On balancing market design

Citation for published version (APA):

Document status and date:
Published: 17/05/2016

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
On balancing market design

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op dinsdag 17 mei 2016 om 16:00 uur

door

Frank Alfred Nobel

geboren te Utrecht
Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr.ir. A.B. Smolders
1e promoter: prof.dr.ir. P.P.J. van den Bosch
2e promoter: prof.ir. M.A.M.M. van der Meijden (TUD)
leden: prof.dr.ir. R. Belmans (KU Leuven)
        prof.dr. J.-M. Glachant (European University Institute)
        prof.dr.ir. G.P.J. Verbong
        prof.dr. S. Pront-van Bommel (UvA)†
adviseur: dr.ir. J. Frunt

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.
A catalogue record is available from the Eindhoven University of Technology Library.
On balancing market design
Proefschrift.
Summary

On balancing market design

The European electrical power system is facing major challenges due to policy decisions. One decision has been to consider electric energy as a commodity, to be traded on markets, including balancing markets and thus subject to (preferably cross border) competition. A second policy change has been to reduce the dependency on fossil fuels and their atmospheric emissions, thus requiring accommodating of ever increasing volumes of solar and aeolic power generation in the power system. Both decisions have had and will have major impact on the electricity transmission infrastructure itself, on the operations of this transmission system, and on the users of this transmission system. Any market design accommodating these policy decisions has to be effective by complying to physical, legal or regulatory requirements.

This dissertation therefore examines the following central research question: Which necessary conditions are required in an electricity market design aimed at power balancing, to achieve its objectives: efficiency and effectivity while satisfying compliancy requirements?

In this dissertation a comprehensive decomposition of power systems is proposed, which allows reasoning about both technical and economical characteristics of differently designed balancing markets. This conceptual, rather than mathematical, model is based on existing legislation and publications by professional organizations. By using this model the responsibilities of different actors with respect to energy and power balancing are distinguished, their behaviours are analysed, and the consequences of their failures are established. These distinct responsibilities are key ingredients in any market design. Balancing market design imposes obligations and opportunities on actors. The premise is that none of these should invite any actor to violate boundary conditions imposed by physical law, nor allow abusive behaviour towards other actors. Hence gaming mechanisms as well as exclusion mechanisms should be avoided.

Key elements of balancing market design are energy settlement elements requiring the identification of several volumes and price components, granularities (imbalance settlement period), and settlement mechanisms. Other key elements are gate closure times, and the real-time feedback of system information to all users. Settlement mechanisms can be pay-as-bid or marginal pricing mechanisms for balancing energy, and market- or cost-based pricing for imbalance. The last choice is closely linked to the financial position of the system operator as a result of its energy settlements, and its cost relating to comply to operational standards. It will be shown that some choices result de facto in exclusion mechanisms, or result in inefficient mark-ups in prices.
An analytical model is proposed to assess coincident energy prices (commodity, balancing energy and imbalance) and their mutual consistencies. It is deduced that under specific boundary conditions imbalance (i.e. balance responsible parties) can compete at least with balancing energy on a real-time energy market. This allows more users to compete in energy balancing by removing exclusion mechanisms. An analytical model is proposed to assess the effectivity and efficiency of actual power balancing. It is deduced that under specific boundary conditions related to imbalance settlement period duration, the effectivity and efficiency of balancing processes and the effects of balancing philosophies in different designs can be demonstrated.

Some key price and volume elements are tested on empirical data from the Netherlands and Germany, and confirm hypotheses on price formation mechanisms.

Finally, in this dissertation suggestions are made for future research, a.o. to examine future European legislation on the presence of exclusion mechanisms or mechanisms that induce inefficient mark-ups in prices.
Preface

The scientific, professional education I was fortunate to have received at the Universiteit van Amsterdam had geology as subject. It was centred on field courses geological mapping, in south eastern Spain, even to the extent that I was allowed to teach undergraduates field mapping skills. So I learned to read maps, to make maps and to like maps, as well as mapping things, a.o. the Prebetics in south-eastern Spain. Thus equipped to look back in geological time, I entered the world of electricity at NV Sep, the Dutch electricity generating board, in 1988. And there in a planning department I became involved in forecasting load and demand, as the electrical power industry is continuously looking forward in time. Planning of course is designing a future. I have been privileged to observe, and at times to participate in great changes affecting this electricity world. Which incidentally showed me the limited value of long term forecasting and models as input to planning. It also introduced me to quite a few people working in and on balancing market design. They have become dear to me, just like the people at the Geological Institute in Amsterdam, just like my past and present colleagues in Arnhem.

I am grateful to my employer TenneT TSO BV for allowing me to wrap up some of my ideas, research, and results, and to the Technical University Eindhoven to invite me to do so. There it was Professor Paul van den Bosch who convinced me that these ideas, research, and results might have some scientific value. He guided me through the simple yet profound process of proper thinking and writing, and not losing faith. And I have to be grateful to my family: Volkert J., Maatje, Karin, Mark & Josse, Josse† and Valentine, and to my colleagues and friends for enabling me.

The ideas expressed in this work are my own and do not necessarily represent those of TenneT TSO BV.
## Summary

4

## Preface

6

1. Introduction to power systems and electricity markets

1.1 Research question  
1.2 General aspects of electricity market design  
1.3 An introductory bottom up analysis  
1.4 Organization of this thesis  

9

2. A conceptual framework for electricity markets

2.1 Domains  
2.2 Actors  
2.3 A decomposed service model  
2.4 Power balancing processes  
2.5 Power balancing and compliancy requirements of system operators  
2.6 Control imperfections and associated risks  
2.7 Conclusions and discussion on a conceptual framework for balancing  

18

3. Balancing variables and performance indicators

3.1 Balancing process performance indicators  
3.2 Conclusions and discussion on balancing variables and performance indicators  

41

4. Prices related to balancing

4.1 Commodity price  
4.2 Commodity and balancing energy prices  
4.3 Commodity and imbalance prices  
4.4 Balancing energy and imbalance prices  
4.5 Conclusions and discussion on prices related to balancing  

52

5. Time units for trading and balancing purposes

5.1 Conclusions and discussion on time units for trading and balancing purposes  

67

6. The financial balance of the system operator

6.1 The financial neutralization of the system operator  
6.2 Conclusions and discussion on the financial balance of the system operator  

71

7. Testing models and design choices

7.1 Commodity price model  
7.2 Balancing energy $\Delta$ price model  
7.3 Imbalance margin $\Delta$ value  

79
7.4 Energy parameter choice for the net system state $\text{ACE}_{\text{OL}}$
7.5 Conclusions and discussion on testing models and design choices

8. Testing design choices 99
8.1 Restricting users 101
8.2 Incentivizing users 101
8.3 Limiting user incentives 105
8.4 Imbalance netting as case example of adapting translation scheme 107

9. Conclusions and discussion on balancing market design 110

A. Appendices 114
Business cycles 114
USA experience and expertise 116
A case example of real-time incentives to users 118
Price caps and floors of day ahead price in the Netherlands 121
System operator tariff characteristics in European countries 122
Financial position gas system operator for balancing 123

B. Bibliography 125
1. Introduction to power systems and electricity markets

Power systems
People need power and electricity satisfies part of these needs. To fulfill these needs electricity has to be supplied to them, i.e. to be generated, transmitted and distributed. In fact, we as individuals and as societies almost everywhere and anytime depend on a continuous electricity supply. To satisfy our need for electrical power, large electrical power systems have been developed as illustrated by the following statistics:

ENTSO-E in figures:
- 41 transmission system operators (TSOs)
- 35 European Countries
- 532 million customers served by the represented power systems
- 1023 GW net generation capacity
- 3310 TWh electricity consumption
- 424 TWh of electricity exchange between member TSOs
- 307,000 km of transmission lines managed by TSOs

Source: ENTSO-E (2014) Factsheet

Continuity and disruptions
Power systems are perpetually under development, but presently they are facing two major transitions, both in Europe and elsewhere. One is the transition towards (international) electricity markets, due to (supra-) governmental policies. The other one is the accommodation of increasing volumes of power injection from intermittent generation, mainly as a consequence of governmental energy policies aimed at decarbonizing the economy. Notwithstanding these transitions, the show must go on. These aspects of continuity in spite of major transitions can be summarized by the terms the European Network of Transmission System Operators for Electricity (ENTSO-E) bears in its banner: Reliable. Sustainable. Connected.

Electricity and markets
Electricity is a manifestation of a global frequency [Hz] and a local voltage [V] in a power system. Power system control aims at maintaining a stable global frequency f [Hz] and a stable local voltage [V] within the power system. A stable frequency at any given reference value, in Europe 50, in USA 60Hz, is a manifestation of global balance between active power injections and withdrawals from different connections into a transmission system as shown in Figure 1-1 in a synchronous area. An alternate current (AC) power transmission system allows but also requires synchronous consumption (withdrawal) and production (injection) of power. The transmission system itself has no buffering capacity.

Electrical energy E [MWh] is the integral over time of active power P [MW] injections or withdrawals at a connection. Electricity markets aim at enabling its participants to trade electrical energy amongst themselves, in predefined time intervals [h], using the transmission system to transfer energy by injection or withdrawal of active [MW], and reactive [MVA] power.
Electricity markets are *markets* that allow *freedom* to their participants as illustrated here:

> "The freedoms which the Treaty guarantees the citizens of the Union — inter alia, the free movement of goods, the freedom of establishment and the freedom to provide services — are achievable only in a fully open market, which enables all consumers freely to choose their suppliers and all suppliers freely to deliver to their customers."

Source: DIRECTIVE 2009/72/EC, 2009

Electricity market design has to align the market participants’ active power injections and withdrawals with their energy trades, taking the real-time physical constraints of the transmission system and of its connections into account.

Designing a market for electricity is not the same as designing a market for products like potatoes. Potatoes are material, they can be counted and weighed. Potatoes do not require a specific dedicated infrastructure to bring demand and supply together. Transport really moves identifiable potatoes from one place to another. With a typical time constant in potatoes of a yearly growth season, and with sufficient storage capacity, no real-time synchronicity is required between consumption and production. Production and consumption of potatoes are decoupled in time.

Electricity on the other hand is immaterial. There is no material transport between power injection and withdrawal. The concept of a power flow is an expression of laws of physics rather than actual electrons changing ownership and moving all the way from a source to a sink.

A stable frequency is a common good shared and enjoyed at all connections but jointly maintained by transmission system operators (ENTSO-E, 2012, AHTor). Assuming a joint responsibility for a common good, without mutually assuring contributions to its joint maintenance as a requirement, is not a robust or sustainable strategy. It may lead to the "tragedy of the commons" and to failure. Neglecting this sensible and fundamental requirement may induce perverse behaviour in some, thus undermining the willingness of others to contribute to its maintenance. Inter alia, public healthcare, a common €-currency, international peace missions are examples.

Power systems and electricity markets, and changes therein, are culminations of a multitude of processes, decisions and developments. Some of these are technical, others are legal. Some span several decades, others are instantaneous. Some are physical, others economical. Some are by choice, others by regulation or physical law. Power systems and electricity markets combine the individual desires of those that use it to consume, to generate or trade electricity with the collective and mutual obligations of those responsible for operating these systems. The identification of users and operators, their desires and obligations, and the assignment of responsibilities and obligations are issues of market design and consequently subject of this thesis.
Assumptions in this research

Key to this research is the decomposition of a power system shown in Figure 1-1.

![Figure 1-1 Decomposition of a power system in a transmission system configuration](image)

Access to the transmission system is assumed to be free and fully granted. Grid control, aiming at maintaining voltage stability is out of the scope of this research. Congestion management to avoid overloading of components of the transmission system are introduced only in identifying actors or boundary conditions. This basic assumption or "copper plate principle" is extreme, and represents one end of a range of potential market designs, allowing maximal freedom to its participants. This choice allows the research to be focused on balancing market design, aiming purely at effective and efficient power balancing, but poses limits to its applicability.

The opposite decomposition is a design with fully centralized central control over the transmission system and over the power generation as shown in the conceptual decomposition in Figure 1-2 of a power system into a power supply system to which loads are connected:

![Figure 1-2 Decomposition of a power system in a power supply configuration](image)

Full central control encompasses business cycles of planning and investment decisions, the commitment, scheduling, and dispatch of both grid and generation. These business cycles are illustrated in the appendix "Business cycles". Such a model allows for integral central optimization of generation and grid assets, but disallows risk management and risk mitigation by market participants, other than consuming (and paying). Nevertheless, fully centralized control obviously can be exercised on a copper plate as well. Designs in which consumption is under central control by rationing are out of the scope of this thesis, as too remote from a level playing field market.

A final assumption is that power deviations $\Delta P$ [MW] are quantifiable and attributable. This restricts direct applicability of this research to at least the synchronous area Continental Europe.

1.1 Research question

A general assumption is that the effectivity and the efficiency of a market design minimally require consistency as necessary condition. Consistency is assessed by determining the presence, or absence, of implicit or explicit exclusion mechanisms. Exclusion mechanisms obstruct participation in and therefore distort markets. The efficiency of markets can also be assessed by the presence, or absence, of perverse incentives to those involved. Such incentives may result in ineffective or inefficient behaviour and/or
inefficient mark-ups in prices. And of course market design has to be effective by complying with physical, legal or regulatory requirements as a necessary condition. With this in mind the main research question that I will address in this thesis is:

“Which necessary conditions are required in an electricity market design aimed at power balancing, to achieve its objectives: efficiency and effectivity while satisfying compliancy requirements?”

This generic research question is decomposed into specific questions:

- Which actors can be distinguished?
- Which processes can be distinguished?
- What process requirements and operational responsibilities apply to whom?

Subsequent questions then are:

- Why should that be so?
- What other choices can be made?

Design variables that I will consider are:

- processes,
- prices,
- delineations of:
  - responsibilities,
  - timings,
  - areas.

There is some evidence that consistency is more than an academic technical-economic issue. One comes from the difficulties encountered in drafting of the European network code on electricity balancing, basically a market design issue, by the transmission system operators themselves (ENTSO-E, 2014, NC EB). The complaint by the European Agency for the Cooperation of Energy Regulators (ACER) in their reasoned opinion on this draft, that this draft was descriptive of present designs rather than prescriptive towards a common design accentuates this point (ACER, 2014). This complaint evidently admits the existence of a multiplicity of existing market designs in Europe. Another one is the conspicuous absence of an attempt to establish a European network code on tariffs, as prerequisite to a European internal electricity market based on a level playing field.

1.2 General aspects of electricity market design

Cramton (2003) proposes principles of market design in electricity markets, a.o.:

- "as simple as possible, but not simpler",
- "address the essential complexities",
- "understand incentives",
- "avoid fundamental flaws".
Avoidance of fundamental flaws requires consistency within and between all aspects of a market design. For responsibility assignments and for granularity consistency dictates: mutually exclusive and jointly exhaustive entities or variables, with no overlaps and no gaps. Multiple granularities require unambiguous hierarchical relations. The vulnerability of inconsistent market design is noted by Boisseleau (2004). As a general rule market design ought to avoid design flaws, or inconsistencies that may open the gates to:

- free riding,
- free loading,
- free lunch,
- moral hazard, or
- exploitation of differences in market design, rather than fostering competition.

### Parameters in electricity markets design

Van der Veen (2012) extensively investigated literature on market design variable space for (multinational) balancing markets, referring a.o. to the work of Vandezande (2011). The 35 balancing design variables identified by van der Veen (2012), can in fact be condensed to a few dimensions. The definitions of processes and the assignment of process responsibilities, determine who is responsible for what process. These assignments should also determine what happens if a responsibility is not met, or if a process fails. Process definitions determine product specifications. Incidentally, the framework guidelines for the network code on electricity balancing actually work the other way around (ACER, 2012). Granularity determines where, and over what periods, these responsibilities for certain processes count, and where, and over what periods, products and prices are firm. Pricing determinations should both allow users to "express directly the underlying economics" (Cramton, 2003) and promote "truth saying" (Stoft, 2002). As prices can obviously influence behavior, pricing should at least not introduce perverse incentives to anyone. They might on the other hand introduce effectivity and efficiency supporting incentives to all.

Abassy & Hakvoort (2009) distinguish two main categories of performance criteria: operational security of supply (effectivity), and incentive compatibility (efficiency). Van der Veen (2012) distinguishes more specifically balance quality as effectivity criterion, and utilization efficiency as efficiency criterion. I will propose quantitative parameters for balance quality effectivity and its efficiency that can be used to develop criteria.

### Models in electricity markets design

Market design assumes models, abstractions of a real 'world'. A 'world' may be the formerly non-market based, central planning and dispatch power supply system, according to the Electricity Act then in vigour (Ministry Of Economic Affairs Of The Netherlands, 1989; Figure 1-2). Or it may be the present electricity market in the Netherlands (Ministry Of Economic Affairs Of The Netherlands, 1998; Figure 1-1) or anywhere else in Europe or beyond. The fact that several designs are possible simply means that there must have been different policies to choose from. The fact that some policy choices are described in legislation means that such choices are political and supported or contested by analyses that are often based on models.
Models are bound by the selection of system boundaries, and valid only for the implicit and explicit assumptions made in deriving these models. A useful model has to be sufficiently simple to perform robust analysis, yet sufficiently expressive to allow for relevant results that can be tested against empirical data. Models are constructs based on theoretical considerations and/or empirical data. Qualitative models try to describe and explain relations, quantitative models aim at predicting outcomes e.g. loss of load probabilities, generation adequacy or future energy prices (De Vries, 2002) or social cost benefit analysis (de Nooij, 2012).

Most models address parts of the whole chain of processes leading to a power system, often with emphasis on central control and optimization in a power supply system configuration as shown in Figure 1-2. Therefore, such models host numerous implicit assumptions on the state of affairs outside the selected system boundaries, taking those for granted. Behavioural preferences (benevolent or malevolent) of users and operators within the system boundaries are seldom taken into account explicitly in optimization studies, in load flow studies and in grid models. Models to assess market functioning may not consider the contingencies and imperfections operators and users continuously are actually dealing with due to less rational, or more malevolent users than implicitly modelled, or due to actual rather than statistical failing of equipment, or imperfect control.

Boisseleau (2004) gives a critical assessment of the use of models in the analysis of day ahead electricity markets, and the inability to cope with the dynamics of power exchanges and their participants, and the inability of such models to cope with high levels of volatility.

Combining results of multiple models may inadvertently combine inconsistent implicit assumptions. De Bruin (2004) identified “five problematic consequences” of the “true-in-the-abstract view” of models:
- “exaggerated stress on mathematical layout”
- “strong insistence on intuition in place of empirical data”
- “an internal development hardly influenced by factors from outside of the theory”
- “clever model building ignoring descriptive adequacy”
- “instrumentalism.”

De Bruin refers to game theory. Boisseleau (2004) expresses similar concerns to models used in power system and electricity markets design analysis. An example of such problematic consequences is that predictive models are rarely tested against empirical data. Testing of predicted loss of load probabilities indeed is almost impossible. Another example is the implicit and intuitive assumption that there is, or should be, a standard market design that only needs some harmonization to create a European wide internal electricity market. Again it is Boisseleau (2004) who critically comments on the European Union lack of details on market design it its encouragement of cross border trade of electricity and of elimination of discriminatory practices. Indeed the European target model offers little more guidance other than ‘energy only’ and ‘flow based implicit’ allocation of limited interconnection between member states.

A final example is found in the supposedly conceptual framework guidelines on electricity balancing (ACER, 2012, FWGL). Here the focus is on exchanges between markets, without going into detail on compatibility.
issues between present market designs. Conceptuality should not imply obscurity. But then again the
references used by me serve policy aims in a political environment.

1.3 An introductory bottom up analysis
To put the statistics of the European power system and market freedom into context I took my own electricity
bills over 2006/2007 and 2013/2014, each one about one year’s worth: 4373 [kWh] in 372 days, respectively
4535 [kWh] in 351 days. One kWh is a thousandth of a millionth of one [TWh], so each bill represents about
0.00000013 % of the yearly consumption quoted above, admittedly a small sample. The consumption of
electrical energy in my home can be attributed to several appliances in use: washing machine, clothes dryer
(not the first billing period), vacuum cleaner, electrical lighting, refrigerator, central heating pump, radio and
television, a lot of battery chargers, and having two sons with computers and limited energy consciousness.

Studying my bills reveals a set of ‘dramatis personae’ or roles performed by actors. Though the roles are
fixed, some of the cast might have changed, albeit for different reasons. First of all there is me, unchanged.
After all I consumed the electrical energy and I had to pay these bills. My supplier receives part of my
payments. It is the company from which I bought the electrical energy that I consumed at my home. In the
intervening years I chose not to select one of its competitors or rather I did not bother to switch, even when
entitled to do so (DIRECTIVE 2009/72/EC, 2009). Then I paid to a regional grid company the transmission to
my home of the electricity I bought. The regional grid company changed its name, but otherwise remained
the same. Its activities and tariffs are regulated by the national regulator Autoriteit Consument en Markt.
They are non-commercial subject to national legislation (Ministry Of Economic Affairs Of The Netherlands,
1998). I could only have switched to another one by moving to another region, as regional grid companies
are regional monopolists (ACM Gebiedsindeling, 1999-recent). My payment to my regional grid company
includes a part they pass through to a national grid company, the national transmission system operator, my
employer. As a monopolist it too is regulated as are the regional grid companies. Inevitably, I paid the state
taxes and levies. Of course, my government is somewhat subject to my voting power.

Normalizing both bills to 365 days, my consumption went up from 4.3 [MWh] in 2007 to 4.7 [MWh] in 2014,
and the bills would be, including value added tax (VAT), € 760 respectively € 991 in 2014, neglecting
inflation. I would then have had to pay my supplier € 335 respectively € 354. This increase would have been
smaller than that of my consumption, indicating a price decrease. To my regional grid company I would have
had to pay in 2007 € 199, most of it energy based, and in 2014 € 223, all of it based on having a connection
of a certain capacity, i.e. all was due to daily capacity [kVA] charges. Nevertheless this indicates a slight tariff
increase. And finally or foremost, the state would have collected € 226 respectively € 414, including VAT.

