

White paper

Market design for low-inertia power systems



Ancillary services
in GB



Contents

List of acronyms	4
Executive Summary	5
1 Introduction to the Main Report	7
1.1 Challenges Associated with Secure Operation of Low Inertia Systems	7
1.2 Comparison of Ancillary Services Markets	8
1.3 Purpose of the White Paper	10
2 Modelling Approach	12
2.1 Condition for Frequency Stability in the Grid	12
2.1.1 Rate of Change of Frequency	13
2.1.2 Frequency Nadir	14
2.1.3 Quasi-steady-state	15
2.2 Ancillary Services for Frequency Containment	15
2.2.1 Inertia	15
2.2.2 Frequency Response	16
2.3 Importance of Coordinating Energy, System Inertia, and Frequency Response	17
2.4 Pricing Ancillary Services	19
3 Who Should Pay for Frequency Containment Ancillary Services?	21
3.1 Relevance of Contingency Size	21
3.2 Cost Allocation	23
3.2.1 Proportional Cost Allocation	24
3.2.2 Shapley value Cost Allocation	26
4 Ancillary Services Market Driven Incentives	29
4.1 Ancillary Services from Renewable Energy Sources	29
4.1.1 Frequency Response from Curtailed Renewable Energy Sources	29
4.1.2 Synthetic Inertia from Renewable Energy Sources	30
4.2 Frequency Response from Electric Vehicles	31
5 Conclusions and Future Work	33
6 References	34
Appendices	36
Appendix 1 - Unit Commitment Model	36
Appendix 2 - Pricing Methodology	37
Appendix 3 - Cost Allocation Methodology	37

List of acronyms

AS	Ancillary Services
BESS	Battery Energy Storage Systems
CCGT	Combined Cycle Gas Turbine
DC	Dynamic Containment
EV	Electric Vehicles
FES	Future Energy Scenarios
GB	Great Britain
GFL	Grid-following
GFM	Grid-forming
IBR	Inverter-Based Resources
NESO	National Energy System Operator
OCGT	Open Cycle Gas Turbines
PFR	Primary Frequency Response
PHES	Pump Hydro Energy Storage
PV	Photovoltaic
q-s-s	Quasi-steady-state
RES	Renewable Energy Sources
RoCoF	Rate-of-Change-of-Frequency
UC	Unit Commitment
UFLS	Under Frequency Load Shedding
V2G	Vehicle-to-Grid

Executive Summary

In 2021, the “Net Zero Strategy: Build Back Greener” set out the UK government’s policies and proposals for decarbonising all sectors of the UK economy [1]. This strategy states a particularly fast decarbonisation of the electricity sector. By 2035, all electricity will need to come from zero-carbon sources, which will require significant investment in low/zero carbon generation, electricity networks, different flexibility technologies and advanced control concepts, to support a cost-effective transition to a zero-carbon system without risking security of supply.

Balancing electricity supply and demand in a low-carbon electricity system is increasing in complexity due to hours with high production of Renewable Energy Sources (RES) which reduce inertia levels [2]. Inertia is an element available in synchronous generators and loads that can resist drops in frequency, giving the grid time to rebalance supply and demand. Frequency must always be close to the nominal operating point (50 Hz in the UK) in order to maintain security of supply (avoid customer disruptions), even when a generation or load outage occurs. The inertia in rotating steam, gas or hydroelectric turbines naturally slows the rate at which frequency changes after a fault, but most RES, such as wind and solar photovoltaic (PV), do not currently provide any inertia. This means power systems are becoming more vulnerable to frequency deviations, making frequency containment Ancillary Services (AS) more valuable during hours with high-RES generation. Great Britain (GB) during the COVID-19 national lockdown [3], is a very informative example of this trend.

In this context, security of supply for low carbon electricity systems is recognised as a growing challenge for system operators. However, current electricity markets do not provide clear incentives for maintaining frequency security [4]. Moreover, consideration of services such as inertia, which depends on the amount of generation synchronised to the grid, challenge current pricing mechanisms[5].

This white paper, *Market design for low-inertia power systems: Ancillary services in GB*, proposes a novel market mechanism that will align the market design with the simultaneous scheduling of energy and frequency containment AS provision in an economic and efficient manner. AS provision will be potentially addressed by a wide variety of providers, including conventional synchronous generators, RES, storage systems, and Electric Vehicles (EVs), which will be incentivised by the proposed future market design. Furthermore, this methodology provides a new framework for allocating the cost of procuring frequency containment AS among the market players that trigger these requirements. The proposed methodology distributes the payments for frequency containment AS among all the market participants that create the need for these services.

This novel market design will provide appropriate incentives for the development/deployment of advanced technologies and control concepts to support a cost-effective transition to a secure low/zero-carbon electricity system, through the provision of appropriate AS. Although at present, electricity consumers directly pay for all AS costs (i.e., costs are socialised), we argue that electricity generators should be responsible for covering these costs, as it is detailed in [6]. This

cost allocation is inspired by the work presented in [7], in which contingency reserve costs are allocated among generation units considering their size and probability of unavailability. This market design is the core new concept, shifting from the historical approach based on socialising AS costs, to an efficient cost allocation concept recognising that the size of the plant is the key driver for these costs.

This white paper concludes by highlighting the potential impact of the proposed market design on advancing technology deployment and supporting a cost-effective transition to a secure low-carbon electricity system.

Acknowledgements:

Development of the fundamental modelling for addressing the market design for low-inertia power systems has been supported by two EPSRC projects and one EU Horizon project: (a) Integrated Development of Low-carbon Energy Systems (IDLES), (b) UK Energy Research Centre and (c) TradeRES. Carlos Matamala was supported by the National Agency for Research and Development (ANID) through Scholarship ANID/PROGRAMA BECAS CHILE DOCTORADO under Grant 2020-72210414.

1 Introduction to the Main Report

1.1 Challenges Associated with Secure Operation of Low Inertia Systems

Power systems with high penetration of Inverter-Based Resources (IBR), such as wind and solar, radically differ from those dominated by conventional synchronous generators in fossil-fuelled, nuclear or hydro-electric power stations. To guarantee electricity supply, frequency security must be maintained, which means that the system's electric frequency must be contained within a narrow band around the nominal value (50 Hz in GB). This frequency is determined by the rotating speed of synchronous generators. They spontaneously release kinetic energy when demand exceeds generation, slowing down their rotation speed and reducing the system frequency. Inertia refers to the buffer of kinetic energy from the rotating masses in synchronous machines (both the generator and the turbine that drives it): a buffer that reduces the rate at which the system frequency would fall.

As most renewables are decoupled from the power grid through power electronics converters, currently these do not naturally provide inertia. This means that in hours with high renewable generation and few synchronous generators providing inertia, system frequency stability is more vulnerable to generation/demand imbalances [2]. This presents a threat to security of supply.

An example of a high renewable operative condition occurred during COVID-19 in GB. During this period, a low demand combined with high renewable outputs created conditions that were not expected until renewables capacity increases to meet emissions targets in coming years. The GB system experienced periods of very high instantaneous production of non-synchronous generation, compromising system stability due to the lack of inertia in the grid. Hence, the GB National Energy System Operator (NESO) had to procure additional services and rapidly create new offerings to prevent frequency imbalances. These additional AS provided between May and July 2020 cost 3 times more than the year before (£302 million compared to £101 million) [3]. Efforts to decrease the needed inertia levels in the system are currently being considered by the NESO [8].

Figure 1 shows the potential change in system operation costs in order to efficiently meet the CO₂ reduction target. Although the overall system operation costs are expected to reduce due to high penetration of low marginal cost low-carbon generation (particularly RES), the costs of AS will potentially increase more than 10 times, relative to 2020 levels in a 'business as usual world'. Furthermore, as can be seen in Figure 2, the economic value of inertia can increase more than 100 times in hours with low demand and high wind generation [9].

However, electricity markets have not yet matured to recognise the relevance of the inertia market. As inertia depends on the commitment of generators, the pricing methodology for inertia provision will inherently need to deal with non-convexities, which can create the need for uplift payments. These are side payments given to generators that fail to recover their costs through energy revenues, which are

provided to ensure that all market participants follow the system dispatch.¹ Convex hull pricing is a methodology conceived to minimise uplift payments [5]. Given its properties, it is applied for pricing the Unit Commitment (UC) model proposed in this work.

In this context, this research work analyses the challenge of procuring AS in GB and demonstrates how the electricity market should evolve as the power system reach the targets for Net-Zero. As the GB system will experience these major balancing challenges much before mainland Europe and the United States, due to its island position and combination of renewable and large-scale nuclear generation, this may guide other electricity systems considering their new challenges associated with the management of frequency stability.

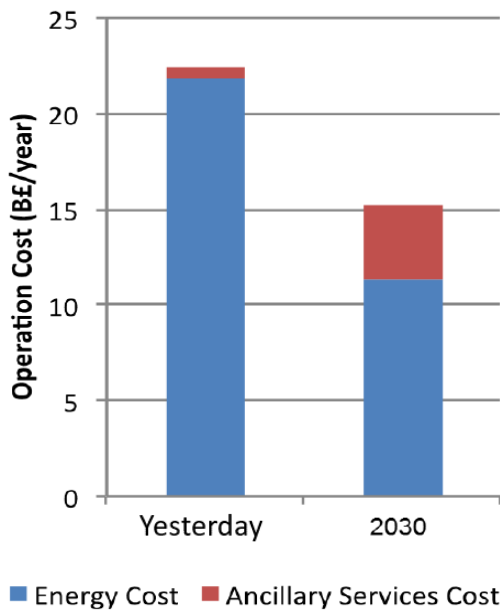


Figure 1: Expected change in system operating costs. Source: [9].

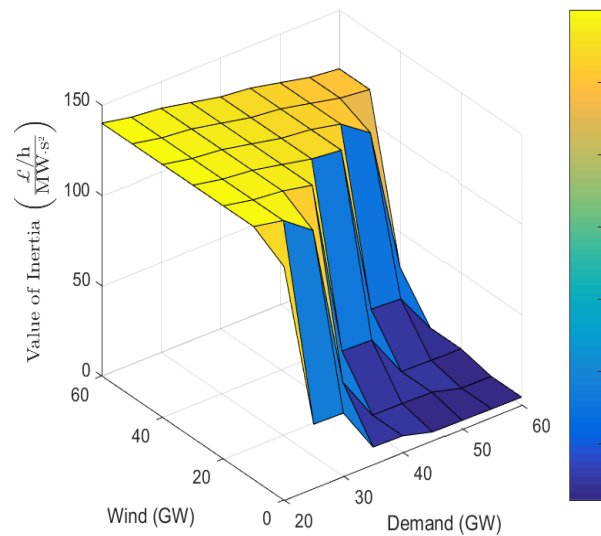


Figure 2: Change in economic value of inertia. Source: [9].

1.2 Comparison of Ancillary Services Markets

In the GB, electricity is settled in “settlement periods”, which are defined as 30-minute periods, defining 48 settlements periods per day. The wholesale market can be traded from years ahead of schedule until one hour before delivery, also known as “Gate closure”. The trading can be done through bilateral contracts between generators and suppliers or through power exchanges with platforms such as EPEX SPOT and Nord Pool. At Gate Closure, planned generation and demand volumes are published, also known as Final Physical Notifications (FPNs).

¹ The electricity markets in the US, where the prices of electrical energy are determined by the marginal cost of providing slightly more power from the stations already operating, frequently make uplift payments to generators that are being scheduled for other reasons, such as to provide reserve or relax a constraint on the transmission system. This could happen when the energy prices it would receive are too low to cover the cost of starting up the unit, for example.

With the FPNs, the NESO can project any imbalances between generation and demand in the system. Any mismatch is settled in the Balancing Mechanism (BM), in which the NESO balance the system in real-time. To participate in the BM, Balancing Mechanism Units (BMUs) submit bid and offer prices at Gate closure. The bid price is the price for flexing down in which a generator would reduce its generation or a demand site would increase its demand (buying electricity in each case), while the offer price is the price for which the BMU will sell electricity, increasing generation or reducing demand. Between Gate closure and the end of the settlement period, NESO determines which bids and offers to accept to balance the system, known as Bid Offer Acceptances (BOAs), and these volumes are paid as bid.

Some of the NESO's actions are needed because it is almost inevitable that some assets will not deliver their promised FPNs and BOAs – generators have outages, and customer demand differs from the forecast. ELEXON, the organisation that manages the Balancing and Settlement Code (BSC)², calculates these imbalances and an imbalance price based on the cost of the most expensive actions taken by the NESO. Participants with a shortfall (insufficient generation or excessive demand) must pay to buy more electricity; those with a surplus of electricity are paid as they sell it back.

Given its relatively low inertia levels, frequency containment AS are defined in some detail in GB. Competitive day-ahead auctions are traded, and the procurement of services is paid as clear. These frequency response services are Dynamic Containment (DC), Moderation (DM), and Regulation (DR). Even though there are plans of co-optimising the procurement of frequency response and reserve products by the NESO, under the Enduring Auction Capability [10], there are no plans of co-optimisation of these services along with energy production and inertia procurement.

