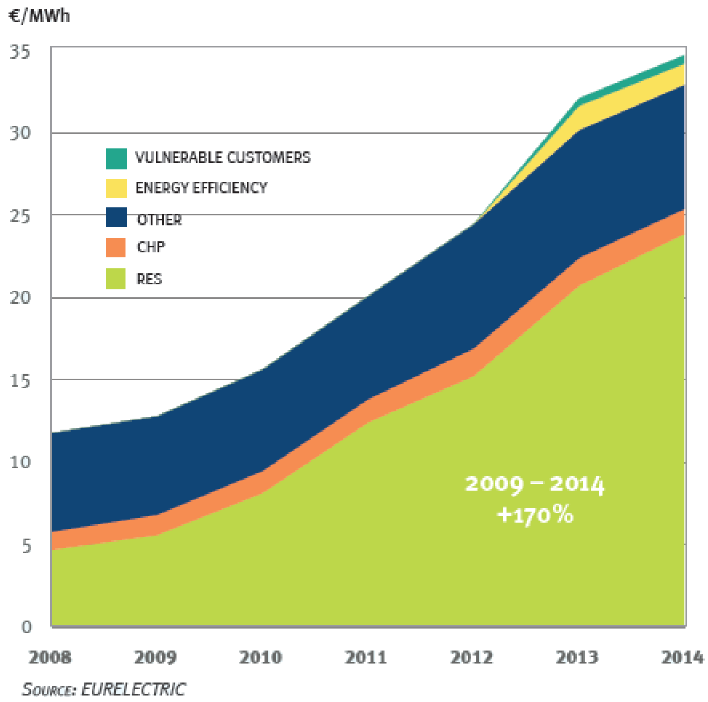

Figure 7. Evolution of policy costs in the European Union; chart from EURELECTRIC (2016)

Policy costs are also the cost item that has been more frequently considered as a residual cost that could not be allocated through an efficient methodology. In some cases, this may be true, since there are cost elements that has no direct cost driver within the electricity supply chain and for which it is impossible to identify beneficiaries (as the institutional costs of system and market operators).

Nonetheless, there are some policy costs, as those related with the support of renewable energy technologies (which, in many cases, account for the largest share of this cost category), that could be assigned efficiently, once again through the methodology of the long-run marginal cost. As explained in MITEI (2016), many jurisdictions have established renewable energy obligations or renewable portfolio standards policies, which require utilities or retailers to produce or procure a percentage of their electricity from renewable sources, or have defined national renewable energy targets expressed as a percentage of electric energy consumption. In these cases, an increase (or a decrease) in electricity demand directly increases (or decreases) the marginal cost of compliance with such policies. For example, with a 20% renewable electricity obligation, increasing total electricity demand by 10 kWh would require an increase of 2 kWh of electricity supplied by renewable electricity sources. A cost-reflective allocation of the cost of renewable support policies, therefore, would entail a volumetric charge calculated as the product of the percentage renewable target and the extra cost of generation from renewable sources.

Batlle et al. (2016) go further in this analysis and shows how both these parameters (the renewable target and the renewable extra cost) can change over time. On the one hand, renewable penetration targets are usually defined as a penetration path, with increasing targets to be achieved each year. Since the renewable LRMC is supposed to be a long-term signal, Batlle et al. (2016) suggest to use the final penetration target. On the other hand, the extra cost of renewable generation varies depending on market price fluctuations and, more importantly, on the learning curve of these technologies. Particularly due to the latter, the renewable extra cost is supposed to decrease over time, until it becomes null when the renewable generation cost achieves the market price (probably in the near future). This trend automatically defines the renewable LRMC and the renewable residual cost, as presented graphically in Figure 8.