In view of my payments and my usage my connection foremost seems to serve convenience. Continuous
access to the transmission system (continuity of supply) is as much as important as is this supply itself. I can
defer washing a couple of days, but for heating I depend on my central heating pump. Malfunction generally
is experienced at very local level: a spent bulb, broken-down equipment, a blown fuse. Once or twice over
the past decades not just my home, but also my neighbours were cut off. This could be traced back to
someone accidentally cutting a cable, and supply was restored swiftly. Such events are annoying but not
threatening. In other societies than in the Netherlands, with households depending on electrical heating the physical dependency on continuous access to the transmission system might be larger. For most consumers of electricity the connection serves a security function, even more so to those that are commercially driven to consume electricity.

Our dependency on continuous supply is ironically best illustrated by analyzing large scale black outs that affect whole countries or counties. These are indeed rare but, when lasting, will have serious consequences for public health and safety, for economy (Rathenau Instituut, 1994; de Nooij, 2012). Large scale disruptions are damaging to society and to individuals. That large scale power systems and their disruptions are more than just technical-economical study objects, and may deeply affect social stability is clear from the following quote:

```
The conflicts of the 1990s led to the disintegration of a unified energy system that stretched from the Adriatic to the Black and Aegean Seas. What was once a single system suddenly was a patchwork of several. Regardless of the frontiers drawn on maps since the conflict erupted, the separate entities still rely on each other for the smooth functioning of their power supplies. In the words of the European Commission:

"Energy Community is about investments, economic development, security of energy supply and social stability; but – more than this – the Energy Community is also about solidarity, mutual trust and peace. The very existence of the Energy Community, only ten years after the end of the Balkan conflict, is a success in itself, as it stands as the first common institutional project undertaken by the non-European Union countries of South East Europe."
```

Source: Energy Community website, emphasis added

1.4 Organization of this thesis
After this introductory Chapter 1 in which general aspects of electricity market design from literature are presented, together with an introduction to, and a bottom-up analysis of an electricity market from an end-user perspective, a conceptual framework for balancing in an electricity market is described in Chapter 2, for power [MW], energy [MWh] and frequency [Hz]. This simplified qualitative model acknowledges two kinds of actors, who interact in various ways, with each other and amongst themselves. The actors are the operators, and the users. Actors do things; that is why they are called actors. So, I explore what the actors must do (responsibilities), what they can do (options), and how they interact. In Chapter 3 I propose a methodology to explore physical results of balancing markets to allow for comparison among different designs. A qualitative price model for balancing energy and imbalance is presented in Chapter 4. I will address the issue of temporal price granularities in Chapter 5. In Chapter 6 I explore the financial side of balancing market designs. In chapters 7 and 8 the conclusions from the previous chapters are put to the test against reality using data in the public domain. In Chapter 9 some conclusions from this research are highlighted.

Use of terms and definitions
References used for this analysis of power systems and electricity markets are full with capitalized terms like "Imbalance", and/or acronyms like "BRP" and "BSP". Capitalized terms and acronyms require clear and unambiguous definitions. Unfortunately different references use slightly different definitions for nearly similar terms, give general terms a very specific definition or vice versa. Imbalance is a general term for
disequilibrium, but as defined, capitalized term relates only to a specific energy volume, over a specific time period, attributable to a specific entity (ENTSO-E, 2014, NC EB). Therefore I avoid the use of acronyms as they may lead to confusion, and avoid the use of capitalized terms. The terms used are described rather than defined when introduced, and where appropriate a reference to an official definition will be given. The term imbalance alone will indicate an energy imbalance over an imbalance settlement period, otherwise I will use the term power imbalance. Another ambiguity arises from the terms control block and control area. Although these terms have specific definitions related to responsibilities, in reality they are actually mostly coincident (ENTSO-E, 2014, SD NC EB). Since I cannot change terms in the references, I am forced to use both. In this thesis the terms control block and control area have the same implication.

**Use of references**

My choice of references is somewhat biased in favor of public reports published by, or under the aegis of, system operators, regulators, or their European representative bodies. Even when such publications and reports are authorized by those entities, which are responsible for effective and efficient functioning of the European electrical power system, they lack the scrutiny of academic peer review of scientific publications. One must be aware that such publications may promote self-interests of those responsible for publishing.

**Use of data**

Even without a legal obligation to provide data on transparency (EU, 2011, REMIT), essential data provision on balancing volumes and prices should be the moral duty of regulated monopolists like transmission system operators. They ought to be accountable to the public, who ultimately are their source of existence, and raison d’être. Nevertheless, there is an astonishing discrepancy between which data are available in the public domain, and which data and how these are used in academic literature. On the other hand one must always be aware that public data, including technical data without financial consequences to the data owner or publisher are not flawless, and always need to be scrutinized by every scholar. Particular vulnerable in this respect are completeness and accuracy of load data, intermittent generation, and installed and available generation capacities, and their marginal cost. Yet there is big data available, although insufficient to feed and test power supply curves empirically. One way to resolve this issue is to enforce the availability of ever more data (EU, 2011, REMIT). Another one is to obtain more information from data already available. Even without full access to complete data on availability, cost and added value of every connection, it will be shown that one can still analyze the functioning of electricity and balancing markets.

In analyzing data I deliberately restrict myself to robust statistics and visual representations, in the spirit (at least so I hope) of respectively John Tukey (1977) and Edward Tufte (1983). My brother Joris † introduced me to their ideas.
2. A conceptual framework for electricity markets

In this Chapter the following concepts in an electricity market are discussed: domains, actors and options. The domains I distinguish are material assets like grids and interconnectors, or an installation or facility that has access through an access point: a connection. The actors I distinguish will come in two flavors: users and operators (Section 2.2). Mutual exclusivity (unbundling) of these roles is prescribed (DIRECTIVE 2009/72/EC, 2009). Finally I discuss the options open to the actors distinguished to fulfill their desires or responsibilities.

ENTSO-E, together with the European Federation of Energy Traders and the European forum for energy Business Information eXchange, has dissected the electricity market on its information requirements. I will utilize this harmonized role model, "identifying all the roles that can be played for given domains within the electricity market". This harmonized role model for electricity markets defines domains and roles, and enables to construct actors, by assigning and combining separate roles.

"A Role Model of this nature shall be the formal means of identifying roles and domains that are used in information interchange in the electricity market. It is important to stress that it is not a model of the electricity market but rather a Role Model related to the information interchange. The necessity for such a model arises from the possibility that a single party in the market may assume multiple roles. This implies that the roles need to be atomically decomposed where necessary in order to satisfy the minimal information flows for a given process required within the electricity market."

- A role represents the external intended behavior of a party. Parties cannot share a role. Businesses carry out their activities by performing roles, e.g. system operator, trader. Roles describe external business interactions with other parties in relation to the goal of a given business transaction.
- A domain represents a delimited area that is uniquely identified for a specific purpose and where energy consumption, production or trade may be determined.
- An actor represents a party that participates in a business transaction. Within a given business transaction an actor assumes a specific role or a set of roles. An actor is a composition of one or more roles and as such does not appear in the model."

Source: ENTSO-E Harmonised Role Model Manual Version 2011-1

This role model is intentionally focused only on information interchange in the electricity market, and not on other responsibilities. These limitations are acknowledged by the ENTSO-E working group Electronic Data Interchange. This clear, but restricted focus implies that this role model cannot offer guidance to the physical objects or material asset in power systems, other than by inference. Nor does it offer specification of products and processes in electrical power systems, such as will be explored in this thesis, using this asset domain/actor model for electricity markets. Compared to the full role model this thesis considers the actor level and not the atomically decomposed roles.
2.1 Domains

In Figure 2-1 the physical domains constituting a power system are mapped:

![Physical domains map](image)

**Figure 2-1 Physical domains map**

**Connection**

A *connection* is the material asset, an installation/facility, in my case the appliances in my house, with access to a grid (ENTSO-E, 2014, NC RFL) by a physical interface. A connection enables the physical exchange of active and reactive power between the installation/facility and the grid. A connection may also serve to separate ownership (of the asset, of the energy) and responsibility (for the asset, for the energy, for the imbalance) between actors on either side of the connection.

**Grid**

To most of us with a connection, the grid is simply that what you are connected to, that what is on the other side of your connection. A *grid* is a spatially contiguous physical infrastructure joining connections and interconnectors, allowing transmission of power among connections, among interconnectors, and among connections and interconnectors. A grid consist of connected elements through which power may flow, like (overhead) lines and (underground) cables, and controllable components like circuit breakers, transformers, phase shifters, or AC-DC converters. A grid enables its physical users to interact physically by injecting or withdrawing power, and its commercial users to interact commercially by trading energy. A transmission grid is a grid with few connections, connecting distribution grids that have the majority of connections. A transmission grid may be connected to other national transmission grids through interconnectors. The transmission grid as used here corresponds to the transmission system described in European legislation (Regulation 2009/714/EC, 2009). The transmission system used in this thesis comprises the transmission and distribution grids and interconnectors.
**Interconnector**

An *interconnector* is a physical transmission line which crosses or spans a border between Member States and which connects the national transmission systems of the Members (Regulation 2009/714/EC, 2009). But the world does not end at Member States borders. Interconnectors may join other, non-Member States too. Interconnectors are either synchronous, passive (AC) or asynchronous, active (DC) elements. Although these types differ in physical, technical and operational aspects, related e.g. to phase angle and frequency, these differences do not affect the outcome of this thesis.

Interconnectors serve dual purposes:
- to enhance system security and stability by allowing the exchange of power,
- to enlarge market access (Supponen, 2011), by allowing energy trading.

### 2.2 Actors

In Figure 2-2 the actors in the power supply system are mapped to the domains from Figure 2-1.

**Figure 2-2 Actors**

#### 2.2.1 Users

The concept of user has two implementations. It applies to physical or grid users that may inject into or withdraw power from a grid. In addition, it applies to commercial or system users that may trade energy with other system users or trade on behalf of grid users. All grid users are, or have to be, represented by system users as will be explained in Section 2.3.2. The reverse is not required. Not all system users need to represent grid users. An example of definitions with relevance to the concept of system user is shown here, with the terms supply, sale and customer clearly having commercial significance:
Note that the Directive uses electricity as a general term without referring to either power or energy. As seen from the operator perspective connections can act as:

- source: injecting power into the grid, and hence into the transmission system (producer, generating),
- sink: withdrawing power from the grid, and hence from the transmission system (customer, load).

Some grid users alternate between injection and withdrawal. Examples of prosumers are household consumers with solar panels, or industries with own generating facilities. Prosumers are usually connected to a grid for their own convenience, but might be able to satisfy their needs without being connected. In such case their connection functions predominantly as supply insurance. Large scale pumped storage facilities switch between withdrawal and injection too, but are generally not referred to as prosumers. These facilities own their existence to, and typically operate in support of grid and/or system. These facilities have no function without access to a grid.

The principal interest of commercial users is energy trade, to buy in case of customers, to sell in case of producers. All users set their energy trade targets themselves according to their ability and desire to trade energy and to inject or withdraw power as grid users. Violation of their own set trade targets is two-sided, too much or too little. The main concern of all users in the end is energy.

Source: DIRECTIVE 2009/72/EC
2.2.2 Operators

Physical users are actors on one side of the connection. The actors on the grid side are operators (Figure 2-2), also with two varieties, grid and system operator. Operators have responsibilities, as shown by the following definitions of the concept of system operator:

" ‘transmission system operator’ means a natural or legal person responsible for operating, ensuring the maintenance of and, if necessary, developing the transmission system in a given area and, where applicable, its interconnections with other systems, and for ensuring the long-term ability of the system to meet reasonable demands for the transmission of electricity."

Source: DIRECTIVE 2009/72/EC

"System Operator: A party that is responsible for a stable power system operation (including the organization of physical balance) through a transmission grid in a geographical area. The System Operator will also determine and be responsible for cross border capacity and exchanges. If necessary he may reduce allocated capacity to ensure operational stability."


The transmission system operator definition above corresponds in this thesis to the term system operator. In words the shorter “system operator” is preferred over “transmission system operator”. The acronym TSO however is embedded to such an extent that this will be used in diagrams. Its unique responsibilities are the organization of physical active power balance, and those related to cross border capacity and exchanges over interconnectors. Other than these responsibilities, a system operator has the same ones as grid operators, also known as distribution system operators (Directive 2009/72/EC, 2009), with DSO as acronym. A system operator is therefore a grid operator with additional responsibilities. However, grid operators are not the main subject of this thesis. The term ‘independent system operator’ is used in USA for a ‘non-profit system operator that can be minimally regulated’ (Stoft, 2002)). This corresponds to a system operator without the responsibilities of a grid operator.

Grid and system operators are the actors that jointly enable all grid users to dispatch (inject or withdraw) active power into the grid and to enable all system users to trade energy according to their desires. Grid have the responsibility to grant grid access to the users within predetermined quality limits with respect to voltage and frequency stability in order not to harm the installations behind the connections (Rebours, 2008), or distort energy markets (Stoft, 2002). System operators have the responsibility to grant system access, including access to interconnectors to system users. Grid and system operators together have to safeguard operational security all the time.

Grid operator

A grid operator is, in a given grid, responsible for:

- a grid as a material asset being in existence (its development, maintenance and exploitation),
- granting its grid users access to its grid,
Grids and grid operators may be organized geographically, and/or hierarchically, by voltage level and/or function. The grid operators’ responsibility amounts to prevent the active power injections into and withdrawals from the grid by grid users from jeopardizing continuity of service by:

- overloading of components of the grid,
- collapse of voltage.

Violation of a control limit for active power occurs only as overloading. Such control limits are physical or contractual. Grid operators have no active power balancing responsibilities. Their main concerns are grid capacity, voltage and reactive power, all of them local topics. According to the copper plate principle I will disregard local congestions, as the prime topic in this thesis is balancing. Therefore not included in this thesis are the following responsibilities of the grid operator:

- keeping the voltage between limits (including the organization of the reactive power balance),
- power quality,
- local congestion management.

Nevertheless local congestion management does impact applicability of some results of this thesis. Likewise the interactions between system operator and grid operators are not included in this thesis.

**System operator**

A system operator is for a given load frequency control area (ENTSO-E, 2014, NC OS), responsible for:

- a transmission grid as a material asset, and where applicable its interconnectors., being in existence, its development, maintenance and (financial) exploitation (Directive 2009/72/EC, 2009),
- granting access to cross border capacities to commercial users,
- keeping the frequency between limits,
- the organization of the active power balance.

System operators are not organized hierarchically, but only geographically. A transmission grid connects grids operated by grid operators, and/or interconnectors. They act as grid operator to those users directly connected to the transmission grid. The system operators’ responsibility amounts to reconcile the active power injections into and withdrawals from all grid users to such an extent that:

- cross border exchange of power and energy is secured, or at least transactional firm,
- an agreed frequency quality is maintained,
- continuity of service is not jeopardized by collapse of the (interconnected) system frequency.

An additional responsibility of the system operator is the financial obligation to settle energy exchanges with other system operators (ENTSO-E, 2014, NC EB).

Violation of the active power balance control target occurs at two sides: too high or too low. The main concerns of a system operator are the active power equilibrium and the cross border exchanges of power and energy, all of them global topics.
2.2.3 Summary of the domain actor model

In the previous sections the conceptual actors in the power supply system have been identified. As much as grids and interconnectors connect physically, at the same time they separate responsibilities between the different actors. A grid separates connections from each other and a grid separates connections from interconnectors and interconnectors separate grids from each other. An additional proposition, that a grid separates interconnectors from each other is valid yet. But with interconnector nodes increasingly drawing attention, e.g. in a North Sea super grid, this proposition may not hold forever as anticipated in the Triffid case study (Twenties project, 2013).

Each connection is served by only one grid operator and only one system operator. A grid operator may serve many connections and is to them a monopolist, in providing grid access to physical users. The system operator is a monopolist in providing system access, including access to interconnectors, to all system users in its control area, whether they are connected or not. The main variables of a market based power system in a transmission system configuration are mapped to the domains in Figure 2-3.

![Figure 2-3 Electrical variables mapped to domains](image)

It is at the connections that users have energy converted into electrical power and injected into the grid, or that power is being withdrawn and energy consumed by grid users. It is into and through the grid that electrical power flows, and that the combined power injections and withdrawals from all connections and interconnectors create a power balance. It is in the synchronously interconnected system that ultimately a global power balance is manifest in a system frequency deviation. And finally a permanent power balance at a constant frequency implies a global energy balance over all connections over all timescales. On the other hand a global energy balance may contain frequency deviations or power imbalances over shorter time scales.
Despite the current plethora of areal definitions and relations in Europe only a country with one single system operator controlling its control block or area, and one single bidding zone (Regulation (EC), 2013, 543/2013) is considered. It is joined to other similar configured countries by interconnectors. This is actually the prevailing configuration in Europe (ENTSO-E, 2014, SD NC EB). This choice has no effect on the results of this thesis, other than respecting Occam’s razor.

2.3 A decomposed service model

According to DIRECTIVE 2009/72/EC (2009) and its predecessors, the trading of energy in a fully unbundled electricity market is the exclusive domain of users. Trading among users has to be facilitated by grid and system operators that jointly have the responsibility to safeguard operational security all the time. The relations between operators and users are an exchange of services. A precursor to a service based decomposition is illustrated by the following quote:

*Although the Directive 96/92/EC, concerning common rules for the internal market in electricity, uses only the term ancillary services (with the meaning of all the services necessary for the operation of a transmission or distribution system), it is useful (from a technical point of view) to distinguish between "Ancillary services" and "System services". "System services" are all services provided by some system function (like a system operator or a grid/network operator) to users connected to the system. "Ancillary services" are services procured by a system functionality (system operator or grid/network operator) from system users in order to be able to provide the system services. Within system services a further sub-classification is sometimes suggested, relating to the specific system function that is supplying the service (i. e. the system operator or the network operator). In practice these arrangements differ widely, depending on the kind of system service provided by the functions in question. As an example there are systems where network losses are provided by the system operator and some other systems where they are provided by the grid operator.

Here it is suggested to define system services as those services from the system that enable system integrity and to define transmission services as those services from the system that enable and assist economic power transfer. This does not provide an unambiguous framework and the assignment of each of the services from the system to one of those categories will depend on the specific arrangements in each system.

The Directive 96/92/EC uses one term for system operator, which shall be responsible for: operating, ensuring the maintenance and developing the grid (Article 7.1), managing energy flows (Article 7.3), ensuring a secure, reliable and efficient electricity system ensuring availability of all the necessary ancillary services (Article 7.3).

From this point of view the System and Grid Operator are one person, the Transmission System Operator (TSO). The advantages of this decomposition are:

- it is possible to define clear interfaces between system operator and system user so that users are separated from difficult problems with power system control and the technical requirements can be defined;
- a fair contribution of the individual users to the system proper and safe operation can be ensured;
- a clear separation is created between the TSO and the generation, especially when they are parts of one company."


Decomposition in directionality not only is useful from a technical point of view, but also serves to highlight some market aspects for these exchanges of services. The provider and recipient of each service are unambiguously identified. Therefore in Figure 2-4 a more precise decomposition is proposed of the relation
between operators and users, distinguishing between grid and system operators, grid and system users, and grid, system and user services.

![Decomposed service model](image)

Figure 2-4 Decomposed service model

User services is in this thesis a more general term for all the services that users may be required to provide to the operators. User services are here not restricted to ‘ancillary services’ as they are in current definitions (DIRECTIVE 2009/72/EC, 2009). The operator services include administrative market processes as accreditation as grid or system user, or imbalance settlement. The grid and system access services enable economic power transfer and the stable frequency service system integrity (Eurelectric, 2000).

Users have freedoms (DIRECTIVE 2009/72/EC, 2009)] or rights. The authority to request to connect to a grid or to use a connection to actually inject into, or to withdraw active power from a grid rests with the grid user. It can be a customer that wants to consume energy at acceptable prices, whenever it wants, and hence withdraw power. Or it can be a generator that wants to sell energy at profitable prices, whenever it can, and hence inject power as grid user. So all users want to trade energy to their advantage as system users, in volumes (how much), as well as in time of delivery (whenever) and, as grid users, dispatch accordingly by injecting or withdrawing power (Boisseleau, 2004). Therefore users need access to grid and system. However access to the grid and the system cannot be unconditional to protect the grid and system’s security and integrity, to prevent operator expenditure, or to prevent user abuse. Conditions to grid access and to system access of users thus should aim at fostering efficient cost and fair distribution of costs and risks among users, increasing social welfare. The common interest of the collective should prevail over the fate of the individual.
2.3.1 Operator services

Grid services

Grid access by a physical connection and a connection agreement is a prerequisite for users to inject power into or withdraw power from a grid. The grid operator delivers the connection service to each grid user by the provision of a physical connection and of course its opposite by disconnecting. Connecting to a grid is a choice, governed by (economic) preference of its user over alternatives. Economic alternatives to permanent connection to a grid are rare. Examples are remote dwellings or villages, ships, or satellites. The connection service is to deliver a physical connection and a connection agreement to the grid user. By connection agreements grid operators allocate physical capacity to grid users, enabling them to exchange energy. A grid operator then has to allow, all the time, each grid users' exchange within the contractual limits of their agreement. At the same time the grid operator has to comply to its standards referring to voltage limits, and safe and secure grid operation.

A common aspect of power systems is that consumers are neither obliged to connect to a grid, nor are they ever forced to consume, although consumption may be curtailed under exceptional circumstances. Without economic alternatives consumers are rather captive grid users, to whom little conditions to grid access apply. Grid access to consumers is rarely refused and only limited physically at the connection by fuses, and contractually limited in the connection agreement.

Grid access cannot be unconditional as that would lead to perverse user behaviour, a fundamental design flaw. But from the user perspective conditions on access are restrictive. Restrictive measures for grid access can be a refusal to connect at all, or less extreme, connection size limitation. Both options are aimed at either reducing the need for deep grid investments or reducing the need for operational measures to relieve congestion. Other restrictive measures influencing freedom of dispatch of grid users are curtailment (e.g. of withdrawal, or of intermittent injection) or commandeering (i.e. of controllable injection).

System services

Without appropriate trades, withdrawal of energy would amount to theft from, and injection of energy to dumping into the power system. So dispatch by grid users has to be covered by trades, with other system users or with operators as explained in Section 2.3.2. A prerequisite to trade is therefore not so much connection to a grid but rather system access or accreditation or permission to trade. System access is prerequisite and thus a condition for grid access. But dispatch of active power is not a prerequisite to energy trade, nor is grid use therefore prerequisite to system access. System access is delivered by accreditation as balance responsible party. Whereas grid access is granted per individual user, balance responsible parties may represent multiple grid users, or not be connected to a grid at all. Examples of system users without connections are power exchanges and traders.

