Energy market investigation

Wholesale electricity market rules

27 February 2015

This is one of a series of consultative working papers which will be published during the course of the investigation. This paper should be read alongside the updated issues statement and the other working papers which accompany it. These papers do not form the inquiry group's provisional findings. The group is carrying forward its information-gathering and analysis work and will proceed to prepare its provisional findings, which are currently scheduled for publication in May 2015, taking into consideration responses to the consultation on the updated issues statement and the working papers. Parties wishing to comment on this paper should send their comments to energymarket@cma.gsi.gov.uk by 18 March 2015.

The Competition and Markets Authority has excluded from this published version of the working paper information which the Inquiry Group considers should be excluded having regard to the three considerations set out in section 244 of the Enterprise Act 2002 (specified information: considerations relevant to disclosure). The omissions are indicated by [✂].

Contents

Page

Introduction

1. This working paper examines three aspects of the design of the GB wholesale electricity markets:

   (a) mechanisms governing the dispatch of wholesale electricity (centralised dispatch versus self-dispatch);

   (b) the influence of the cash-out rules,[1] as proposed in the reforms, known as the electricity balancing significant code review (EBSCR), on market participants'[2] short-run costs; and

   (c) the interaction between the investment incentives provided by two separate regulatory mechanisms designed to ensure that sufficient capacity is available at time of system stress, the capacity auctions and the pricing of cash-out (and more specifically the pricing of Short Term Operating Reserve (STOR), known as reserve scarcity pricing (RSP), which is a proposed part of the EBSCR).

2. We assess the degree to which these aspects of the design of the GB wholesale electricity market might preclude a well-functioning market. At this stage:

   (a) we do not find that centralised dispatch has significant competitive advantages over self-dispatch, but intend to investigate certain specific issues further;

   (b) we intend to investigate further whether the reformed cash-out rules will minimise generators' short-run costs;

   (c) we find pro-competitive arguments in favour of the capacity market, but we raise the possibility that Ofgem's cash-out reforms, especially RSP, introduce a risk of overpayment for capacity, which we intend to investigate further.

Self-dispatch versus centralised dispatch

3. The current dispatch mechanism in force in Great Britain, introduced by the NETA/BETTA reforms,[3] was designed as a self-dispatch wholesale electricity market. This contrasts with the system that it replaced, the pool,[4] which was

[1] If a market participant generates or consumes more or less electricity than they have contracted for, they are exposed to an imbalance price, or 'cash-out', for the difference.
[2] That is, generators and suppliers.
[3] Legal and regulatory framework working paper.
[4] Legal and regulatory framework working paper.

centrally dispatched. This section considers the impact of each design on competition.

4. In a centralised dispatch system, generators and flexible loads[5] tell the system operator (SO) the prices at which they are willing to supply to the system and the prices at which they are willing to reduce consumption. These bids come with detailed technical information of constraints in plant operation. The SO determines what it believes to be the least cost way of matching supply and demand and communicates a planned running order to each participant, often one day ahead of time. In determining the running order of plant, the SO also determines the system price in each period that is consistent with that running order. Centralised dispatch exists in the Australian national electricity market (NEM)[6] and in some form in most deregulated markets in the USA.

5. Under a self-dispatch system, buyers and sellers of electricity contract ahead of time for their anticipated demand at prices that are bilaterally negotiated or determined through demand and supply matching on public exchanges. Generators and suppliers prepare operating plans for their anticipated physical behaviour or that of their customers. The parties communicate their anticipated physical behaviour and their contractual position to the SO. The SO takes central control of balancing supply and demand close to real time, at a point known in the industry as 'gate closure'. In this sense, the system is not truly self-dispatched – there is always some point at which central control is asserted. After the fact, discrepancies between what parties physically did (actual delivery or offtake) and their contractual positions are 'cashed-out' at prices determined administratively by the SO.[7]

6. In Great Britain, the SO[8] receives notification of the physical and contractual position of each party one hour prior to operation, on which basis it will assess whether the system is at risk of imbalance. It will intervene if it predicts a discrepancy between the amount of electricity produced and demanded during a certain settlement period. The SO has the obligation to balance the system at minimum cost. The role of the SO in determining the operating plans of plants is shared to a greater degree with plant operators in a self-dispatch system compared to a centrally dispatched system. Internationally, GB is unique in having taken self-dispatch to this level. Most international

[5] 'Flexible loads' are consumers that have the flexibility to reduce consumption at short notice in response to market signals.
[6] In the Australian NEM, dispatch is determined 5 minutes ahead of time, rather than one day.
[7] Effectively, any physical shortfalls or excesses compared to contract are 'made up', or balanced, by contracts with the SO in the process known as cash-out.
[8] The exact definition and duties of an SO vary from system to system. In the GB system, National Grid Electricity Transmission plc. carries out the SO role. Under the pool, centralised dispatch was carried out by National Grid.

systems are hybrid, in which firms may either announce their operating plans to the SO or allow the SO to dispatch them.

7. In order to assess whether self-dispatch systems impede effective competition, we consider in turn the main arguments that may inform our assessment:

   (a) self-dispatch increases incentives to vertical integration;

   (b) self-dispatch reduces technical efficiency;

   (c) self-dispatch reduces price transparency;

   (d) self-dispatch increases transaction costs for new entrants and smaller players; and

   (e) self-dispatch precludes locational pricing.

Self-dispatch increases incentives to vertical integration

8. Some have argued[9] that self-dispatch created an incentive for parties to ‘contract with themselves’ – effectively, to vertically integrate – which in turn might lead to competition concerns. The reason proposed is that the designers of NETA/BETTA were worried that participants would not have sufficient incentive to enter bilateral contracts ahead of time and that this would increase the burden on the SO. The designers therefore introduced a mechanism whereby electricity purchases or sales that were not covered by bilateral contracts would be settled at a price that was **by design** unattractive: finding oneself in imbalance would be costly.[10] SSE has argued that this design decision was compounded by the fact that at the introduction of NETA, there was material uncertainty about the manner in which energy trading would take place and the commercial consequences of this highly significant change in market operation.

9. These arguments are plausible partial explanations for the attractions of vertical integration at the time of NETA’s introduction. However, these factors do not appear to apply in the current market conditions. We have found that:

[9] For example, Dieter Helm (2014) *The return of the CEGB?* The argument has been picked up by the Institute for Public Policy Research (September 2014) *A new approach to electricity markets*.
[10] The design philosophy that sought to make bilateral contracting more attractive by design is confirmed and described by Professor Stephen Littlechild (January 2012) *Response to Ofgem’s consultation on electricity cash-out issues*.

(a) near-term bilateral and exchange-based energy markets are liquid, so there is no real fear that parties will be unable to contract with third parties in the run-up to gate closure;[11]

(b) levels of self-supply, especially in near-term markets, are low, which suggests that vertical integration is not a substantial advantage in this respect; and

(c) The cash-out rules which made uncontracted positions unattractive by design have been gradually taken away, with the latest reforms (described in more detail below) making cash-out a 'fair' market with a single price, therefore reducing any incentive to integrate vertically in order to avoid cash-out.

10. We therefore do not believe that a self-dispatch system in the current market conditions provides significant incentives to vertical integration. The fact that about 30% of generation and 12% of supply are in the hands of non-vertically integrated firms suggests that any such remaining incentives are not insuperable. Furthermore, current trends may indicate that the level of vertical integration in the GB wholesale electricity market is decreasing:[12]

(a) E.ON and Centrica have both announced substantial moves towards de-integration; and

(b) the first capacity market auction attracted a number of generation-only bids for new entry.