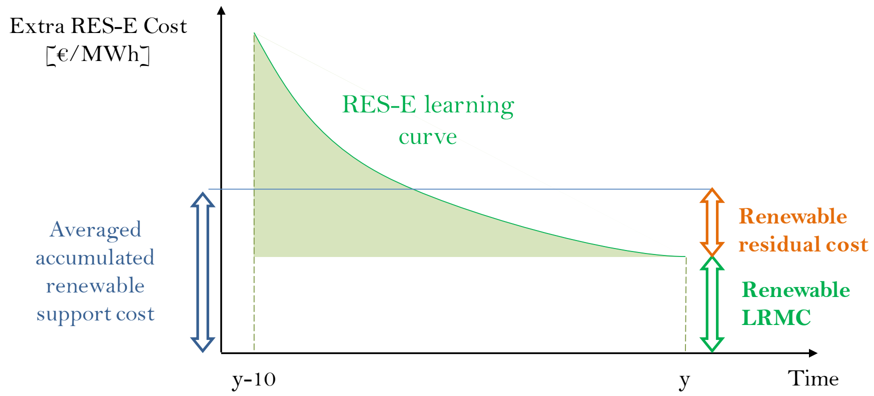

Figure 8. Renewable LRMC evolution and impact on the renewable residual cost; chart from Batlle et al. (2016)

Batlle et al. (2016) also highlight the importance of adequately allocate renewable support costs among energy users. In many countries, the power sector has historically born most of the national emission-reduction burden. If the renewable support cost is fully recovered through electricity tariffs, electricity consumers are clearly subsidising the consumption of other energy sources, which are not required to achieve any reduction target. This may lead to inefficient decisions, for example favouring standard internal-combustion-engine cars over plug-in electric vehicles. In order to avoid such undesired effect, these authors recommend the renewable support burden to be borne by all energy consumers, according to their final energy consumption, or to the total carbon emissions provoked by each energy sector.

2.4 Residual costs and grid defection

Residual costs can be defined as the difference between the recognised costs of a certain activity and the revenues collected through the application of an efficient allocation methodology. In the electricity sector, there are many cost items that can be encompassed, entirely or partially, in this category: residual network costs, residual renewable support costs, subsidies for vulnerable customers, economic support to islands or rural areas with high costs of service, institutional costs (system and market operators), and interests on tariff deficits.

These costs must be recovered through complementary charges on the top of the system of prices and charges defined through the application of efficient allocation methodologies. However, the latter are supposed to convey the most efficient signal for the operation and expansion of the power sector. Therefore, the basic recommendation for the allocation of residual costs is to minimise distortions of the already defined economically efficient signals.

Historically, this has been achieved through the application of the so-called Ramsey-pricing theory[3], or inverse-elasticity rule. The idea is that the complementary charge should modify as little as possible the behaviour resulting from the application of efficient prices and charges. Higher complementary charges should be therefore applied to those agents who change their behaviour the least in response to price changes, i.e., who are less elastic.

However, until recently, Ramsey-pricing has been applied to electricity tariffs through very rough estimations of electricity demand elasticity. Domestic consumers have been considered as almost completely inelastic in the short term and not very reactive to price increases in the long term, so they have been charged a significantly higher share of residual costs. As already mentioned, this line of thinking is no longer valid in a context of deployment of DERs. The latter increase the elasticity of electric demand not only in the short term, but also, and more dramatically, in the long term, as it will be analysed in the next sections. The enhanced elasticity of demand is certainly a positive element in liberalised power sectors, where the ability of demand of responding to price signals is essential to improve the efficiency of the market outcomes. However, problems arise when distributed energy services and the enhanced elasticity they bring are introduced in a system where tariff design still reflects the old regime.

2.4.1 Inefficiencies related to the format

In most jurisdictions, residual costs are nowadays recovered through a volumetric charge[4] ($/kWh). This represents a very inefficient economic signal, because energy consumption is not the cost driver for the cost items grouped in the policy costs. By reducing energy consumption, an end-user can reduce its contribution to the coverage of policy costs, but the latter will not diminish as a consequence of such demand reduction.

The typical example of this situation can be found in those systems where net metering coexists with volumetric charges. Figure 9 presents an illustrative example for a typical 3-kW residential connection in Italy. A typical household consumes 2 700 kWh/year. By installing distributed generation, this household can become a prosumer and reduce its consumption to 1 700 kWh/year. Through this demand reduction, the prosumer stops paying a large part of network costs and policy costs. However, the latter do not decrease and will have to be allocated among the rest of consumers.