System operators must grant system users access to cross border capacities (Regulation (EC) 714/2009, 2009) allowing energy trading over interconnectors, enlarging market access. But at the same time interconnectors enhance system security and stability by allowing the exchange of active power. These purposes compete with each other. Besides, interconnectors take time and capital to be built as is shown in
appendix "Business cycles: Interconnectors". Rules aimed at maximizing market access, firmness of transfer capacity, allocation and use of transfer capacity, and remuneration of investment in interconnectors is the subject of the EC Regulation on capacity allocation and congestion management (Regulation (EC), 2014, CACM). This subject is outside the scope of this thesis, but the main point is that cross border capacities are made available by system operators to their users in limited volumes.

Each active power injection or withdrawal over a connection constitutes in itself a local power imbalance. Without any power injection or withdrawal the system would be stable, albeit without frequency. A black out is prejudicial to users, as it prevents grid users to exercise their access rights, as well as harming system users in market access rights.

A stable frequency at a set value is a pervasive system parameter throughout a synchronous area. It requires a global power balance, meaning that the sum of power injections and withdrawals, including losses, in the transmission system equals the sum of imports and exports over all its interconnectors. If not, frequency will increase in case of more power injection than withdrawal, or the frequency will decrease in the reverse situation. Power imbalances require active control measures and processes performed by the system operator. They aim at a stable, set, system frequency, and aim at interconnector power flows conform to agreed firm energy trades (Section 2.4). This system service is delivered continuously to all grid users by complying with internationally agreed requirements and standards.

Glachant & Saguan (2007) or Vanderzande (2011) mention the absence of economic storage as causing the need for system balancing. It is the absence of infinite inertia within the transmission system itself, or connected to the transmission system, that necessitates active power balancing. Storage facilities connected to the grid may of course affect the number and volume of power and energy balancing requirements and reserves.

**2.3.2 User services**

**Balance responsibility**

All market trades and transactions enabled through system access occur among users themselves. In principle these trades are unlimited, although access to interconnectors is limited by limited capacity entitlements, and physical grid access by grid users is limited by their grid access conditions. To manage the risk of unlimited system access by users in combination with the responsibilities of the system operators (safeguarding operational security), users are obliged to accept financial responsibility for their imbalances (ACER, 2012, FWGL EB). This requires accreditation as balance responsible party. Balance responsibility is basically an administrative service to the system operator. It includes the obligation for the balance responsible party to, per discrete time interval or imbalance settlement period (ISP):

- notify the system operator of its *trade position*,
- accept the *allocation* of realized values of actual injections and withdrawals from those connections for which the balance responsible party has declared balance responsibility,
- accept *adjustments* due to balancing service provision, and ultimately,
- accept financial imbalance settlement.
Imbalance is an energy product. All its components are energy volumes in [MWh], established for each imbalance settlement period for each balance responsible party. The trade position is the net volume of all energy bought from and sold to any other balance responsible party. The allocated volume is the net volume of all energy injected in and withdrawn, over all the connections the balance responsible party has assumed responsibility for. Allocation can be based on actual metered values, or on standardized consumption profiles. The adjustment for each balance responsible party is, the net volume of all energy requested to withdraw or inject by the system operator due to balancing service provision. Balancing responsibility enforces balance responsible parties to match their energy trades in one or more of only the following ways:

- transactional: buy equals sell,
- physical: dispatch equals trade (net injection and withdrawal equals net buy and sell),
- financial: imbalance settlement.

Mandatory imbalance settlement imposes a financial risk onto the balance responsible parties for their energy imbalance. But it relieves them from any responsibility for their own or anyone else's, power balance. On the reverse side of imbalance settlement the system operator is the one and only counterpart to all balance responsible parties. Imbalance settlement, like system access provision is a monopolistic function of the system operator. Delegation of imbalance settlement to another entity than the system operator does not change this character.

**Balancing service provision**

Operators need access to control power and energy resources to perform their assigned control processes. Unbundling (DIRECTIVE 2009/72/EC, 2009) disallows operators to own and operate assets providing energy to users. Recovery of expenditure by access tariffs would distort a level playing field. To what extent this includes ownership of balancing power resources is not perfectly clear. To illustrate the unclarity by analogy, reactive power compensation for voltage control is a local requirement for grid operators. Not every local requirement may have local connections available to offer economically attractive sources. Therefore capacitors and coils providing reactive power compensation are considered grid components, and therefore admissible to be owned and operated by grid operators. Opinions on owning and operating active power storage facilities (flywheels, batteries, pumped storage, hydrolysis) differ. They can for the time being owned and operated by users or by operators. Nevertheless, assuming that system operators cannot, in principle, invest for this specific purpose, then they are forced to rely on users to provide the means to perform the control processes aimed at a stable, set, frequency, and agreed interconnector power flows. Balancing service provision comprises balancing reserve (capacity) provision, and balancing energy delivery to the system through real-time activation of balancing reserve capacity.

Mandatory balancing services may be enforced by technical requirements to appliances, or by connection requirements for an installation/facility. Mandatory balancing service provision is restrictive from the user perspective. Stoft (2002) provides a critical discussion on mandatory self-provision of spinning reserves. Marked based balancing service provision provides an opportunity for commercial trades from balance service provider to the system operator. This requires accreditation as balancing service provider, another
system access service that can only be provided by the system operator. The reverse perspective of balancing services provision is that the system operator is the only counterpart to all balancing services providers. The balancing services market is a monopsony.

2.3.3 Governing principles in the decomposed service model

In a liberalized power system users can, or ought to be allowed to, trade energy as a commodity on an open energy market (DIRECTIVE 2009/72/EC, 2009). Operators are monopolies in the services they deliver to their users. These services are access provision, including accreditation services, and a stable, set frequency. System operators are also monopolists in imbalance settlement they mandatory require from balance responsibility parties. And finally, system operators are monopsonies with respect to the balancing services they require from their users. These governing principles are shown in Figure 2-5.

Figure 2-5 Governing principles

Frequency, as a common good shared between synchronously interconnected system operators, or the responsibility of a single system operator for non-interconnected systems, has no market value in itself. The (shared) responsibility for its stability is governed by non-market based principles. The thus inherent “free rider” problem from the user perspective is discussed in Stoft (2002). The principles of solidarity and non-intervention hitherto practiced between system operators are currently formalized in a more robust design, a.o. to prevent such free riding behavior among system operators:

"...and the requirements under the Electricity Regulation and the Electricity Directive, such as the need for establishing objective, fair, transparent and non-discriminatory rules for balancing in a cost-reflective way, and for creating appropriate incentives for network users and transmission system operators (hereinafter referred to as “TSOs”) for more efficient balancing (see e.g. Articles 15(7), 37(6) and 37(8) of the Electricity Directive)."

2.4 Power balancing processes

In the previous sections the conceptual actors in the power supply system have been identified, and their relation to the infrastructure and to each other established. Users and operators interact by exchanging services as described in Section 2.3. The prime objective of all operators is to provide continuity of their services to its users. For grid operators this includes taking responsibility for reactive power equilibrium in the grid, for system operators this means taking responsibility for the active power equilibrium between injections and withdrawals, at a sufficient level to maintain a stable voltage respectively frequency within set ranges. In theory such continuity is passively guaranteed in an electrical power system by an infinite alternate current grid with an infinite [MW/mHz] power-frequency characteristic K (with a non-zero frequency) and [Ω] impedance. In practice none of these conditions is satisfied, so active power control has to be organized and exercised to enable consumers to consume when they want, producers to sell when they can, and grid and system operators to safeguard operational security all the time. Active power control by grid operators is not a balancing responsibility, but only aimed at not exceeding capacity limits. Grid operators do not have active power balancing responsibilities.

The main active power control processes for stable power system operation are laid down in the ENTSO-E policies (ENTSO-E, Policy 1). They have been described by Cigre (1999). Processes are structured, measured sets of activities designed to result in a specific outcome for a particular customer or market; a process is thus a specific ordering of activities across time and space, with a beginning and an end, and clearly defined inputs and outputs. A process is a structure for action (Davenport, 1993). Cigre (1999) uses the keywords: contain, restore, replace in their process descriptions:

"The System requirement for an active Power Reserve Service is to:

- Contain system frequency deviations caused by overall generation availability or in demand;
- Restore system frequency to within statutory or operational limits within acceptable timescales;
- Provide a sufficient level of active power reserve to accommodate the impact of cumulative generation losses or unexpected increases in demand, and to allow active reserves associated with the containment and restoration of system frequency to be replaced within specific timescales, following changes in generation availability or system demand;
- Provide a level of system gain (ΔP/Δf) which is compatible with the desired level of continuous frequency control;
- Enable tie line power flow control within contracted, agreed or accepted levels of deviation."

Source: Cigre, 1999, emphasis added

Cigre (1999) acknowledged that at least in and across Europe, these processes are not uniformly performed, nor are they using identical products. The process based approach has been further developed by ENTSO-E (ENTSO-E, 2012, AHTor). The main conclusions from this report related to power and frequency control are:
This report also maps extensively correspondences between processes and products in different designs across ENTSO-E’s members (ENTSO-E, 2014, AS Enquiry). The processes in different synchronous areas are more universal than the products used by different system operators. The development of European network codes is supposed to lead to more harmonization and integration in and across Europe. It is considered beneficial to reach the European target model in time, even without exactly knowing its specifications. These network codes, or guidelines, will enter into force in the remaining part of this decade, which poses a considerable, additional challenge to system operators and users.

The universal power control processes distinguished and their objectives, including appropriate time limits are prescribed as the following processes (ENTSO-E, 2013, NC LFC-R), represented in Figure 2-6.:

- **frequency containment process ('braking'):**
  - stabilize the system frequency by activation of frequency containment reserves,

- **frequency restoration process ('reversing'):**
  - regulate the frequency restoration control error towards zero within the time to restore frequency (15 minutes in continental Europe),
  - and progressively replace the activated frequency containment reserves by activation of frequency restoration reserves (for continental and northern Europe).

An additional process is:

- **reserve replacement process ('releasing'):**
  - progressively restore the activated frequency restoration reserves,

*The system frequency is a representative value for the rotation speed of the synchronised generating units. System frequency is a common property with an equal value in the whole synchronous area and the responsibility of maintaining it within the agreed limits is shared by all TSO’s in that area. The actual system frequency value is a consequence of the power balance between generation and load resulting from all simultaneous events and actions of all system users, system inertia system static characteristics and activation of operational reserves. The following aspects have to be taken into account:

- Stable system frequency is a common good for all system (and grid) users, but no (tradable) commodity.
- Deviations from the nominal frequency value may jeopardize normal operating conditions.
- Both size and duration of the system frequency deviations must be limited.

**Responsibility for defining and providing stable system frequency is assigned to all TSOs in the synchronous area:**

- Frequency Containment is a joint responsibility distributed among all TSOs in the synchronous area.
- Frequency Restoration is a local responsibility only of the imbalanced TSO).

Source: ENTSO-E, 2012, AHTor, emphasis added in italics.
The principles of the frequency containment process execution contain a.o. that "the overall frequency containment reserve activation is characterized by a monotonically decreasing function of the frequency deviation" $\Delta f = f - f_0$ [mHz] (ENTSO-E, 2013, SD NC LFC-R). With actual and target frequency as input, the frequency containment process is feedback control. It has a well-defined start: a frequency disturbance, and a well-defined end: a stabilized frequency offset from the target frequency $f_0$. Simultaneous counter-activation of automatic, proportional frequency containment reserves within a synchronous area is unlikely. All reserves react to the same frequency, but response and reaction times may vary between different resources. Nevertheless, as frequency deviations may vary faster than the actual activation or deactivation of frequency containment reserves, frequency containment is not a perfect control process.

The frequency restoration process too has a well-defined start, an actual power exchange $P$ [MW] between synchronously interconnected system operators offset from its set value ($P_{set}$, [MW]) taking into account the contribution from the frequency containment process $P_{err} = P - P_{set} + k \Delta f$ [MW], and formerly known as control error or ACE (ENTSO-E, 2013, NC LFC-R). It also has a well-defined end: $P = P_{set} + k \Delta f$ [MW], or more realistically, within bounds around $P_{set} + k \Delta f$ [MW]. The frequency error $\Delta f$ [mHz] is global and affects all interconnected system operators, whereas in continental Europe the $P_{err}$ [MW], and ACE are local for each system operator. Frequency containment and frequency restoration therefore may locally counteract.

The factor $k$ [MW/mHz] describes how the local system is considered to react to a frequency deviation $\Delta f$ [mHz], including activation of local frequency containment reserves. It must be understood that for continental Europe, $k$ [MW/mHz] is in practice an administrative, fixed value, agreed upon by the European system operators. All real-time deviations from this value, including non-delivery of frequency containment, will therefore automatically be accounted for in the value of the area control error.
Using actual P and scheduled $P_{set}$ power exchange values [MW] as input, the automatic frequency restoration process is feedback control. Manually activated frequency restoration reserves are additional to automatic activated frequency restoration reserves, and/or to prevent further activation of more expensive resources for the frequency restoration process. Deactivation is either part of the product definition itself, or has to be instructed. Direct activated frequency restoration reserves have a power profile attached and therefore the process is power control. Schedule activated frequency restoration reserves are energy volumes per imbalance settlement period and without a power profile attached and is therefore energy control rather than power control. Prerequisite for functionality in the frequency restoration process is that the imbalance settlement period, nor the activation time, exceeds the time to restore frequency.

Frequency restoration has a clear replacement function as well. It is to deactivate locally the globally activated frequency containment reserves by restoring the frequency to its set value $f_0$ [Hz]. The system operator of the area from which the frequency disturbance originated has to do this by activation of its frequency restoration reserves in its own control block. For systems with only frequency as control target, like Scandinavia and the British Isles, frequency restoration and frequency containment are supplementary.

The start of the reserve replacement process is not a clearly defined actual event, but an expectation (by the system operator) that frequency restoration reserves will be or remain activated in some future time frame, in a specific direction and extent, and need to be deactivated. The reserve replacement is feed forward control. It is therefore a proactive process, to take effect after the conclusion of the frequency restoration process, i.e. after the time to restore the frequency restoration control error. And therefore it is energy control rather than power control by the system operator, as it does not affect the frequency restoration control error.

2.5 Power balancing and compliancy requirements of system operators

The frequency responsibility of system operators with a common system frequency is mutual. It involves compliancy to requirements that are written in e.g. the Nordic system operations agreement for Scandinavia, and for the synchronous area continental Europe the policies (ENTSO-E, Policy 1), formerly known as UCTE policies. In future they will be harmonized, standardized and formalized in European network codes. But compliancy requirements were not invented to serve the codification process. Compliancy requirements have long, venerable track records, and contain decades of cumulative engineers experience and know-how. Not surprisingly there are close analogies to safety regulation and requirements in the airspace and airline industry:

* Chain Robbins, Safety Engineering Group Supervisor at Boeing, has put it this way: "One of the most unpleasant things about this business is the day you suddenly realize that many of the safety codes the Air Force and Industry have were generated out of tragedy -- someone killed, someone mangled for life. You might say one of the objectives of the safety movement, which got under way around 1911, is to generate codes from tests, studies of human reactions, statistical data, near misses, everything we can get, to prevent future tragedies from ever happening."*

The present compliancy requirements in Europe seem rather effective to safeguard system security. In terms of the decomposed service model system security is translated more specific in continuity of grid and system services. Willingly violating compliancy requirements almost certainly compromises continuity of services as in Italy, 2003 (UCTE, 2004) or India, 2012. But compliancy by itself does not guarantee system security as witnessed by blackouts in New York in 1977, and in the Californian crisis throughout 2000-2001. System and grid operators know, or ought to know, that compliancy in the end only warrants limited protection against black-outs. One might hope that regulatory authorities and governments share this wisdom.

Compliancy targets to the obligatory frequency containment and frequency restoration processes, aimed at delivering the system service stable frequency come in two categories:

- performance requirements (real-time energy delivery service; "doing"):
  - physical,
  - transactional (energy delivery settlements),
- capability requirements (security insurance service; "having").

Both performance and capability requirements have temporal components. For performance the time requirements (ENTSO-E, 2013, NC LFC-R) are both the time to start the frequency containment respectively frequency restoration processes, and to complete either process. Being feedback control processes a control target is not just $\Delta f = 0 \text{ [mHz]}$, or $P_{err} = 0 \text{ [MW]}$, but secondary performance indicators are derived from the distributions of $\Delta f$, respectively $P_{err}$. Compliancy in performance can be established a posteriori. The capability requirement on capacity requirements of the system operator can be satisfied a priori, e.g. a year ahead or a day ahead, or at the balancing gate closure time. Or it might have to be proven a posteriori. Unfortunately this question is not explicitly addressed in any framework guideline. Equally missing is a clearly appointed entity responsible for enforcing compliancy and ensuring the necessary remuneration.

In case of frequency containment, the performance criteria are based on the frequency itself. Those are a joint responsibility of the synchronously connected system operators, and not of an individual user or system operator. No element of the frequency containment process, neither the joint dimensioning nor its distribution (ENTSO-E, 2013, NC LFC-R), nor its actual performance, can be attributed to the behaviour of single or joint balance responsible parties or users, unless by contractual arrangement between the system operator and frequency containment reserve provider. For frequency restoration the performance requirements obviously bear relation to the actual or potential power imbalances. These are intrinsically not the responsibility of a
balance responsible party or of individual users, as their responsibility relates to energy imbalances. They are the responsibility of the individual system operator.

The size of the capability requirement for the frequency restoration process at each load frequency block level is related to (ENTSO-E, 2013, NC LFC-R):

- "the distribution of consecutive system imbalance values (i.e. joint balance responsible parties imbalances per imbalance settlement period), with a resolution of not more than the time to restore frequency, and
- the largest imbalance that may result from an instantaneous change of active power of a single power generating module, single demand facility, and single HVDC interconnector or from a tripping of an AC-Line within the Load Frequency Control Block, the dimensioning incident, both positive and negative to be considered".

It is for compliancy to these dimensioning rules, the largest result is to be used, that frequency restoration reserves are procured by the system operator. This is a capability requirement. In addition system operators are required to "have" permanently:

- communication and information systems to communicate with:
  - system users,
  - connected users,
  - other operators:
    - grid operators
    - other transmission system operators,
  - common inter transmission system operator functions
- SCADA functionality to perform real-time control of own assets,
- emergency procedures (ENTSO-E, 2014, NC OS)
- black start capability.

All capabilities and other "have" requirements are permanent, and are not related to the actual joint or individual imbalances of all individual balance responsible parties per single imbalance settlement period. Only the performance requirements bear relation to the joint imbalances of all individual balance responsible parties per imbalance settlement period. In addition the performance requirements of the system operator include the actual power fluctuations within and between each consecutive imbalance settlement period. Some energy imbalances are caused global frequency deviations, or by ramping procedures in scheduled power exchange over interconnectors. These cannot be attributed to balance responsible parties nor to individual users. Nevertheless these deviations result in actual power flows and energy exchanges among system operators that must be accounted for, including financial settlement to avoid free riding behaviour of system operators.

2.6 Control imperfections and associated risks

The mapping shown in Figure 2-3 of electrical variables to domains leads to a logical assignment of control responsibilities to actors. Their control performance indicators are shown in Figure 2-7 for respectively:
- energy: the imbalance \([\text{MWh}]\): trade position – allocation + adjustment, attributable to the balance responsible party for each discrete imbalance settlement period,
- power: the frequency restoration control error \(P_{\text{err}}\) [MW], attributable to the system operator at each point in time,
- frequency: the frequency deviation \(\Delta f\) [mHz] from standard \(f_0\), attributable to the joint, synchronously interconnected system operators, at each point in time.

**Figure 2-7 Control performance indicators**

The distribution of responsibilities among actors has a profound impact on the consequences of imperfections, or failure for each to meet its responsibilities as shown in Figure 2-8. These risks serve as basis for some of the performance criteria proposed ENTSO-E (2013, NC LFC-R). Users, by their balance responsible parties will aim at least at financial optimization taking into account the consequences of their imbalance costs, affecting the efficiency criteria of Abassy & Hakvoort (2009). System operators aim at compliancy to their requirements, affecting the operational security criteria of Abassy & Hakvoort (2009).

The effects of failure differ considerably. Failure by grid users affects the individual user, and may jeopardize the operator side. Failure at the operator side may affect its users collectively and may affect other operators. A critical aspect of market design therefore is what failures are anticipated, both physically and financially, how they affect users and operators, and how they are dealt with by means of contingency measures. A commonly used and fairly robust requirement is based on an \(n-1\) security criterion. Violation of such requirement violates compliancy, but does not have direct physical consequences.
A global power imbalance results in an increasing frequency deviation. When a frequency deviation is not contained in due time the power system may eventually collapse, due to disconnection of generation and tripping of lines by overloading, a chain reaction. These actions provide self-protection to the generators and lines. Although not operating as intended, they will not be damaged. System collapse through operator failure is not taken lightly by regulatory authorities or governments, as it affects the electorate in a rather visible way: a black out. Operator failure obviously jeopardizes the operator’s license to operate.

In our conceptual model where operators and users interact by exchange of services, a black-out is an obvious infringement of grid access rights for all affected grid users. This differs from the former central dispatch model in the Netherlands described in the former Electricity Act (Ministry Of Economic Affairs Of The Netherlands, 1989). There it would represent an infringement of the consumers’ right to be supplied. Contingency planning for not normal operating conditions, including black-outs, is therefore an integral part of the operator obligations. But planning and using rolling black outs as preventive means to achieve compliancy, like effectively occurring in Italy June 2003 or considered (but not occurring) in Belgium, for the 2014-2015 winter, i.e. sacrificing access rights of some users to safeguard those of others is also a political risk. A core question of such contingency planning is to what extent infringement of grid access rights should lead to, or necessitates, suspension of system access rights and obligations (ENTSO-E, 2014, NC ER).