Self-dispatch reduces technical efficiency

11. In most commodity markets, prices consistent with technical efficiency[13] are discovered through bilateral competition or, sometimes, through bids and offers on exchanges: a less efficient provider cannot offer a price that is attractive to buyers. In the case of homogeneous commodities bought in wholesale markets by sophisticated agents, this process is likely to work quite well. Traditionally in electricity systems, technical efficiency was achieved centrally without the use of market mechanisms by calculation of cost-minimising operational plans. Some deregulated electricity markets retained the centralised calculation of optimal operation but required firms to compete by offering attractive input values to the centralised calculation (centralised dispatch). All electricity systems maintain some degree of centralised dispatch

---

[11] Liquidity working paper.
[12] We have not concluded under any of our theories of harm that current levels of vertical integration per se cause any harm for competition.
[13] 'Technical efficiency' refers to the property of minimum-cost production for the economy as a whole.

decision – in the minimal case that Great Britain operates, this relates to residual balancing energy needs only.

12. The evidence we have seen suggests that bilateral trading is leading to close to technically efficient operation of the system. Several parties have shared with us their modelling approaches based on cost minimisation by the SO and their close fit to actual prices. We have reviewed these models in the context of our work on unilateral upstream market power[14] and we find that their results are convincing. If bilateral contracting were leading to systematic technical inefficiency, we would expect to see this in systematic deviations of forecast and actual prices. We do not see these in the model calibration results. Our own wholesale price modelling[15] suggests that day-ahead prices are well forecast by a cost-minimising assumption.

13. InterGen has suggested that some of its combined-cycle gas turbine (CCGT) plant runs less frequently than less efficient plant owned by some of the large vertically integrated companies, and suggests that this is evidence of technically inefficient operation. There are many possible explanations for the data that InterGen highlights. One possible explanation could be the different treatment of maintenance and other fixed costs: if these are included in InterGen's price offers but not in the comparators', technically similar plants could end up with very different usage.

14. Another possible explanation suggested by InterGen is that the vertically integrated firms are better at forecasting the costs of balancing that are spread over each half-hour and therefore better able to forecast the profitability of a plant operating or not in a given half hour. It could be that InterGen is forecasting higher balancing costs than it actually incurs, leading it to be more conservative in its offer of capacity to the market than would be comparable plant, leading InterGen to run its plants less often.

15. The InterGen example may suggest that self-dispatch puts a greater burden on each party to forecast imbalance costs than would a centralised dispatch, and this may contribute to some technical inefficiency. However, it should be noted that most centralised dispatch systems still need some method for recovering the costs of unexpected changes to operating plans.[16] The benchmark against which to compare the GB system would therefore not be one of perfectly efficient centralised dispatch with no uncertainty, but rather central dispatch plus a mechanism for resolving uncertainties in real time.

---

[14] Market power in generation working paper.
[15] Market power in generation working paper.
[16] The Australian national electricity market does not require an imbalance market at all because its gross pool is almost identical to a real-time market, with prices set every 5 minutes with no concept of 'gate closure'.

16. We have not concluded on whether the data pointed to by InterGen amounts to a systematic inefficiency due to bilateral contracting rather than centralised dispatch. We will continue to investigate this particular case.

17. We asked National Grid to consider possible sources of savings that might be seen from reverting to centralised dispatch. It concluded that there would not be substantial savings from the point of view of balancing the system. It also commented that in moving from the pool to NETA, it found generation asset owners were now more reluctant to switch plants off than National Grid had been as central dispatcher under the pool. National Grid hypothesised that plant owners may be able to factor in the additional maintenance costs implied by frequent starts and stops more accurately than could the SO under centralised dispatch rules, and that self-dispatch may in this sense be more technically efficient.[17]

Self-dispatch reduces price transparency

18. The extreme case of a centralised dispatch system, like the Australian NEM mandatory gross pool, requires all supply-side and all demand-side parties to submit bids and offers. The market generates a market price that is based on all anticipated market activity and is publicly available. In other systems – like ERCOT and NordPool – bidding into the pool is not mandatory but produces cleared prices based on the bids and offers that are submitted.

19. One advantage claimed for centralised dispatch has therefore been the public availability of a market price based on a substantial proportion of physical trades. Part of the value of a mandatory market comes from the fact that everyone can be confident that the price is the result of supply and demand matching in the whole market. If it is the result of a market whose functioning is regulated as a public good rather than a private good, it can provide a firm and trustworthy reference price for policy.

20. We have found that for most purposes prices are transparent in the GB wholesale electricity market.[18] The N2EX and APX exchanges publish day-ahead electricity auction prices. Approximately 40% of electricity goes through these auctions. The conclusions of our foreclosure and market power in generation working papers suggest that parties do not have the ability or incentive to make this price systematically diverge from a competitive spot market price. This suggests that the price signal from these auctions is likely to be robust. Prices of individual trades in the forward market are available for

[17] See National Grid (January 2015) *Would it be more efficient/less costly for National Grid to manage all dispatching?*
[18] Liquidity working paper.

a modest fee from Trayport, the screen-based trading software provider that most traders use.

21. Real-time cash-out prices are made public, as are the balancing mechanism bids that went to determine those prices. The reforms to cash-out that are anticipated in the next three years – and particularly the move to a single cash-out (see paragraphs 46 to 49 below) – render the cash-out price in most periods a good measure of a real-time spot market price. In this sense, there will be, post-reform, a market price based on the real-time, mandatory centralised matching of supply and demand that applies to the whole market.

22. In relation to price transparency, the difference between the Australian NEM, characterised as a mandatory gross pool, and the GB system after the proposed EBSCR reforms, is very slight and rather technical in nature. Generators are mandated to bid into the GB balancing mechanism just as they are mandated to bid into the NEM in Australia. In both systems, these bids are used by the SO to build a supply curve and to generate a price used in almost real-time purchases and sales. In both the GB system and the NEM, most trading occurs outside the real-time market in forward and futures markets, both in brokered bilateral trades and on exchanges. This occurs, as it does in many markets, because of parties' desire to add predictability to cash-flows. Future trading adds a layer of obscurity to purchase costs, but there is no difference in that respect between the Australian NEM and the GB system. In practice, the difference in terms of price formation between the exemplars of 'self-dispatch' and 'mandatory gross pools' tend to disappear.

23. For all these reasons, we do not believe that there would be a large advantage to competition from the point of view of increasing price transparency by reverting to centralised dispatch.

Self-dispatch increases transaction costs for new entrants and smaller players

24. A separate advantage claimed for a centralised dispatch system is that it provides a simple route to market for energy: a generator knows that it can sell its output by bidding into a pool; a supplier can buy energy from the gross pool. In the case of a mandatory pool, the entire market participates, so the depth of the market is maximised.

25. Under a self-dispatch system, parties are responsible for finding generators or suppliers with which to trade. This requires, in-house or outsourced, teams of buyers and sellers and may be more complex than participating in a pool.

26. However, even in centralised dispatch systems with gross pools, most of the trading takes place in the forward markets that lead up to bidding in the gross

pool. This arises from the corporate need for prudent risk management.[19] The relevant comparison to assess transaction costs should not, therefore, be between having no need for a trading team versus needing a full trading team, since both self- and centralised dispatch systems typically require participants to have trading teams.

27. Once again, the difference between centralised dispatch systems and self-dispatch systems starts to become less clear the closer we look at the details of each, as each system requires mechanisms that are characteristic of the other system. For instance, centralised dispatch systems require a procedure for dealing with unanticipated changes to production or demand – they need some sort of balancing mechanism – while self-dispatch systems need some form of central control in real time (ie it becomes a centralised system close to real time). The nub of the difference is the number of opportunities that generators and suppliers have to contract with the SO for energy. In the GB system, this happens in the balancing mechanism after gate closure; in centrally dispatched systems, this also happens then, but in some cases also before that, typically a day before.[20]

28. Participation in spot markets in Great Britain involves low transaction costs. The day-ahead auctions allow day-ahead trading on a very similar basis to that which would be provided by a gross pool.[21] Moreover, the reforms to the cash-out regime (especially the elimination of 'dual pricing'; see paragraphs 46 to 49) mean that reliance on the centrally cleared balancing mechanism for energy will no longer be unattractive by design. This will provide a further low transaction cost option for buying or selling electricity.

29. In light of the above, our initial view is that there would not be significant transaction cost difference between the self-dispatch system in Great Britain and a centralised dispatch alternative.