---

[3] A literature review on Ramsey-pricing theory can be found in Baumol and Bradford (1970)

[4] Actually, this does not apply only to residual costs, but also to network costs; in fact, in many jurisdictions, the entire tariff is designed as an additive volumetric charge.

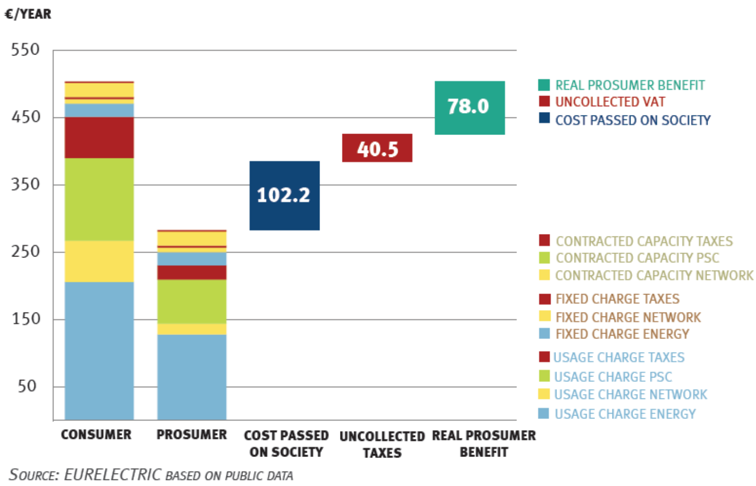

Figure 9. Illustrative example of net metering inefficiencies in Italy; chart from EURELECTRIC (2016)

Furthermore, a volumetric charge to recover residual costs is also highly inefficient if demand is elastic, since it distorts the efficient signal conveyed by the marginal price of electricity and "dilute" the differences introduced by temporal and spatial granularity. Therefore, residual costs would be better recovered through a fixed charge, expressed as a lump sum that could be computed on a yearly basis and billed in monthly instalments. However, this solution has two negative implications:

- Consumers would pay the same charge, irrespective of their energy and capacity demand, and this may raise equity issues (see section 3.1).
- If the fixed charge does not consider the long-term elasticity, it may result in inefficient grid defections, as analysed next (see next section).

2.4.2 Long-term elasticity and grid defection

Distributed energy resources increase the long-term elasticity of end-users, who can make investment decision in response to electricity prices. An extreme instance of this long-term elasticity is represented by grid defection. The combined effect of decreasing costs of both domestic distributed generation (rooftop photovoltaic above all) and small-scale batteries (and/or an onsite gen-set) is reducing the cost of supplying a kWh through a stand-alone system, and this cost is getting closer (at least in the same order of magnitude) to the cost of supplying the same kWh through the grid, when the classic tariff design is taken into consideration[5]. However, this apparent competitiveness stems, most of the times, from an

[5] This is not completely true. A proper economic assessment should consider also the cost of non-served energy. A stand-alone system (a properly-sized rooftop PV panel and a battery) has a loss-of-load probability much higher than a modern interconnected power system. Depending on the value assigned to non-served energy, this lower reliability would affect the economic comparison, reducing the competitiveness of stand-alone systems.

improper allocation of residual costs. A grid defection, in this context, would be beneficial for the end-user, but would be inefficient from a system-wide perspective.

In order to avoid inefficient grid defections, Batlle et al. (2016) propose the application of thresholds to residual cost allocation. Figure 10 compares the tariff for grid supply, represented as the summation of generation costs (and other costs related to competitive activities), long-run marginal costs (including network and renewable LRMCs), and residual costs, with the cost of two stand-alone systems. Stand-alone system 1 is a theoretical and extremely cheap system that supplies electricity at a cost lower than the summation of efficient generation, network, and RES-E support costs. Apart from some exceptional cases (isolated or very unreliable interconnected systems), such a scenario cannot be found in practice with the current prices of photovoltaic panels and batteries and it is also quite unlikely for the near future. However, if a stand-alone system with these characteristics existed, it would produce at a cost lower than the overall marginal cost of producing electricity from the grid. In this case, grid defection would not be detrimental for the power system, since it would be fully cost-efficient. No tariff threshold is to be applied in this case.