Contingency measures to prevent or minimize black-outs or loss of load are seldom taken into account in social cost benefit analyses (de Nooij, 2012). Apart from inflicting damage and cost upon society, black outs, or power cuts, cause political damage. And so, in modern civilization, black outs can and will be used to enforce policy changes upon non-friendly regimes: with graphite or blackout bombs as in Serbia in 1999,
attacks on power lines or gas boycotts. And so, in modern war and conflict, the next casualty, after truth, can easily be frequency stability or the access rights of power system users.

2.7 Conclusions and discussion on a conceptual framework for balancing

The conceptual framework presented here allows mapping the responsibilities of different actors with respect to the physical components of a power system. It describes and maps the interaction between operators and users as the exchange of services. In this thesis I concentrate on the responsibilities of users and system operator in maintaining and restoring the active power and energy balance in a power system.

There is a non-trivial difference between the balancing responsibilities of the system operator and that of balance responsible parties, representing the users:

- the responsibility of the system operator is an obligation to comply to power and frequency control requirements and refer to power [MW],
- the responsibility of the balance responsible parties is a financial obligation to settle its energy imbalance [MWh].

Different responsibilities lead to different consequences. For the system operator they are physical and relate to power balancing processes. Non compliancy may affect the power balance and ultimately jeopardize system security. As so often with public service obligations probably the easiest or cheapest way for the public service supplier to comply with its public service obligation is to restrict usage or quality of the service. However, reduced quality of service increases the risk of failure and thus may violate (firm) access rights of users and may lead to subsequent damage to users, to society (Rathenau Instituut, 1994) and to other operators. System operators therefore require access to power balancing services. Balancing services can be provided commercially by balancing service providers, representing grid users and resulting in operational expenditure by the system operator. Or they can be provided mandatory at individual users' expense.

For balance responsible parties the consequences are financial. A crucial design factor then is to whom balance responsibility applies. This has been decided on European level as a non-exemption rule (ACER, 2012, FWGL NC EB). All dispatch from each connection has to be assigned to a balance responsible party and no connection or any system user is to be exempted from balance responsibility. Otherwise the system operator will have to accommodate the imbalance risk from the exempted connections, which incidentally is still the case in some European countries. The ability of users to manage their imbalance risks (and opportunities) is determined by their freedom to adapt their notified positions, by their freedom of dispatch, and by economic signals to dispatch. Limitation by design of either may act as an exclusion mechanism to some users to control their imbalance risk. It would deny them opportunities to compensate for uncertainties in forecasted load, intermittent generation, or sudden unavailability of conventional generation, whether at their connection or their portfolio level, or at system level.
3. Balancing variables and performance indicators

In the previous Chapter the responsibilities and qualitative relations between users and operators have been established. This Chapter will explore quantifiable aspects, taking into account asymmetries. One asymmetry is that a continuous global power balance at a stable, predefined frequency automatically results in a global energy balance. But a global energy balance does not imply a continuous global power balance. Nor does a global energy balance, in an electricity market where balance responsible parties trade electrical energy amongst themselves, indicate that each balanced responsible party is without imbalance.

Directionality
Unlike potatoes that have a clear and uniform directionality, from the potato field to the consumer's mouth, and most commonly are regarded as positive entities, the directionality and signs in electricity balancing require more attention. The components of the open loop area control error all have a magnitude and a direction. This direction may be explicitly given in terms as upward or downward, import or export, shortage or surplus, or it may be implicit, and then may require specification by a sign. For terms like import or export a reference point is required. As a general rule the perspective is that from the actor to whom a variable is attributed. For the (open loop) area control error this is the system operator, for the imbalance this is the balance responsible party. These rather innocent statements are not trivial. An additional complicating factor is the existence of negative prices, that in energy settlement between actors may reverse the transfer of money with respect to the transfer of energy. Therefore directionality and signs should be defined and implemented carefully in European legislation, transparency platforms, and reporting. To avoid ambiguities arising from implicit directionality, variables are absolutized in formulae where deemed appropriate.

Another asymmetry arises due to temporal resolution. Power imbalances are continuous real-time variables that in practice are measured over short time cycles, usually in the seconds range. Power variables can be integrated over any time period into energy variables. The imbalance components are energy variables over a given time period, a reverse transformation into a unique set of power variables is not possible. Logically the energy imbalance of a balance responsible party over a discrete imbalance settlement period cannot be resolved into power imbalances for each point in time. Therefore balance responsible parties are not held responsible for the power they inject or withdraw from the grid. Yet, in this Chapter I will show that energy variables can be used to assess the effectivity and the efficiency of power balancing.

Power variables
In the synchronous area continental Europe the area control error (ACE or $P_{err}$) of a system operator consists of the power error and the frequency error: $\Delta P + k\Delta f$ [MW] (ENTSO-E, 2013, NC LFC-R). A non-zero area control error of a system operator, as shown in Figure 3-1 results in a power flow that is not directed towards any specific other system operator, but towards the collectivity of system operators in continental Europe. It is the responsibility of the system operator to control a non-zero area control error back to zero, within the time to restore frequency (ENTSO-E, 2013, NC LFC-R). Therefore the system operator must operate the frequency restoration process and may operate a reserve replacement process. These processes result in set-point values $S_{FRR, RR}$ for activation by the system operator of balancing power. The open loop area control error $ACE_{OL}$ then is the value of the area control error in the absence of any control processes, and thus the hypothetical joint power error of the users (Havel et al., 2008).

40
The open loop area (block) control error $\text{ACE}_{OL}$ is defined (ENTSO-E, 2013, NC LFC-R) as the difference between the frequency restoration control error of this control block or ACE and the:

- frequency restoration reserve activation within this control block $S_{(FRR)}$,
- replacement reserve activation within this control block $S_{(RR)}$,

and the exchange (or interchange) of this control block with other control blocks of the:

- imbalance netting power exchange,
- frequency restoration power interchange,
- replacement power interchange.

For system operators that do not perform the reserve replacement process, $S_{(FRR, RR)}$ equals $S_{(FRR)}$. In this Section the exchange components are not considered. These choices do not affect the results or reasoning. In Section 8.3 the imbalance netting process and its impact on performance and efficiency of the system operator is shown.

**Energy variables**

As the open loop area control error $\text{ACE}_{OL}$ [MW] represents the hypothetical joint power error of the users, then $\int \text{ACE}_{OL}(t)dt$ [MWh] per imbalance settlement period resembles the net joint energy error of those users. And that energy error corresponds to the net joint imbalance of all balance responsible parties. Prerequisite is that all injections and withdrawals, including grid-losses, are covered by balance responsibility.

Each balance responsible party will have for each imbalance settlement period an imbalance that is the net sum of the following variables in [MWh]:

- the trade position that reflects the net energy traded by a balance responsible party with other balance responsible parties,
- the allocation that reflects the net energy injected or withdrawn, aggregated over all connections that the balance responsible party has assumed balance responsibility for, and
- the adjustment that reflects the net balancing energy $\int S_{(FRR, RR)}(t)dt$ traded with the system operator, from all balancing service providers that have assigned this balance responsible party to accept such adjustment.

\[
\text{ACE}_{OL} = \text{ACE} - S_{(FRR, RR)}
\]

\[
\text{ACE} = \Delta P + k\Delta f
\]
Imbalances are netted energy volumes per balance responsible party, per imbalance settlement period. For a balance responsible party each imbalance ≠ 0 [MWh] reflects either a surplus $Q_{\text{surplus}}$ or a shortage $Q_{\text{shortage}}$ [MWh] to be settled financially with its system operator. Calculating imbalances over multiple or longer imbalance periods will reduce or more formally not increase the imbalance per balance responsible party through cancelling out of successive $Q_{\text{surplus}}$ and $Q_{\text{shortage}}$ (Frunt, 2011). Likewise, the net joint imbalance over all imbalance responsible parties $\sum(Q_{\text{surplus}}, Q_{\text{shortage}})$ will not increase either over multiple or longer imbalance settlement periods.

A prerequisite to equate $\int \Delta \text{OL}(t)dt$ to $\sum(Q_{\text{surplus}}, Q_{\text{shortage}})$ per imbalance settlement period is that all activated balancing energy $\int S_{\text{FRR, RR}}(t)dt$ [MWh] is adjusted to the appropriate balance responsible parties. Energy from activated frequency containment reserves, as part of $\int k\Delta(t)dt$ [MWh] is in general not settled nor adjusted on the imbalance of any balance responsible party (ENTSO-E, 2014, WG AS Enquiry). Notable exceptions for synchronous area continental Europe are France and Poland. Actual net energy volumes per imbalance settlement period from activated frequency containment reserves ought to be small however because:

- in general as the frequency ought not to deviate from its nominal value, and
- the volume of frequency containment reserves is relatively small, presently 3000 MW in the entire synchronous zone continental Europe (ENTSO-E, 2012, AHTor), < 1% of the peak load.

A second source of deviation is that whereas market trades over interconnectors are transactional firm in energy volume, system operators apply ramping procedures to manage power control in the frequency restoration process. These result in energy transactions between system operators, and do not affect balance responsible parties in their imbalances. Deviations from scheduled flow over HVDC links to other synchronous zones do affect the power and energy equilibrium in each synchronous zone but may not be attributable to any balance responsible party. A trip of an HVDC interconnector executing a trade flow between synchronous areas will immediately affect the frequency and power balance at both ends of the link. The resulting energy imbalances are in principle not attributable to any balance responsible party.

To assess and compare the operations of users in energy balancing with the system operator in active power balancing, the requirements and the actions of the system operator require integration of power variables into energy as shown in Table 1. This allows incorporation of energy from the schedule activated manual reserves for frequency restoration, or from reserve replacement (ENTSO-E, 2012, AHTor) into $\int S_{\text{FRR, RR}}(t)dt$. The integration period or schedule length then must conform to the imbalance settlement period.

Absolute net energy values enable the construction of performance indicators for single and multiple imbalance settlement periods, and allow for decomposition of the results of the frequency restoration process. Mean absolute values as an indicator of dispersion are more robust to outliers than conventional standard deviations. Besides, absolute net variables can be added, unlike standard deviations, and therefore are more practical to assign value to.
At all times $t$ the area control error target is $\text{ACE}(t) = 0$ [MW], meaning that over any time period $\int \text{ACE}(t) \, dt \to 0$ [MWh], and that over multiple periods $\sum \int \text{ACE}(t) \, dt \to 0$ [MWh]. An appropriate control performance indicator then is $\sum |\int \text{ACE}(t) \, dt|$. It can be decomposed in:

- $|\sum \int \text{ACE}(t) \, dt|$: the absolute net systematic component of $\sum |\int \text{ACE}(t) \, dt|$
- $\sum |\int \text{ACE}(t) \, dt| - |\sum \int \text{ACE}(t) \, dt|$: the random component of $\sum |\int \text{ACE}(t) \, dt|$

A system operator may activate balancing power in either directions $S_{\text{FRR, RR} \uparrow}(t)$ and $S_{\text{FRR, RR} \downarrow}(t)$ [MW]. When this occurs within the same imbalance settlement period it results in balancing energy in both directions. Counter-activation may occur sequentially or simultaneously. Simultaneous counter-activation may be the result of legacy effects: automatic, or of direct activated frequency restoration reserves that cannot be deactivated immediately, or of schedule activated reserves that, once activated, cannot be annulled. Sequential counter-activations may occur due to a change in the direction of the area control error. This will happen more often over longer periods. The probability of sequential counter-activation within an imbalance settlement period will increase with increasing imbalance settlement period length. This means that the volume of counter-activation, $0.5 \times \left( |S_{\text{FRR, RR} \uparrow}(t)| + |S_{\text{FRR, RR} \downarrow}(t)| \right) - \left( |S_{\text{FRR, RR} \uparrow}(t)| - |S_{\text{FRR, RR} \downarrow}(t)| \right) [\text{MWh}]$, will increase with longer imbalance settlement periods (Beune & Nobel. 2001; Frunt, 2011).

### Table 1 Power and energy variables in balancing process

<table>
<thead>
<tr>
<th>Power [MW]</th>
<th>Net Energy/ISP [MWh] $t = \text{start, end ISP}$</th>
<th>Absolute Energy/ISP [MWh] $t = \text{start, end ISP}$</th>
<th>$\sum$ Absolute Energy [MWh] $t = \text{start, end ISP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ACE}(t)$</td>
<td>$\int \text{ACE}(t) , dt$</td>
<td>$\int \text{ACE}(t) , dt$</td>
<td>$\sum \int \text{ACE}(t) , dt$</td>
</tr>
<tr>
<td>$S_{\text{FRR, RR} \uparrow}(t)$</td>
<td>$\int S_{\text{FRR, RR} \uparrow}(t) , dt$</td>
<td>$\int S_{\text{FRR, RR} \uparrow}(t) , dt$</td>
<td>$\sum \int S_{\text{FRR, RR} \uparrow}(t) , dt$</td>
</tr>
<tr>
<td>$S_{\text{FRR, RR} \downarrow}(t)$</td>
<td>$\int S_{\text{FRR, RR} \downarrow}(t) , dt$</td>
<td>$\int S_{\text{FRR, RR} \downarrow}(t) , dt$</td>
<td>$\sum \int S_{\text{FRR, RR} \downarrow}(t) , dt$</td>
</tr>
<tr>
<td>$\text{ACE}_{\text{OL}}(t)$</td>
<td>$\int \text{ACE}_{\text{OL}}(t) , dt$</td>
<td>$\int \text{ACE}_{\text{OL}}(t) , dt$</td>
<td>$\sum \int \text{ACE}_{\text{OL}}(t) , dt$</td>
</tr>
</tbody>
</table>

### Sign conventions, $\text{ACE}_{\text{OL}}$ and proxies for $\text{ACE}_{\text{OL}}$

The following sign conventions are used:

- $\text{ACE} < 0$ [MW]: flow towards the system operator, import, system operator shortage
- $\int S_{\text{FRR, RR} \uparrow}(t) \, dt < 0$ [MWh/ISP]: absorption by the system operator due to activation of balancing energy, downward regulation
- $\sum Q$ [MWh/ISP] < 0: Net market imbalance, market shortage

Under perfect frequency restoration $\sum Q + \int S_{\text{FRR, RR} \uparrow}(t) \, dt = \int \text{ACE}(t) \, dt = 0$ [MWh/ISP]

With no exemption from balance responsibility $\int S_{\text{FRR, RR} \downarrow}(t) \, dt = 0$ [MWh] $\sum Q - \int \text{ACE}(t) \, dt = 0$ [MWh/ISP]

Therefore: $\int \text{ACE}_{\text{OL}}(t) \, dt = \int \text{ACE}(t) \, dt - \int S_{\text{FRR, RR} \downarrow}(t) \, dt = \sum Q$ [MWh/ISP]
The net volume of balancing energy per imbalance settlement period ($\int S_{FRR, RR\uparrow}(t)dt + \int S_{FRR, RR\downarrow}(t)dt$) is together with the net joint imbalance $\sum Q$ proxy for $\int ACE_{OL}(t)dt$. This assumes no exemptions from balance responsibility, and fast and accurate frequency restoration. Following the same logic as for individual or net joint imbalance, the values of $|\int ACE(t)dt|$, $|\int ACE_{OL}(t)dt|$ and $|\int S_{FRR, RR}(t)dt|$ over longer imbalance settlement periods will all decrease, thus reducing insight in performance of power control by the system operator. But for very short imbalance settlement periods, i.e. shorter than the time to restore frequency, correspondence may decrease as well due to:

- the reactive and unaccounted for activation of frequency containment reserves,
- the reactive nature of frequency restoration; no immediacy is possible, nor is it required (ENTSO-E, 2013, NC LFC-R),
- the inexact nature of frequency restoration: limited ramp rate of control power, control inaccuracies,
- the potential legacy of automatic or direct activated, i.e. non-scheduled frequency restoration reserves], and
- ramping restrictions or ramp rate limitations (ENTSO-E, 2013, NC LFC-R) for cross zonal commodity exchange.

All these causes result in, and are attributable to exchanges between system operators. The area control error $|\int ACE(t)dt|$ [MWh] results in an unintended exchange of energy, and therefore is the object to be minimized by the responsible system operator through the frequency restoration process. Obviously the real-time power requirements formulated in the frequency restoration process have to be honoured. The frequency restoration error has to be controlled back within time to restore frequency, and no overshoots are to be induced on purpose to reduce an energy imbalance.

Finally a remark has to be made on the meaning and usage of “set-point value" ($S_{FRR, RR}$ [MW]) in the process scheme shown in Figure 2-6. This represents signals originating from the system operator to the balancing service providers. The requests for activation of frequency restoration reserve power are supposed to be honored by them. But actual power delivery will not coincide exactly with the set-point values requested, neither in time nor in volume. This is due to the delay time allowed for the frequency containment and frequency restoration processes, and to the limited number and ramping capabilities of balancing power sources. These limitations are also apparent in the requested set-point values themselves. Their potential rate of change is slower than that of the frequency restoration control error. This feed-back control input is "noisy" compared to the frequency restoration process.

It is impossible to make an unconditional distinction between the active power dispatched to match the energy position, on behalf of balance responsibility, or the dispatch of balancing power $S_{FRR, RR}$ on request of
the system operator, on behalf of balancing service provision. There is no potential arbitrage between balancing energy, and imbalance, including adjustment, if all have the same value. In that case it makes sense to use the requested power volumes $\int S_{FRR,RR}(t) dt$ [MWh] as basis for the calculation of the balancing energy to be settled with the balancing service provider, and to be adjusted to the imbalance of the imbalance responsible party. This accelerates and increases robustness of settlement processes by removing complexity and arbitrage potential. Measurements of activated frequency restoration reserves, be it automatic or direct activated can and probably can be used for quality assessment, but need not be used for volume determination.

3.1 Balancing process performance indicators
In a frequency restoration process with input $\int ACE_{OL}(t) dt$ and output $\int ACE(t) dt$, two trivial performance lines can be distinguished as elucidated in Figure 3-2:

- perfect control: $\int ACE(t) dt = 0$, for all $\int ACE_{OL}(t) dt$: $\int ACE_{OL}(t) dt \leq \int S_{FRR,RR}(t) dt$
- no control: $\int ACE(t) dt = \int ACE_{OL}(t) dt$, equivalent to $\int S_{FRR,RR}(t) dt = 0$

![Figure 3-2 A control performance diagram](image)

The no control line and its mirror, define counterproductive areas in Figure 3-3: $|\int ACE(t) dt| > |\int ACE_{OL}(t) dt|$:
Figure 3-3 Counterproductive areas identified (red) in the control performance diagram

An indicator for effectivity, measuring the remaining distance to $\int \text{ACE}(t)dt = 0$ [MWh] is:

$$0.5 \times (\sum |\int \text{ACE}_{OL}(t)dt - \int \text{ACE}(t)dt| - |\sum |\int \text{ACE}_{OL}(t)dt| - |\int \text{ACE}(t)dt|)|$$ [MWh].

Since $\int \text{ACE}_{OL}(t)dt - \int \text{ACE}(t)dt = \int S_{(FRR, RR)}(t)dt$, and $\int \text{ACE}_{OL}(t)dt$ can be equated to $\sum (Q_{\text{surplus}}, Q_{\text{shortage}})$, this indicator can be rewritten as:

$$0.5 \times (\sum |S_{(FRR, RR)}(t)dt| - |\sum |\sum (Q_{\text{surplus}}, Q_{\text{shortage}})| - |\sum |\int \text{ACE}(t)dt||)$$ [MWh].

In case $\sum |\int \text{ACE}_{OL}(t)dt| < \sum |\int \text{ACE}(t)dt||$, indicative of overshooting the target this extends to:

$$0.5 \times (\sum |S_{(FRR, RR)}(t)dt| - |\sum |\sum (Q_{\text{surplus}}, Q_{\text{shortage}})| - |\sum |\int \text{ACE}(t)dt|| + |(\sum |\sum (Q_{\text{surplus}}, Q_{\text{shortage}})| - |\sum |\int \text{ACE}(t)dt||$$ [MWh].

$\sum |\int \text{ACE}_{OL}(t)dt|$, or $\sum |\sum (Q_{\text{surplus}}, Q_{\text{shortage}})|$ is the required energy for perfect (energy) control.

This means that $|\sum |\sum (Q_{\text{surplus}}, Q_{\text{shortage}})| - |\sum |\int \text{ACE}(t)dt||$ is the unavoidable component of $\sum |\int \text{ACE}(t)dt||$, due to imperfect control. The counter-activation volume is:

$$\sum 0.5 \times ((|S_{(FRR, RR)}(t)up(t)dt| + |S_{(FRR, RR)}(t)down(t)dt|| - |(|S_{(FRR, RR)}(t)up(t)dt| - |S_{(FRR, RR)}(t)down(t)dt||)$$ [MWh].

**Correlations of ACE and open loop ACE_{OL}**

Figure 3-4 shows on the horizontal axis the $\int \text{ACE}_{OL}(t)dt$ distribution, a user result, and on the vertical axis the $\int \text{ACE}(t)dt$ distribution as indicator of performance quality requirement of the system operator. The graph may act as a balancing quality performance indicator (ENTSO-E, 2012, AHTor; Van der Veen, 2012).
Figure 3-4 An idealized performance diagram with shading according to expected ISP values density

It shows an $\int \text{ACE}(t) \, dt$ reduced with respect to $\int \text{ACE}_{OL}(t) \, dt$. The vertical distances from the no control line show the efficacy of the frequency restoration process; the vertical dispersion of the data points show the accuracy of the frequency restoration process. An asymmetric distribution of $\int \text{ACE}_{OL}(t) \, dt$ (biased market error) or a positive autocorrelation of $\int \text{ACE}_{OL}(t) \, dt$ at lag 2 (persistence of market error), are indicative of energy control by the system operator rather than only power control. It therefore might indicate absence or insufficiency of incentives or information to balance responsible parties to support system balance. Full reduction of $\int \text{ACE}_{OL}(t) \, dt$ to an $\int \text{ACE}(t) \, dt$ of 0 [MWh], the perfect control line in Figure 3-3 points at a very fast, perfect, frequency restoration process performed by the system operator. Data points parallel to the no control line point to saturation of available frequency restoration reserves (ENTSO-E, 2012, AHTor).