Self-dispatch precludes locational pricing

30. Fully fledged locational pricing[22] is typically associated with centralised dispatch systems. In order to compute prices at each of thousands of points on the network in the Texas ERCOT market, for example, the SO computes the incremental cost of energy at each location. In order to do this, it requires

---

[19] Liquidity working paper.
[20] This is not true of the Australian NEM, which might be characterised either as operating a centrally dispatched system or as operating only a mandatory gross balancing mechanism with no gate closure.
[21] The day-ahead auctions require parties to post collateral for their trades. This may be a substantial cost, but we have not found evidence that it is an undue cost. A day-ahead pool would also need to have some insurance mechanism against a party's inability to make good on its commitment.
[22] Locational pricing in the electricity market in Great Britain working paper.

detailed information on generator and supplier bids as well as network constraints.

31. However, it is possible that such prices – or the prices of a smaller number of zones – could be computed in Great Britain from the current residual balancing operation of the SO. This is an aspect of this debate that we will continue to investigate.

Our current views

32. For the reasons set out above, especially with the proposed move to a single price for cash-out, we do not believe that the self-dispatch system in Great Britain, when compared with alternative dispatch systems, reduces price transparency or increases transaction costs. Although we accept that self-dispatch may have been a contributing factor in creating the current wholesale market structure at the time of the NETA reform, we do not believe that self-dispatch provides incentives for vertical integration any longer. It is possible that self-dispatch systems reduce technical efficiency and that more centralised dispatch may make locational pricing easier to implement.[23] We intend to investigate these two aspects further.

Cash-out and the balancing mechanism

33. In this section, we describe and discuss the relationship between cash-out and balancing following the proposed EBSCR (which has yet to be finally agreed by Ofgem or implemented) in the GB wholesale electricity market. In particular, we examine how the proposed reformed cash-out rules use competition to minimise short-run costs.

34. Where more electricity is generated than consumed, or vice versa, system frequency may rise or fall to a degree that requires central intervention (an imbalance). This can be the consequence of an unforeseen peak or an unexpected fall in supply or demand (eg due to weather conditions or technical failures in the system or by particular power plants). In order to prevent imbalances, the GB system of maximum self-dispatch becomes a centralised mechanism close to real time. This is required by the nature of the coordination problem involved in maintaining an interconnected grid that is safe and stable at reasonable cost. For this purpose, National Grid intervenes directly by running the balancing mechanism and, at times of great system

[23] Locational pricing in the electricity market in Great Britain working paper.

stress, by calling upon contracted capacity providers in the form of various contracts, the largest of which is Short Term Operating Reserve (STOR).

35. Under the balancing mechanism, National Grid collects up to an hour before any settlement period (ie until 'gate closure') information from generators and suppliers that it uses – among other sources – to forecast levels of generation and consumption of electricity across Great Britain for any half-hour period (a 'settlement period'). Generators and suppliers must also notify before gate closure their contracted delivery and offtake of electricity on a portfolio basis to the central imbalance settlement system administered by ELEXON. National Grid will assess whether the system is at risk of imbalance and will accept offers or bids to sell electricity to or buy electricity from the system if it predicts a discrepancy between the amount of electricity generated and the amount of electricity consumed during a certain settlement period.[24]

36. Any generators and suppliers who contributed to an imbalance (eg through a failure to comply with their contracted delivery or offtake) are then charged an imbalance price ('cash-out'). In any settlement period, a generator's revenues[25] from the production of electricity is equal to its revenues under contract plus or minus the cash-out price multiplied by the uncontracted quantity. If the generator has overproduced relative to contract, this last is a positive; if it has underproduced relative to contract, it is a negative. Similarly, a supplier's purchase costs for electricity are equal to costs under their contracts plus or minus the cash-out price multiplied by uncontracted quantity. If the supplier has consumed more than contracted, this last is positive; if it has consumed less than contracted, it is negative.

37. As one of several additional tools to maintain the capability of achieving an energy balance on the electricity transmission network (given the potential for various eventualities that will cause imbalance between approximately 4 hours ahead of time to real time), National Grid maintains (STOR) contracts whereby the counterparty undertakes to make available a contracted level of power when instructed by National Grid.[26] STOR contracts are procured via a competitive tender process with three tender rounds per year. National Grid pays an availability payment to STOR service providers, which is paid to the provider regardless of whether they produce, and a utilisation cost in case of

[24] For example, this can be as a consequence of an erroneous forecast, the failure by a party to comply with its commitment, a transmission issue or a combination of these factors.
[25] This abstracts from other, ancillary, services that the generator can supply to the system operator and from balancing market revenues. A more complete description would be this: Generator Revenues = Contract Revenues + Balancing Service Revenues + Cash-out Price × Imbalance Volume, with Imbalance Volume = Energy Produced – (Energy Bilaterally Contracted + Energy Contracted in the Balancing Mechanism). The details are given in ELEXON (2014) *Imbalance pricing guidance*.
[26] The other tools are smaller in capacity and more limited in their uses. Details can be found at National Grid (nd) *What are reserve services?*

actual delivery. STOR providers agree to make available capacity to National Grid and face contract penalties (depending on terms) if the capability cannot be made available: for example, because they are producing energy for other parties. In practice, this means that STOR capacity is reserved for use by National Grid, although the contracted parties retain commercial discretion to use the capacity elsewhere.This discussion will lead in the next section to an examination of the combined impact of the capacity market and cash-out rules on longer-term investment incentives.

An outline of balancing and cash-out

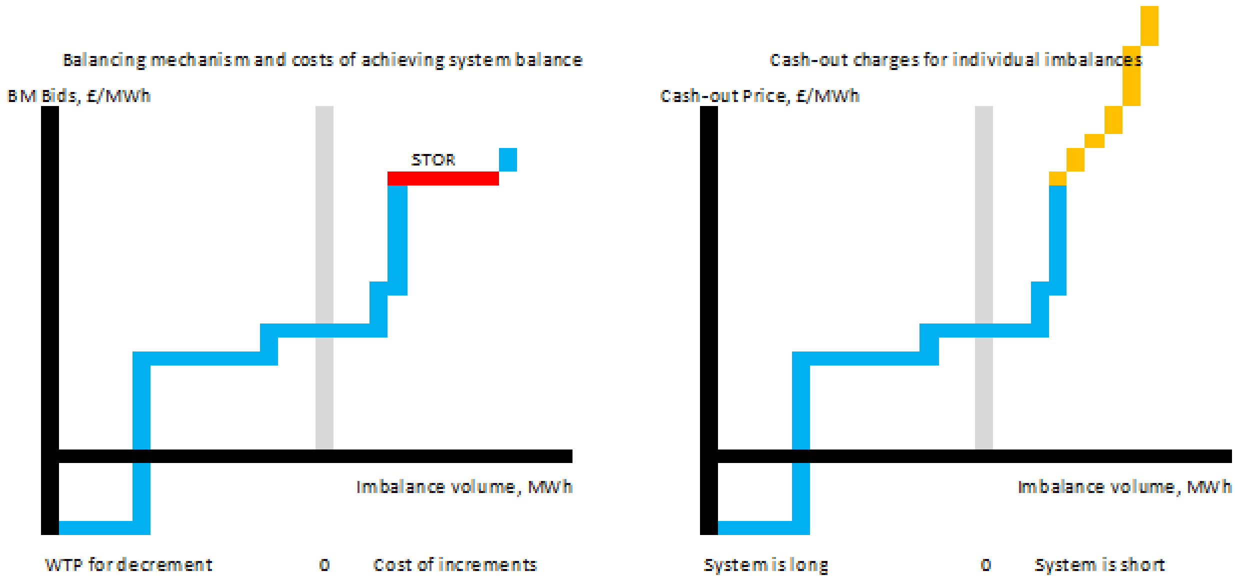
38. The balancing mechanism is the process by which National Grid procures and rewards the energy that it needs in real time (and up to 90 minutes before that) to balance supply and demand. The cash-out mechanism relates to the conceptually separate question of what parties are individually charged or paid for electricity that they use or produce without having had a contract to do so. We explore in this section how the pricing of cash-out and the likelihood of an imbalance influence the behaviour of market participants before and after gate closure, and how the mechanisms outlined above are using competition to minimise market participants' short-run costs.