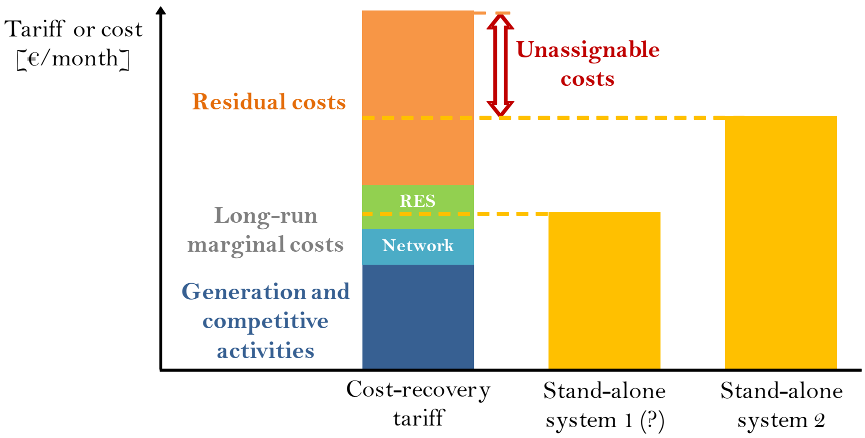

Figure 10. Stand-alone systems cost compared to tariff for grid supply

A completely different situation is depicted in Figure 10 for stand-alone system 2. This system is producing at a cost higher than the overall long-run marginal supply cost from the grid. Therefore, a grid defection from this user would be inefficient from the economic point of view, since the electricity produced by this stand-alone system would be more expensive than the one withdrawn from the grid. Inefficient grid defection is caused, in this case, by an improper allocation of residual costs. In this context, it must be also remarked that, as soon as grid defections start taking place, tariffs must be readjusted in order to fully recover residual costs, causing an increase in electricity bills for remaining customers and worsening the problem, an extreme version of the so-called “death spiral” for electric utilities.

The cost of stand-alone system 2 then must become a threshold[6] that is not to be exceeded by the inappropriate allocation of residual costs. The share of residual costs beyond such a threshold should be treated as unassignable costs.

2.4.3 How to recover unassignable residual costs

Such unassignable costs must be recovered, in order to guarantee the financial stability of the power sector and adequate funding for public policy objectives, but these revenues cannot be recovered through conventional components of electricity tariffs. Different alternative options have been proposed in literature for the collection of these costs:

- Move part of the residual costs to the state budget and collect them through conventional taxes. As already mentioned, renewable support costs permit to achieve objectives that go beyond the electricity sector and could be included in the state budget. MITEI (2016) states that also the residual costs of electricity networks may be paid by taxpayers.
- Embed unassignable residual costs in real-estate taxes, proportionally to the property tax currently paid. The real-estate tax is used in this proposal because it is considered as a good proxy of the wealth of the household and of its electricity consumption. Therefore, this solution would allow to charge more residual costs to end-users with higher consumption, but without affecting the efficient economic signals and without the risk of grid defection.
- Introduce a specific exit fee for grid defection, by which grid defectors pay their share of unassignable costs. If the fee is conceived as a lump sum, it should be calculated as the summation of the expected shares of unassignable costs along a predefined period of time. This alternative is difficult to be applied in practice (especially as regards the calculation of the exit fee) and its implementation may be more than contentious for legal reasons.

Each of these alternatives has pros and cons and each system needs a tailored solution. The selection of the methodology for the allocation of residual costs will dramatically affect the potential for electrification of the energy sector. The latter is claimed by many experts as the main strategy to reduce greenhouse gas emissions (NREL, 2017), but an inefficient electricity tariff design would definitely hamper this transition.