**Autocorrelations of ACE and open loop ACE<sub>OL</sub>**

In Europe the time to restore frequency is 15 minutes. This means that calculated over discrete 15 minute periods, $\int \text{ACE}(t) \, dt$ over ISP<sub>i-1</sub> may be positively correlated to $\int \text{ACE}(t) \, dt$ over ISP<sub>i</sub>, but certainly should be uncorrelated to $\int \text{ACE}(t) \, dt$ over ISP<sub>i-2</sub>, as shown in Figure 3-5. Persistence of $\int \text{ACE}(t) \, dt$ outside target range values (ENTSO-E, 2013, NC-LFC-R) over subsequent multiple discrete 15 minute periods is evidence of non-compliance to one of the performance criteria for the frequency restoration process and therefore should be avoided.
The expected auto correlation of 15 minute ISP ACE values, lag 2, shading according to expected ISP values density.

The reason is that a frequency disturbance may occur anywhere in any imbalance settlement period, and the frequency restoration is required to be concluded in the next one but prior to the one after the next. Shorter imbalance settlement periods, say 5 or 10 minutes (not practiced in Europe) would mean that each frequency restoration process could continue over several consecutive periods.

For longer imbalance settlement periods, 1 hour or 30 minutes, the time to restore frequency of 15 minutes means that the system operator has to apply a reserve replacement process, essentially energy control. To avoid counter-activation of already activated reserves by the system operator in such case balance responsible parties cannot be incentivized to support system balance themselves. The efficacy of market incentives is indicated by the decay of the autocorrelation of \( \int ACE(t) dt \), ISP\(_{i-2}\) over successive lags. Slow decay, or convergence of the autocorrelation coefficient \( \to 0 \) over increasing lags as shown in Figure 3-6, indicates the presence of a “slow component” \( \Delta E_O \) as part of \( \int ACE_{OL}(t) dt \) as proposed by Havel et al. (2007). It points to little or ineffective incentives to the balance responsible parties to return their net, joint imbalance to 0 [MWh]. This is of course all right if \( \int ACE_{OL}(t) dt \) is already close to 0 [MWh]. The length of the imbalance settlement period is critical. Imbalance settlement periods like 1 hour, e.g. used by Havel et al. (2007), require more complex decomposition.
Figure 3-6 The expected auto correlation of 15 minute ISP ACEOL values, lag 2, with no market incentives, shading according to expected ISP values density.

Persistent and large $\int \text{ACE}_{OL}(t) \, dt$, ISP(i-2) [MWh/ISP] values, or slow decay of autocorrelations for $\int \text{ACE}_{OL}(t) \, dt$ lag 2 for imbalance settlement periods equal or longer than the time to restore frequency are indicative of a system operation exercising energy control, even if $\sum \int \text{ACE}_{OL}(t) \, dt \to 0$ [MWh]. In such a case the system operator de facto takes over responsibility for energy control from the balance responsibilities, regardless of a reserve replacement process being implemented or not.

3.2 Conclusions and discussion on balancing variables and performance indicators

With only a few variables the balancing efforts required from and performed by the system operator can be compared to its actual balancing results, at least in energy terms. Necessary are its public time series, per imbalance settlement period, per direction of $\int \text{ACE}(t) \, dt$, of $\int \text{FRR,RN}(t) \, dt$, and of $\sum Q$, the net imbalance of balance responsible parties. This allows the calculation of the required net absolute energy parameters per imbalance settlement period.

Another requirement to make the efforts and results of the system operator comparable to the efforts and results of the users is that there are no connections exempted from balance responsibility. Otherwise, there will be no correspondence between the net joint imbalance of the balance responsible parties and the net volume of balancing energy.

The imbalance settlement period length as design parameter is decisive for other design choices. With increasing imbalance settlement period lengths:

- individual and joint imbalance volumes of balance responsible parties will decrease/will not increase,
- absolute balancing energy volumes will increase,
- net balancing energy volumes will decrease/will not increase,
- the difference between net balancing energy volume and net imbalance volumes will decrease.
A second important conclusion is that with imbalance settlement periods of 30 minutes or more a reserve replacement process by the system operator is required. Indeed all systems with an imbalance settlement period > 15 minutes seem to have such a reserve replacement process (ENTSO-E, 2013, WG AS enquiry). Operating a reserve replacement process requires either the absence of incentives to balance responsible parties to restore system balance, or an explicit restriction to do so. In both cases it restricts the freedom to dispatch by grid users. It also requires more effort by the system operator. The North American Electric Reliability Corporation concludes that backup supply, i.e. to back up operating reserves (reserve replacement) is: *not a service but a purely commercial function*. Their conclusion is given in appendix "USA experience and expertise".
4. Prices related to balancing

In the previous Chapter quantifiable energy volumes related to balancing processes between users and operators have been identified, and their relations and significance has been explained and discussed. European designs distinguish several electrical energy products, with different pricing mechanisms (Glachant & Saguan, 2007). In this Chapter a qualitative analysis will be used to explain and discuss prices for these energy volumes:

- the commodity, traded and settled on the whole sale market between balance responsible parties,
- balancing energy, settled between balancing service providers and their system operator, and
- imbalance, settled between balance responsible parties and their system operator.

These settlements will impact the involved actors financially. In distinguishing imbalance and commodity European market designs differ significantly from USA designs (Glachant & Saguan, 2007). Non-European designs depend on real-time pricing of power. It is a single, time varying price that includes both the commodity and the balancing cost incurred by the system operator. In such designs there are no separate commodity and imbalance volumes, and logically no separate commodity and imbalance prices.

Commodity prices are a market result established prior to delivery per market time unit, synchronized with and overlapping with one or a sequence of imbalance settlement periods. Prices for balancing energy and imbalances are established per imbalance settlement period by the system operator. For balancing energy on a single buyer market, respectively as monopolist for imbalance, in real-time or a posteriori. The relation between market time units and imbalance settlement periods is discussed in Chapter 5. For convenience the analysis will focus on imbalance settlement periods.

Because the balance responsible party can choose in its actual dispatch to substitute delivery of the commodity, or delivery of imbalance, the following prices need to be compared to analyse any potential for arbitrage and perverse incentives for the users:

- between commodity and balancing energy prices, for the balancing service provider,
- between commodity and imbalance prices, for the balance responsible party, and
- between balancing energy and imbalance prices, for the balancing service provider.

The financial result of the settlements with balance responsible parties and balancing service providers will have an impact on the system operator, as will the energy settlements amongst the system operators resulting from their operational responsibilities (ENTSO-E, 2014, NC EB). The financial balance of the system operator is discussed in Chapter 6.
4.1 Commodity price

Commonly featuring in power system analysis is the relation of the cost to the level of power supply. Its classical shape is a monotonically non-decreasing function as shown in Figure 4-1. Both cost, and usually cost gradient as well, do not decrease with increase in power level.

![Figure 4-1 A supply cost function for power](image)

A power supply cost function of a power system is derived from theoretical long or short run marginal costs of all power generating equipment in that power system. They are sorted in ascending order on a merit order list of committed units (de Vries, 2004; Boisseleau, 2004). This representation is simplified and conditional, because a power supply curve is not static. It may rise and fall vertically according to fuel prices. It may stretch or shrink horizontally according to availability of power supply. Another drawback is that on an energy market the prices refer to energy [€/MWh] and not to power [€/MW]. This requires substituting power [MW] by energy [MWh] per time unit. And since in an electricity market demand equals supply, the power supply axis can be substituted by energy demand per time unit, e.g. a year, a day, a market time unit, or an imbalance settlement period as shown in Figure 4-2

![Figure 4-2 A supply cost function for energy demand per imbalance settlement period (ISP)](image)
In competitive energy markets competition will drive down market prices to efficient cost levels, i.e. close to true marginal cost. Furthermore, as long as the time unit is sufficiently short, the supply cost function for energy demand approaches the supply cost function for power due to smaller power fluctuations within shorter time units (Frunt, 2011). Selecting an energy demand for a particular time unit on the supply cost function for that particular time unit will define an operating point OP for the power supply system (Figure 4-3), with its coordinates shown in terms of energy quantity $Q_{\text{OP}}$ [MWh] and energy price $P_{\text{OP}}$ [€/MWh], the lowest price according to the power supply curve.

![Figure 4-3 An operating point (OP) on a supply cost function for energy demand](image)

Over time, i.e. multiple imbalance settlement periods and given identical supply cost functions, the operating point will move along the supply cost function, according to varying $Q_{\text{demand}}$. Accurate knowledge of the power supply curve and of the expected demand is the concept for centralized commitment and scheduling. In that case there are no imbalance settlement periods or market time units required per se. Some temporal resolution nevertheless will be used to organize the subsequent commitment and dispatch of power generating equipment.

In an energy market the need for accurate information on all power supply and power supply costs that are required to establish a supply cost function for all time units, and the need for information on total load required to perform central commitment and scheduling is no longer centralized as a system operator's responsibility. The energy balance, supply equals demand, is left to the users. If each withdrawal is covered by injection and if each injection is covered by withdrawal, than all demand equals all supply.

Although this very simple model neglects all kinds of inaccuracies, imperfections, including congestions, and dynamics, the concept of a supply cost function and an operating point is useful to assess the pricing of balancing energy by balancing service providers. Likewise it still can be used to identify price peaks on real electricity markets (Lange et al., 2002; Brennan, 2003). The imperfections admitted above are in line with the criticism expressed by Boisseleau (2004). He states that power exchanges insufficiently grasp the dynamics of energy markets.
4.2 Commodity and balancing energy prices

In European market designs, the system operator is responsible for the organization of the physical active power balancing, and therefore operates a balancing market for at least the frequency restoration process. On this single buyer market, balancing service providers offer in their balancing energy bids their willingness to inject or withdraw power or energy on request by the system operator. They announce this willingness for specific volumes, at specific imbalance settlement periods, and set a specific price for the energy thus delivered or withdrawn. A balancing bid is thus an *option* to the system operator. The option gives the system operator the right to activate balancing power or energy from the balancing service provider, in a specific direction, up to a certain volume, and according to some pre-set rules adhering to the defined products. These rules, or product definitions can be:

- according to a real-time variable signal for automatic frequency restoration reserves, or
- to a predetermined power profile for direct activated frequency restoration reserves, or
- as an energy product over the imbalance settlement period without a power profile attached for schedule activated frequency restoration reserves.

In contrast to the conceptual power supply curve for the commodity in Section 4.1, that is exclusively injection based, bids of balancing energy reflect *relative* injection or withdrawal. Balancing energy bids are in principle directionally impartial to the kind of resources that are offered.

Another difference concerns the meaning of price. Whereas in the power supply curve model the price relates to the trade of energy from generation to consumption, bid prices for balancing energy reflect transactions with the system operator:

- On the relative injection side the price the balancing service provider wishes to receive from the system operator to inject more, respectively withdraw less energy.
- On the relative withdrawal side bid prices reflect the price the balancing service provider wishes to pay to the system operator to withdraw more, respectively inject less energy. These bid prices may be negative, in which case the balancing service provider receives both energy and money from the system operator.

At some time prior to the imbalance settlement period of delivery, balancing energy bids become firm. This is the balancing energy gate closure time, after which the system operator will sort all balancing energy bids for each imbalance settlement period in increasing order of bid price on a merit order list for relative injection, and in decreasing order for relative withdrawal. Such merit order lists act as supply curves for balancing power and energy, to be activated by the system operator as required. An operating point $Q_{OP}$ from the power supply curve for an imbalance settlement period, as shown in Figure 4-4, acts as a starting point for merit order lists for that imbalance settlement period.
The actual values of coordinate $Q_{OP}$ [MWh] of the system for each imbalance settlement period are neither relevant for the system operator, nor for the individual balancing service provider or balance responsible party. Balance responsible parties and balancing service providers will only consider their own position at imbalance settlement period of delivery, and their controllable means of relative injection or withdrawal at their disposal. For the system operator $Q_{OP}$ is simply the starting point on the merit order list dividing relative injection from relative withdrawal, the actual value of $Q_{OP}$ in [MW] or [MWh/ISP] does not matter.

The total volume of balancing energy bids of a single, or of all balancing service providers together, will generally be less than the total range contained in a supply cost function starting from coordinate $P_{OP}$. Glachant & Saguan (2007) remark that "inter-temporal constraints on generation (cost or speed of starting up, or shutting down, plants; cost or speed of adjusting output) can impede the ability of certain plants to contribute to adjustments in generation for purposes of balancing", and that "Not all technologies are equally able to respond to short-term signals". And time to restore frequency is 15 minutes (ENTSO-E, NC LFC-R, 2013). Cramton (2003) observes that "the supply curves (i.e. merit order lists) become steeper as options vanish closer to real-time", because e.g.:

- for spinning conventional generators a fast reduction or increase in power output will reduce fuel efficiency compared to a steady state output. In stored hydro a similar effect is less pronounced,
- in addition non-convexities, like fixed costs, at the relative injection side due to start-up costs for non-spinnable generation, or rebound costs for load shedding, or for relative withdrawal, stop costs for spinning generation may add mark-ups, or mark-downs to prices on the power supply curve.

The coordinate $P_{OP}$ [€/MWh] on the conceptual power supply function is the price at which the whole system is considered to be functioning and thus serves as the opportunity value of energy at the time of delivery. In an efficient market based design rational balancing service providers will not offer to receive less than the cost function for additional relative injection. Nor will they propose to pay more than the cost function for relative withdrawal (Figure 4-5; red areas). In their bid prices the balancing service providers will take into account (their) perception of this opportunity value, and a margin: \[\text{price } \Delta \geq \text{bid price} - P_{OP} \ [\text{€/MWh}]\] for

---

**Figure 4-4 Balancing merit order list coordinates**

\[\Delta = \text{bid price} - P_{OP} \ [\text{EUR/MWh}]\]

- relative withdrawal
- relative injection

$Q_{OP}$ = starting point
merit order list [MW]
additional injection and price $\Delta \leq \text{bid price} - P_{\text{OP}} \text{[€/MWh]}$ for additional withdrawal or reduced injection. This means that merit order lists of balancing energy bids will occur in the green areas in Figure 4-5.

Inverting the green area for relative withdrawal to $\Delta = P_{\text{OP}} - \text{bid price} \text{[€/MWh]}$ allows to show in Figure 4-6 the price $\Delta$'s contained in the balancing energy bid prices in both directions as a positive number. This is the margin the balancing service provider wants to receive. This operation facilitates the perception of negative prices for commodity or balancing energy. Although commodity and bid prices can be negative, the margin $\Delta$ in the green area in Figure 4-6 is positive for the balancing service provider, regardless of the sign of the underlying parameters. Using price $\Delta$'s allows for comparison between different designs, and assessing their development over longer periods.
The model in Figure 4-6 shows some trivial properties for potentially expected price Δ's on the balancing energy merit order list:

- none are expected to be negative,
- expected convergence to [0 €/MWh] at the starting point,
- an increase away from the starting point, after all the balancing energy bids are sorted per direction, that can be less pronounced on the relative withdrawal side than on the relative injection side.

**Merit order price model**

The balancing energy bids on a merit order list are ordered per direction, up or down, according to their bid price Δ [€/MWh] for each bid Bi [MW]. Therefore the price Δ [€/MWh] at a certain volume B [MW] of balancing energy bids on a merit order list can be considered to be related to that volume. For positive bids for relative injection this relation can be modelled as a monotonically non-decreasing function e.g.:

\[ \text{price } \Delta(B) = B^y + z \text{ [€/MWh]} \]

with y representing a scarcity coefficient, and z as constant mark-up. For negative bids the model function would be:

\[ \text{price } \Delta(B) = (-B)^y + z \text{ [€/MWh]} \]

The following analysis will be based on positive balancing energy bids. For each bid with size Bi [MW] the maximum volume of balancing energy [MWh] that can be activated per imbalance settlement period ISP [h] is Bi * ISP [MWh]. So the relation between the actually activated volume of balancing energy \( \int S_{FRR, RR}(t)dt \) [MWh] and the volume of all bids activated in that direction can be expressed by a utilization factor \( u = \frac{\int S_{FRR, RR}(t)dt}{\sum (B_i \times \text{ISP})} \) [-]. Thus \( 0 < u \leq 1 \). For negative bids a similar utilization factor \( 0 < u \leq 1 \) results.

The utilization factor may differ per activated bid Bi, and from one imbalance settlement period to the next. The utilization factor will in general be \( u < 1 \) for automatic and direct activated manual frequency restoration bids, due to partial or over time varying activation. Full saturation of all activated bids on a merit order list in one direction, or \( u = 1 \), as shown in Figure 3-4 is in the absence of counter-activation actually a clear indicator of imperfect power balancing by the system operator.

The mark-up z is the composite result of the bidding strategies of the balancing service providers. These can be influenced by reducing any risk factors perceived by them. A major factor affecting risk perception of balancing service providers is the settlement scheme for balancing energy:

- pay marginal, or
- pay-as-bid.

The bid price function per imbalance settlement period \( \Delta(B) = B^y + z \text{ [€/MWh]} \) results in a balancing energy price Δ function \( (\int S_{FRR, RR}(t)dt/u)^y + z \text{ [€/MWh]} \). This can rewritten for each imbalance settlement period, and per direction as \( 1/u^y \times (\int S_{FRR, RR}(t)dt)^y + z \text{ [€/MWh]} \). Under pay marginal the revenue for the balancing service providers is per imbalance settlement period the rectangular area defined by the balancing energy volume \( \int S_{FRR, RR}(t)dt \) [MWh] and the price of the most expensive bid that has been activated. This results in a balancing energy revenue function:\n
\[ \frac{1}{u^y} \times (\int S_{FRR, RR}(t)dt)^{y+1} + z \times \int S_{FRR, RR}(t)dt \text{ [€/ISP]} \]
Under pay-as-bid the revenue function is the anti-derivative of this conceptual balancing energy price function: \( \frac{1}{u^y} \cdot \frac{1}{y+1} \cdot (\int S_{(FRR, RR)}(t) dt)^{y+1} + z \cdot \int S_{(FRR, RR)}(t) dt \) [€/ISP], disregarding the constant that appears in the anti-derivative.

The difference between these revenue functions is the producer surplus:

\[
\frac{1}{u^y} \cdot (\int S_{(FRR, RR)}(t) dt)^{y+1} + z \cdot \int S_{(FRR, RR)}(t) dt - (\frac{1}{u^y} \cdot \frac{1}{y+1} \cdot (\int S_{(FRR, RR)}(t) dt)^{y+1} + z \cdot \int S_{(FRR, RR)}(t) dt) = (1 - \frac{1}{y+1}) \cdot \frac{1}{u^y} \cdot (\int S_{(FRR, RR)}(t) dt)^{y+1} [€/ISP].
\]

This can be simplified and rewritten as: \( \frac{y}{y+1} \cdot \frac{1}{u^y} \cdot (\int S_{(FRR, RR)}(t) dt)^{y+1} [€/ISP]. \)

Under pay-as-bid this producer surplus is retained by the system operator, and therefore lost to the balancing service providers. Pay-as-bid introduces an additional risk to balancing service providers, that of loss of producer surplus. Rational balancing service providers will cut their losses by collectively reducing scarcity coefficient \( y \) as a result of their bids and bidding strategy, compared to bidding under marginal settlement. By Increasing relatively low mark-up \( \Delta \)'s in their individual bid prices, balancing service providers accomplish a lower scarcity coefficient \( y \) and compensate for lower revenues as a consequence of this lower scarcity coefficient. Therefore the mark-up \( z \) will be larger under pay-as-bid than under pay marginal. This bidding strategy is known as "guess the strike price" (Stoft, 2002, Tierney et al., 2008), and will be apparent in the price \( \Delta \) model, as illustrated in Figure 4-7:

- a potential discontinuity between the merit order lists in each direction,
- no price \( \Delta \)'s approaching 0 [€/MWh].

![Figure 4-7 Effect pay-as-bid on price \( \Delta \) model](image)

**4.3 Commodity and imbalance prices**

Each balance responsible party is financial responsible for settlement of its own imbalance. Its imbalance will be a surplus \( Q_{\text{surplus}} [\text{MWh}] \) if the balance responsible party sold less (position) than actually injected (allocated volume), or bought more (position) than actually withdrawn (allocated volume). Or it will be a shortage \( Q_{\text{shortage}} [\text{MWh}] \) in reverse situations (Chapter 3). Of course it can be 0 [MWh] if position equals allocated volume, indicating perfect energy balancing by the balance responsible party. In case there is no allocated volume in the absence of connections the balance responsible party has taken up balance
responsibility, e.g. a power exchange, this means that in their position buy equals sell. The net joint imbalance of all balance responsible parties per imbalance settlement period \( \sum Q \) [MWh] is the net joint energy imbalance or \( \int \text{ACE}_{\text{OL}}(t) dt \) [MWh] can be a surplus, or a shortage or nil. An individual imbalance of a balance responsible party can be either reducing to the net joint imbalance. The total volume of reducing imbalance \( Q_{\text{reducing}} = 0.5 \times ((|Q_{\text{surplus}}| + |Q_{\text{shortage}}|) - (|Q_{\text{surplus}}| - |Q_{\text{shortage}}|)) \) [MWh], that of aggravating imbalance \( Q_{\text{aggravating}} = (|Q_{\text{surplus}}| + |Q_{\text{shortage}}|) - Q_{\text{reducing}} \) [MWh]. Without reducing imbalances the absolute net joint imbalance would have been larger. With less aggravating imbalances it would have been smaller. In case the net joint imbalance is 0 [MWh], individual imbalances are neither aggravating nor reducing, so all imbalances are neutral.

The potential to arbitrage in imbalance between trading on previous commodity markets (affecting position) or dispatching in real-time (affecting allocation), is the margin between the (expected) imbalance price and the commodity price, as both prices apply to the same imbalance settlement period. Arbitrage by controlling its own imbalance requires awareness by the balance responsible party of this potential, and resources that it can control (Beune & Nobel, 2001). But no arbitrage potential should allow abusive behaviour of balance responsible parties. So next I will explore different imbalance pricing systems with potential beneficial or perverse incentives to the balancing responsible parties.