GB system inertia is mostly provided as a by-product of energy provision and is not defined as an open, transparent, and competitive market. However, since 2020, NESO has launched the Stability Pathfinders, which create competitive tenders to provide long-term contracts to units that provide inertia with reduced carbon emissions. So far, the winners of these tenders have been synchronous condensers, already operating in the system and Battery Energy Storage Systems (BESS) units with Grid-forming (GFM) inverters that can provide synthetic or virtual inertia, which will likely start operating in 2024. The NESO has confirmed that the Stability Pathfinders project has finished, and it is now exploring through the Stability Market Design project the best options to define the inertia market³.

In the case of the US, the provision of energy and AS are co-optimised by independent system operators (ISOs), thus allowing more efficient scheduling compared to the separated stages used in GB and the rest of Europe. However, the market for frequency containment AS is less developed in the case of the US as most of the systems are interconnected through AC links, thus avoiding the risks associated with low-inertia operation.

² BSC can be found in <https://www.elexon.co.uk/bsc-and-codes/>

³ Stability Pathfinder and Stability Market Design are detailed at: <https://www.nationalgrideso.com/industry-information/balancing-services/pathfinders/noa-stability-pathfinder>

In most US ISOs, the frequency containment AS market can be classified as regulation (up and down) and spinning and non-spinning reserve [11], among others with slower dynamics. A more developed market for AS operates in the Electric Reliability Council of Texas (ERCOT) system. Given that ERCOT is not linked through AC interconnections to the rest of the country, frequency containment AS cannot be imported from neighbour power systems, and inertia can only be provided locally. As in the other US systems, ERCOT also co-optimises the energy and AS markets, while the AS market covers services such as regulation (up and down) and responsive reserve, providing fast-responding, short-duration AS. Meanwhile, slowing-responding-longer-duration AS are defined as non-spinning reserve and ERCOT Contingency Reserve Service (ECRS).

A more developed market for frequency containment AS can be found in Australia. According to the Australian Energy Market Operator (AEMO), Frequency Control Ancillary Services (FCAS) can be either regulation or contingency services. The dispatch engine of AEMO, known as National Electricity Market Dispatch Engine (NEMDE), performs a co-optimisation that minimises the total cost of energy plus FCAS. NEMDE determines the clearing price for the FCAS markets in each region and trading interval. This price settles the payments for FCAS providers [11][12], [13].

However, in comparison with GB and ERCOT, a key difference of the market design established by AEMO is the way that FCAS payments are recovered from market participants. Generators pay for Contingency Raise Services (loss of the largest generator) and additional Rate of Change of Frequency (RoCoF) Requirements, while customers pay Contingency Lower Services (loss of largest load/transmission element) [14].

The mechanisms to recover Contingency Raise Services and additional RoCoF Requirements payments make responsible generators pay for the share of the frequency containment costs that they create, which differ in the two main systems under AEMO operation. In the case of the National Electricity Market (NEM), AEMO defines a proportional cost allocation rule [14] that will be discussed in section 3.2.1. In the case of the Wholesale Electricity Market (WEM), a “runway” allocation method is used [13], which is similar to the Shapley value methodology, providing more suitable cost allocation properties, as will be discussed in section 3.2.2.

1.3 Purpose of the White Paper

This white paper sets out a fundamentally new proposition for future electricity market design to incentivise cost-effective operation of low-carbon electricity systems with reduced levels of system inertia. This is a significant scientific challenge, as currently there is a lack of appropriate mathematical models to provide evidence regarding the development of cost-effective market mechanisms that would incentivise affordable real-time balancing of demand and supply. This is a critical issue for maintaining frequency stability due to the massive reduction in system inertia in low carbon electricity systems.

This research has practical relevance, given the ambitious emission targets set by electricity system in the UK, which makes market reform for enabling renewable integration a critical research area of growing importance.

The main content of this white paper is presented as follows: Section 2, *Modelling Approach* describes the nature of the problem considering the dynamics of the power system in the economic optimisation; Section 3, *Who Should Pay for Frequency Containment Ancillary Services?*, proposes a new market design framework that would distribute the total payments for AS among all generators or loads that create the need for these services; Section 4

Ancillary Services Market Driven Incentives, puts in place the cost-effective incentives for market players to provide AS, such as inertia and frequency response, in the most efficient way possible and proposes alternatives for AS provision for power systems with high-RES penetration; Section 5, *Conclusions and Future Work*, wraps up key topics addressed in this white paper and highlights the future work.

2 Modelling Approach

2.1 Condition for Frequency Stability in the Grid

If supply and demand fall out of balance, such as when a power plant suddenly disconnects from the grid due to a failure, then the system frequency is affected as can be seen in Figure 3. For the GB system, maintaining a consistent frequency of 50 Hz is important because deviating too far from the standard 50 Hz could inflict damage to equipment, create the instruction of Under Frequency Load Shedding (UFLS), or even a blackout in the system.

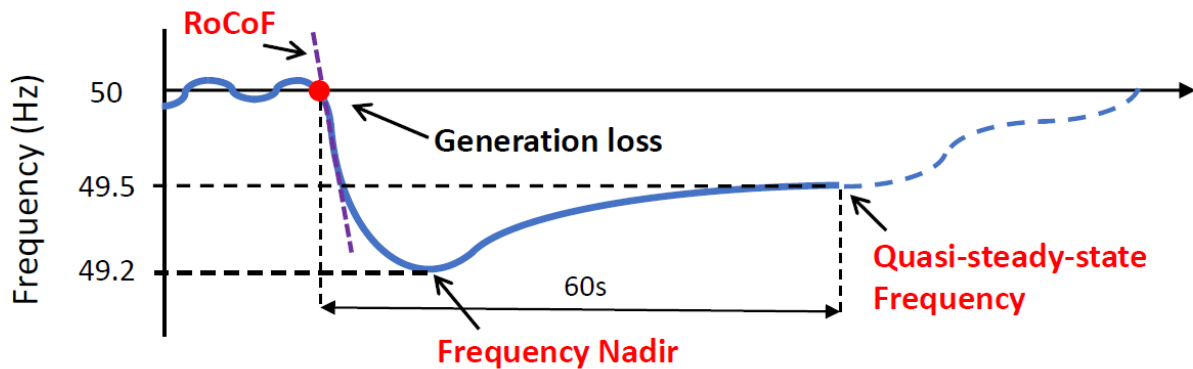


Figure 3: Frequency stability.

The key modelling challenge lies in incorporating the differential-equation-driven frequency dynamics, defined by the swing equation⁴ [15], into an algebraic-equation-constrained optimisation problem such as the UC that needs to be solved for 30 min periods, as shown in Figure 4. Furthermore, the constraints for frequency security must be computationally efficient since they may be used in computationally intense problems like Stochastic Unit Commitment models.⁵ The full methodology can be seen in Appendix 1 - Unit Commitment Model.

The relevant boundaries for frequency stability are shown in Figure 3. The three main constraints to respect after an outage are: (a) Rate of change of frequency, (b) Frequency nadir and (c) Quasi-steady-state, described in the next section.

⁴ The swing equation describes the time-evolution of frequency deviation after a unit's outage.

⁵ Stochastic unit commitment models are optimised across a range of possible scenarios for the near-term future (such as for different levels of demand), greatly multiplying the size of the problem to be solved.

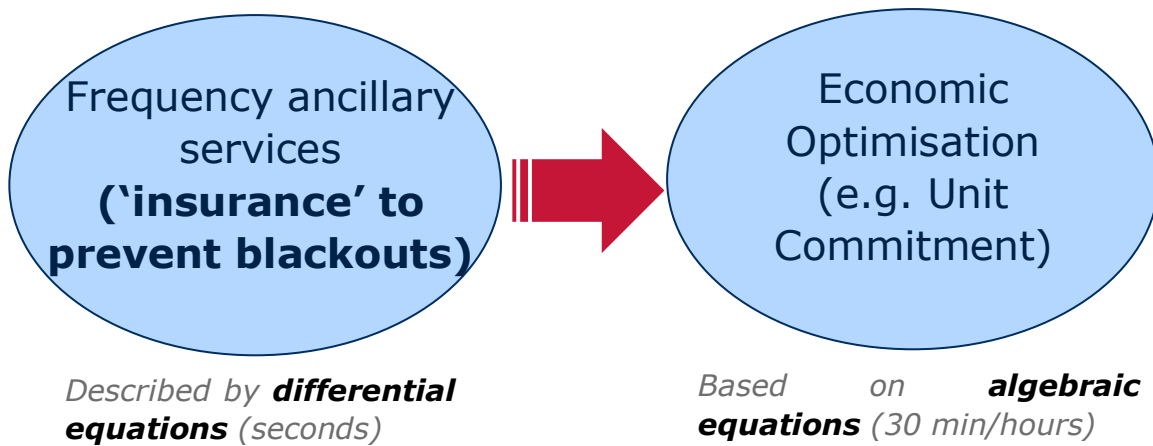


Figure 4: Challenge for the economic analysis of frequency stability.

2.1.1 Rate of Change of Frequency

Frequency levels must not fluctuate by more than 1 Hz per second, as a rapidly changing frequency could damage other generators connected to the grid. The RoCoF is greatest, and hence the RoCoF constraint is defined, at the very start of the contingency, as can be seen in Figure 3. Given their natural delays, frequency response services cannot support RoCoF. Therefore, RoCoF directly depends on the post-contingency inertia level in the system – the less inertia, the greater the RoCoF. The maximum allowable RoCoF thus defines the minimum level of inertia that should be available in the system. Each operating generator contributes a fixed (but individual) amount towards meeting this constraint, giving a linear equation that can be easily derived from the swing equation [15] and is not computationally demanding for the UC problem.

Figure 5 shows the post-contingency frequency dynamics for a 1.8 GW generation loss considering high and low system inertia. As can be seen, operating conditions with a lower inertia level implies a faster frequency decline, which could potentially exceed the RoCoF limit.

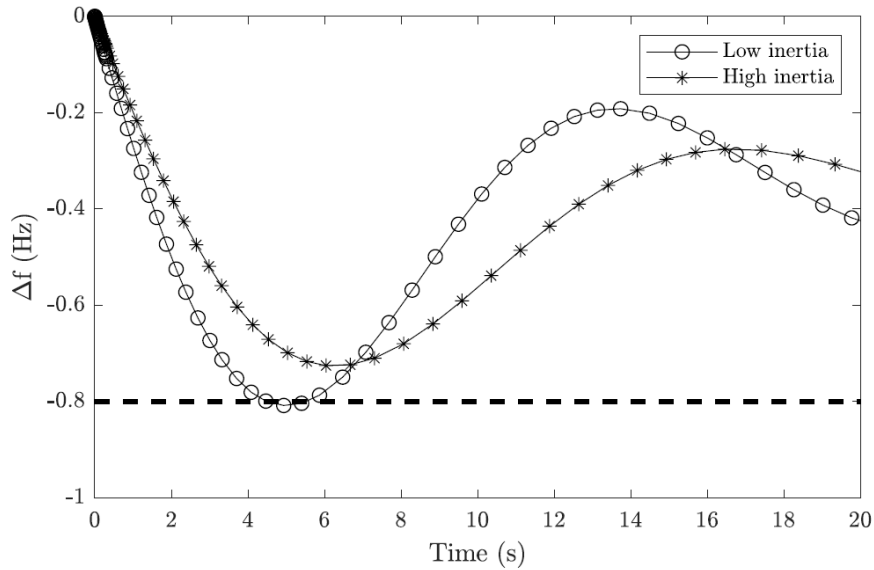


Figure 5: Contingency comparison for low and high inertia operating conditions.
Source: Imperial's modelling analysis.

In hours with high renewable generation, maintaining minimum levels of inertia implies synchronising more expensive out-of-merit units operating at their minimum stable generation. RoCoF standards are now under review and could be relaxed to meet the zero carbon ambition by 2025⁶ [8].

2.1.2 Frequency Nadir

As illustrated in Figure 3, frequency must not fall below a certain minimum level to avoid the activation of UFLS. This minimum level is called the nadir point and, in the case of the GB system, is 49.2 Hz. It is essential to keep the system frequency above that level, otherwise generators might be damaged and would disconnect to protect themselves, leading to uncontrolled blackouts. UFLS implies the disconnection of demand that would restore the generation-demand balance and, therefore, bring the frequency back towards the desired 50 Hz. Currently, the cost of UFLS is high for society, as (almost) randomly⁷ disconnecting consumers could cause great inconvenience and represents one of the last resource options. Given the transition to a digitalised energy paradigm, however, more flexible demand could provide a strong contribution at low cost.

Maintaining frequency above the nadir point involves a coupled provision of AS, such as inertia and different frequency response services. In the GB, one of the main challenges is to provide market incentives for the provision of AS, and correct prices will be key for guiding the market outcomes. In the UC problem, even though the nadir constraint is non-linear, it can be represented as a convex constraint [16]. This

⁶ According to the Frequency Risk and Control Report (2023), RoCoF limits could be relaxed to operate the system with lower inertia levels by 2025.