39. The relationship between the pricing under balancing mechanism and under the cash-out rules as reformed following the proposed EBSCR (which is anticipated to come into force in a gradual way over the next three years) is shown schematically in Figure 1 below.[27]

[27] This diagram abstracts from many detailed elements of the relationship between balancing mechanism and cash-out, for example: 'tagging' of actions; cost recovery and the Residual Cash-flow Reallocation Cash-flow (RCRC) 'beer fund'; and several additional sources of fast response. The diagrams also abstract from the auction design of the balancing mechanism, which, as a pay-as-bid auction, will not reveal balancing costs in the way assumed in the diagrams. These complications are not central to the arguments that follow.

FIGURE 1

Balancing mechanism and cash-out prices



Source: CMA analysis.

40. The left-hand diagram shows the cost of achieving system balance in the context of the balancing mechanism auction process. The very short-run supply curve for wholesale electricity is represented as the blue curve. The grey line represents the contracted demand curve (ie the expected point of system balance just before gate closure).

41. On the assumption that parties are aiming to be balanced and that near-term markets are liquid, National Grid would not be expected at gate closure either to need to buy or sell energy (ie under these idealised assumptions, the grey line would represent both the aggregated contracted position and the physical position of all market participants).[28] However, there are always unexpected events on the supply and demand sides between gate closure and delivery which may cause an imbalance, requiring National Grid to buy or sell electricity through the balancing mechanism. The extent of aggregate imbalance will determine the actions that need to be taken (and therefore the short-run marginal cost of National Grid's intervention to balance the system). The short-run marginal cost of energy for balancing is given by the point of intersection of the actual demand for electricity (ie the out-turn imbalance) and the blue and red curve on the left-hand diagram. The blue portions of the curve represent actions that generators bid into the mechanism, and the red

[28] We relax some of these assumptions in some of what follows. Until now, it has been thought that parties prefer to over-contract and the expectation in most periods is therefore that the system will be long rather than balanced. We consider further below the degree to which this behaviour might be the result of the proposed new rules for cash-out pricing.

portion of the curve represents capacity available to National Grid under STOR contracts. Individual parties may or may not themselves be in balance. For small parties, the probability of imbalance is independent of the overall system balance (while the imbalance of large parties is likely to cause overall system imbalance).[29] If a party has under-generated or over-consumed compared to its contracted volume, it will be charged for that shortfall of energy at 'system buy price'; if a party has over-generated or under-consumed compared to its contracted volume, it will have to sell that extra energy at 'system sell price'. These cash-out prices are derived largely from the weighted average prices of the offers and bids accepted by National Grid through the balancing mechanism. This is shown in the right-hand diagram.[30]

42. Under the proposed rule changes, the single cash-out price (which means that the system sell price will be equal to the system buy price) will be set, as shown in the right-hand diagram, by the intersection of actual demand (ie overall required out-turn system balancing quantity) and the blue supply curve. This follows the supply curve for the balancing mechanism over most of the range. However, when the system is short at high levels of demand and STOR comes to be used, the proposed rules will introduce a wedge with the balancing market cost (in yellow in the diagram). This is known as reserve scarcity pricing (RSP) and is discussed at some length in paragraphs 77 to 85.

43. In extreme cases where National Grid is not able to balance the system by increasing supply through the balancing auction and STOR contracts, under the proposed EBSCR reforms it will force some consumers to consume less energy (ie there will be blackouts or brownouts), and the cash-out price will be set administratively at £6,000/MWh.[31]

---

[29] In what follows, this is the definition of a 'small' or 'large' party.
[30] The cost of actions is not always reflected in cash-out prices and the SO goes through a complex 'tagging' procedure to determine which actions are properly energy imbalances rather than locational or other system-related effects. We abstract from these features of the mechanism in our analysis.
[31] Strictly speaking, the exact procedure by which this occurs is slightly different. The EBSCR proposals do not include automatically setting the price to VoLL or RSP. Rather, one or more of the actions in the stack of actions used to calculate the cash-out price are to be repriced, at either VoLL or RSP, as appropriate. This stack of actions then undergoes a number of steps to remove certain actions – known as flagging and tagging – before PAR is applied and the price is determined. The impact of flagging and tagging – in particular of NIV tagging and SO constraint flagging – may mean that prices will not rise to VoLL even when there are blackouts or brownouts. See ELEXON (2014) *Imbalance pricing guidance*.

Description of the components of the EBSCR

44. The proposed EBSCR reforms will modify cash-out pricing on four major aspects, each of which has been briefly touched on already and is described at greater length below.[32]

(a) A move to **single cash-out price**.

(b) A move to making the cash-out price in all periods equal to the cost of the 1MWh most costly action in the balancing mechanism (known as 'price average reference volume of 1MWh', or **PAR1**), which is a narrowing of the base for the calculation from the previous 500MWh (itself a narrowing from the original design, which was a simple average cost of all balancing actions).[33]

(c) A move to re-price STOR actions (typically periods of tight short-run margins due either to high demand or to supply disruptions)[34] to the probability of lost load (a measure of how stressed the system is, known as 'loss of load probability' (LOLP)) multiplied by £6,000/MWh (the putative 'value of lost load' (VoLL)),[35] if this is greater than their utilisation price. This is known as 'reserve scarcity pricing', or **RSP**.[36]

(d) A move to price disconnection or voltage reduction actions equal to the VoLL.[37]

45. The short-run incentive properties of each of these reforms are described in more detail below.

46. **Single cash-out price** is the proposed rule by which there is to be a single price for contractual imbalances. For example, if the system is short and a generator is producing more than contracted, it will receive the same price for

[32] Ofgem told us that 'EBSCR consists of an integrated package of several elements', the implication being that it should be assessed as a whole. In the sections that follow, we seek to assess each individual aspect of the package, noting interrelationships between aspects.

[33] Under the proposed reform, the move to PAR1 is gradual. The intention of the proposed changes is to introduce PAR50 initially, and further narrow PAR to 1MWh ahead of winter 2018/19.

[34] Periods of tight margins are periods when STOR is likely to be used. However, STOR is also used outside of very tight periods. The SO has discretion to use a STOR plant over a balancing mechanism plant when it is more efficient to do so. STOR may even be used when the system is overall long. RSP, however, is likely to set cash-out prices only in periods when the margin is tight.

[35] LOLP measures the probability that the system will suffer an interruption. At times of high demand, the system loses resilience in that a power station breaking down could lead to an inability to find enough replacement capacity rapidly enough. This is the sort of situation when LOLP rises to being close to 1. LOLP is typically calculated by simulating the system. The VoLL represents the willingness to pay for an incremental MWh at times of system stress – it is the amount that the consumer of the last MWh is willing to pay to avoid being cut off. The value cannot be measured directly in any sense and is typically estimated once and for all using survey techniques.

[36] As noted above, the cash-out price calculation is applied to the stack of actions.

[37] There is a transitional period during which this will be set to £3,000/MWh and it will settle at £6,000/MWh

its electricity as that paid by a supplier who has not contracted enough electricity. This rule is replacing the current dual cash-out price rule whereby actors who were long when the system was short (or vice versa), therefore contributing to the rebalancing of the system, were effectively penalised – or at least not rewarded – for doing so.[38]

47. The current system was designed with the fear that parties might not have sufficient incentives to try to balance their supply and demand positions through bilateral contracts ahead of gate closure. The unattractive charge for beneficial imbalances was designed to encourage parties either to contract ahead of time or to participate in the balancing mechanism (BM) but not to rely on cash-out as a market of last resort (ie taking a long or short physical position voluntarily). This no longer appears to be a significant concern in the proposed EBSCR.

48. There is some evidence that the reform will be beneficial to smaller generators, to renewable producers and smaller suppliers who tend to be more reliant on cash-out than the large vertically integrated players. To a first approximation, we can consider that a small player aiming to be in balance will randomly find itself long or short with the same probability;[39] in the long term, under single cash-out, any losses made when contributing to the overall system imbalance should be offset by gains made when helping to solve it. Relying on cash-out as a market of last resort is no longer loss-making **by design**.