Commodity (whole sale) prices are established between users on an open energy market, or series of markets from long-term, day ahead to intraday markets (Boisseleau, 2004). The prices reflect conditions expected for, but prior to the imbalance settlement period of delivery (Section 4.1). Imbalance prices are monopolistically determined by the system operator and when consistent with scarcity, reflect actual conditions at the imbalance settlement period. Compared to the expected conditions the actual condition includes changes occurring after establishing the market price. These include the real-time actions by the system operator to keep its power balancing responsibility. The margin \( \Delta [\text{€/MWh}] \) between commodity and imbalance price defines the potential to arbitrage between commodity and imbalance. This potential thus may influence trading and dispatching behaviour of the balance responsible parties, and the grid users they represent. A margin \( \Delta > 0 [\text{€/MWh}] \) signifies that the balance responsible party pays more to the system operator in imbalance settlement of a shortage (buy from the system operator), and receives less for a surplus (sell to) than by performing such a transaction at commodity prices: the balance responsible party “loses” in imbalance. With a margin \( \Delta < 0 [\text{€/MWh}] \) the balance responsible party would be a “winner”. With a margin \( \Delta \approx 0 [\text{€/MWh}] \) the balance responsible party would neither win nor lose.

Analysing imbalance prices without reference to commodity prices therefore obscures the arbitrage potential for balance responsible parties. Furthermore, conversion to margin \( \Delta \)'s facilitates the perception of negative prices for commodity or imbalance. Irrespective of the signs of these prices, each margin \( \Delta > 0 [\text{€/MWh}] \) is a penalty to the balance responsible party. For imbalance margin \( \Delta < 0 [\text{€/MWh}] \), i.e. a benefit to the balance responsible party, there is no incentive for balance responsible parties, or their users, to trade or to dispatch according to their trades. The system operator would in such a virtual experiment eventually become both source and sink for all commodity and for all imbalances thus removing the rationale for separate imbalance settlement. The system operator thus de facto takes on the role of balance responsible party for all users, as
used to be the case in the monopolistic central dispatch systems in Europe prior to liberalization. In European market designs nowadays there is a separate imbalance settlement, and thus there has to be a margin $\Delta > 0 \text{ [€/MWh]}$, at least on average, between the imbalance and commodity price.

It is a necessary condition to have, at least on average, a cost or $\Delta \text{ [€/MWh]}$ attached to the imbalance of balance responsible parties. It is in the interest of individual balance responsible parties to keep their balancing costs [€], the result of their margin $\Delta$'s applied to their imbalances, minimal. The necessary condition that the margin $\Delta > 0 \text{ [€/MWh]}$ may be assigned to each imbalance, aggravating or not, to guarantee an average margin $\Delta > 0 \text{ [€/MWh]}$. Lifting the condition that each margin $\Delta > 0 \text{ [€/MWh]}$ results in $\Delta_{\text{aggravating}} \cdot |Q_{\text{aggravating}}| + \Delta_{\text{reducing}} \cdot |Q_{\text{reducing}}| > 0 \text{ [€/MWh]}$, or $\Delta_{\text{reducing}} > - \Delta_{\text{aggravating}} \cdot |Q_{\text{aggravating}}|/|Q_{\text{reducing}}|$. And since $|Q_{\text{reducing}}/Q_{\text{aggravating}}| < 1$, the necessary condition of an average imbalance margin $\Delta > 0 \text{ [€/MWh]}$ is met with $\Delta_{\text{reducing}} \geq - \Delta_{\text{aggravating}} \text{ [€/MWh]}$.

### 4.3.1 The imbalance price as control mechanism

The imbalance price as control mechanism may affect the net joint imbalance of all balance responsible parties per imbalance settlement period, and thus the net volume of balancing energy to be activated by the system operator. It may also affect the distribution of net joint imbalance volumes over multiple imbalance settlement periods, to be used by the system operators as dimensioning requirement for frequency restoration reserves (ENTSO-E, 2013, NC LFC-R).

Several concepts for imbalance pricing as control mechanism are possible and, indeed, are practiced. The main imbalance pricing options are, with main consequences in italics:

- **a)** for all imbalances the imbalance margin $\Delta > 0 \text{ [€/MWh]}$: each imbalance is penalized and all balance responsible parties will lose,
- **b)** for $Q_{\text{aggravating}}$: $\Delta_{\text{aggravating}} > 0 \text{ [€/MWh]}$ and for all other imbalances: $\Delta = 0 \text{ [€/MWh]}$: imbalance is never attractive and balance responsible parties may lose, and at best cannot win,
- **c)** for $Q_{\text{aggravating}}$: $\Delta_{\text{aggravating}} > 0 \text{ [€/MWh]}$, for $Q_{\text{reducing}}$: $\Delta_{\text{reducing}} \geq - \Delta_{\text{aggravating}} \text{ [€/MWh]}$: system support is rewarded at some balance responsible parties may lose, all others win.

Under all options (a, b, c) $\Delta_{\text{aggravating}} > 0 \text{ [€/MWh]}$, so penalizing to the balance responsible party. As it cannot be known in advance which imbalance, shortage or surplus, eventually will be aggravating, this functions as a priori knowledge and thus as a feed forward control to the balance responsible parties. Each balance responsible party's individual imbalance should be priori be minimal, and a posteriori, i.e. in real-time, aggravating imbalance should be unintentional. For controllable injection this means that dispatch needs to conform net sales, for non-controllable injection it means that sales should match expected injection, and for load it means that buying should match expected withdrawal.

The general formula for the energy volume over which the system operator collects the margin $\Delta_{\text{aggravating}} \text{ [€/MWh]}$ from balance responsible parties is: $|Q_{\text{aggravating}}| + \alpha \cdot |Q_{\text{reducing}}| \text{ [MWh]}$, with $\alpha = \Delta_{\text{reducing}} / \Delta_{\text{aggravating}} \text{ [-]}$. The parameter $\alpha$ reflects the pricing options for dual or single imbalance pricing (ENTSO-E, 2014, AS Enquiry).
I will consider only 2 dual pricing options $\alpha = 1$ and $\alpha = 0$ and the single pricing option $\alpha = -1$:

a) $\alpha = 1$: all balance responsible parties are incentivized to reduce their individual imbalance, regardless of the net joint imbalance,

b) $\alpha = 0$: those balance responsible parties with aggravating imbalances are more strongly incentivized to reduce their imbalances than the other balance responsible parties to reduce theirs,

c) $\alpha = -1$: incentivizes all balance responsible parties to reduce the net joint imbalance.

Dual pricing (options a, b)

Under dual pricing the collective objective for all balance responsible parties is to minimize their imbalances $\left| \sum O_{\text{surplus}} \right| + \left| \sum O_{\text{shortage}} \right|$ [MWh/ISP]. For dual pricing to function as an incentive balance responsible parties require awareness or feedback, on their own actual imbalance. This awareness may vary between balance responsible parties, depending on the portfolio of connections they are balance responsible for. A balance responsible party with only its fleet of a few large, conventional generators is in a totally different position as balance responsible parties with a large number of small scale consumers, and intermittent generation dispersed all over the system. In contrast to the first type the latter type balance responsible parties will not know their actual imbalance sign (surplus or shortage) in real-time, unless an additional layer of real-time metering and flexible communication systems are implemented. It requires real-time access to all connections of the portfolio of each balance responsible party, addition of all data per portfolio, and in case of the system operator providing the feedback, returning the proper values to the proper balance responsible party. Such a system must be very flexible as grid users have freedom in selecting their balance responsible parties. Providing each balance responsible party with its own actual position, whether it is done by each balance responsible party for itself, or by the system operator to each balance responsible party as a system service, is no trivial matter in terms of telecommunication and data handling. Not knowing their imbalance in real-time the grid users will not know in which direction to use any flexibility that they might have available in real-time (Beune & Nobel, 2001). This means that under the dual pricing part of the available flexibility in the system might be effectively sterilized. Dual imbalance pricing implicitly includes an exclusion mechanism.

Single pricing (option c)

The incentive from single pricing aims at supporting the system, and minimizing the joint net imbalance of all balance responsible parties $||\sum O_{\text{surplus}}| - |\sum O_{\text{shortage}}||$ [MWh/ISP]. To a balance responsible party under single pricing it is as beneficial to reduce an aggravating imbalance as it is to increase a reducing imbalance. This incentive may actually lead to an increase in $|\sum O_{\text{surplus}}| + |\sum O_{\text{shortage}}|$ [MWh/ISP] rather than minimization. Single pricing offers equal incentives to all balance responsible parties. Informing close to real-time on relevant, directional information from $\int \text{ACE}_{OL}(t) \, dt$, or $\int \text{S}_{FRR, RR}(t) \, dt$ (Chaves Avila, 2014) by the system operator offers equal opportunity to all balance responsible parties to know in which direction they might use any flexibility available to them. This removes an exclusion mechanism inherent to dual pricing. Providing system information already available to the system operator itself is less expensive, and more robust than trying to provide real-time information on imbalances of all balance responsible parties. That real-time system information feed-back to users may have effect on their real-time behaviour is demonstrated in appendix "A case example of real-time incentives to users". Single imbalance pricing potentially allows all
users through their balance responsible parties to benefit from reducing system imbalance. Prerequisite for effective incentives is timely, i.e. real-time feedback.

However, single pricing incurs a risk that because there will be winners and losers, and that by reacting on feedback, losers may become winners and vice versa. This potential reversal of winners and losers may, as a potential objective to balance responsible parties, result in overshooting rather than just reducing the net joint imbalance \( \sum Q \) [MWh] of the balance responsible parties. This effect of single imbalance pricing may lead to oscillation of the system imbalance and therefore can be regarded as a perverse incentive.

Single imbalance pricing also may impact local congestion management negatively. Winners and losers alike may be incentivized to dispatch connections at their disposal at the ‘wrong’ side of a congestion, that is temporarily relieved by re-dispatch. In such a case single imbalance pricing may give perverse incentives to balance responsible parties to counteract the re-dispatch and jeopardize grid security.

### 4.4 Balancing energy and imbalance prices

Balancing energy requested by the system operator from a balancing service provider generates additional turnover for the balancing service provider. Since this request is not accounted for in the market trades and thus in the position of the affected balance responsible party or parties it requires an adjustment to these balance responsible parties. When the balancing service provider delivers the requested volume, adjustment of the imbalance by the requested volume of balancing energy will not result in any additional imbalance. Or, reversely, without adjustment, delivering the requested balancing energy would result in an imbalance. But creating an imbalance by delivering less than the requested volume should not offer to the balancing service provider a profitable alternative. That would deteriorate the quality of delivery and hence fulfilment of the power balancing requirements of the system operator. Inconsistent pricing between balancing energy and imbalance might create such a perverse incentive.

A price inconsistency between balancing energy and imbalance is the difference between the \( \Delta \)'s [€/MWh] of the balancing energy and of the applied adjustment, that is settled at the imbalance price. A price \( \Delta \) for balancing energy less than margin \( \Delta_{\text{aggravating}} \) for non-delivery is a risk for the balancing service provider. They might include this risk as a mark-up in their balancing energy bid prices.

The reverse means that the balancing service providers will earn on non-delivery. This gaming opportunity can be remedied through ex-post verification by the system operator and additional penalties for balancing service providers for non-delivery. But this would increase complexity and add the cost of verification that must be recovered. In such a case the balancing service providers might include the penalty risk as a mark-up in their balancing energy bid prices.

Using the same, marginal, prices for both balancing energy and imbalance has a sobering effect on balancing energy bidding strategy of balancing service providers, most visible at the start but also at the end of the merit order list in each direction (Figure 4-7). This last effect of marginal pricing is most obvious at the shortage side (upward regulating) and can be explained by the following chain of reasoning:
A bid is an option, expressing the premium or $\Delta \ [\text{€/MWh}]$ that the balancing service provider minimally wishes to receive to accept an adjustment of a requested volume of balancing energy. However, by using the marginal price of balancing energy as the imbalance price, the system operator turns this original individual premium of a balancing service provider into a collective imbalance price risk for all balance responsible parties, including that of the balancing service provider. So the balancing service providers with large generators will have to consider the volume settled in balancing energy against the imbalance volume risk in case of a trip. As the expected volume of the latter might greatly exceed the former, bidding 'excessively' high prices is a risky strategy for a balancing service provider. Once he trips all bids will be activated and he will be facing a large volume of imbalance to be settled against these 'excessive' prices, as can be grasped from this example:

The balancing service providers in the Netherlands were forewarned, by the system operator, to these risks prior to change over to the new market arrangements in January 2001. Nevertheless, at the very beginning, some balancing service providers did use this extreme bidding strategy. This led to disastrous results in their imbalance settlement, that were softened by extraordinary decimations of some imbalance prices, thus deviating from balancing energy prices.
These events are still visible on the TenneT website:

These interventions (actual balancing energy prices ranged from -99000 to 5000 [€/MWh]) were considered justifiable, in order to appease all balance responsible parties, including those of the balancing service providers, and to prevent the learning experience becoming too expensive. In subsequent days balancing service providers duly refrained from such extravagant bidding strategies, and this they have done ever after. For downward regulating this risk is less prominent as presently no events within balance responsible parties' responsibility resulting in large immediate surpluses.
4.5 Conclusions and discussion on prices related to balancing

The Δ price model for prices for balancing energy and for imbalances, decomposes these prices in an opportunity component and a margin Δ [€/MWh] that can be related to activation of balancing energy.

In their bid prices the balancing service providers express both their perception of an opportunity value \( P_{OP} [€/M\text{Wh}] \) and their required margin \( \Delta > 0 [€/M\text{Wh}] \) for absorbing or delivering balancing energy. This mark-up should minimally cover their expected marginal cost of injecting or withdrawing additional. In addition it will contain mark-ups to manage perceived risks. Efficient market design should aim at minimizing such mark-ups to the marginal cost, by reducing uncertainty and by fostering competition. Any uncertainty perceived by the balancing energy provider on the opportunity value may \( P_{OP} \) result in larger mark-ups incorporated in their bid price \( \Delta \)'s. To achieve market conformity for bid prices for balancing energy it follows as a necessary condition that it is prerequisite to have day ahead and intraday market information publicly available, before balancing energy gate closure time. Each balancing service provider is responsible to incorporate such information into their final bid prices. Therefore a late balancing gate closure time, close to the imbalance settlement period of delivery will increase efficiency by reducing mark-ups in balancing energy bid prices. Late balancing energy gate closure times furthermore facilitate balancing energy bids from intermittent generation, and demand side resources to participate in balancing service provision. Early balancing gate closure times act as exclusion mechanisms to some users, reducing competition and efficiency.

Balancing service providers will furthermore reflect in their bid prices their expectations of the balancing energy price itself as mark-up components in their price \( \Delta \). Settling balancing energy at marginal prices allows the balancing service providers to retain the producer surplus, thus effectively to collect a market based scarcity rent. In pay-as-bid settlement of balancing energy scarcity rent will be collected as individual mark-ups in the balancing energy bid prices. Pay-as-bid leads to a different merit order list with a different relation between scarcity and bid prices, compared to marginal settlement.

The Δ price model for prices for imbalance reveals the incentives, and force of these incentives imposed on the balance responsible parties to manage their imbalance risk. These incentives depend on the imbalance pricing scheme for imbalances reducing or aggravating the joint net imbalance \( \sum Q [\text{MWh}] \) of all balance responsible parties per imbalance settlement period. On basis of the factor \( \alpha = \Delta_{\text{reducing}} / \Delta_{\text{aggravating}} \) balance responsible parties can be identified as 'losers', 'non-winners' and 'winners'.

Permanent dual pricing (\( \alpha = 0, \alpha = 1 \)) reduces the user's ability to manage the energy balance of the system, and thus functions as an exclusion mechanism. Pure single pricing (\( \alpha = -1 \)) leads to perverse incentives for balance responsible parties. Inconsistency between balancing energy and imbalance prices leads to perverse incentives for balancing service providers. Inconsistency between imbalance and commodity prices leads to perverse incentives for balance responsible parties. All these perverse incentives jeopardize the stability and/or the efficiency of the frequency restoration process, and therefore increase the burden of responsibility of the system operator.
5. Time units for trading and balancing purposes

Electrical power is a real-time variable, electrical energy requires a time interval over which its value is established. Such a time interval can be as much as a year, as witnessed by my bills. An energy only market requires time intervals over which to establish energy products to be traded or settled. These fixed length, fixed time units define the temporal resolution of prices for such energy products and thus have to be mutually exclusive and jointly exhaustive. The time units need not be identical for different energy products, like commodity and imbalance. But when time units for different products are not identical, a clear and consistent hierarchy is required between them.

In the coupled European day ahead markets for commodity trading all cross zonal processes by system operators and nominated electricity market operators (power exchanges) are based on "market time units" (Regulation (EC), 2014, CACM). These processes (e.g. capacity calculation, coupling, matching and scheduling) are out of the scope of this thesis. A market time unit means the period for which the market price is established, or the shortest possible common time period for different markets across interconnectors, if their market time units are different (Regulation (EC), 2013, 543/2013). This means that if one bidding zone has a market time unit of 2h and another has one of 3h, the market coupling between them occurs on a 6h market time unit.

ACER requires the imbalance settlement period to be consistent with "program time unit", defined as: time unit used for scheduling and programs (ACER, 2012, FWGL EB). It is understood that this program time unit corresponds to the market time unit. ENTSO-E formulated this requirement as: all boundaries of market time periods shall coincide with boundaries imbalance settlement periods (ENTSO-E, 2014, NC EB, 2014). This demonstrates the difficulties in maintaining exact and consistent terminology during parallel development of several European network code drafts and regulations. "Market time unit" appears in Regulation (EC) (2013, 543/2013) and the present draft Regulation (EC) (2014, CACM), "market time period" was used in earlier drafts.

An imbalance settlement period consisting of multiple market time units would be consistent. But this would invite balance responsible parties to arbitrage between differently priced market time units within a single imbalance settlement period. Price driven shifting in time of actual injection or withdrawal within an imbalance settlement period will not increase imbalance to be settled, but might increase the volume of counter-activation of frequency restoration reserves. Therefore the reverse hierarchical condition is required in which a market time unit consists of 1 to n imbalance settlement periods, to discourage such practices.

The participating countries in the current north-western European day ahead price coupling area are the Netherlands, Denmark, Finland, Norway, Sweden, Great Britain, Belgium, France, Germany (with Austria) and Luxemburg. It is the current imbalance settlement of the Nordic countries that prevents the market time unit in this north-western European price coupling area to be shorter than one hour.
Independent from the imbalance settlement period length are parameters or variables as e.g.:

- the 30 sec time for full activation of frequency containment reserves,
- the 30 sec reaction time allowed for activated automatic frequency restoration reserves,
- the 15 min time to restore frequency,
- the volume weighted average price of activated balancing energy [€/MWh].

A dependency of the dispersion of short term energy imbalances within the imbalance settlement period on its length was demonstrated by Beune & Nobel (2001) from 15 minutes to 1 hour for a Dutch load curve. This dependency has been confirmed by Frunt (2011). The dispersion of open loop power fluctuations within an imbalance settlement period gives an indication of power balancing execution and requirements and hence balancing power requirements of the system operator. The distribution of open loop area control errors over, not within, discrete 15 minute periods \( \int \text{ACE}_{OL} \, [\text{MWh/15min}] \) is prescribed by ENTSO-E (2013, NC LFC-R) as a dimensioning parameter for frequency restoration reserves, regardless of actual imbalance settlement period length.

The Dutch daily load curve is comparable to that of the other European countries. The conclusions of Frunt (2013) are therefore assumed to be generally applicable. Whereas he modelled a range of imbalance settlement period lengths from 8 seconds to 2 years, I will not consider in this thesis imbalance settlement periods longer than the maximum 1 hour market time unit in Europe.

A decrease in imbalance settlement period length will ceteris paribus result in a decrease of:

- the volume of counter-activation of balancing energy [MWh],
- the marginal price \( \Delta \) for balancing energy per direction [€/MWh],
- the power fluctuations within the imbalance settlement period,
- the allocated volumes per imbalance settlement period per direction per balance responsible party [MWh].

Vice versa, longer settlement period lengths will result in potentially more counter-activation of balancing energy, and in larger marginal price \( \Delta \)'s of balancing energy per direction. Both mechanisms affect the expenditure on balancing energy activation (volume \( \times \) price \( \Delta \), [€]) of the system operator. The marginal price effect can be mitigated by applying pay-as-bid rather than pay marginal for balancing energy. The average balancing energy price under pay-as-bid is reduced compared to marginal pricing. Applying this reduced price in imbalance pricing reduces the incentives to the balance responsible parties. The volume of counter-activation can be reduced by giving balance responsible parties less incentives to balance the system in energy, e.g. by not having single imbalance pricing.
By such design the system operator then becomes more involved in maintaining the energy balance, rather than only restoring the power balance. With less risk for counter-activation the system operator is enabled to do this by (de-)activating reserves that then may include replacement reserves in predominantly one direction. Deactivation of frequency restoration reserves by activation of replacement reserves in the same direction will in such a case not lead to counter-activation. But the notion that energy from activated replacement reserves might be cheaper than that from frequency restoration reserves only makes sense under a pay-as-bid settlement for energy from replacement reserves.

It is concluded that longer imbalance settlement periods are less compatible with marginal pricing for balancing energy, and less compatible with single imbalance pricing.

Over a fixed time interval like a calendar day or year, a decrease in imbalance settlement period length will, ceteris paribus result in an increase in:
- the number of imbalance settlement periods,
- the number of allocations to balance responsible parties,
- the number of merit order lists for balancing energy bids,
- the total imbalance volumes of balance responsible parties: \( \sum(\left| Q_{\text{surplus}} \right| + \left| Q_{\text{shortage}} \right|) \) [MWh]

The fixed 30 [sec] maximal delay time for frequency restoration and full activation for frequency containment (ENTSO-E, 2013, NC LFC-R) will cause a discrepancy between activation request and delivery. These will be manifest at the beginning and at the end of the imbalance settlement period. It means that if an adjustment for activation of frequency containment or frequency restoration reserves is based on the requested volumes in one imbalance settlement period that the actual delivery may occur in the next one. Even perfect delivery within these boundary conditions may result in imbalance in both imbalance settlement periods. Whereas 30 seconds is 3.3 % of 15 minutes, it is 10 % of 5 minutes. This mismatch is therefore larger with shorter imbalance settlement periods. Since allocated volumes are smaller as well over shorter imbalance settlement periods, the relative mismatch will increase even more. From the balancing service provider perspective this poses an additional risk, to be taken into account in bid prices. The issues related to volume determination on delivered versus on as requested as a remedy is discussed in Chapter 3.