⁷ Essential users such as hospitals should be protected from being selected for UFLS, but the events of 9 August 2019, when UFLS took place and a million customers lost power after two power stations failed in quick succession, showed that two hospitals (both with back-up generation) had been disconnected.

convex constraint allows the computing of prices to make market players recover their operational costs in the short term.

2.1.3 Quasi-steady-state

As shown in Figure 3, the quasi-steady-state (q-s-s) represents the stable value, normally smaller than the system nominal frequency, that frequency must reach after all AS are provided. Specifically, maintaining an acceptable q-s-s implies that the amount of all frequency response services must be at least the same as the loss of generation. Q-s-s is represented as a linear constraint, which is suitable for its representation in the UC problem.

2.2 Ancillary Services for Frequency Containment

Frequency containment AS are scheduled to maintain the system's frequency within secure ranges; AS will only come into play in the event of a frequency drop to ensure grid stability, as shown in Figure 6. A more specific description is given as follows. The complete model description is shown in Appendix 1 - Unit Commitment Model.

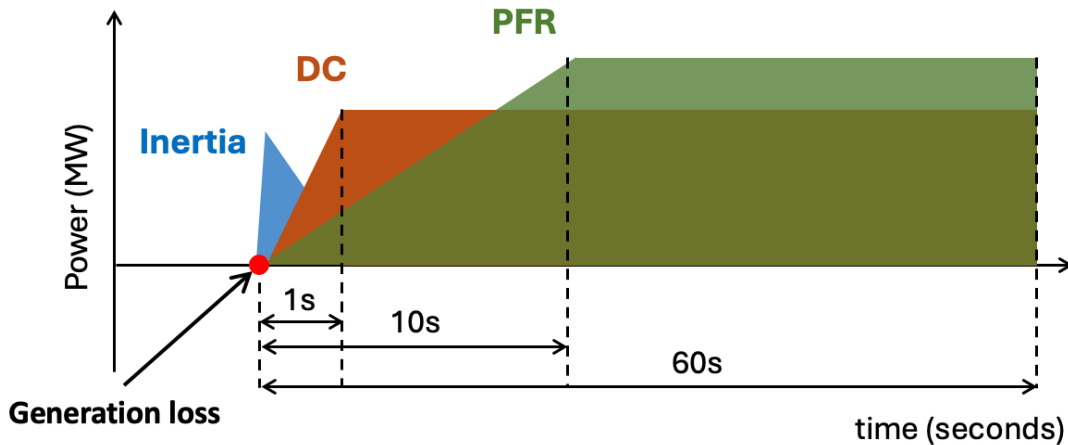


Figure 6: Frequency containment AS, including inertia, Dynamic Containment (DC) and Primary Frequency Response (PFR). Source: [3].

2.2.1 Inertia

Refers to the kinetic energy stored in rotating masses, proportional to its moment of inertia and the square of its angular velocity. It is essential in maintaining stability. If there is a sudden loss of generation causing an imbalance between generation and load, the inherent physical behaviour of the rotating machines is to draw on this mechanical kinetic energy to restore the balance.

2.2.1.1 Synchronous Inertia

Synchronous inertia is the prime source of system inertia available in power systems. It is stored as kinetic energy in the rotating masses of synchronous machines, such as thermal or hydropower plants. These masses are controlled to rotate at a certain speed to produce power injection of 50 Hz to the grid. As only thermal or hydropower plants provide synchronous inertia, it is challenging to maintain high levels of it if their output is displaced by high volumes of nonsynchronous RES.

2.2.1.2 Synthetic Inertia

Synthetic inertia can be provided by wind turbines if equipped with GFM converters, where the kinetic energy stored in their rotating masses can be used for this service. Specifically, synthetic inertia can be provided by few seconds, extracting the rotating mass' kinetic energy. After this, there is a "recovery period" that slows down the turbine, therefore deviating it from the pre-fault power injection. As stated in [17], the recovery period would exist during the q-s-s timeframe; thus, it does not influence the nadir constraint, as it occurs after. The power delivered to the grid by a wind turbine equipped with GFM inverter after a sudden imbalance can be seen in Figure 7. The specific aspects of synthetic inertia and recovery effect of wind turbines are included in the modelling.

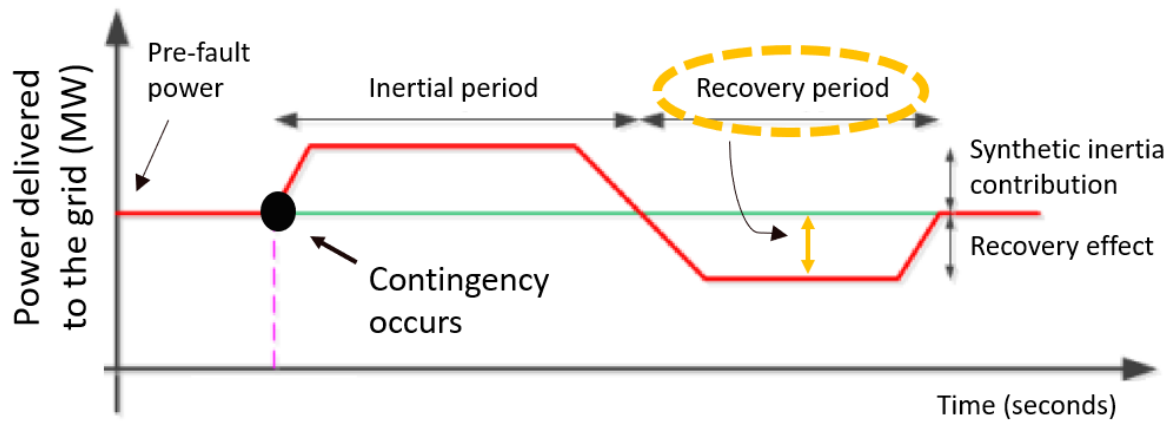


Figure 7: Synthetic inertia and recovery effect of wind turbines [18].

2.2.2 Frequency Response

Complementary to inertia, frequency response services are needed to maintain grid stability in low-inertia systems.

2.2.2.1 Primary Frequency Response

Primary Frequency Response (PFR) was traditionally provided by synchronous generators, which require a control mechanism known as a governor to provide an additional fuel injection to deliver extra power to the grid. Nowadays, batteries can also provide a similar service. PFR should fully act in the order of 10 seconds after the event [19]. This timing constraint can be met by synchronous generators, such as thermal gas plants and Pumped Hydro Energy Storage (PHES), that are already synchronised to the grid.⁸

EVs, with the appropriate control infrastructure in the distribution and transmission system, could potentially provide PFR to the system when charging and/or discharging. In this case, EV charging strategies such as smart charging or Vehicle-to-

⁸ Generators that have not been synchronised can provide non-spinning reserve, starting up on slightly slower timescales (a few minutes) to allow the units that have provided PFR to reduce their output to a level that allows them to provide PFR again when next required.

Grid (V2G) will play a key role in defining the availability of the service. Further details can be found in section 4.2.

In this white paper, we consider PFR as a service that can be fully delivered in 10 seconds and maintained for several minutes after the outage. Compared with the Dynamic Response Services defined by the NESO⁹, PFR is similar to Dynamic Moderation (DM), given that one is also considered for large imbalances, but needs to be fully delivered in a shorter timescale.

2.2.2.2 Dynamic Containment

Dynamic Containment is a Dynamic Response Service that reacts in less than one second after a loss of generation or demand. BESS are the primary providers of Dynamic Containment, but it could be obtained from any source that can provide frequency response in less than 1 second. IBRs such as wind turbines operating with curtailed output could potentially provide a fast power injection in the order of 1 second following the power imbalance [17].

2.3 Importance of Coordinating Energy, System Inertia, and Frequency Response

As inertia is a by-product of energy production from synchronous machines, defining a new market for it was not relevant in the past – far more was provided than required. However, with the large-scale integration of RES, conditions are changing, and there is a need to define a market for frequency containment AS, in which all units can get incentives for their AS provision, along with their energy production.

However, the provision of energy and AS is currently performed in separate markets in GB and many other places with a similar market design. First, energy is traded, and after this, if the energy market does not ensure that the system will remain stable and safe, then AS are procured separately. This differs from the market design of the organised wholesale markets run by ISOs in the US, where energy and AS are procured together and co-optimised. The NESO in GB is also active in developing advanced models that would co-optimize the provision of frequency response [10].

As shown in Figure 8, net energy demand levels and the need for Dynamic Containment interact. In this case, the need for Dynamic Containment is lower during high net demand hours. When overall demand is high or renewable output low, more output has to come from thermal units, increasing the levels of inertia in the system, which decreases the need for other frequency containment AS.

⁹ <https://www.nationalgrideso.com/industry-information/balancing-services/frequency-response-services/new-dynamic-services-dcdmdr>

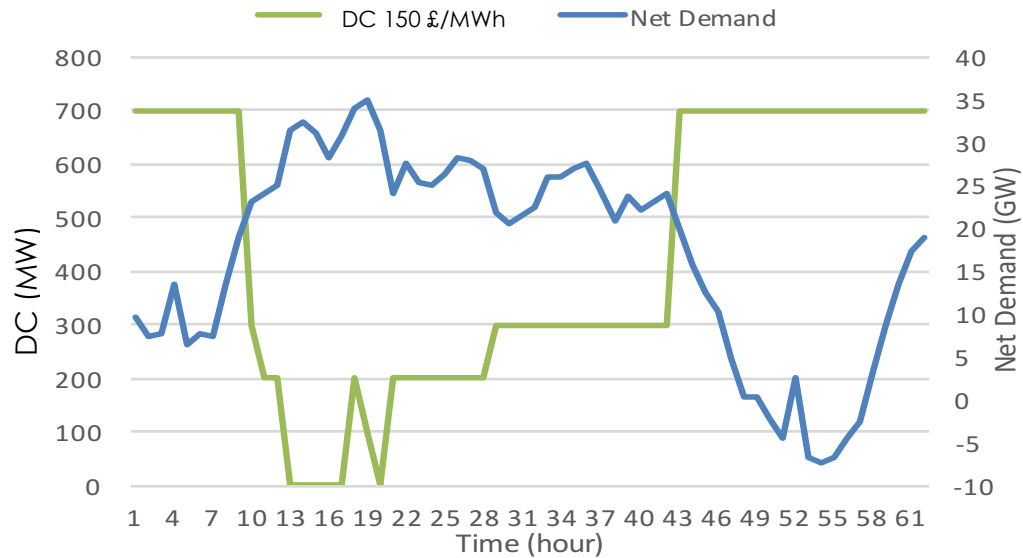


Figure 8: Interaction between the net energy demand and the Dynamic Containment (DC) market.
Source: Imperial's modelling analysis.

There is also an interaction between different services inside the AS market. As seen in Figure 9, Dynamic Containment can effectively reduce the need for inertia in the system. Even though there is always a minimum needed level of inertia because of the RoCoF constraint, the nadir constraint will also increase the need for inertia if there is insufficient frequency response in the system.

The interactions presented in Figure 8 and Figure 9 show that a co-optimisation that considers all components, including energy and the variety of AS for frequency containment, should be performed in order to obtain the most economically efficient system operation, while maintaining frequency stability.

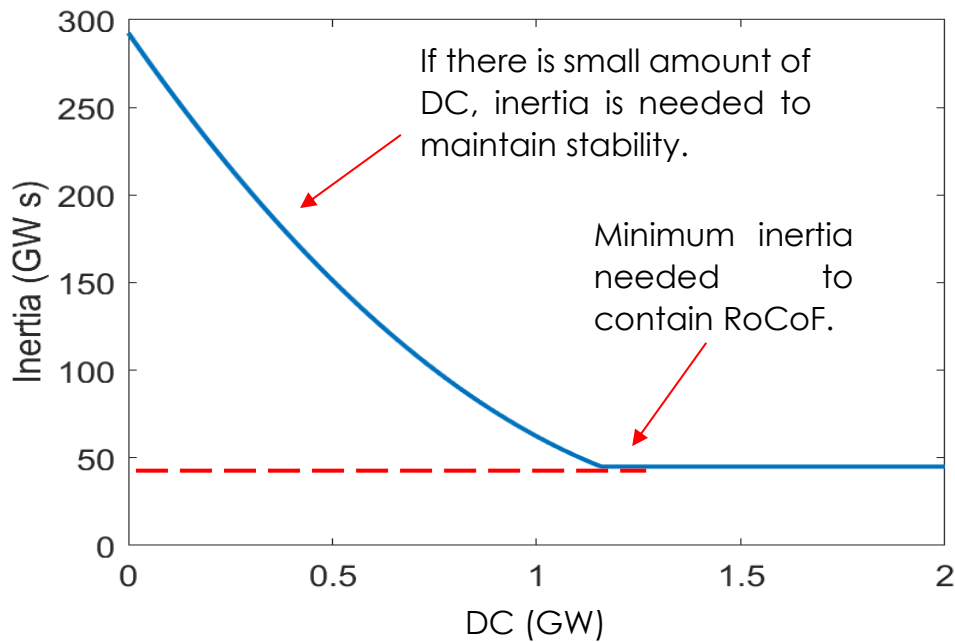


Figure 9: Interaction among services in the AS market: inertia and Dynamic Containment (DC).
Source: Imperial's modelling analysis.