2.5 A potential roadmap for transition

The guidelines presented in the previous sections represent a dramatic change of paradigm in tariff design. Furthermore, each power system has its own characteristics and its own regulation and some of the recommendations expressed here may not be applicable or may result in very reduced benefits. In this sense, MITEI (2016) proposes a list of recommendations ordered

[6] Batlle et al. (2016) propose the application of the "marginal threshold". In fact, the cost of a stand-alone system varies depending on many factors, but the threshold should be unique and the lower stand-alone cost should be considered. Moreover, these authors remark how the threshold should be subject to frequent revisions, since the cost of a stand-alone system may evolve rapidly in the next decade.

according to some sort of rate between their expected benefits and their expected implementation costs. Such list is resumed hereunder:

- Remove residual costs from the volumetric component of the tariff and charge these costs through a fixed charge determined through some proxy of the end-user wealth, always keeping in mind the need to avoid inefficient grid defections.
- Smart meters, whose rollout is almost completed in many power systems, allow to easily expose customers to hourly or sub-hourly energy prices and this exposure could be highly beneficial, especially in capacity-constrained systems.
- Extend wholesale energy prices to all voltage levels of the distribution network through loss factors (this can be done even if a uniform price is computed in the wholesale market). Loss factors would be time-dependent and, at time of scarcity, they may significantly increase the energy price at the end of a feeder, signalling the comparative benefits of DER installation.
- Smart meters also allow for an easy application of coincidental peak capacity charges for firm capacity and for responsibility in network investment.
- Calculate nodal prices at the transmission level and consider their application not only to generation but also to price-responsive demand and DERs in general.
- Introduce detailed locational signals at the distribution level, capable of guiding DER installation towards those zones of the distribution network where they would be beneficial.

3. Advanced regulatory issues

3.1 Distributional effects

Section 2 presented guidelines for designing efficient tariffs that reflect, as much as possible, cost-causality principles. These guidelines should be used to reform the actual tariff design, which, in many jurisdictions, recover all costs through a single additive volumetric charge, regardless of the actual cost drivers. However, as anticipated in section 2.4.1, these changes in the tariff format may have an impact on how electricity costs are distributed among different classes of consumers.

In particular, changing the format of the residual-cost charge, from a volumetric to a fixed one, is expected to result in higher electricity bills for low-income consumers. In fact, the latter are supposed to have lower electricity demands than wealthier households and the volumetric charge allows them to pay a lower share of residual costs. Through a fixed charge, in principle, all end-users pay the same amount, irrespective of their incomes. This may cause a sharp increase in the bill of low-income consumers.

In order to avoid this effect, without scarifying the efficiency of the tariff design, the new system of prices and charges may be complemented through "equity" measures, which can be applied during a transitional period or permanently. MITEI (2016), for example, proposes to complement the new tariffs with means-tested rebates for low-income consumers; such rebate could be provided as a lump sum, thus not distorting the efficient economic signals.

Another alternative would be to introduce an "uneven" fixed charge. Residual costs would still be recovered through a fixed charge, but the lump sum would not be the same for all end-users. Following the line of thinking already presented in section 2.4.3, a proxy of wealth, as the real-estate tax, could be used to define a fixed charge that grows proportionally to the wealth of the household. If setting a charge of the electricity bill through external parameters is not possible, another option is to link the fixed charge to the historical consumption. This way, the fixed charge would be proportional to electricity consumptions (thus, somehow, to wealth) exactly as it was with a volumetric charge, but without affecting the economic signal and without creating wrong incentives for the installation of distributed resources[7].

3.2 Long-term signals for distributed energy services

Distributed energy resources require investments that, from a household perspective, may be considered as capital-intensive. This requires a careful analysis of the risk involved in this investment and on the risk perception of the agents affected by it. An optimal and fully-efficient tariff design may provide end-users with a whole set of economic signals which should guide their behaviour both in the short and in the long term. Such system of efficient prices and charges may indicate the convenience of carrying out a certain investment in distributed energy resources. However, if end-users are risk averse, they may decide not to invest even if the expected value of such investment is positive.

Sources of risk for DER investment come, among other factors, from the lack of complete information (an end-user cannot predict the behaviour of the Distribution System Operator, or DSO, and of other agents in the same distribution grid), from regulatory risk, or from the long-term volatility of the short-term economic signals conveyed by prices and charges. In order to hedge these risks and to prompt investments that are beneficial from a system-wide perspective, it may be necessary to introduce mechanisms that produce long-term investment signals. Some services that can be provided by a DER may already rely on a long-term mechanism, as with firm capacity; in this case, distributed energy resources should be able to participate in such mechanisms and hedge their risk through a long-term contract.