49. Some small suppliers rely, even under current rules, to a much greater extent on cash-out than do the larger firms. This is plausibly because the transactions costs of being involved in the on-the-day bilateral markets are high. The move to a single price will make cash-out relatively more attractive for these parties.[40]

50. **PAR1** is a rule change by which the calculation for the cash-out price outside times of system stress will be determined by the average cost of the last 1MWh of balancing actions taken. This contrasts with the current rule by which the price is determined by the average of the last 500MWh of actions

---

[38] This is achieved in the current system by those in 'helpful' imbalance being charged (if short) or paid (if long) an administrative price (the 'market index price') that was designed usually to be more (if short) or less (if long) than the corresponding payment or charge incurred in the balancing mechanism.
[39] A small player's own imbalance will not have a significant effect on system imbalance, hence the 'fair bet' involved in cash-out.
[40] This is confirmed in Ofgem (2014) *Further analysis to support Ofgem's updated impact assessment*, Figure 3, which shows smaller suppliers benefiting from EBSCR.

taken (PAR500).[41] PAR1 is described as making the cash-out price 'more marginal'.

51. **RSP** is a proposed rule change for cash-out prices in times of low short-term capacity margin, which will directly affect cash-out pricing when STOR contracts are called upon by National Grid. The SO typically seeks to contract under STOR ahead of time for between 2.2 and 2.3GW of capacity.[42] As mentioned above, STOR providers are paid an availability payment by the SO and also, when called upon by National Grid to deliver electricity, a utilisation payment intended to cover its operating costs when it actually produces.

52. The RSP rule is expected to lead to a substantial net increase in cash-out prices, especially at times of very tight short-term margin. Under the current system, there is a formula that averages out availability costs over periods when STOR was used historically. The cost of availability is thus reflected, to some extent, in the cash-out price, but not necessarily when STOR was used. When the RSP rule bites, the cash-out price is raised administratively beyond the BM utilisation cost.[43]

53. **VoLL** is an administratively set price applied to cash-out payments if blackouts or brownouts occur for reasons of energy imbalance. This move to set prices to £6,000/MWh in those circumstances is a natural extension of the thinking behind RSP: if the LOLP is 100%, the cash-out price will be equal to the VoLL.

54. The overall EBSCR package is designed to improve balancing incentives and so provide adequate signals for short- and long-run operational efficiency. The analysis presented in paragraphs 77 to 90 argues that the main impact for the RSP and VoLL rule changes will be in providing incentives for capacity investment.[44] We consider in paragraphs 96 to 102 the interaction between the capacity market and the cash-out rules which have both been designed to provide adequate investment incentives for capacity and flexibility.

Use of competition to minimise market participants' short-run costs

55. We have considered two arguments concerning the move to PAR1 suggesting that the cash-out rules as reformed by ESBCR will not minimise short-run costs.

[41] See note 36 above for the phasing of the reform.
[42] National Grid (nd) *STOR market information report: tender round 24*.
[43] Subject to the flagging and tagging process mentioned above.
[44] Currently, demand disconnections in themselves have no impact on prices at all.

56. Stephen Littlechild has argued that PAR1 was not necessarily 'more marginal' because balancing actions are not necessarily simply incremental – they may be sequential. They may even be forward-looking and reflect expected imbalances in periods outside the period in which the action is taken. One solution to making balancing prices more clearly reflective of incremental energy costs is to reduce the settlement period from 30 minutes to something shorter. The Australian NEM and ERCOT in Texas, for example, both use 5-minute intervals for imbalance price calculation.

57. George Yarrow submitted that one of the original rationales for using an average price over a large number of actions was that this made the price less easy to manipulate; in the reformed cash-out pricing, with the cash-out price being calculated on the basis of actions amounting to 1MWh, it would be possible for a generator to learn that it tended to be a price setter in certain circumstances and might therefore be able to change its BM bids to take advantage of what would, in effect, be a lower price-elasticity of demand.

58. We have sought some clarification on both of these criticisms of PAR1 and will continue to investigate these issues. We consider that the principle of marginal cost pricing of balancing electricity is right from the point of view of competition, and we are seeking clarifications as to its specific implementation.

Interaction between the capacity market and the cash-out rules

59. In this section we describe a market failure identified by Ofgem and the Department of Energy and Climate Change (DECC), known as the 'missing money problem', as well as the measures that have been implemented to address, among other things, this market failure. We then examine the interaction of two regulatory mechanisms addressing this problem: the capacity market introduced by DECC, which will, from 2018, directly remunerate capacity for being available, regardless of energy produced; and the cash-out arrangements, with the reforms proposed by Ofgem under the EBSCR. We then seek to assess whether the coexistence of these two mechanisms might give rise to overpayment to capacity providers.

The 'missing money problem' in 'energy-only' electricity markets

Definition

60. Revenues to generators may come either:

(a) almost exclusively from sales of energy ('energy-only markets'); or

(b) from two distinct sources of revenues: that is, revenues from sales of energy and revenues paid to generators for making capacity available regardless of actual delivery (ie electricity and capacity markets).

In Great Britain, the NETA/BETTA market was originally designed as an 'energy-only' electricity market. Internationally, we see both energy-only markets (eg ERCOT in Texas and NordPool in Scandinavia) and markets with capacity mechanisms (eg PJM in the north-eastern United States).

61. The theory of a well-functioning competitive energy-only electricity market[45] is that generators will fully recover sunk capital costs (eg the costs to build generation capacity) at very occasional peak times[46] – once every 20 years,[47] perhaps.[48] At these times, demand is high enough to give the owners of this generation capacity the ability to produce electricity (or actively to interrupt its demand) to earn a price far in excess of short-run marginal cost.

62. The theory of energy-only markets is that the promise of very occasional, very high rents in periods of extreme demand is sufficient to reward and incentivise the owners of generation capacity and that the capacity which meets demand at very occasional peak times provides sufficient capacity margin for the SO to balance the system safely and efficiently in the normal course of events.

The missing money problem in Great Britain

63. In practice, there is considerable doubt that events in the electricity market would ever unfold quite as the theory requires.

64. Extreme demand periods in Great Britain are likely to be in the depths of a very cold winter when weather amounts to a national emergency.[49] There is a critique of the theory of energy-only markets that energy companies would plausibly not believe that they would be allowed to charge extreme prices in these extreme circumstances; they might not even wish to, given the damage to reputation that the appearance of such 'profiteering' would cause.

---

[45] RE Bohn, MC Caramanis, FC Schweppe (1984) Optimal pricing in electrical networks over space and time, *RAND Journal of Economics* 15(3), pp360–76.
[46] For peaking plant, the only opportunities to recover sunk costs are in such periods. For other plant, some contribution to sunk costs will come in 'ordinary' periods when there are plant with higher operating costs setting market price.
[47] Twenty years is used as an example. In traditional, centrally planned electricity systems, engineers would often define adequacy standards in terms of being able to withstand a 'once in 20 years' winter'. It should be emphasised that in real (rather than theoretical) electricity markets, peaking plant can earn revenues at other times, for example by supplying essential system stability services unrelated to energy supply.
[48] If there is demand responsiveness, then voluntary demand reductions will also occasionally – and possibly much less rarely – lead to prices being set above short-run marginal cost.
[49] Alternatively, prices could rise to extreme levels in periods of ordinary demand when a large-scale supply outage had occurred: for example, a nuclear shutdown. A sufficiently catastrophic supply-side event would probably also count as a national emergency.

65. But if owners of generation capacity, especially peak capacity, do not charge extreme prices in extreme demand periods, and if they are competing fiercely on price at other times, then they are unlikely to recover sunk capital costs fully. Therefore, continues the critique of energy-only markets, they cannot be expected to invest adequately to provide sufficient capacity margin for the SO to balance the system safely and efficiently. This is the phenomenon widely known in this industry as the 'missing money problem'.