2.4 Pricing Ancillary Services

The pricing mechanism should be set considering the constraints that guarantee frequency security, i.e., RoCoF, frequency nadir and q-s-s, and the energy supply. These prices are quantified from the whole-system optimisation using corresponding Lagrange multipliers,¹⁰ where the full derivation of AS prices is shown in Appendix 2 - Pricing Methodology. The market is assumed to run on a paid-as-clear basis, where all sources (generators, batteries and so on) providing a given service are paid the price offered by the most expensive source required to do so.

The resulting market prices for a co-optimised energy and AS market are shown in Figure 10, for an operational week with the data presented in Table 1, which represents the 'Leading the Way' scenario within National Grid ESO's Future Energy Scenarios (FES) for 2030.

As can be seen, energy and AS prices show opposite trends: when energy prices are high, AS prices are low (and less relevant). This is because high energy prices are driven by the dispatch of thermal units, which also provide inertia as a by-product of energy. Even though it is still necessary to schedule a certain amount of AS in these hours, the additional cost of AS becomes very low for the optimisation (for example,

¹⁰ In Lagrangean optimisation, the multiplier on each constraint measures the marginal cost of slightly tightening that constraint (the multiplier on the constraint that energy demand equals supply gives the marginal cost of generating an additional MWh) and hence the price that a generator should receive for providing the service that relaxes it. Where this kind of optimisation is used within an electricity market, as in the US, generators offer prices for each service from each unit and these are used as the costs within the optimisation.

generators running at less than full capacity can (quite) cheaply provide PFR, reducing significantly the corresponding prices.

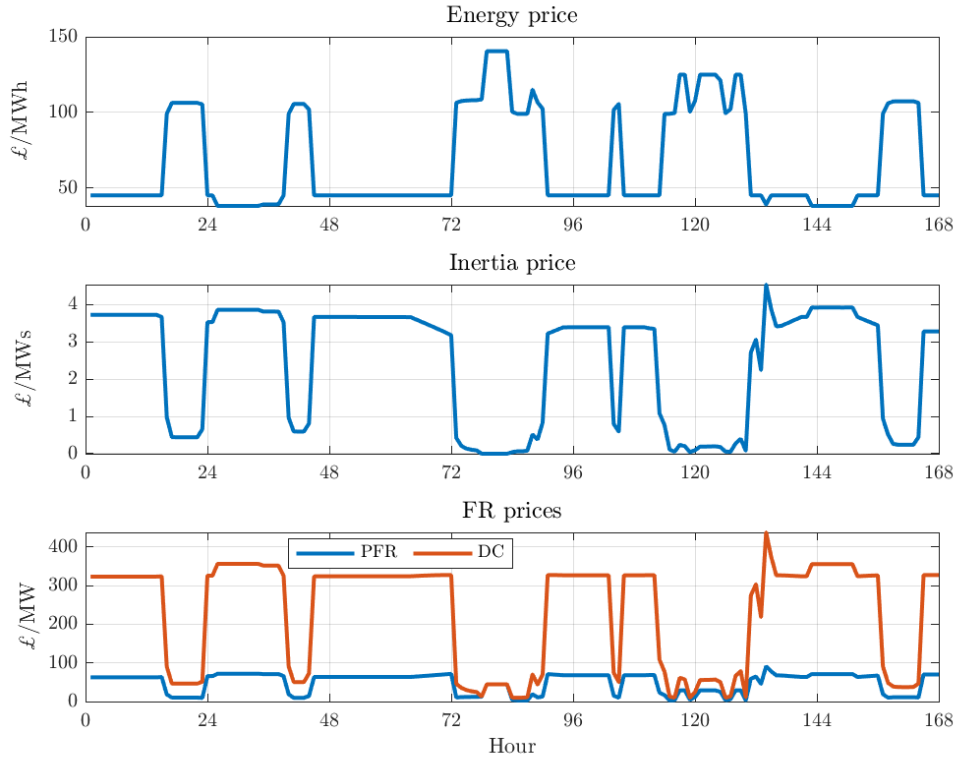


Figure 10: Energy and AS market prices. Source: [6].

Table 1: Characteristics of Generators, RES and Storage, based on the 'Leading the Way' scenario within National Grid ESO's FES for 2030. Source: [6].

	Big Nuclear	Nuclear	CCGT	OCGT	Offshore Wind	Onshore Wind	Solar PV	PHES	BESS
Installed capacity (GW)	1.8	3.2	25	1	50.4	30	42	4.8	20
Power range (MW) ($P_g^{\text{msg}}, P_g^{\text{max}}$), ($0, P_r^{\text{max}}$), ($-P_s^{\text{max}}, P_s^{\text{msg}}, P_s^{\text{max}}$)	1800, 1800	1600, 1600	500,1000	50,100	0,1800	0,600	0,250	-400,0,400	-100,0,100
Inertia constant (H_g, H_s) (s)	5	5	5	5	N/A	N/A	N/A	5	N/A
Response type	N/A	N/A	PFR	PFR	N/A	N/A	N/A	PFR	EFR
Response capacity (MW)	N/A	N/A	$30\% \cdot P_g^{\text{max}}$	$30\% \cdot P_g^{\text{max}}$	N/A	N/A	N/A	$20\% \cdot P_s^{\text{max}}$	$5\% \cdot P_s^{\text{max}}$
Energy price offer λ_g^E , λ_r^E, λ_s^E (£/MWh)	78	78	99	222	47	45	39	60	50
Inertia price offer λ_g^H , λ_s^H (£/MWs)	1	1	2	5	N/A	N/A	N/A	1	N/A
FR price offer λ_g^{PFR} , $\lambda_s^{\text{PFR}}, \lambda_s^{\text{EFR}}$ (£/MW)	N/A	N/A	3	7	N/A	N/A	N/A	5	10
Fuel costs (£/MWh)	8	8	90	137	0	0	0	0	0

3 Who Should Pay for Frequency Containment Ancillary Services?

This work has highlighted the importance of the AS market and its relevance for a cost-effective transition to the decarbonised electricity system. This section goes further with the analysis of the nature of this market, while identifying the core components that trigger its necessity.

The AS market presents a core difference from the traditional energy market. In the case of the energy market, the need for energy production is driven by consumer demand, for which different tariffs apply. In the case of the AS market, frequency-containment AS are procured because of the need to deal with the largest generation/demand loss in the system. If there were no significant losses in the system (e.g., an ideal electricity system with only very small size generation units), there would be no need to schedule significant amounts of frequency containment AS, such as inertia and frequency response.

However, no specific tariffs apply to the triggers of the AS market. In contrast, the costs are commonly socialised to consumers, thus not providing market signals that should create incentives for the reduction of AS costs and support more cost-effective transition to zero carbon energy future.

In this context, one of the core questions that arises is:

- Who should pay for frequency containment AS?

The methodology presented in this section is inspired by the concept proposed in [7], which is further developed in [6]. In [7] the allocation mechanism is based on the capacity of generating units and their probability of unavailability. The allocation considers a proportional cost allocation concept based on the expected undelivered power generation, as the main driver for the mechanism.

On the other hand, [6] highlights the importance of contingency size in determining the appropriate volume of AS required to manage various system outages. It also introduces a cost-allocation methodology designed to distribute the total costs of frequency-containment AS across market participants, based on their contributions to the demand for these services. Two different mechanisms are considered to achieve this: 'Proportional' and 'Shapley value', which will also be discussed in this section.

3.1 Relevance of Contingency Size

As shown in Figure 11, the most relevant variable for frequency security is the size of the contingency, the amount of generation that might be lost, P^{Loss} . In any circumstance, the system operator would secure the grid against the largest loss (1.8GW in GB, the capacity of a big nuclear unit), where, in this example, this contingency would stress the system to its stability limit in low inertia hours, leading to the maximum acceptable frequency deviation. In this context, the second largest loss (1.4GW, that could be represented by a regular-size nuclear unit or an offshore wind

farm), would be managed by using the AS procured to cover the largest loss, and the same AS will be used in smaller outage conditions.

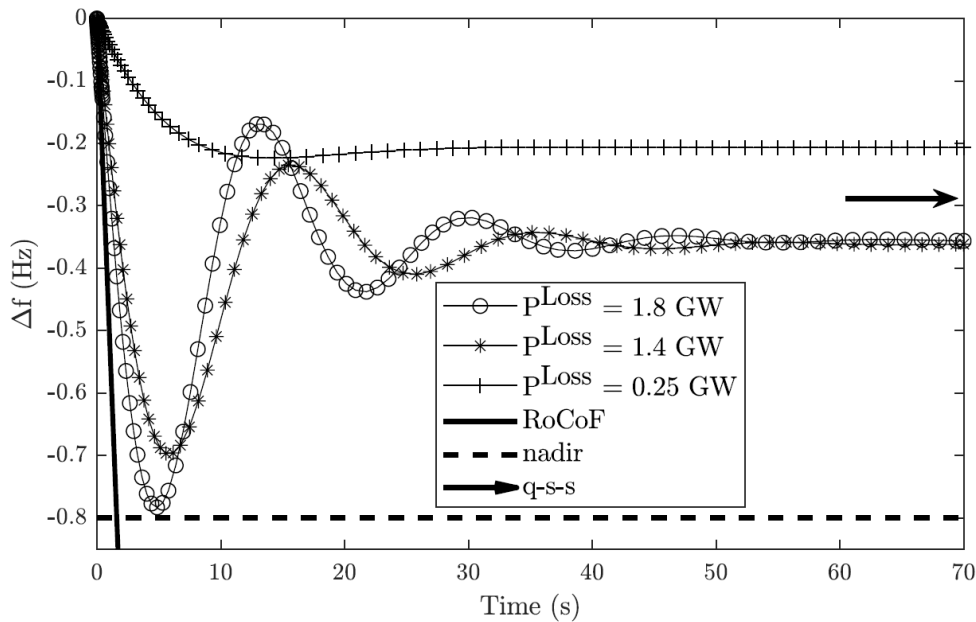


Figure 11: Schematic of system frequency following three different generator contingencies.
Source: [6].

To study the market impact that each generating unit would have in the AS market, Figure 12 presents week-long generation profiles per unit and technology, and the total AS market (quantity per price of frequency containment AS) that each unit would create if the AS were scheduled to contain (only) its outage, based on the data presented in Table 1.

As shown at the top of Figure 12, "Dispatch per technology" represents the whole market dispatch including all units from each technology. Note that nuclear units are differentiated between Big Nuclear, which represents the biggest unit in the system, and all the regular-size nuclear units, as also described in Table 1.

In the middle of Figure 12, "Dispatch per unit" represents the single dispatch per each type of unit in the system, where this dispatch profile is used to calculate the AS market per unit.

Finally, at the bottom of Figure 12, in the "Total AS market size per unit", the Big Nuclear plant (in blue) defines the actual size of the AS market (as it is the largest loss, with a constant generation of 1.8GW). Other big units, such as regular-size Nuclear and Offshore Wind plants, would also create requirements in the AS market. Given the significant difference in the corresponding generation capacities, smaller units, such as Combined Cycle Gas Turbines (CCGTs) and Open Cycle Gas Turbines (OCGTs), would create a very small AS market. Note that the Big Nuclear unit represents the only "real" AS market cost, as it is the biggest dispatch each hour, so the NESO will consider the security for that level of possible outage. The other AS market costs are "artificial"; as their dispatch is smaller, the NESO has already considered security for

these dispatch levels as well. The artificial costs are calculated here to distribute the payments using the cost allocation methodologies that will be presented in the next section.

Although electricity consumers directly bear frequency containment costs (i.e., the costs are socialised), in the proposed market concept, generators (including interconnectors) would be responsible for covering these costs. This methodology is described in the next section.