However, many DER investments are particularly beneficial at the distribution level. In current power systems, at this hierarchical level, there is no long-term signal available (even if some pilot projects are being launched; see NYISO, 2017). This situation may create a double source of uncertainty; on the one hand, the distribution system operator cannot predict accurately the installation of DER and, therefore, cannot plan the gird expansion efficiently; on the other hand,

[7] As already mentioned, with a volumetric charge, a wealthy household can install a distributed energy resource, decrease its energy demand and reduce the share of residual costs it is paying. This could not happen with a fixed charge based on historical consumption, because the fixed charge would not decrease when current demand decreases.

end-users cannot predict investment decisions from other agents in the network, while these decisions may have a dramatic impact on the value of the DER. In order to eliminate this uncertainty and provide agents with long-term signals, it may be envisaged the organisation of periodic auctions for forward network capacity options. Through this mechanism, the DSO may communicate to end-users the marginal cost of forthcoming network expansions, thus creating incentives for network users to reveal their willingness to pay for forward options to use network capacity or to avoid such charges by installing DERs. In this context, DERs would commit to some sort of firm call option which network utilities can exercise at periods of network congestion, up to the contracted firm capacity quantity.

4. References

Batlle, C., Chaves-Avila, J.P., Rodilla, P., Mastropietro, P., Gomez, T., Perez-Arriaga, I.J, 2016. Regulated and Policy Charges and Electricity Bills for a Distributed Future: Efficient Price Signals for Increasingly Elastic End-Users. Working Paper MITEI-WP-2016-09. Cambridge, MA. September 2016.

Baumol, W., Bradford, D., 1970. Optimal Departures from Marginal Cost Pricing. American Economic Review, vol. 60, iss. 3, pp. 265-283.

Caramanis, M., Ntakou, E., Hogan, W., Chakrabortty, A., Schoene, J., 2016. Co-Optimization of Power and Reserves in Dynamic T&D Power Markets with Non-dispatchable Renewable Generation and Distributed Energy Resources. Proceedings of the IEEE, vol. 104, no. 4.

ENA, Electricity Networks Association, 2017. A Guidance Paper for Electricity Distributors on new pricing options. Working document published in August 2017.

EURELECTRIC, 2016. Retail Pricing for a Cost-effective Transition to a Low-carbon Power System. Report published in June 2016.

FSR, Florence School of Regulation, 2005. A Study on the Inter-TSO Compensation Mechanism. Report released on October 2005.

IRENA, International Renewable Energy Agency, 2017. Adapting Market Design to High Shares of Variable Renewable Energy. IRENA report, published in May 2017.

MITEI, Massachusetts Institute of Technology Energy Initiative, 2016. Utility of the Future: An MIT Energy Initiative Response to an Industry in Transition. Report developed in collaboration with IIT-Comillas, published in December 2016.

NREL, National Renewable Energy Laboratory, 2017. Electrification & Decarbonization: Exploring U.S. Energy Use and Greenhouse Gas Emissions in Scenarios with Widespread Electrification and Power Sector Decarbonization. Technical report NREL/TP-6A20-68214, published in July 2017.

NYISO, New York Independent System Operator, 2017. Distributed Energy Resources Roadmap for New York's Wholesale Electricity Markets. Report published in January 2017.

Reneses, J., Rodríguez, M.P., Pérez-Arriaga, I.J., 2013. Electricity tariffs. Chapter of the book *Regulation of the Power Sector*, 2013, Springer, Pérez-Arriaga, Ignacio J. (Ed.). ISBN 978-1-4471-5033-6.

Rubio-Oderiz, F.J., 1999. Methodology of Transmission Network Cost Allocation under a Competitive Regulatory Framework. Doctoral dissertation, Universidad Pontificia Comillas, Madrid, Spain.

Williams, P., Strbac, G., 2001. Costing and Pricing of Electricity Distribution Services. Power Engineering Journal, vol. 15, iss. 3, pp. 125-136.