66. Notwithstanding the possible missing money problem, Great Britain witnessed a considerable amount of new investment in CCGT in the early years of the 21st century.[50] However, the NETA/BETTA system has been in existence for a short period of time relative to the expected frequency of extreme events, and the system has never been tested in terms of extreme conditions (and therefore potential for extreme prices). We therefore do not know whether investors could, within the system, have hoped to recover sunk costs. Moreover, it is not clear that the system in its early days was sufficiently competitive to engender a missing money problem: less than competitive prices in ordinary times could allow market participants to recover fixed costs even in the presence of a missing money problem.

67. We have some direct evidence from company corporate documents that at least some generators believe that there has been a missing money problem. For example, a paper presented to the SSE board in 2013 included a comment that prices may not rise despite tightening system margins because of fear of rent extraction at peak times:

> A key risk in this assumption is the likelihood of system shortages duly materialising – but failing to translate into higher distress prices. The experience of December [2012] is that even the last CCGT on was reluctant to extract sufficient rent to make a meaningful contribution to its fixed costs of remaining open. Owners of CCGTs are nearly all vertically integrated utilities with a cautious approach to regulatory obligations and interpretation.

68. It is plausible that investments in capacity might be inadequate if potential investors share the view that the missing money problem is material.

69. As policy to decarbonise electricity production developed in the late 2000's, it became clear that investors in thermal generation would have increasing challenges in recovering sunk capital costs. Low carbon generation mostly has very low short-run marginal costs. In an energy-only market, increased renewable capacity brought on to the system through subsidy is likely to make

[50] Descriptive statistics: generation and trading working paper.

thermal generators more and more reliant on increasingly infrequent periods of system stress to earn a positive margin. The falls in peak demand due both to recession and to energy efficiency measures have exacerbated the problem for investors in thermal plant. A missing money problem has therefore become more and more significant.

70. While investigating the potential for the exercise of unilateral market power by generators,[51] we have been told by generators that they are very reluctant to bid above cost in any period in part because of regulatory concerns. They have cited Transmission Constraint Licence Condition (TCLC)[52] and the European regulation REMIT, both of which potentially impose severe penalties for specific types of market manipulation. It is possible that under both of these instruments, generators will be genuinely unsure whether a price rise might be caught under them or justified by capacity shortages.[53] The risk and uncertainty this creates might exacerbate the missing money problem.

DECC's capacity market and Ofgem's cash-out reform as reactions to the missing money problem

71. DECC and Ofgem each responded to the increased challenge of the missing money problem with reforms. DECC has addressed the problem directly with the capacity market (CM), a mechanism that will, from 2018, directly remunerate capacity for being available, regardless of energy produced. Ofgem has proposed to reform cash-out arrangements through the EBSCR,[54] with the RSP element of the reform, described in paragraph 52, above, particularly relevant to this issue (see paragraphs 77 to 85, below). One of the four aims of the EBSCR is to 'incentivise an efficient level of security of supply'.[55]

72. The capacity market is discussed in more detail in our capacity working paper. We believe that it is plausible, and we assume for the purposes of this paper

---

[51] Market power in generation working paper.
[52] The TCLC is discussed in our locational pricing in the electricity market in Great Britain working paper.
[53] The TCLC is worded in such a way that it is clearly aware of the danger that it might dampen investment signals. 'Opportunities arising from the acceptance of large offers in the Balancing Mechanism could provide important investment signals for new generation in import-constrained areas, potentially leading to ongoing reductions in constraint costs associated with import constraints. Hence the removal of this signal might be problematic. In contrast, further generation investment in export-constrained areas would tend to exacerbate constraint costs' (DECC (2011) *Modifying the conditions of electricity generation licences*). In response to this, the TCLC prohibits decrement bids in the BM that are not cost-reflective. However, at a time when an energy-only market should be pricing incremental capacity away from short-run marginal cost, a decrement bid in the BM should also be priced in this way. If as well as this the peak generator is uncertain whether it might be caught under TCLC, it might restrain itself from a rent-accruing bid when in theory it should not.
[54] Ofgem lists another three high-level objectives of the EBSCR: to improve the efficiency of balancing the system; to ensure compliance with EU rules; and to complement the DECC CM.
[55] Ofgem (2014) *Electricity balancing significant code review – final policy decision*, paragraph 1.7.

that DECC's CM addresses any missing money problem of the sort described above as it is designed to provide payments for adequate capacity availability.

73. We examine below in more detail the ways in which the proposed reforms to cash-out arrangements might also incentivise investments in capacity levels and flexibility, and whether the EBSCR plausibly might solve the missing money problem.

74. Finally, we will assess whether the CM and the EBSCR might interact and overlap in seeking to solve the missing money problem, and whether there may be concerns that they might lead to overpayment for capacity.

How do the elements of the EBSCR incentivise investment?

75. Move to **single cash-out price**. Under the current dual cash-out rule, capacity that contributes to meeting demand when needed but that was, for whatever reason, not contracted to do so does not receive the market price for its output but instead a price that is likely to under-reward energy. In this sense, dual cash-out is one institutional feature in the current arrangements which contributes to missing money. However, the magnitude of this effect might be quite small because most energy (in periods not affected by the RSP rule change) will be rewarded outside of cash-out,[56] and at prices that reflect expected demand levels at gate closure.

76. The move to **PAR1** outside periods when RSP bites (ie in 'ordinary' periods) is expected to raise cash-out prices when the system is short, and to lower them when the system is long.[57] The impact on revenues and investment

---

[56] If parties aim to be in balance (as we would expect for small parties under single cash-out and risk-neutrality), then only situations in which parties find themselves unexpectedly long when the system is short will contribute to these additional revenues. We would not expect the move to a single cash-out price on its own to have significant knock-on effects on forward prices because, by definition, the cash-out price deals with unexpected demand, while forward prices will reflect expected demand.

[57] Remember that the BM is a pay-as-bid (PAB) auction. In a PAB auction, bids rise with cost, but not nearly as steeply as they would under a uniform price auction. The intuition is that if a low-cost bidder knew for certain what the PAR1 action would be, it would be optimal for them to bid just below that price rather than at cost. Uncertainty will add some slope to the bid function. Therefore, there should be some price difference between PAR1 and PAR500, but the difference should not be nearly as dramatic as it would be in a uniform price.

incentives is likely to be slight for a risk-neutral[58] or risk-averse[59] party that aims to be in balance at gate closure.

77. The move to **RSP** has a clearer investment incentive property. STOR capacity can be thought of as being, among other things, the last available increment of capacity on the short-run supply curve. Any capacity purchased by the SO under STOR contracts cannot be used to supply the energy market in a 'STOR window' (a period when STOR is used). Therefore, if STOR is used for energy balancing, it is almost certain that suppliers and generators as a whole will have been short: some energy is needed that could not have been contracted for in the open market.[60] This is why this element of the reform will have a clear impact on wholesale electricity prices: since there are periods when STOR contracts will be expected to be used for energy balancing, the forward electricity price for those periods (or for longer periods expected to include some of these periods with some frequency) will reflect any impact of STOR on prices.[61]

78. If supply and demand are in a configuration in which parties can be almost sure that STOR will be used for energy balancing purposes, then the likelihood is high that cash-out prices will be set at LOLP × VoLL (ie above short-run marginal cost) for some of their energy.[62] This is the crucial difference in the argument that follows between the effects of PAR1 on its

---

[58] Consider the case of a small risk-neutral supplier with deep pockets. (It is 'small' in the sense that whether it is contractually long or short has only a very slight influence on the probability that the system as a whole is long or short). The supplier forecasts its demand ahead of gate closure and contracts accordingly. Some of the time, it will have over-forecast demand and will be long in cash-out; at other times it will have under-forecast demand and will be short in cash-out. When it is long and the system is short, PAR1 prices will earn it more than PAR500 prices would have done. When it is long and the system is long, PAR1 prices will earn it less than PAR500 prices would have done. When it is short and the system is short, PAR1 prices will cost it more than PAR500 prices would have done. When it is short and the system is long, PAR1 prices will cost it less than PAR500 prices would have done. However, on average it is neither long nor short. A small risk-neutral supplier, therefore, will not be affected by a move from PAR500 to PAR1 – it has no greater incentive to invest in capacity (or to contract for someone else's capacity) under one pricing rule than under the other.