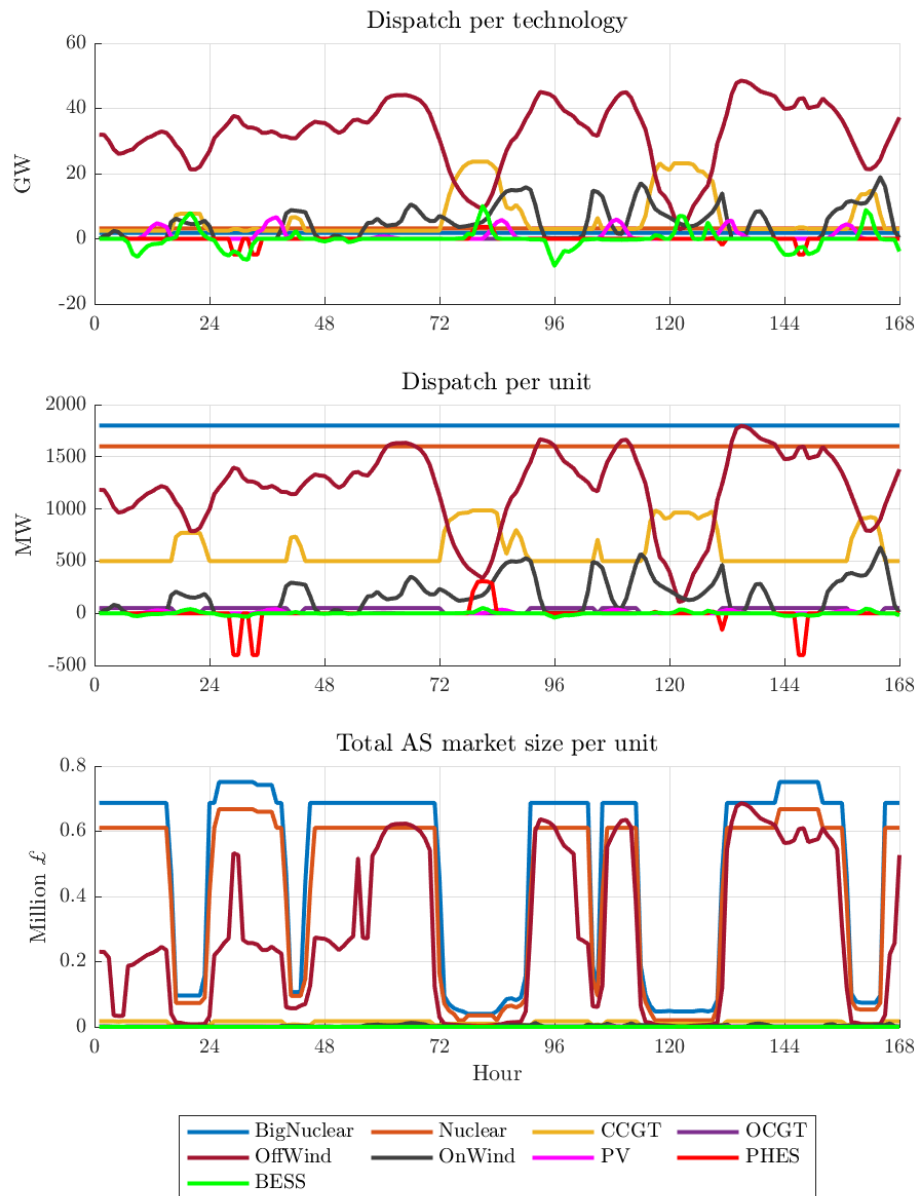


Figure 12: Dispatch per technology and per unit, and frequency containment AS market costs per unit (Ω). Source: [6].

3.2 Cost Allocation

Modelling all possible outages is the key for assessing the impact that each unit would have on the need for frequency containment AS, as is shown in Figure 13, which shows

a simple example with three units: Big Nuclear, Wind Offshore and CCGT. Further details of the modelling are described in Appendix 3 - Cost Allocation Methodology. This is done considering the dispatch for each hour and the size of the AS market required to deal with an outage from that unit, as shown in Figure 12. Once the size of the market that each unit would create is identified, the AS cost allocation for each market player can be quantified.

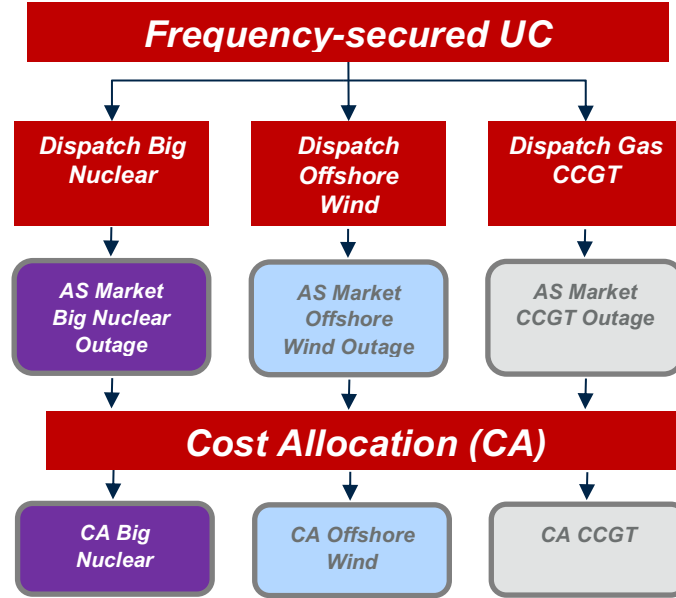


Figure 13: AS cost allocation methodology.

The methodology, analysis and case study presented in this section is based on [6] and is based on the GB system, considering the characteristics of generators, RES and storage, based on the 'Leading the Way' scenario for 2030 presented in Table 1. The summary of weekly AS revenues is presented per technology in Figure 16. The results show that Dynamic Containment represents 45% of the AS market, followed by inertia and PFR with 37% and 18%, respectively. Technologies such as BESS, CCGT and PHES get the largest share of the AS market, with 45%, 27%, and 15% of the total, respectively.

To allocate the corresponding costs of the AS market among market players (and not just socialising them among demand), two different cost allocation methodologies are developed in this section: (a) Proportional and (b) Shapley value.

3.2.1 Proportional Cost Allocation

Proportional cost allocation is one of the simplest ways of allocating AS costs. It assumes that each unit should pay an amount proportional to the stand-alone AS market, Ω_i^{Loss} , that it would create. The stand-alone AS market represents the total market volume that each unit i would create if it was the unit that the system operator hedged for, in terms of frequency security (i.e., $\Omega_{i,t}^{\text{Loss}} = \lambda_{i,t}^{\text{H}} * H_{i,t} + \lambda_{i,t}^{\text{PFR}} * \text{PFR}_{i,t} + \lambda_{i,t}^{\text{DC}} * \text{DC}_{i,t}$ represent the price multiplied by the quantity of frequency containment AS necessary to secure each unit i outage, at each hour t).

A simple example of proportional cost allocation is presented in Table 2 and Figure 14. As presented, the total AS market cost, Ω_{BN}^{Loss} (represented by the biggest plant, a Big Nuclear unit in this example), is distributed between the three market players, considering their relative proportion of the AS market. It can be observed that the total AS market, defined by the Big Nuclear plant in this example, is recovered considering this cost allocation.

Moreover, with this cost allocation φ_i , each market player i pays less than its stand-alone cost Ω_i^{Loss} (or 'total AS market size per unit' as shown in Figure 12), having $\varphi_{BN} \leq \Omega_{BN}^{Loss}$, $\varphi_W \leq \Omega_W^{Loss}$, $\varphi_G \leq \Omega_G^{Loss}$. However, when pairs of market players are analysed, the sum of the cost allocation of the couple is not always smaller than the maximum stand-alone cost between the couple. For instance, if the sum of the cost allocations for Wind (second-largest loss) and Gas (smallest loss) are higher than the maximum stand-alone cost of either plant, i.e. $\varphi_W + \varphi_G > \max\{\Omega_W^{Loss}, \Omega_G^{Loss}\}$, the Wind and Gas units would subsidise the Big Nuclear plant's cost allocation, in the sense that they would be paying more than the stand-alone cost of securing the system for the Wind plant (biggest plant between Wind and Gas), and the Big Nuclear plant would be paying less than the incremental cost of adding security against it to a system secured against the loss of Wind. This demonstrates that this cost allocation might not be considered fair for all market players.

Table 2: Proportional cost allocation

Market player	Cost Allocation φ [£]
Big Nuclear	$\varphi_{BN} = \Omega_{BN}^{Loss} \cdot \frac{\Omega_{BN}^{Loss}}{\Omega_{BN}^{Loss} + \Omega_W^{Loss} + \Omega_G^{Loss}}$
Wind	$\varphi_W = \Omega_{BN}^{Loss} \cdot \frac{\Omega_W^{Loss}}{\Omega_{BN}^{Loss} + \Omega_W^{Loss} + \Omega_G^{Loss}}$
Gas	$\varphi_G = \Omega_{BN}^{Loss} \cdot \frac{\Omega_G^{Loss}}{\Omega_{BN}^{Loss} + \Omega_W^{Loss} + \Omega_G^{Loss}}$

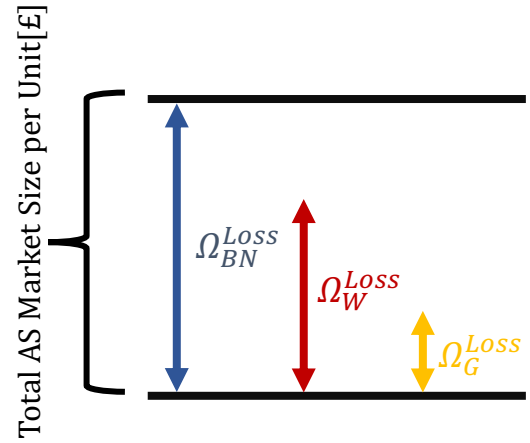


Figure 14: Proportional cost allocation.

Even though proportional cost allocation is a very straightforward methodology, the potential to create these cross-subsidies is a disadvantage. As shown in the example, this means that smaller units would end up paying for a higher level of AS than they need, hence subsidising larger generation units.

Big Nuclear plant will activate constraints such as RoCoF and nadir more strongly than the other plants, demanding more inertia and frequency response services. Subsequently, this will lead to higher prices and larger quantities of AS.

In the case of the GB system, making Big Nuclear units pay a similar amount compared to regular-size Nuclear power plants, as shown in Figure 17, sends inappropriate cost signals. Also, if the size of the largest unit increases (it is expected

that some offshore wind connections would be larger than nuclear), the cost of AS to smaller plant (e.g. Gas-plant etc.) would increase, which could be considered unfair.

3.2.2 Shapley value Cost Allocation

The Shapley value is a payoff vector, which gives each player their weighted marginal contribution. It comes from cooperative game theory, where it was applied to the task of dividing a reward earned by a group of players between them, but it can also be used for allocating costs.

This idea can be seen in the simple example in Table 3 and Figure 15. In this example, the mechanism divides the total AS market into three levels. In the lowest level, determined by Ω_G^{Loss} , the three units would pay the same amount. As all these units need that level of the AS market, the three units pay one-third of it, as it can be seen in the cost allocation function φ . For the second level, $\Omega_W^{Loss} - \Omega_G^{Loss}$, the Gas unit would not pay for it because it exceeds the necessary AS market for the Gas plant, so this level is distributed among the Big Nuclear and Wind plants. Finally, only the Big Nuclear power plant would pay for the upper level, determined by $\Omega_{BN}^{Loss} - \Omega_W^{Loss}$, as the other two units will not need that much AS¹¹.

In the Shapley value case, the total AS market cost is also recovered by the cost allocation mechanism; furthermore, each player pays less or the same than its stand-alone cost, having $\varphi_{BN} \leq \Omega_{BN}^{Loss}$, $\varphi_W \leq \Omega_W^{Loss}$, $\varphi_G \leq \Omega_G^{Loss}$ as in the proportional case. However, the key difference is that each pair of players (or any other subset, in a more complex situation) would prefer this cost allocation given that the sum of the cost allocation of the couple will always be smaller than the maximum stand-alone cost of either plant. For instance, in the case of the Wind and Gas units from the example in Table 3 we will have $\varphi_W + \varphi_G \leq \max\{\Omega_W^{Loss}, \Omega_G^{Loss}\}$. This means that Wind and Gas will just care about their own levels of security; thus, if the Big Nuclear plant increases its dispatch (or even its size with a new investment), they will continue paying an amount related to their own needed levels of security. This property means that Shapley value cost allocation will avoid cross-subsidies as no subset of players will pay more than the maximum Ω^{Loss} of that subset.

¹¹ In a more general example, with several units of each type, the cost of securing against the loss of each type of plant is divided equally among all the units of that size or smaller.

Table 3: Shapley value cost allocation

Market player	Cost Allocation φ [£]
Big Nuclear	$\varphi_{BN} = \frac{\Omega_G^{Loss}}{3} + \frac{\Omega_W^{Loss} - \Omega_G^{Loss}}{2} + \frac{\Omega_{BN}^{Loss} - \Omega_W^{Loss}}{1}$
Wind	$\varphi_W = \frac{\Omega_G^{Loss}}{3} + \frac{\Omega_W^{Loss} - \Omega_G^{Loss}}{2}$
Gas	$\varphi_G = \frac{\Omega_G^{Loss}}{3}$

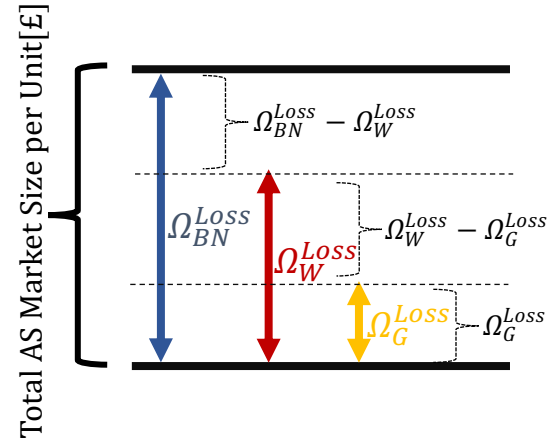


Figure 15: Shapley value cost allocation.