5.1 Conclusions and discussion on time units for trading and balancing purposes
The required consistency (ACER, 2012, FWGL EB) between market time units and imbalance settlement periods is hierarchical. A market time unit can consist of one or more imbalance settlement periods. The reverse configuration may lead to non-efficient price arbitrage by balance responsible parties.

The longer the imbalance settlement period, the less a market design becomes compatible with marginal pricing for balancing energy, and the less it becomes compatible with single imbalance pricing.

Imbalance settlement periods longer than the time to restore frequency, result in an energy balancing responsibility for the system operator, in addition to its power balancing responsibility. An energy balancing responsibility of the system operator increases the potential risk and therefore the cost of counter-activation.
This can be suppressed by limiting the balancing incentives to balance responsible parties through the imbalance price or prices.

Very short imbalance settlement periods with single imbalance marginal imbalance pricing potentially may result in improved feedback to and from the balance responsible parties. At the same time it renders such incentives less effective by reducing the imbalance volumes per imbalance settlement period. Any delay-time to produce the feedback information and reaction time of the balance responsible party will further reduce the effectiveness of such incentives.

The subject of this thesis is not the search for the optimal imbalance settlement period length. It points out that both longer and shorter imbalance settlement periods have some beneficial and adverse consequences. With a time to restore frequency of 15 minutes (ENTSO-E, 2013, NC LFC-R) an imbalance settlement period of 15 minutes is a reasonable compromise that allows for pay marginal settlement of balancing energy and requires no reserve replacement process. The major difference of course is that the 15 minutes time to restore frequency is a continuously moving time window, whereas 15 minute imbalance settlement periods are discrete, fixed time intervals.
6. The financial balance of the system operator

In this Chapter the financial position of the system operator is analysed with emphasis on its expenditure related to its compliancy requirements, and the impact of the diverse energy settlements the system operator is involved in. Grid operation costs, like grid reinforcements, re-dispatch and grid losses are out of scope in this analysis. These costs are explicitly excluded from the imbalance price (ACER, 2012, FWGL EB, 2012), and thus from imbalance settlement.

Per imbalance settlement period the net financial result for the system operator from the settlements of balancing energy and imbalances is:

$$(|Q_{aggravating}| \times \Delta_{aggravating} + |Q_{reducing}| \times \Delta_{reducing}) - ([\int S_{(FRR, RR)}(t)dt] \times \Delta_{UP} + [\int S_{(FRR, RR)}(t)dt] \times \Delta_{DOWN})$$ [€]

Under perfect balancing the net volume of balancing energy settled with the balancing service providers equals the net volume of imbalance settled with the balance responsible parties within a single imbalance settlement period. Under perfect balancing this net volume of balancing energy is transferred between balance responsible parties and balancing service providers at $P_{OP}$ [€/MWh] with all the $\Delta$'s [€/MWh] remaining with the system operator. Under imperfect balancing the system operator is exposed to $P_{OP}$ [€/MWh] over the difference between the net volumes of balancing energy $\int S_{(FRR, RR)}(t)dt$, and of imbalances $\Sigma Q$ [MWh]. Physically this difference will be accounted for by a frequency deviation, and/or by an exchange of energy with other system operators.

Financial settlement of energy exchanged with other system operators as required in (ENTSO-E, 2014, NC EB) obviously will affect the financial position of the system operator. It follows that the price differential between the internal commodity price $P_{OP}$ [€/MWh], and the external settlement prices used with the other system operators should not be beneficial to the system operator to avoid a perverse incentive not to activate balancing energy. The settlements among system operators (ENTSO-E, 2014, NC EB) only have distributional effects; no money is generated or destroyed. Still the impact on each system operator may be positive or negative.

The system operator fulfils its balancing compliancy requirements as a system service, delivered to its users. This public service obligation should not lead to its bankruptcy, nor should the system operator profit from it to avoid using perverse risk taking strategies, or abusing its monopoly. The system operator should thus be financially disinterested in honouring its compliancy requirements. Financial disinterest means that system operator requires adequate and secure remuneration of expenditure for compliancy requirements, and predetermined, dedicated allotment of financial surpluses resulting from its balancing act.

Expenditures and remunerations of a system operator will have distributional effects amongst the actors involved (Abassy & Hakvoort, 2009). These actors are its balancing service providers, its balance responsible parties and the users these represent. But it may impact as well on other system operators. It therefore will affect indirectly their balancing service providers, their balance responsible parties and their grid users as well. And finally it will impact on their owner, or on their access tariff payers.
Balancing energy expenditure

The balancing energy expenditure of the system operator is per imbalance settlement period:

\[
(|\int S_{\text{FRR, RR}}(t)dt| * \Delta_{\text{UP}} + |\int S_{\text{FRR, RR}}(t)dt| * \Delta_{\text{DOWN}}) \ [\text{€}]
\]

The balancing energy price \(\Delta_{\text{UP}}\) and \(\Delta_{\text{DOWN}}\) \([\text{€}/\text{MWh}]\) are volume weighted averages. With marginal pricing the average equals the marginal price \(\Delta\ [\text{€}/\text{MWh}]\) in that direction without volume weighting required. With consistent pricing the expected expenditure of the system operator is \(\geq 0\ [\text{€}]\).

Counter-activation of balancing energy within an imbalance settlement period results in a net energy transfer of \(\Sigma 0.5 \times (|\int S_{\text{FRR, RR}}(t)dt| + |\int S_{\text{FRR, RR}}(t)dt| - |\int (S_{\text{FRR, RR}}(t)dt| - |\int (S_{\text{FRR, RR}}(t)dt|)) \ [\text{MWh}]\) among balancing service providers at \(P_{\text{OP}}\ [\text{€}/\text{MWh}]\). The system operator absorbs all the \(\Delta_{\text{UP}}'s\) and \(\Delta_{\text{DOWN}}'s\) \([\text{€}/\text{MWh}]\) over this transfer. This expenditure impacts on the system operator financially, regardless whether the balancing energy settlement mechanism is pay-as-bid or pay marginal.

Imbalance settlement remuneration

The imbalance remuneration of the system operator is per imbalance settlement period (Section 4.3.1):

\[
(|Q_{\text{aggravating}}| * \Delta_{\text{aggravating}} + |Q_{\text{reducing}}| * \Delta_{\text{reducing}}) = \Delta_{\text{aggravating}} * |Q_{\text{aggravating}}| + \alpha * |Q_{\text{reducing}}| \ [\text{MWh}],
\]

with \(\alpha = \Delta_{\text{reducing}} / \Delta_{\text{aggravating}}\) and \(\Delta_{\text{aggravating}} > 0\ [\text{€}/\text{MWh}]\).

The volumes exposed to \(\Delta_{\text{aggravating}}\ [\text{€}/\text{MWh}]\) decrease from \(\alpha = 1\) to \(\alpha = -1\). Therefore, the system operator collects with identical margin \(\Delta_{\text{aggravating}}\ [\text{€}/\text{MWh}]\), more remuneration from balancing responsible parties from \(\alpha = 1\) to \(\alpha = -1\). The imbalance volume \(Q_{\text{reducing}}\) is transferred among balance responsible parties at \(P_{\text{OP}}\ [\text{€}/\text{MWh}]\). The margin \(\Delta's\ [\text{€}/\text{MWh}]\) over these volumes are collected by the system operator.

In case no balancing energy has been activated at all, there is no balancing energy price definable in either direction. Consequently the imbalance price then cannot be derived from the balancing energy settlement, and must be derived by other means. For instance by referring to a previous market, or by referring to the centre of the merit order list for balancing energy bids, as that is the latest market where multiple offers and bids are collected. Under pay marginal for balancing energy, and close to real-time balancing energy gate closure times, this midpoint ideally should converge to \(P_{\text{OP}}\ [\text{€}/\text{MWh}]\), resulting in imbalance margin \(\Delta \approx 0\ [\text{€}/\text{MWh}]\) in either direction.

6.1 The financial neutralization of the system operator

Imbalance does result in a net cost to the balance responsible parties. An objective of market design is to keep costs at efficient levels, with little mark-ups. Different designs reflect to different philosophies regarding the purpose of imbalance settlement, with efficient cost and the system operator’s financial disinterest as boundary conditions. Glachant and Saguan (2007) explicitly distinguish:

- a real-time market arrangement based on user prices, or
- a penalty based balancing mechanism based on costs.
These design options result in different financial outcomes and therefore incentives to balance responsible parties, balancing service providers and system operators.

6.1.1 Real-time market
In a real-time market the imbalance of balance responsible parties is the real-time product of an energy only market. It is the conclusion to a sequence of long term, day ahead and intraday energy markets. This allows, invites, and basically incentivizes active participation and competition between imbalance and balancing energy. It is competition by users that forces imbalance margin $\Delta$'s down to efficient levels. The system operator performs a residual power balancing role without energy optimization responsibilities. Necessary conditions for a real-time market are consistency of relations between commodity, balancing energy and imbalance prices and scarcity, and timely information feed-back to users. These consistencies require a late gate closure time for balancing energy bid prices. The admittance of non-procured bids and close to real-time gate closure times for balancing energy foster competition from intermittent resources and loads. Both will inevitably lead to dynamic merit order lists, i.e. one for each imbalance settlement period per direction.

Consistent pricing between balancing energy and actual real-time scarcity requires merit order activation, including that of automatic frequency restoration reserves, per imbalance settlement period. It requires a close to real-time gate closure time for balancing bids and marginal price settlement for balancing energy in order to have balancing service providers express their own preferences and costs in their bid prices, without adding inefficient mark-ups. Single, marginal imbalance pricing requires an imbalance settlement period not exceeding the time to restore frequency as explained in Chapter 5. With an imbalance settlement period of 15 minutes actions of the system operator in power balancing will affect only the current imbalance settlement period and may continue in the next one, but certainly not the one after that.

One effect of pay marginal settlement of balancing energy is that all producer surplus of the merit order list is retained by the balancing service providers, and none by the system operator. The marginal price is a better approximation of the full value of requested balancing energy than the individual, or average bid-price. Another side effect is that the development of marginal prices can be established in real-time. This practical feature allows for faster imbalance price determination and information feed-back.

The use of additional components to the imbalance price, and thus to the margin $\Delta \ [\text{€/MWh}]$ is rejected for a real-time market by Glachant & Saguan (2007) on the economic consequences they deduce from numerical simulations:
- a distortion of the forward price,
- an asymmetric shift in the welfare of market participants,
- an increase in the system operators' revenues,
- inefficiencies.

The increase in the system operator's revenues is interpreted as an increase in remuneration, rather than an increase in profits.
Additional components added to the imbalance price, and hence to the imbalance margin $\Delta \ [€/MWh]$ jeopardize the consistency between balancing energy and imbalance prices. Constant additional components will not make imbalance prices more consistent with scarcity. In fact they will make imbalance margin $\Delta$'s less consistent with scarcity, therefore less incentive compatible.

The joint net imbalance $||\sum Q_{\text{surplus}}| - |\sum Q_{\text{shortage}}||$ [MWh] of the balance responsible parties corresponds to the net activated balancing energy but does not correspond to counter-activation of balancing energy over that imbalance settlement period. Some combinations are inevitable. There might be counter-activation of balancing energy to be settled with $||\sum Q_{\text{surplus}}| - |\sum Q_{\text{shortage}}|| = 0$ [MWh], or imbalances to be settled with no balancing energy to be settled, but only exchanges with other system operators. Using the same price for settlement of balancing energy and imbalance will therefore not lead to exact cost recovery by the system operator, not even for the balancing energy itself due to these volumetric aspects. A similar conclusion with respect to non-exact transmission cost recovery using marginal cost is drawn by Büchner (1996). Therefore an a posteriori mechanism is required to neutralize the system operator financially.

**A posteriori neutralization mechanisms**

An open financial position of the system operator must be closed in order to guarantee its financial disinterest in fulfilling its power balancing compliancy requirements. The European regulation on financial neutrality of the gas system operators in their balancing duties in appendix "Financial position gas system operator for balancing" requires this open financial position to be neutralized though access tariffs.

A comparable solution is ordained for another source of operator income from international congestion management. System operators collect congestion rents from users by granting them access rights to socialized, i.e. non-merchant, interconnectors. EC regulation (Regulation (EC) 714/2009, 2009) clearly limits to what purposes congestion income of system operators may serve, as a clear example of an a posteriori neutralization.

\* Any revenues resulting from the allocation of interconnection shall be used for the following purposes:
(a) guaranteeing the actual availability of the allocated capacity; and/or
(b) maintaining or increasing interconnection capacities through network investments, in particular in new interconnectors.

If the revenues cannot be efficiently used for the purposes set out in points (a) and/or (b) of the first subparagraph, they may be used, subject to approval by the regulatory authorities of the Member States concerned, up to a maximum amount to be decided by those regulatory authorities, as income to be taken into account by the regulatory authorities when approving the methodology for calculating network tariffs and/or fixing network tariffs.

The rest of revenues shall be placed on a separate internal account line until such time as it can be spent on the purposes set out in points (a) and/or (b) of the first subparagraph. The regulatory authority shall inform the Agency of the approval referred to in the second subparagraph.\*


A similar solution is applied for power balancing in the Netherlands (ACM, SysteemCode). Any financial result, and this implies positive or negative, of the Dutch system operator from settlement of imbalance with
balance responsible parties and of balancing energy with balancing service providers accrued over a year is to be absorbed by next year’s access tariffs. Since the cost of procurement of balancing reserves is remunerated through access tariffs, allotment of a surplus revenue to the access tariff will at least make balance responsible parties pay some of security insurance cost, and avoid the creation of a “beer-fund” for the system operator (Glachant & Saguan, 2007).

The conclusion is that a posteriori financial neutralization through access tariffs is an acceptable, and even a prescribed mechanism for system operators honouring public service obligations other than electricity balancing. There is no compelling reason to deviate from this policy line for electricity balancing.

6.1.2 Penalty based balancing mechanism
A balancing mechanism as distinguished by Glachant & Saguan (2007) aims at secure cost remuneration for the system operator. In such a design imbalances are used to distribute costs over the balance responsible parties. Imbalance is not the final product on a sequence of long term, day ahead and intraday energy markets. Basic issues then are the nature of the costs to be recovered, and the accuracy of cost remuneration through imbalance settlement.

Costs to be remunerated in a balancing mechanism
In addition to the real-time operational expenditure on net balancing energy per imbalance settlement period, the system operator may be exposed to energy settlement expenditures on counter-activation within an imbalance settlement period. Furthermore it faces energy settlements among system operators (ENTSO-E, 2014, NC EB), that may be positive or negative. Dimensioning and securing of frequency containment reserves is unrelated to actual imbalances of the balancing responsible parties in synchronous areas with multiple system operators ENTSO-E, 2013, NC LFCR). Other capability requirements of the system operator such as frequency restoration dimensioning and reserve procurement, black start capability, communication and information systems, SCADA functionality etc. are all system operator expenditures. System operators are legally or regulatory obliged to comply to such requirements and settlements, whether or not there is a real-time power deviation or energy imbalance. Such compliancy requirements therefore refer to security insurance (Tractebel, 2009). The beneficiaries of security insurances are all their own grid users, and by mutual interest, the interconnected transmission system operators, with their grid users. Security insurance is not related to real-time imbalances of the balance responsible parties.

Potential capacity remuneration mechanisms as proposed by de Vries (2004) may also be added to the system operator responsibilities and costs. These certainly are not related to the actual imbalances of balance responsible parties or to net real-time energy delivery by the system operator. Real-time energy delivery services expenditure should be recovered by imbalance settlement, in agreement with Tractebel (2009). But they then propose some procurement cost to be recovered through imbalance settlement as well.
The same conclusion is drawn by Vandezande et al. (2010), although it is admitted that there is a rationale to socialize part of the reserve’s cost. An argument to incorporate reserve procurement cost for the frequency restoration process in imbalance settlement is that it would add the long run marginal cost to the short run marginal cost from settlement of balancing energy. This would reveal the ‘true value’ of reserves in the imbalance price (ENTSO-E, 2014, LTMD). This argument immediately loses credibility if balancing energy is settled at pay-as-bid, thus withholding the producer surplus from the balancing service providers, required to cover their long term marginal cost. It is inconsistent to protect the balance responsible parties from exposure to marginal prices, but subsequently expose them to other non-energy market based costs in the imbalance price.

**Imbalance volumes as cost bearer**

Exact remuneration of security insurance expenditure per imbalance settlement period is impossible due to the volatility of imbalance volumes per imbalance settlement period. It would obviously fail if there is no imbalance to be settled. Remuneration has to be accomplished over multiple imbalance settlement periods. And this reduces the applicability of the imbalance price as a feed-back mechanism to the balance responsible parties, compared to a real-time balancing market. The smaller their actual imbalances, the larger this cost based penalty component will become.

The system operator cost can be attributed to relevant imbalance volumes derived from the imbalance margin Δ model: $|Q_{aggravating}| + \alpha \cdot |Q_{reducing}|$ [MWh], with $\alpha = 1, 0, or -1$.

Over any number of imbalance settlement periods $|Q_{aggravating}| + \alpha \cdot |Q_{reducing}|$ reduces, for $\alpha = 1$, to $\sum |Q_{surplus}| + |Q_{shortage}|$ [MWh]. In this design the volumes of aggravating respectively reducing imbalances need not be established. The design is independent from imbalance settlement period length. There is no consistency between balancing energy and imbalance prices, nor is there consistency between imbalance prices and real-time scarcity. It is essentially an imbalance tariff, without real-time incentives to the balance responsible parties. It is therefore exclusively the responsibility of the system operator to keep its operational balancing energy expenditure down to efficient levels through optimization.

"Two main recommendations need to be taken into account here. Firstly, with respect to harmon[iz]ing real-time market designs, a distinction should again be drawn between security insurance and real-time energy delivery services. The former should be procured using capacity payments only, and their costs should be socialized, whereas the latter should preferably be remunerated solely through energy payments. Capacity payments can be allowed for a transitional period, but should ultimately be phased out. The costs of procuring these services should be passed on to imbalanced Balance Responsible Parties (BRPs) via the imbalance settlement. This imbalance settlement should be cost-reflective and market-based, implying that no other components such as power exchange prices or penalties are included in the real-time energy price.

An additive component is however needed to settle possible capacity payments. To limit the impact of this additive component on overall real-time energy prices, the volume of real-time energy delivery services contracted using capacity payments should be limited – and abolished in the long run – so that real-time energy prices are based mainly on balancing services procured in real-time, rather than dominated by the capacity component.”

Source: Tractebel (2009)
For $\alpha = -1$ the relevant imbalance settlement volume becomes $|Q_{\text{aggravating}}| - |Q_{\text{reducing}}| \text{[MWh]}$ per imbalance settlement period. Calculating $\Delta_{\text{aggravating}} \text{[€/MWh]}$ by dividing the balancing energy expenditure of the system operator ($|\int S^{\text{frr}, \text{rr}, \text{up}}(t) dt| \times \Delta_{\text{UP}} + |\int S^{\text{frr}, \text{rr}, \text{down}}(t) dt| \times \Delta_{\text{DOWN}} \text{[€]}$ by $(|Q_{\text{aggravating}}| - |Q_{\text{reducing}}|) \text{[MWh]}$ for each imbalance settlement period should result in financial neutrality for the system operator over balancing energy expenditure and imbalance settlement. However $\Delta_{\text{aggravating}} \text{[€/MWh]}$ may become extremely high with a joint net imbalance $|Q_{\text{aggravating}}| - |Q_{\text{reducing}}| \text{close to 0 [MWh]}$. This causes huge distributional effects between balance responsible parties. Having the largest imbalance prices when the balance responsible parties are jointly well balanced is not consistent with prices reflecting real-time scarcity. Of course, these effects can be tweaked by additional, nonmarket based rules, but only at the expense of neutralization over multiple rather than per individual imbalance settlement periods. This may be practical but is contradictory to the design principles.

Although the other options (a, b) with $\alpha = 1$, $|Q_{\text{aggravating}}| + |Q_{\text{reducing}}|$, or $\alpha = 0$, $|Q_{\text{aggravating}}|$, are not fully immune to small divisors, they will result in less huge redistributions among balance responsible parties. Transferring prices instead of costs of balancing energy to the imbalance settlement completely avoids this issue.

6.2 Conclusions and discussion on the financial balance of the system operator

Transferring system insurance remuneration from access tariffs to an imbalance mechanism will reduce access tariffs. This might be welcomed by regulators and governments alike. It would also make producers, presently often exempted from access tariffs, pay a share of the system security insurance cost. This would correct somewhat for highly discriminating tariff structures in Europe, as shown in appendix "System operator tariff characteristics in European countries". But it is still a poor alternative to harmonizing the presently exemption based access tariff structures across Europe.

* Shifting costs between capacity [tariffs] and energy payments [imbalance settlement] therefore implies distributional effects between grid users and balancing responsible parties. Finally, the end customer will end up paying the additional costs – either via grid charges or via the supplier who will tend to pass through higher imbalance charges to the final customer."

Source: Frontier Economics, 2011

This statement reiterates the distributional effects mentioned by Abassy & Hakvoort (2009) and is undisputedly true. But exposure to dynamic imbalance prices in a real-time balancing market can affect the actual volume of individual imbalances of balance responsible parties, and therefore the aggregate imbalance volume. Thus it may affect the volume of balancing energy to be activated by the system operator, and ultimately the volume of balancing capacity required, when not determined by the dimensioning incident. Imbalance price driven imbalance management by balance responsible parties can induce competition among balance responsible parties and balancing service providers. It is the users that have the responsibility and opportunity to keep their energy balancing cost, their margin $\Delta$’s at efficient levels. But ultimately it is a political choice whether to include a real-time balancing market as the final part of
the successive energy markets with non-regulated prices, from long term, day ahead and intraday markets. The alternative is to design a separate balancing mechanism with nationally regulated imbalance tariffs. In that case it is through optimization by the system operator that the balancing energy component of an imbalance tariff is minimized. Tariffs for imbalances require national regulatory oversight and are inconsistent with a real-time balancing market as already noted by Glachant & Saguan (2007).