[59] Consider now the case of a small risk-averse supplier. The move from PAR500 to PAR1 increases the cost of being short when the system is short and of being long when the system is long. If that supplier wants to avoid such anticipated losses, it might increase its demand for flexible capacity options in the run-up to gate closure. This might seem like an increase in demand for capacity and therefore an additional incentive to invest in capacity. But this is wrong: the demand is for protection against losses that may be incurred under the cash-out mechanism, not for additional capacity. An increase in demand for capacity simply due to risk-aversion would not lead, on average, to the actual increased use of capacity in satisfying electricity demand. In that sense, the demand could be satisfied just as well by a purely financial insurance contract as by a physical plant. (A deep-pocketed financial firm could agree to swap the uncertain flow of revenues to and from cash-out, which, for the (small) firm aiming to be in balance will be zero in expectation, for a fixed premium payment.) The impact of risk-aversion on actual investment incentives is therefore not clear. It might be, in the extreme case, that the whole of the increased demand for insurance were supplied by financial contracts and that the move to PAR1 would have no investment incentive enhancement at all. Alternatively, if financial insurance is unavailable or too costly, it might be that the cheapest way to insure oneself would be through capacity investment.

[60] Strictly speaking, there is nothing to stop the SO from using STOR for energy balancing actions in a period when the system is overall long if that is the efficient thing to do. But these will be rare cases and do not invalidate the general point about the system when STOR is used as the last available increment of supply.

[61] One useful way to conceptualise STOR and RSP is to consider STOR as the last increment of capacity in the merit order; RSP is then the rule which determines how that capacity is offered to the market.

[62] Any one party may still be long, but in aggregate we know that they cannot be when STOR is used.

own and PAR1 alongside STOR. When price is set outside RSP periods, there is spare capacity on the system and there is no reason for market participants to be systematically short. Competition to supply energy therefore leads to prices set at short-run marginal cost whether that is in the BM or in the bilateral market ahead of gate closure. But when STOR is likely used someone will almost certainly be left short. RSP changes the incentives around the remuneration of incremental capacity in such cases.

79. Imagine two generators with some uncontracted incremental capacity in such a situation of system stress, just a few hours before gate closure. STOR is almost certain to be used.[63] Where will competition between the two generators set the price at which they offer to sell their energy in the bilateral market before gate closure?

80. They have three alternatives to selling the energy forward: they can offer it in the BM; they can sell it bilaterally; or they can produce without a contract and 'spill' energy into cash-out (ie intentionally go long in cash-out).

81. The BM is structured as a pay-as-bid tender auction, and STOR is priced in that auction at its utilisation price (approximately its short-run marginal cost). Therefore, in order to sell output for sure in a typical period in which STOR is called on, the incremental generators would have to offer energy for less than the STOR utilisation cost.[64] This will be lower than the cash-out price. RSP will therefore have no effect on the earnings of incremental capacity if that is offered in the BM.

82. However, if STOR is almost certain to be used, then at least one supplier will know that it will be short and be exposed to high cash-out prices. Such a supplier would be willing to pay up to the RSP-set price in the bilateral market. Owners of capacity in this scenario should understand their residual monopoly power and offer to contract at the RSP-set price.

83. Alternatively, incremental capacity could go long into cash-out. The cash-out price in this scenario will be at the high price set by RSP.[65] This is clearly a preferable option for the owner of capacity compared to selling in the BM.

84. The effect on investment incentives of this price impact are clear:

---

[63] This scenario abstracts from many uncertainties that in practice may effect where a generator chooses to sell its output. It does this in order to provide a very clear case of the way in which the high prices of RSP can be transmitted into higher revenues for owners of capacity – and hence incentives to invest.
[64] There is nothing in the market rules to stop a generator bidding higher than STOR utilisation costs into the BM. However, if a generator wants to be dispatched from BM bids for sure when STOR is used – because they earn a positive margin at STOR utilisation costs – then they must not bid above the STOR utilisation price.
[65] Subject to the flagging and tagging process mentioned above.

(a) It will reward capacity when STOR is used over and above the utilisation cost of STOR and right up to the level of VoLL in case of extreme stress (ie where LOLP is close to 100%).

(b) If STOR is used too frequently, parties who have been exposed to LOLP × VoLL prices will find it cheaper to build new plant than to continue to be exposed to these prices.

85. One way of thinking about what RSP is doing is that it is committing the owner of the last units of peak capacity – the SO under its STOR contracts – to price that capacity based on the VoLL rather than its utilisation cost. In the case in which STOR were priced at utilisation cost, the missing money problem would be institutionalised. When demand is such that STOR must be used for energy balancing, the current system (before the introduction of RSP) ensures that the price of the last units of electricity is determined by utilisation cost, which is exactly what the theory of energy-only markets says should not happen.

Does the EBSCR incentivise adequate flexibility?

86. There are some suggestions in the Ofgem's impact assessment of the EBSCR that its rationale is as much to do with encouraging sufficiently flexible capacity as with simply ensuring the correct level of capacity.[66]

87. The CM requires plants that receive payments from it to be available at 4 hours' notice ahead of gate closure. This provides some degree of investment in flexibility.[67] Plants that are certified on the system need to have a mandated level of flexibility in terms of an ability to provide second-by-second responses. Equipment manufacturers – a small number of companies worldwide – will design generating capacity with some relatively standard levels of flexibility. Nevertheless, there are some investment decisions to be made regarding the rate at which plant can change energy output and also the minimum levels of energy that they can produce. There can be a trade-off, for example, between flexibility and efficiency.

88. We therefore need to assess whether the above-mentioned elements of the EBSCR reform will make a difference to the investments in flexibility for

---

[66] For example, the introduction of Ofgem (2014) *Electricity balancing significant code review – final policy decision* states that 'we have long standing concerns that cash out prices are not creating the correct signals for the market to balance, and in particular are not correctly signalling the value of flexibility and peaking generation'. One of the goals of the EBSCR is 'balancing efficiency', and the adequate provision of flexibility will be a very important aspect of balancing efficiency.

[67] An inflexible plant – one, for example, that could not start up in 4 hours – would need to believe that it was very likely that it would have been operating in any case if it is to receive income from the CM and not risk penalties for non-delivery. See our capacity working paper.

investors benefiting from the CM. Consider the case of an investment that would allow a plant to become ready to produce in 210 minutes. The CM only requires the plant to be available in 240 minutes. What is the extra 30 minutes of flexibility worth, and how does the EBSCR contribute to that value?

89. The additional investment could be the difference between being able to supply or not in a given half hour if the demand for that capacity were only revealed between 240 and 210 minutes. The investor would need to consider the net present value of profits from those events in order to assess investment profitability. We would expect RSP to have an impact on this flow of profits, because in its absence prices for flexible demand between 240 and 210 minutes would be lower for delivery periods when STOR was expected to be used. The magnitude of this additional incentive for flexibility is likely to be small since it applies only to demand that is revealed between 4 hours and 1 hour before delivery **and** in periods when STOR is likely to be used.

90. It is possible that flexibility will become of greater value because of the intermittency of renewables. In order to earn revenue from more intermittent demand, plant owners will need to be able to supply energy on a more flexible basis. That will increase the number of circumstances, outside system stress, in which generators will have an opportunity to recoup their investment. Higher profit flows in periods when STOR is used will also contribute to the additional remuneration for flexibility. However, in a world of frequent intermittency, flexibility will be needed to participate in energy markets much of the time, and not just at times of system stress. Within that context, the incremental impact of the EBSCR reform on incentives for investment in flexibility is likely to be small.[68]

Is RSP effective in solving the missing money problem?

91. RSP satisfies some of the characteristics needed to solve the missing money problem:

   (a) National Grid is committed by RSP to supplying energy at times of low capacity margin to those who are short at a price above utilisation cost, which is a requirement for the proper functioning of an energy-only market.