This mechanism defines a cost allocation that ensures that market players do not pay more than what they require to satisfy their needs, avoiding smaller units paying for the AS levels that are bigger than their requirement, and therefore avoiding cross-subsidies.

In the case of the GB system, in the balance per unit in Figure 17, the difference between the Big Nuclear and the average Nuclear plant cost allocation is sizeable: Although the difference in size between these plants is only 200 MW, Big Nuclear would pay twice as much as the regular-size Nuclear plant.

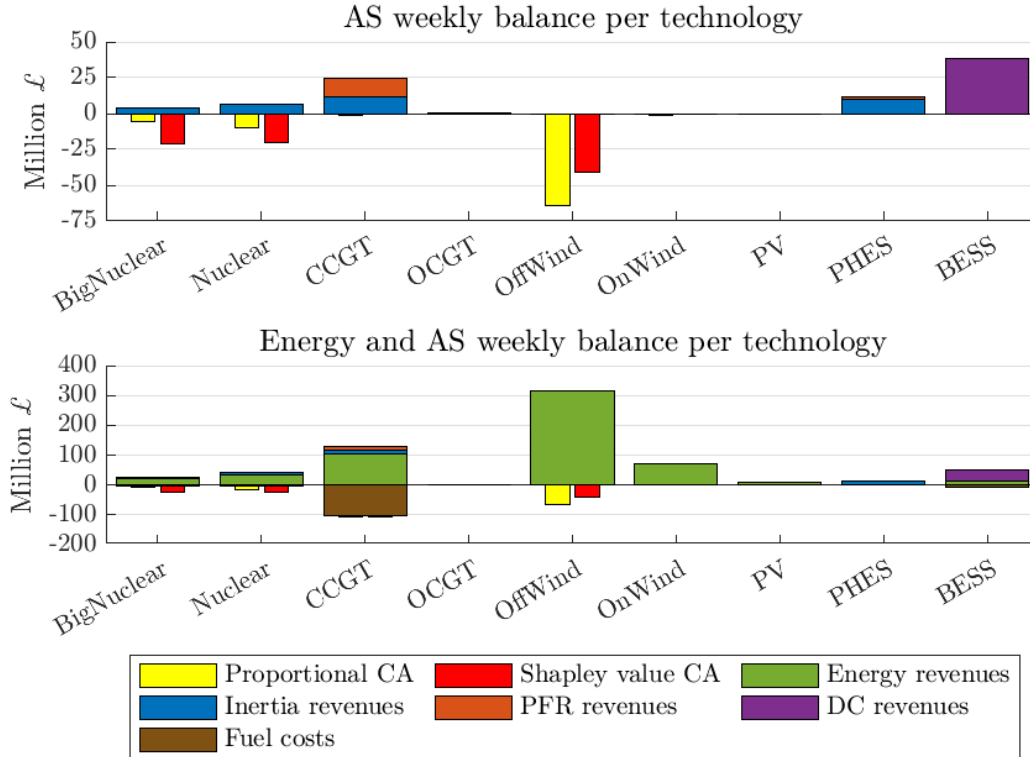


Figure 16: Energy market, AS market (inertia, Dynamic Containment (DC), and Primary Frequency Response (PFR)) and cost allocation (CA) per technology (include all units from the same technology). Source: [6].

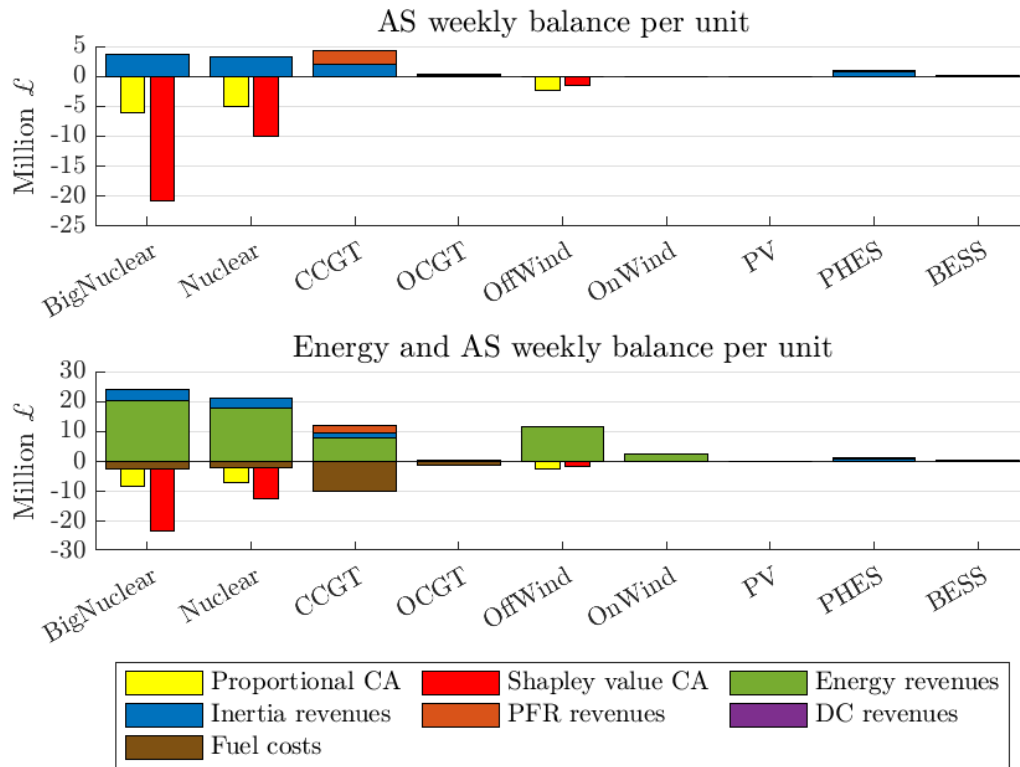


Figure 17: Energy market, AS market (inertia, Dynamic Containment (DC), and Primary Frequency Response (PFR)) and cost allocation (CA) per unit. Source: [6].

A more detailed analysis for the Shapley value cost allocation can be seen in Figure 18, which shows that this methodology will strongly depend on the inertia levels in the system. As frequency containment AS costs are higher in hours with low inertia, the allocated costs to big units will be higher than in hours with high system inertia.

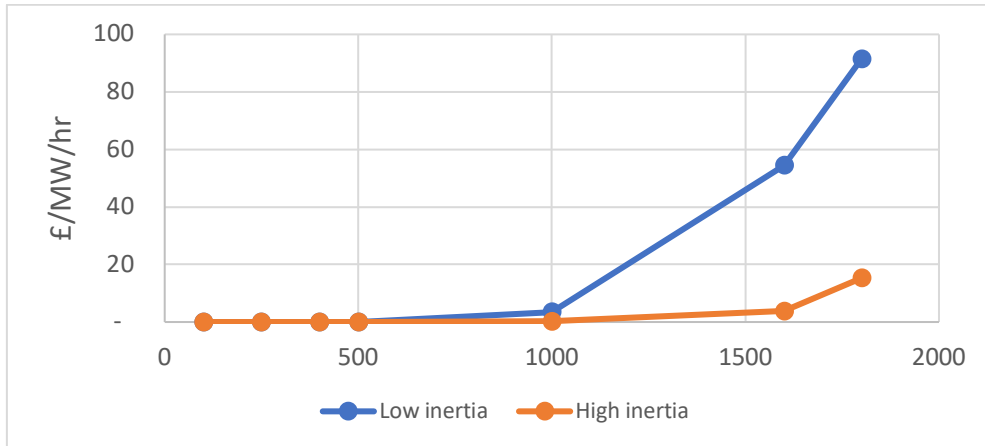


Figure 18: Cost allocation for frequency containment AS, considering low and high system inertia.

Regarding the market mechanism for the cost allocation, we consider that to recover the costs of the pay-as-clear AS market, NESO could allocate tariffs in £/MW (i.e., the cost allocation) to generators, interconnectors and loads large enough in a similar way as is shown in Figure 18. With this arrangement, big units will make the decision of their future dispatch, balancing their earning for energy and AS but also the tariffs for the deployment of AS necessary to manage the security of their operation.

4 Ancillary Services Market Driven Incentives

As highlighted, the present electricity markets have not provided clear incentives to enhance frequency security cost-effectively, and do not reward market participants for investing in AS provision. This section presents a brief description of different options for AS provision in power system with high-RES penetration.

4.1 Ancillary Services from Renewable Energy Sources

RES are expected to largely replace conventional synchronous generators in most systems, in order to meet energy decarbonisation targets. Given that most RES such as wind and solar are decoupled from the grid through power electronics converters, at present they do not naturally contribute to system inertia. Nevertheless, IBRs have the capability to contain frequency variations if appropriately controlled. Two main categories of grid-connected inverters have been defined, depending on their design: Grid-following (GFL) and Grid-forming (GFM) converters.

Currently, virtually all installed RES are based on GFL converters. A key feature of GFL control is the Phase-Locked Loop used for tracking the voltage phase angle of the grid and then synchronising to it. It is also used to estimate frequency, but has an inherent measurement delay which means this type of converter cannot provide frequency support at the time when the frequency decline starts. That contrasts with the synchronous inertia of conventional generation, which does slow frequency decline from its start, when the RoCoF constraint is tightest and inertia-related services are most valuable.

On the other hand, GFM inverters, like synchronous generators, can provide frequency support at the very instant of the contingency, as was explained in section 2.2.1.2 and shown in Figure 7. This is the main difference with GFL inverters that do not support the RoCoF constraint and do not provide inertia-related services [17].

In this section, prices for AS are quantified from the whole-system optimisation using the Lagrange multipliers corresponding to the amount of each kind of inertia and frequency response AS, rather than from expectations of the bids that might be submitted for the provision frequency containment AS.

4.1.1 Frequency Response from Curtailed Renewable Energy Sources

Wind turbines equipped with GFL inverters can provide Dynamic Containment, as long as the control of the inverters includes a droop characteristic, which increases the power output of the turbine in the event of a frequency drop in the grid.

Prices for inertia and frequency response services for one day's operation with 30 GW wind capacity, where 15% of the wind fleet has the capability to provide Dynamic Containment, are presented in Figure 19. As demonstrated, the price of frequency response drops to zero for the last hours of the day, which also coincides with high renewable curtailment.

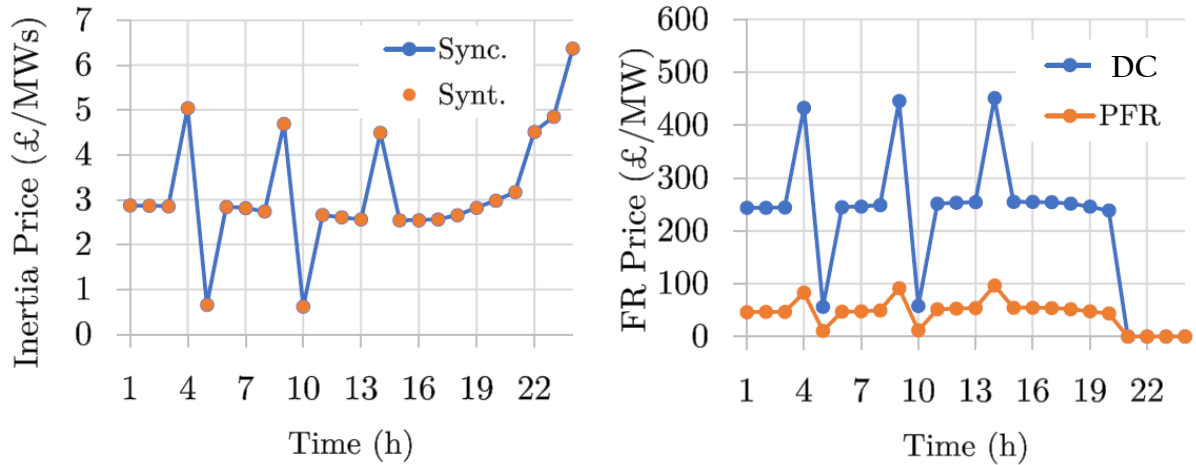


Figure 19: Prices for inertia (synchronous and synthetic) and frequency response (Dynamic Containment (DC) and Primary Frequency Response (PFR)). 24 hours with 15% of wind capacity can provide Dynamic Containment. Source: [17].

4.1.2 Synthetic Inertia from Renewable Energy Sources

In this case it is assumed that 30% of the wind fleet is equipped with GFM inverters and therefore have the capability to provide synthetic inertia, while the remaining 70% of wind only provides energy.

In this context, Figure 20 presents one simulated day. As demonstrated, in the last hours of the day, the price of synthetic inertia drops to zero for high wind output, while the price of synchronous inertia is still positive. The reason for this difference is the recovery effect of synthetic inertia: once the penetration of synthetic inertia in the system is sufficiently high, the necessary frequency response reserves to compensate for the recovery effect in the quasi-steady-state constraint becomes very high (see Figure 7 in section 2.2.1.2). After this point, adding more synthetic inertia does not lead to a decrease in system operating costs, hence the price drops to £0/MWs for this service.

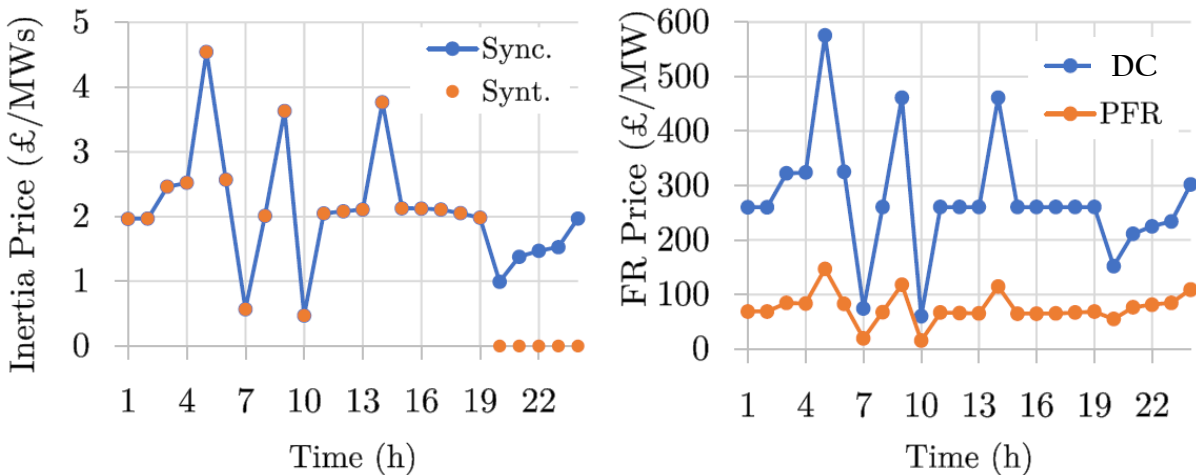


Figure 20: Prices for inertia (synchronous and synthetic) and frequency response (Dynamic Containment (DC) and Primary Frequency Response (PFR)). 24h with 30% of wind capacity GFM-equipped. Source: [17].

4.2 Frequency Response from Electric Vehicles

A rapid uptake of EVs will shift energy demand for transport away from the combustion of fossil fuels. In the UK, up to 37.4 million EVs should be on the roads by 2040, according to National Grid ESO's FES [20]. Such large penetration of EVs will create an important opportunity in supporting frequency control. Given this context, it will be relevant to quantify the flexibility value that EVs could provide to support the electricity grid balancing by providing frequency response. The questions that arise are:

- How can market participants take advantage of different charging approaches to add competency to the frequency containment AS market?

To model EV capability to provide services, three different charging approaches will be considered:

- **Unmanaged charging:** In this approach, EVs immediately charge to full when plugged in.
- **Smart charging:** This methodology allows the charge of an EV to adjust its charging behaviour in real time to avoid high-demand periods on the system. This approach represents a unidirectional power flow from the grid to the vehicle, demand shifting, and frequency response (in the same timescales of PFR) via demand alleviation.
- **V2G:** With this approach EVs can deliver bidirectional power flow, arbitrage, and greater PFR provision via fast switching from charge to discharge.

In this work, EVs can provide PFR in Smart charging and with the V2G approach. AS prices are computed to quantify the potential charging costs and revenues that flexible EVs could capture. The main results are shown in Figure 21, presenting the impact of the three different charging methodologies: Unmanaged, Smart charging, and V2G. Additionally, we consider different levels of EV penetration (25%, 50%, 75%, and 100%), in which the 100% penetration level is represented by the Consumer Transformation scenario within the FES of National Grid ESO for 2035.

As shown in Figure 21, AS prices are lower in the Smart charging and V2G scenarios. This reduction of prices is more substantial in the V2G case, given that EVs can provide a larger amount of PFR. This higher PFR provision is based on the assumption that these units can swiftly switch from charging to discharging, providing a higher PFR volume while operating in charging mode. This can be directly reflected in a more remarkable price reduction, given that there is less need for thermal units to be synchronised just to provide PFR, allowing a more efficient dispatch.

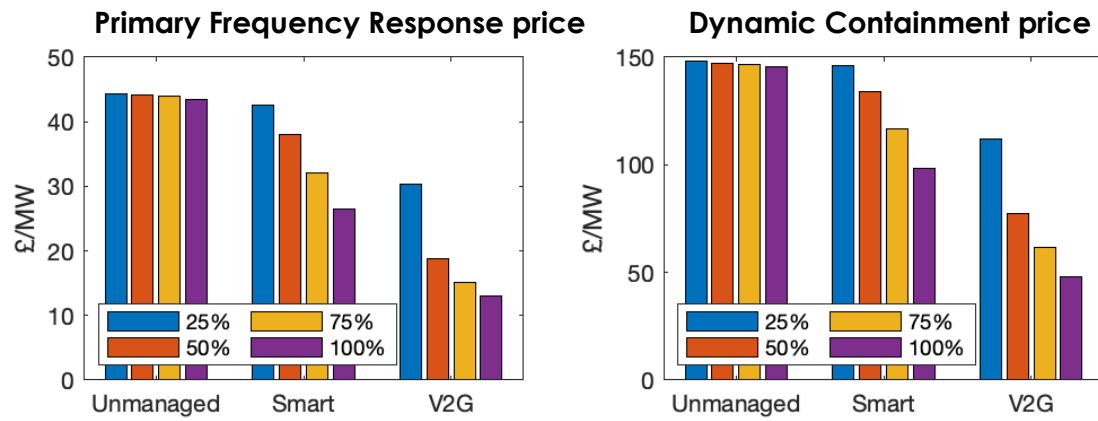


Figure 21: AS market prices for different levels of EVs participation. Source: Imperial's modelling analysis.

5 Conclusions and Future Work

This white paper has proposed a novel market design for future low-inertia electricity systems. It defines a methodology that would provide incentives to market players to participate in this AS market while at the same time making players that cause the need for AS accountable for efficiently bearing these costs.

The provision of AS is enabled by conventional generation, such as thermal and PHES units, but it also accounts for IBR technologies that will be key to decarbonising the electricity grid, such as RES, BESS, and EVs. We have shown that RES with the appropriate control mechanisms can provide Dynamic Containment and synthetic inertia, decreasing AS system costs. EVs can provide frequency response by smart charging or a V2G approach, which can significantly reduce AS prices, particularly with bi-directional V2G.

The pricing structure for AS provision is derived, ensuring that market players can recover operational costs in the short term through their AS provision. Notably, this white paper reveals an inverse relationship between energy and AS prices. Higher energy prices are driven by the commitment of thermal plants, which will increase inertia levels, resulting in lower AS prices during these peak hours. Conversely, higher AS prices appear in high RES/low demand hours, coinciding with lower energy prices.

A fundamentally new methodology for AS cost allocation is proposed in this white paper. We propose shifting from the historical approach based on socialising AS costs to an efficient cost allocation concept, recognising that the size of the plant is the key driver for the cost of AS. This approach acknowledges the market players responsible for triggering the need for AS and distributes costs among them. Two allocation methodologies, Proportional and Shapley value, are explored, with the latter avoiding cross-subsidies among market players, making it an appropriate cost allocation paradigm.

This market design will also provide appropriate incentives for the deployment of advanced technologies and control concepts that would support a cost-effective transition to a secure low/zero-carbon electricity system, through the provision of appropriate AS.

For future work, AS related to loss of demand should be considered – specifically, disconnection of interconnectors that export energy from the GB would cause a significant frequency increase that would also need to be managed. In this context, the AS market should include appropriate procurement of different services and allocate the corresponding costs appropriately. Regional frequency markets should also be considered when trading AS between different zones. This will be relevant in the GB system, considering the differences between Scotland and England's generation mixes. In Scotland, as its generation will be highly based on IBRs, low inertia levels will add more value to frequency containment AS in the north of the GB. Lastly, a comprehensive understanding of the market design proposed in this white paper within a fully decarbonised electricity systems that consider new technologies such as hydrogen generation and consumption, flexible demand side resources, etc., will be imperative.

6 References

- [1] Department of Business Energy & Industrial Strategy, "Net Zero Strategy: Build Back Greener," London, 2021.
- [2] B. Mohandes, M. S. El Moursi, N. Hatziargyriou, and S. El Khatib, "A Review of Power System Flexibility with High Penetration of Renewables," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3140–3155, Jul. 2019, doi: 10.1109/TPWRS.2019.2897727.
- [3] L. Badesa, G. Strbac, M. Magill, and B. Stojkovska, "Ancillary services in Great Britain during the COVID-19 lockdown: A glimpse of the carbon-free future," *Appl Energy*, vol. 285, Mar. 2021, doi: 10.1016/j.apenergy.2021.116500.
- [4] F. Billimoria, P. Mancarella, and R. Poudineh, "Market design for system security in low carbon electricity grids," Oxford, United Kingdom, Jun. 2020. doi: 10.26889/9781784671600.
- [5] B. Hua and R. Baldick, "A convex primal formulation for convex hull pricing," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3814–3823, Sep. 2017, doi: 10.1109/TPWRS.2016.2637718.
- [6] C. Matamala, L. Badesa, R. Moreno, and G. Strbac, "Cost Allocation for Inertia and Frequency Response Ancillary Services," *IEEE Transactions on Energy Markets, Policy and Regulation*, pp. 1–11, 2024, doi: 10.1109/TEMPR.2024.3355291.
- [7] G. Strbac and D. Kirschen, "Who Should Pay for Reserve?," *The Electricity Journal*, vol. 13, no. 8, pp. 32–37, Oct. 2000.
- [8] National Grid ESO, "Frequency Risk and Control Report," 2023. Accessed: Jan. 24, 2024. [Online]. Available: <https://www.nationalgrideso.com/industry-information/codes/security-and-quality-supply-standard-sqss/frequency-risk-and-control>
- [9] Pöyry Management Consulting and Imperial College London, "Roadmap for Flexibility Services to 2030," 2017. Accessed: Nov. 14, 2023. [Online]. Available: <https://www.theccc.org.uk/publication/roadmap-for-flexibility-services-to-2030-poyry-and-imperial-college-london/>
- [10] N-SIDE, "EAC Market Design Report," 2023. Accessed: Feb. 16, 2024. [Online]. Available: <https://www.nationalgrideso.com/industry-information/balancing-services/enduring-auction-capability-eac>
- [11] E. Ela, R. Hytowitz, and U. Helman, "Ancillary Services in the United States Technical Requirements, Market Designs and Price Trends," Palo Alto, California, 2019. [Online]. Available: www.epri.com
- [12] AEMO, "Guide to Ancillary Services in the National Electricity Market," Oct. 2023.
- [13] AEMO, "Wholesale Electricity Market Design Summary," Sep. 2023.
- [14] AEMO, "Settlements Guide to Ancillary Services Payment and Recovery," Oct. 2023.

- [15] P. Kundur, *Power System Stability and Control*, 1st ed. McGraw-Hill Education, 1994.
- [16] L. Badesa, F. Teng, and G. Strbac, "Optimal Portfolio of Distinct Frequency Response Services in Low-Inertia Systems," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4459–4469, Nov. 2020, doi: 10.1109/TPWRS.2020.2997194.
- [17] L. Badesa, C. Matamala, Y. Zhou, and G. Strbac, "Assigning Shadow Prices to Synthetic Inertia and Frequency Response Reserves from Renewable Energy Sources," *IEEE Trans Sustain Energy*, vol. 14, no. 1, pp. 12–26, Jan. 2023, doi: 10.1109/TSTE.2022.3198324.
- [18] NERC Inverter-Based Resource Performance Task Force (IRPTF), "Fast Frequency Response Concepts and Bulk Power System Reliability Needs," NERC, pp. 1–23, 2020.
- [19] H. Chavez, R. Baldick, and S. Sharma, "Governor rate-constrained OPF for primary frequency control adequacy," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1473–1480, 2014, doi: 10.1109/TPWRS.2014.2298838.
- [20] National Grid ESO, "Future Energy Scenarios," 2023. [Online]. Available: www.nationalgrideso.com/future-energy/future-energy-scenarios

Appendices

Appendix 1 - Unit Commitment Model

The UC model, developed by Imperial College London, represents an optimisation model that simultaneously schedules energy and AS, considering RES uncertainty, while balancing generation and demand supply and respecting frequency constraints in the event of any credible contingency, as defined in the section Modelling Approach.

Nomenclature:

Sets

$g, \mathcal{G}$	Index and set of generators.
$r, \mathcal{R}$	Index and set of RES.
$s, \mathcal{S}$	Index and set of storage.
$i, \mathcal{M}_t, \mathcal{N}_t$	Index, subset, and set of total dispatched units at hour t .
$t, \mathcal{T}, t^{\text{end}}$	Index, set of time horizon, and final time.