[68] If there is a market failure in the provision specifically of flexibility – perhaps a distinct missing money problem relating just to flexibility – then it is not clear that a generalised payment like RSP, which does not distinguish between payments to flexible and inflexible plant, will be the best instrument.

(b) The pricing formula based on LOLP will spread out earnings over time[69] and will not rely on a small number of extremely infrequent high prices; this may reduce uncertainty of cash flows (ie risk) for investors in capacity and should therefore lower their capital costs relative to a 'pure' energy-only market.

92. However, it is not clear that RSP constitutes a good solution to the missing money problem. There are four levels of criticism of RSP:

(a) alternative solutions to the problem may be preferable;

(b) aspects of the ESBCR reform may render it ineffective;

(c) it may be poorly designed in its detail; and

(d) it creates the possibility, together with the CM, of overpayment for capacity.

We assess these four levels of criticism of RSP in turn below.

Alternative solutions to the problem may be preferable

93. The capacity market is an alternative solution to the missing money problem. In its essence the difference between the two solutions is whether the government chooses the level of capacity required (the CM) or the price that capacity will earn (RSP). Different international systems have chosen different approaches and there appears to be little consensus as to which is the better solution to the missing money problem.

Aspects of the ESBCR reform may render it ineffective

94. Some of the mechanism by which RSP prices translate into investment incentives, especially those involving spilling into cash-out, might be thought to go against the grain of the self-balancing aspects of the electricity market design, thus increasing the possibility of regulatory intervention. This could exacerbate the problem of missing money. It might be thought that incremental energy ought, by design, to be offered in the BM, and that cash-out is intended to deal with unanticipated rather than planned imbalances, as in the case in which a capacity owner chooses to go long into cash-out to achieve the RSP-set price. If this perception of 'going against the grain' is shared, it might be that generators would feel that there is some risk of reform

---

[69] The LOLP formula effectively makes prices above system short-run marginal cost more frequent but less extreme than would a pure energy-only market, hence smoothing revenues of peak capacity owners over time.

of the system. The degree to which a future regulator could be committed to these prices and this mechanism may be in doubt.

RSP may be poorly designed in its detail

95. Even if RSP did work perfectly, the value of lost load that has been used in the RSP and forced demand reduction modifications is of £6,000/MWh (moving up from £3,000/MWh as part of transitional arrangements). This is well below the level of £17,000/MWh that DECC uses to determine the level of capacity demand in the CM. If DECC is right in this value, then the EBSCR in some ways can be seen as institutionalising missing money in that it sets a level for the most extreme prices that is too low adequately to reward the right level of investment.

RSP, together with the CM, creates the possibility of overpayment for capacity

96. Ofgem and DECC have both stated that the CM and the EBSCR in general are complementary. The argument is that the contribution of EBSCR reforms towards solving a missing money problem is that bidders in the CM will anticipate these potential additional revenues, displacing revenues they would otherwise seek through the CM. This should lower the clearing price for all capacity in the CM, and lower prices would then be passed through to consumers' bills. In the extreme case in which the EBSCR reforms are believed to solve any missing money problem completely, prices in the CM should fall to zero.

97. Within the context of its assessment of the capacity market reform under state aid rules, the European Commission received a submission raising concerns regarding overcompensation caused by the coexistence of the CM and payments under STOR. In response to this submission, the UK government noted that capacity providers could not benefit from both long-term STOR contracts and CM contracts, and that concerns regarding overcompensation would not be present in the annual STOR auctions. This is because the STOR auction for annual contracts occurs after the CM auction has taken place, and therefore providers would be able to factor their CM revenues before bidding in the annual STOR auctions, resulting in no overcompensation. The European Commission accepted that the CM had been designed to be consistent with the reform of electricity cash-out arrangements.[70]

98. Ofgem and DECC do not, as far as we have seen, distinguish between the impacts of the different elements of the EBSCR in creating this offsetting

---

[70] European Commission (2014) Letter to the UK government in relation to State Aid S.A.35980 (2014/N-2) – United Kingdom – electricity market reform – capacity market, paragraph 131.

payments effect. Our argument above suggests that this might apply to RSP and – to a small extent – to the single cash-out price, but probably not to PAR1. We consider the impacts of RSP only in what follows.

99. Under the optimistic view of the mechanisms, RSP and the CM are offsetting, so that any additional revenue that generating capacity earns through RSP leads to a reduction in the revenue it receives through the CM. Ofgem, in its final decision, points to the possibility that the two mechanisms together might lead to a transfer to consumers from the owners of inflexible capacity.[71] However, it is not clear that this effect leads to the right incentives for investment in inflexible plant: to the extent that revenues are lowered for inflexible plant, the interaction may lead to the wrong signals for investment in flexibility.[72] The size of this transfer to consumers is likely to be modest, since the market for exclusively flexible energy and capacity is small. Hence, revenues that accrue **only** to flexible plant due to RSP are likely to be modest.

100. Ofgem argued that increased revenues due to EBSCR being offset against CM bids is a good thing not only from the point of view of system efficiency but also because it means that the CM could be a temporary mechanism, to be replaced when the electricity market has settled down in its new, decarbonised configuration, with a return to an energy-only market augmented by the EBSCR. However, the CM is expected to continue to operate well into the next decade, so any benefits from easing a possible eventual transition away from the CM ought to be balanced against possible costs of having both systems working in parallel.

101. The greatest of these costs is the risk that the two mechanisms might come to generate overpayments for capacity rather than offsetting payments. If generators believe that RSP prices might not come to be allowed due to regulatory intervention, they are likely to discount anticipated revenues from RSP when determining their bids into the CM.[73] But if the mechanism then turns out to allow high prices and payments, then there will have been an overpayment: the higher price for capacity in the CM auction together with the high prices allowed by RSP. The CM has the ability to offer 15-year contracts.

---

[71] The argument suggested is the following: flexible plant can participate in some last-minute markets in which inflexible plant cannot participate; RSP increases revenues in those markets (as well as in other markets); so flexible plant will discount larger energy market revenues than inflexible plant in preparing their CM bids; since flexible plant are likely to be price-setters in the CM, they will set a lower clearing price and this will be a consumer benefit compared to what would have been the case without the contribution to solving the missing money problem that RSP represents. This argument points to a real net consumer benefit.

[72] Inflexible plant is already ruled out of near-term markets, and this may be a sufficient signal for investment in flexibility. The transfer to consumers should be considered a net benefit only if the incentive to provide flexibility was too low in the absence of RSP.

[73] It might be thought that the possibility of discounting higher RSP payments will be all the greater given the slightly unorthodox mechanism by which generators can turn the higher cash-out prices created by RSP into revenues.

Thus, even if parties eventually learn to trust RSP payments and do eventually discount them from CM bids, there could still be substantial ongoing costs from earlier rounds of CM auctions in which prices were set too high because of overpayments.[74]

102. If there is overpayment – or the appearance of it – the commitment to maintain the system as it is will be all the harder. This reduces the degree to which RSP contributes to solving the missing money problem.[75] We intend to investigate the degree to which overpayment might occur by trying to assess whether bidders in the CM have taken into account higher RSP payments in their bidding strategies. If they have not, we would consider the risk of overpayment to be high.

103. Combining the Ofgem/DECC view of the offsetting qualities of RSP and the CM with an acknowledgement of the risk of overpayment leads to the view that, at best, RSP and the CM are not both necessary; and, at worst, that they lead to overpayments.

---

[74] The opposite worry is also present: that the interaction of the two mechanisms leads to under-remuneration of capacity. Imagine that generators are overly optimistic about the degree to which RSP solves the missing money problem. They will then discount their CM bids too much and eventually lose money. This would eventually raise the cost of capital of investment and lead to higher consumer prices. The fundamental issue is that by having two overlapping mechanisms to solve missing money, an avoidable source of uncertainty is introduced into investment decisions.

[75] Some of the US markets have both CMs and RSP-like provisions. However, there are mechanisms in these markets to adjust CM payments downwards in view of high RSP payments. This sort of provision avoids overpayment risks in these designs.