Parameters

Δf_{max}	Maximum admissible frequency deviation at the nadir (Hz).
$\lambda_g^E, \lambda_r^E, \lambda_s^E$	Energy price offer of generator g , RES r , and storage s (£/MWh).
λ_g^H, λ_s^H	Inertia price offer of generator g , and storage s (£/MWs).
$\lambda_g^{\text{PFR}}, \lambda_s^{\text{PFR}}, \lambda_s^{\text{DC}}$	FR price offer of generator g , and storage s (£/MW).
f_0	Nominal power grid frequency (Hz).
H_g, H_s	Inertia constant generator g , and storage s (s).
P_t^D	Total system demand at hour t (MW).
$\text{RoCoF}_{\text{max}}$	Maximum admissible RoCoF (Hz/s).
$T_{\text{DC}}, T_{\text{PFR}}$	Delivery time of DC and PFR (s).

Variables

λ_t^E	Market-clearing energy price (£/MWh).
$\lambda_t^{\text{DC}}, \lambda_t^{\text{PFR}}$	Market-clearing DC and PFR price (£/MW).
λ_t^H	Market-clearing inertia price (£/MWs).
μ_t^{nadir}	Dual variables for the SOC nadir constraint.
$\mu_t^{\text{q-s-s}}$	Dual variable for the q-s-s constraint.
μ_t^{RoCoF}	Dual variable for the RoCoF constraint.
$\varphi_{i,t}^{\text{Pro}}, \varphi_{i,t}^{\text{SV}}$	Proportional and Shapley value cost allocation for unit i (£).
DC_t, PFR_t	Aggregate headroom for providing DC and PFR (MW).
H_t	Aggregate synchronous inertia (MWs).
$P_{g,t}, P_{r,t}$	Power generated by generator g and RES r (MW).
$P_{s,t}^{\text{cha}}, P_{s,t}^{\text{dis}}$	Charge and discharge power of storage s (MW).
P_t^{Loss}	Largest power infeed (MW).
$PFR_{g,t}, PFR_{s,t}, DC_{s,t}$	Headroom in generator g and storage s for providing FR (MW).
$y_{g,t}$	Binary variable. Commitment of generator g .
$y_{s,t}^{\text{cha}}, y_{s,t}^{\text{dis}}$	Binary variables. Charging and discharging state of storage s .

The objective function minimises operational costs of the system.

$$\min \left\{ \sum_{t \in \mathcal{T}} \left[\sum_{g \in \mathcal{G}} \left(\lambda_g^E \cdot P_{g,t} + \lambda_g^H \cdot P_g^{\text{max}} \cdot H_g \cdot y_{g,t} + \lambda_g^{\text{PFR}} \cdot PFR_{g,t} \right) + \sum_{r \in \mathcal{R}} \lambda_r^E \cdot P_{r,t} \right. \right. \\ \left. \left. + \sum_{s \in \mathcal{S}} \left(\lambda_s^E \cdot P_{s,t}^{\text{dis}} + \lambda_s^H \cdot P_s^{\text{max}} \cdot H_s \cdot (y_{s,t}^{\text{cha}} + y_{s,t}^{\text{dis}}) + \lambda_s^{\text{PFR}} \cdot PFR_{s,t} + \lambda_s^{\text{DC}} \cdot DC_{s,t} \right) \right] \right\}$$

The balance of generation $P_{g,t}$ and demand P_t^D is considered with the following constraint:

$$\sum_{g \in \mathcal{G}} P_{g,t} + \sum_{r \in \mathcal{R}} P_{r,t} + \sum_{s \in \mathcal{S}} (P_{s,t}^{\text{dis}} - P_{s,t}^{\text{cha}}) = P_t^D : (\lambda_t^E)$$

Total provision of inertia:

$$H_t = \sum_{g \in \mathcal{G}} H_g \cdot P_g^{\max} \cdot y_{g,t} + \sum_{s \in \mathcal{S}} H_s \cdot P_s^{\max} \cdot (y_{s,t}^{\text{dis}} + y_{s,t}^{\text{cha}}) : (\lambda_t^H)$$

Total provision of Primary Frequency Response (PFR) and Dynamic Containment (DC):

$$PFR_t = \sum_{g \in \mathcal{G}} PFR_{g,t} + \sum_{s \in \mathcal{S}} PFR_{s,t} : (\lambda_t^{PFR}) \quad \forall t$$

$$DC_t = \sum_{s \in \mathcal{S}} DC_{s,t} : (\lambda_t^{DC}) \quad \forall t$$

Frequency constraints are accounted considering the maximum loss of the system P_t^{Loss} . RoCoF constraint is accounted as:

$$H_t \geq \frac{P_t^{\text{Loss}} \cdot f_0}{2 \cdot \text{RoCoF}_{\max}} : (\mu_t^{\text{RoCoF}}) \quad \forall t$$

Nadir constraints are represented as a second-order cone:

$$\left\| \begin{bmatrix} \frac{1}{f_0} & \frac{-T_{\text{DC}}}{4\Delta f_{\max}} & \frac{-1}{T_{\text{PFR}}} & 0 \\ 0 & \frac{-1}{\sqrt{\Delta f_{\max}}} & 0 & \frac{1}{\sqrt{\Delta f_{\max}}} \end{bmatrix} \begin{bmatrix} H_t \\ DC_t \\ PFR_t \\ P_t^{\text{Loss}} \end{bmatrix} \right\| \leq \begin{bmatrix} \frac{1}{f_0} & \frac{-T_{\text{DC}}}{4\Delta f_{\max}} & \frac{1}{T_{\text{PFR}}} & 0 \end{bmatrix} \begin{bmatrix} H_t \\ DC_t \\ PFR_t \\ P_t^{\text{Loss}} \end{bmatrix} : (\mu_{1,t}^{\text{nadir}}, \mu_{2,t}^{\text{nadir}}, \mu_{3,t}^{\text{nadir}}) \quad \forall t$$

Quasi-steady-state defines that the total amount of frequency response provided should be bigger than the power loss:

$$DC_t + PFR_t \geq P_t^{\text{Loss}} : (\mu_t^{\text{q-s-s}}) \quad \forall t$$

Appendix 2 - Pricing Methodology

A relaxed version of the optimisation model that avoids non-convexities, such as the commitment of generators, is considered to derive AS prices. Prices are calculated following the Karush-Kuhn-Tucker stationarity conditions of the Lagrangian function, as explained in [17]. These prices for AS depend on the dual variables of the frequency security constraints given in Appendix 1 - Unit Commitment Model.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial H_t} = 0 &\Rightarrow \lambda_t^H = \frac{\mu_{3,t}^{\text{nadir}} - \mu_{1,t}^{\text{nadir}}}{f_0} + \mu_t^{\text{RoCoF}} \quad \forall t \\ \frac{\partial \mathcal{L}}{\partial PFR_t} = 0 &\Rightarrow \lambda_t^{PFR} = \frac{\mu_{3,t}^{\text{nadir}} + \mu_{1,t}^{\text{nadir}}}{T_{\text{PFR}}} + \mu_t^{\text{q-s-s}} \quad \forall t \\ \frac{\partial \mathcal{L}}{\partial DC_t} = 0 &\Rightarrow \lambda_t^{DC} = \frac{(\mu_{1,t}^{\text{nadir}} - \mu_{3,t}^{\text{nadir}}) \cdot T_{\text{DC}}}{4\Delta f_{\max}} + \frac{\mu_{2,t}^{\text{nadir}}}{\sqrt{\Delta f_{\max}}} + \mu_t^{\text{q-s-s}} \quad \forall t \end{aligned}$$

Appendix 3 - Cost Allocation Methodology

To calculate the influence that each unit has in the frequency containment AS market, we run the relaxed version of the optimisation model as many times as the

number of dispatched units. In each case, the value of P_t^{Loss} is set equal to the dispatch of each unit in the system, as shown in Figure 22. Even though this consideration is hypothetical, given that the NESO only focus on the biggest unit, and not on smaller units, this methodology allows us to calculate the impact that each unit would create in the AS market. Considering this, we calculate the total AS market size per unit as shown in the following equation:

$$\Omega_{i,t}^{Loss} = \lambda_{i,t}^H \cdot H_{i,t} + \lambda_{i,t}^{PFR} \cdot PFR_{i,t} + \lambda_{i,t}^{DC} \cdot DC_{i,t} \quad \forall i \in \mathcal{N}_t, \forall t$$

when considering:

$$P_{i,t}^{Loss} = P_{i,t} \quad \forall i \in \mathcal{N}_t, \forall t$$

All $\Omega_{i,t}^{Loss}$ terms are shown in Figure 12 for the case study presented in this white paper. Then these $\Omega_{i,t}^{Loss}$ are considered for the cost allocation methodology φ mentioned in this white paper, either proportional or Shapley value, as can be seen in Figure 22.

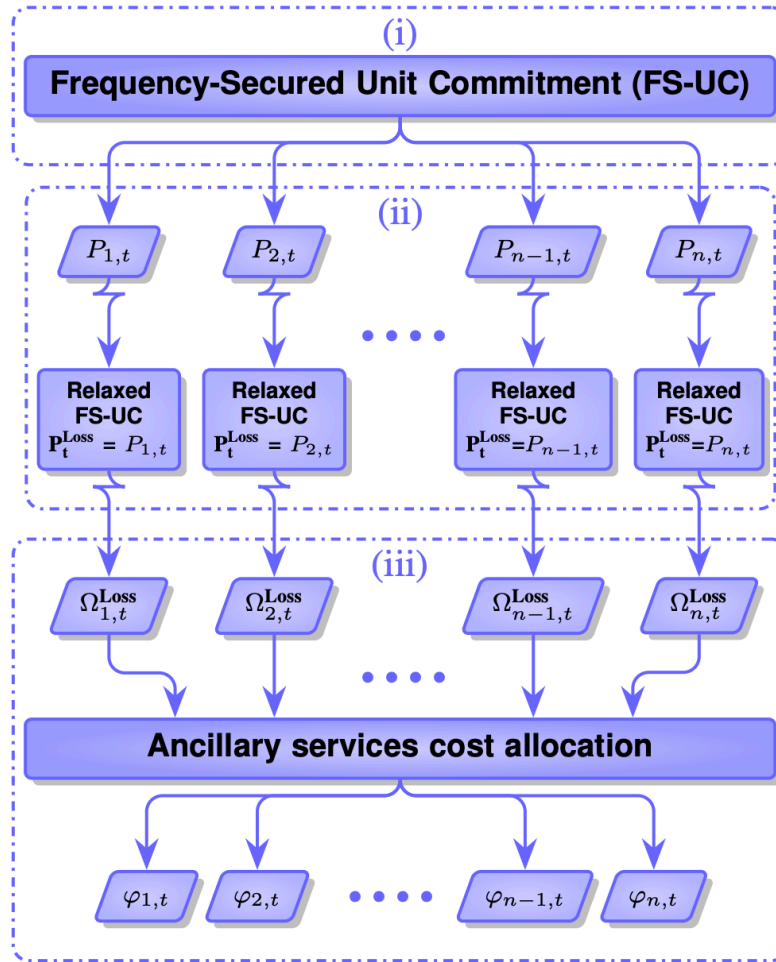


Figure 22: Frequency containment AS cost allocation methodology. Source: [6].

The proportional cost allocation is calculated as (where $\mathcal{N}_t$ represents the set of all dispatched units in hour t , and unit n is the one with the largest dispatch, which therefore determines the amount and cost of the AS required):

$$\varphi_{i,t}^{\text{Pro}} = \Omega_{i,t}^{\text{Loss}} \cdot \frac{\Omega_{n,t}^{\text{Loss}}}{\sum_{i \in \mathcal{N}_t} \Omega_{i,t}^{\text{Loss}}} \quad \forall i \in \mathcal{N}_t, \forall t$$

For the Shapley value case, if the units are indexed in order of increasing dispatch in each hour, the cost allocation is determined by:

$$\varphi_{i,t}^{\text{SV}} = \sum_{\mathcal{M}_t \subseteq \mathcal{N}_t \setminus \{i\}} \frac{|\mathcal{M}_t|!(|\mathcal{N}_t| - |\mathcal{M}_t| - 1)!}{|\mathcal{N}_t|!} [\Omega_t^{\text{Loss}}(\mathcal{M}_t \cup \{i\}) - \Omega_t^{\text{Loss}}(\mathcal{M}_t)] \quad \forall i \in \mathcal{N}_t, \forall t$$

The standard Shapley value formula calculates each player's weighted marginal impact of adding that player to all possible subsets of players that do not include it. However, this computation can be intractable even for few dozens of players.

Nevertheless, in the case of the frequency containment AS market, the system operator determines the AS market for the biggest dispatched unit each hour. This can be represented as an *Airport problem* [6].

The Airport problem is a cost allocation game in which N airlines must share the cost of a runway, which is determined by the largest type of aircraft to land there. This approach allows the computation of the Shapley value in a simple form as described in [6] and shown in the next equation:

$$\varphi_{i,t}^{\text{SV}} = \sum_{k=1}^i \frac{\Omega_{k,t}^{\text{Loss}} - \Omega_{k-1,t}^{\text{Loss}}}{n + 1 - k} \quad \forall i \in \mathcal{N}_t, \forall t$$