The rationale suggested in this thesis to socialize security insurance cost through grid tariffs, and hence not to incorporate these in the imbalance price (cf. Vandezande, 2010) is based on the following arguments:

- **Beneficiary**: stable frequency is delivered to, and enjoyed by all grid users.
- **Causation**: frequency deviations are caused by grid users (and/or operator faults);
  - system users are energy responsible, not power responsible
- **Effectivity**: without imbalance there will be no remuneration for reserve dimensioning compliancy requirements, and for intra imbalance settlement period counter-activation;
  - adding components will distort balancing energy markets (Glachant and Saguan, 2007)
- **Conceptuality**: mandatory, mutual security insurance by system operators is not a market issue.
7. Testing models and design choices

In this Chapter the models and design choices for price formation, described in Chapter 3 are tested against empirical data from the Netherlands and Germany. These models describe the formation of the commodity price $P_{OP}$ [€/MWh], and the prices for balancing energy and imbalance in relation to this commodity price. The system imbalance can be determined using several approaches that are compared with specific attention given to sign conventions. Conclusions from these tests and comparisons are given at the end of this Section

7.1 Commodity price model

The conceptual model for the commodity value $P_{OP}$ [€/MWh] in Section 4.1 is based on the intersection of an operating level and a power supply curve, neither of which is known exactly. The operating level is usually equated to loads as published by system operators, though consumption by a user from generation behind a connection is not often accounted for in published loads.

Grid and system operators need metered data on the energy volumes exchanged between connections and grid for tariffication and imbalance settlement purposes, but functionally do not require additional data. Historically direct consumption from generation, without exchange to the grid used to be small, and metering of netted exchange would suffice. With increasing volumes of electrical energy being generated, stored and consumed locally by small scale consumers this practice is not sustainable. In addition, monitoring and accounting for governmental energy policy purposes may require additional data, and therefore more insight in activity and (private) equipment at the connection. Smart metering can only partially provide a solution to these data requirements. Installing separate, directional meters at prosumer’s connections is an improvement, but cannot account for local consumption from local batteries. This accounting problem, however small presently, represents one of the challenges from the reform of a historically unidirectional power supply systems as in Figure 1-2, into a more bidirectional system as shown in Figure 1-1.

A supply curve can be constructed bottom up. This requires for all timestamps a complete overview of all sources connected to the grid each source the cost of generating power at each output level. One of the aims of the Transparency Regulation is to collect and disseminate source data (Regulation (EU), 2011, REMIT). As proxy for actual generator cost data, which are exclusive to the generation source owner, source type data (technology, fuel type) are requested for reporting (Regulation (EU), 2011, REMIT). A required withdrawal then can be dispatched over these sources according to some optimization criterion. This of course describes the practice in a central dispatch system. Though apparently exact, solvable and deterministic, this exercise in practice requires significant assumptions regarding completeness and accuracy of data on availability, capacity values (name plate capacity is a static property, actual capacity dependent on numerous conditions), and actual costs for fuel, operation and maintenance, and start and stop costs. And in interconnected systems the volumes and directions of cross zonal trades have somehow to be taken into account. Intermittent generation and inflexible load have to be taken into account somehow as well.

An alternative approach is to reconstruct a supply curve by confronting whole sale market prices to loads per time stamp, disregarding any qualitative information on sources and disregarding differences in resolution in
both time series. It is also conventional to regard day ahead market prices as a continuous time series, subject to the same time resolution, persistency, regularities and contingencies that govern published loads.

In an open energy market users should be able to hedge their volume and price risks by entering long term markets that are run prior to the day ahead market (Boisseleau, 2004). Long term markets are driven amongst other by the wish of producers to secure sales and hence to secure income, and by consumers to secure supply and hence avoid imbalance (Glachant & Saguan, 2007), and for both by their expected value on the day ahead market (Boisseleau, 2004). The mere existence of long term markets and products, like forward physical contracts or contracts for differences is sufficient evidence to this desire. This means that day ahead markets are *residual* markets, unless participants are forced to trade all expected withdrawals and all available injections on that market. Such a restriction (no forward physical contracts) on market participants points essentially to a power pool concept, enabling central scheduling and dispatch.

Clearing prices and volumes are assigned for each market time unit through matching all bids and offers for each market time unit of the day of delivery. This matching is performed on the day preceding the day of delivery by a market operator or power exchange, and results in a single price for all cleared bids, for each market time unit, for each bidding zone (Regulation (EC), 2014, CACM). Bids (and offers) to be matched may come from all suppliers, generators, traders, brokers and industrial end-users that are accredited at the power exchange for a bidding zone. All bids represent portfolios, varying from no connections for pure traders, to one or more connections for balance responsible parties, up to entire bidding zones in case of market coupling. Bidders take access to other bidding zones into account in their bidding strategy. Block bids, whether floating or fixed, allow bidders to account for start or stop costs of generators. They do so knowing that only one clearing price and volume will result for each market time unit. In European electricity day ahead markets, operated by nominated electricity market operators or power exchanges (Regulation (EC), 2014, CACM), market access increasing liquidity is offered to users to exchange trade positions across congested interconnectors on a day ahead basis.

It is therefore only day ahead markets that therefore can practically guarantee for each market time unit of an entire day one and only one $P_{OP}$ [€/MWh]. All market time units are cleared simultaneously. Market participants will adapt already traded positions to their updated expectations on injections and withdrawals, in order to manage their imbalance risk.

Without perfect foresight bids on the day ahead market may be inaccurate; they also might be erroneous. Buying or selling at market value will be expressed by using bid price limits in the orders instead of own preferences. Inaccuracies, errors, non-economy driven caps and floors explained in appendix "Price caps and floors of day ahead price in the Netherlands", and non-market based imbalance prices (Glachant & Saguan, 2007) may affect day ahead prices. Any contingencies happening after clearing of the day ahead market such as unexpected demand, outages, curtailment, will not affect the day ahead price, and consequently will not be visible in the day ahead price for that market time unit. Market trade volumes between bidding zones established by day ahead market coupling are firm, and the resulting energy trade flows over interconnectors become the responsibility of the transmission system operators on either side of
the interconnectors, just as physical power flows are their responsibility. These energy trade flows can only be changed by subsequent intraday trades by users, or by curtailment by the transmission system operators on either side of the interconnectors.

Changes in demand and supply after day ahead clearing are due to not yet expected demand or (un)availability of supply sources. Day ahead forecasts of intermittent generation and of demand, traded on day ahead markets will be followed by intraday forecasts with increasing accuracy. This means that 'demand' on intraday markets includes additional demand not previously covered, but also volumes of supply already traded, volumes that sellers are unable to inject. Likewise 'supply' on intraday markets also includes volumes already bought, volumes that the buyers are unable to withdraw.

With increasing volumes of intermittent generation entering the electricity market, day ahead prices will inevitably lose some significance as proxy for real-time commodity value $P_{OP}$. Differences between increasingly accurate forecasts and trades on previous markets will have to be traded intraday, or will be exposed to imbalance settlement. Intraday trading may start after clearing of the day ahead market, well before the day of delivery and therefore actually and confusingly, on the day ahead. Intraday day trades differ from day ahead markets in not necessarily covering each imbalance settlement period, but also in potentially yielding multiple prices for the same imbalance settlement period, by having several gates for each market time unit. Intraday trading in general represents smaller trade volumes, traded between smaller numbers of participants.

Load patterns are regular; as shown by their autocorrelations patterns. Published aggregate hourly loads for several consecutive years in the Netherlands (NL), France (F) and Norway (NO) all display high short term persistency, meaning that loads look like their immediate preceding ones. Load patterns of all three countries, shown in Figure 7-1, contain multiple seasonalities: daily, weekly and yearly cycles.

![Figure 7-1 Autocorrelations of hourly loads [MWh/h] in the Netherlands (NL), France (F), and Norway (NO), 2006 - 2011](image)
The daily cycle for the Netherlands has a larger amplitude than its yearly cycle, in contrast to France and Norway. For those countries the differentiation of loads within and between day and week is substantially less, the yearly cycle basically therefore reflects an energy consumption pattern. The weekly cycle of daily maxima and minima, observable in all three countries, conforms to a week of 5 working and 2 non-working weekend days. Assuming a working day is more similar to another working day than to a weekend day, and that a weekend day is more similar to another weekend day then to a working day, to most of us a common experience, a week might be coded like 1111100, with 1 representing a working, and 0 a weekend day. Identity or a lag of 168 hours (7 days) then results in 7 matches (5 * 1/1, and 2 * 0/0). A lag of 24 or 144 hours (1 or 6 days) results in 5 matches (4 * 1/1, 1 * 0/0) and 2 mismatches (1/0, 0/1); the other lags, 2, 3, 4 or 5 give 4 mismatches (0/1 or vice versa) and 3 matches (1/1). The number of matches per number of days lagged then becomes:

..., lag 3d 3, lag 1d 7, lag 1d 5, lag 2d 5, lag 3d 3, lag 3d 3, lag 4d 3, lag 5d 5, lag 7d 7, lag 8d 5, ....

Similar daily and weekly cycles for NL are described elsewhere by application of wavelet theory to a single week load profile (Frunt, 2011).

The autocorrelations of day ahead prices (APX website) and loads for the Netherlands as shown in Figure 7-2 display similar daily and weekly cycles, and thus are correlated.

Autocorrelations of hourly loads and day ahead prices in the Netherlands, 2006 - 2011

This is consistent with the conceptual commodity price model in Section 4.1, which assumes that prices and loads are positively correlated, so similarity in load and price autocorrelations is to be expected. But the lack of a yearly periodicity in day ahead prices disproves a fixed relation between price and load over the years. Loads span 2 orders of magnitude at most; prices span 5 orders of magnitude, requiring a log scale representation that reveals positive and negative outliers, away from the main body of data.
The median absolute deviation statistic on non-transformed day ahead prices in the Netherlands for 2001 can be used to declare outliers or spikes (De Lange et al. 2002), with median absolute deviation (MAD) defined as the median of the series \( |price_i - \text{median}(price)|, \ |price_{i+1} - \text{median}(price)|, \ldots, |price_n - \text{median}(price)| \) [€/MWh].

Absence of spikes (no-spike) is considered as a 'normal' or 'equilibrium' regime with:

\[
|price_i - \text{median}(price)| \leq 2 \times (\text{MAD}/0.6745) \quad [\text{€/MWh}].
\]

A hi-spike price regime is declared for:

\[
price_i > \text{median}(price) + 2 \times (\text{MAD}/0.6745) \quad [\text{€/MWh}].
\]

And a lo-spike price regime is declared for:

\[
price_i < \text{median}(price) - 2 \times (\text{MAD}/0.6745) \quad [\text{€/MWh}].
\]

The 0.6745 term normalizes the MAD statistic to one standard deviation of a Gaussian distribution. A single MAD calculated over the entire period results in constant regimes and static boundaries: an upper boundary of a no-spike regime of 79 [€/MWh], and a lower one of 0.85 [€/MWh], which both seem at odds with the graphical representation of prices and loads (Figure 7-4). Regimes declared per day, as all prices for each day are established simultaneously, result in dynamic boundaries that would illustrate the contextual nature of such filtering. A drawback of MAD filtering is that it "takes a symmetric view on dispersion" (Rousseeuw and Croux, 1993), a kind of dispersion that non-transformed commodity prices definitely do not display. The distribution of all day ahead prices in the Netherlands from November 1999 to June 2014 is heavily right-tailed with a skewness of 13.6 [-], which reduces to 7.6 [-] for 2006-2014. For 2006-2014 the skewness of the load distribution is just 0.08 [-]. Transformation to log(price) reduces the skewness of price distribution to -3.75 [-]. This negative skewness and the finite price resolution of two decimals in [€/MWh] over-enhances the prominence and visibility of lo-spike prices. Three hourly prices ≤ 0 [€/MWh], from -0.08 to 0 [€/MWh] I have floored at 0.01 [€/MWh]. But any larger future incidence of negative day ahead prices will invalidate the approach of log transformation of whole sale prices. Negative commodity prices do not invalidate the \( \Delta \) price models for balancing energy and imbalance proposed in Chapter 4, since these models do not require log transformation.

In Figure 7-3 are the distributions shown of log(day ahead price) for the Netherlands per calendar-month from November 1999 until June 2014. Shown are the interquartile range (25\textsuperscript{th} to 75\textsuperscript{th} percentile) and medians as boxes, and the adjacent values, the distance to the most extreme prices within 1.5 times the interquartile range from the 25\textsuperscript{th} or 75\textsuperscript{th} percentile as lines (whiskers); values beyond are plotted individually.
I will assume that the MAD statistic and the 'adjacent values' rule in the standard functionality of STATA correspond sufficiently to accept the notion of spike regimes for commodity prices in the Netherlands for November 1999 till June 2014. Terms like 'outlier', 'spike' and 'normal' all are contextual, and therefore arbitrary. The choice of the temporal resolution over which to determine prices spikes, and the method to declare a price spike, and the application of data transformation to the price or not, are all arbitrary.

Yet, from Figure 7-3 it is quite clear that spike regimes might be distinguished for commodity prices in the Netherlands, and that these regimes do evolve. The absence of a low-spike regime 1999 and 2000 is due to binding price fixes existing at that time (de Lange et al., 2002). A (monthly) hi-spike regime, indicated by the dots above the adjacent values seems to lose its prominence after 2008, in incidence and severity, thus reducing both scarcity rent collected by the sellers on the day ahead market (Brennan, 2003), and reducing congestion rent collected by transmission system operators.

The no-spike or equilibrium regime is not stationary over, or within years, as illustrated by the boxes or interquartile ranges per month in Figure 7-3. But both amplitude and irregularity of changes in monthly and yearly price distributions] are inconsistent with them being caused only by seasonal demand variations given a static supply curve. This is illustrated in Figure 7-4 comparing prices and loads directly.
Prices and loads are positively correlated, but neither do maximum prices correspond to maximum loads, nor maximum loads to maximum prices; the same is true for minimum loads and prices. Applying the MAD rule over all log transformed prices results in no-spike regime boundaries of 14.5 respectively 110 [€/MWh], that are more consistent with the actual distribution than those based on non-transformed prices, 0.85 respectively 79 [€/MWh].

Applying the linear regression function from Figure 7-4 to yearly average loads calculated from national consumption figures (CBS, Statline) allows to access years prior to 2006 for which no reliable direct load data are available, and to model yearly mean prices, that can be compared to mean observed prices.
That observed means and medians differ from 2000 to 2007 is due to a larger proportion of hi spike prices, as already noted. Hi spike prices may and do affect the average non-transformed price. Prices in the lo-spike regime are up to some 10 €/MWh lower than the nearest prices in the no-spike regime, and therefore will hardly affect averages of non-transformed prices. The observed yearly means follow a trend similar to the modelled means as shown by thin black lines, based on linear regression. The year to year variation around the trend in observed means is larger than the variation round this trend in modelled means. This observed year to year variation can be explained by assuming that over periods of several months all prices are higher or lower than can be explained by this very simple supply curve model. That all but one of the modelled mean prices exceed the observed means is due the national consumption figures comprising a larger observation base than the load figures.

7.2 Balancing energy Δ price model
In Section 4.2 I concluded that transactions in balancing energy with the system operator should not be less attractive than energy transactions with other market parties. This single buyer market would otherwise not attract balancing service providers. The Δ's for the balancing service providers are therefore expected to be positive whatever the direction of balancing energy provision. On the other hand, effective competition through an incentive compatible bidding system (Stoft, 2002) ought to keep balancing energy price Δ's as small as possible. Incentive compatible bidding requires a close to real-time balancing gate closure time for updating bids and bid prices. This allows balancing service providers to incorporate more actual information and hence to reduce the level of uncertainty in their bids, thus reducing potential mark-ups in their bid prices. This is the same argument as the one identifying "the temporal position of the gate closure" as "a key parameter of the design of the balancing arrangement" (Saguan & Glachant, 2007). Close to real-time gate
Closure times are particularly important to enable balancing service provision from renewable energy sources and from demand, a requirement from ACER (2012, FWGL EB, 2012).

7.2.1 The Netherlands
I have used public data on whole sale day ahead market prices (APX website) as $P_{OP}$ and balancing energy volumes and prices (TenneT website) to test the model from Chapter 4.2 for balancing energy prices in the Netherlands over 2012 in Figure 7-6. In the Netherlands the gate closure time for balancing energy bids for frequency restoration is one full clock hour prior to the imbalance settlement period of delivery. Bids are per imbalance settlement time period (15 minutes) and the merit order list is open for non-procured bids. Settlement of balancing energy is one marginal price per direction.

![Figure 7-6 Balancing energy price $\Delta$ versus $|\text{ACE}_{OL}(t)|dt$, 2012](image)

The horizontal MW axis shows per imbalance settlement period over 2012 $|\text{ACE}_{OL}(t)|dt$ in [MW] per imbalance settlement period. Net system shortage is to the right of the origin, net surplus to the left. At any imbalance settlement period there may or may not be settlement of balancing energy between system operator and balancing service providers in either direction, up or down. In each imbalance settlement period all transactions are settled at one (marginal) price per direction. In case of counter-activation within a single imbalance settlement period there thus may be two prices and price $\Delta$'s, one for each direction. The vertical [$€/MWh$] axis shows the average volume weighted $\Delta$ [$€/MWh$] per direction for each imbalance settlement period:

$$\Delta = ([Q_{up}] \times (\Delta_{up}) + [Q_{down}] \times (\Delta_{down})) / ([Q_{up}] + [Q_{down}]) [€/MWh].$$

For $([Q_{up}] + [Q_{down}]) = 0$ [MWh], $\Delta$ has been forced to 0 [$€/MWh$].

The cut off at $[600]$ [MW] allows more details to become visible close to the origin. Although the individual data points are rather dispersed, the trend of these empirical data, shown as a (white) 96-point moving average line conforms to a model expected for pay marginal clearing as proposed in Section 4.2:

- $\Delta$'s are on average close to 0 [$€/MWh$], when $|\text{ACE}_{OL}(t)|dt$ is close to 0 [MW].
• in both directions increasing positive Δ's with increasing system imbalance \( \int \text{ACE}_\text{OL}(t) \, dt \),
• an asymmetry in these gradients. The gradient in case of a market shortage that has to be recovered by upward balancing energy is steeper than the gradient in case of a market surplus.

The distribution of Δ's around the trend-line is highly skewed: positive deviations are larger than negative deviations, due to physical resources for frequency restoration being limited in size and/or size and ramp rates. This ramp rate required for automatic frequency restoration may therefore exceed the ramp rate limitations of individual bids, thus requiring using more bids of the supply curve simultaneously but partially. Moving thus farther out on the model Δ curve than the actual \( \int \text{ACE}_\text{OL}(t) \, dt \) will result in higher Δ's for that \( \int \text{ACE}_\text{OL}(t) \, dt \).

The average \( \Delta \) is about 42 [€/MWh] balancing energy; the volume weighted average \( \Delta \) is 66 [€/MWh]. This means that each [MWh] balancing energy that is settled with the system operator earns the balancing service provider 66 [€] over the same transaction on the market.

Nevertheless about 12 % of the data in Figure 7-6 shows negative Δ’s, predominantly at smaller \( \int \text{ACE}_\text{OL}(t) \, dt \) values. In such cases the market value \( P_\text{OP} \) may be 'wrong' (traders’ responsibility). Alternatively the balancing energy bid prices \( (P_\text{OP} + \Delta) \) may be 'wrong' (balancing service providers' responsibility). And finally the balancing energy price Δ model might be wrong (our responsibility).

Firstly, negative average Δ’s per imbalance settlement period occur at relatively small net volumes of balancing energy in Figure 7-7. As in merit order list activation there is at least a qualitative positive relation between the volume of balancing energy and the number of bids activated, negative Δ's therefore are considered to occur at the beginning of the merit order list in each direction. Even with negative Δ’s the small volumes represent relatively small financial losses to the balancing service providers. In a pay marginal settlement of balancing energy, balancing service providers are relatively protected against adverse consequences of their choices e.g. for negative Δ’s, since such consequences will only be apparent at small volumes of balancing energy. In pay-as-bid such risks will be incorporated in mark-ups resulting in higher bid prices.
Secondly, negative Δ's occur in Figure 7-8 at low $P_{OP}$ [€/MWh] in combination with downward regulation, and at high $P_{OP}$ [€/MWh] for upward regulation. Plotting per direction results in a larger number of negative values in Figure 7-8 compared to the netted results in Figure 7-7.

Balancing service providers, even when permitted to update their bid prices with positive Δ's, do not always do so. This suggests that they perceive that $P_{OP}$ is too low, or too high, i.e. is not representative of the marginal system cost $P_{OP}$ in an 'equilibrium' regime but rather established in a 'jump' regime (Section 7.1). This reasoning is supported by the correlation of $\Delta s < 0$ [€/MWh] to $P_{OP}$: the farther $P_{OP}$ is from average $P_{OP}$,
the more negative Δ’s may become. A possible explanation for this seemingly anomalous bidding behaviour, especially at low P_{OP}, might be that P_{OP} at the day ahead stage is established taking into account (floating) block bids, through which producers may sell below marginal cost at some hours, provided that they will recoup their losses in other hours.

Figure 7-9 for 2013 displays a similar pattern as for 2012. This is no surprise as this pattern has established itself quite soon after 2001.

Figure 7-9 Balancing energy price Δ versus ∫ACE_{OL}(t)dt, 2013

The proportion of negative Δ’s has fallen to 9%. Average Δ for balancing energy in 2013 is about 46 [€/MWh] balancing energy. The volume weighted average Δ is 75 [€/MWh]. This is a rather spectacular increase in the cost of flexibility. Over 2010 a volume weighted 27 [€/MWh] is reported for the Netherlands in 2010 (TenneT, 2011). Increasing scarcity of flexibility in a predominantly fossil fuelled system may have several causes e.g. like driving out flexible gas fired plant from the power supply curve by intermittent generation, or abuse of market power by balancing service providers. But increasing scarcity as a phenomenon is not that difficult to demonstrate by analysing real world data.

7.2.2 Germany
A study commissioned by TenneT compared the balancing markets in Germany and the Netherlands (TenneT, 2011). In both countries the imbalance settlement period is 15 minutes, and the system operators activate balancing energy bids as required from a merit order list. Essential differences are that the German design the merit order list contains procured balancing energy bids only, and that energy bid prices are firm before the day ahead market closes. In Germany the energy bid price resolution is peak/off peak during a whole week for automatic frequency restoration bids, and 6 periods of 4 hours per day for manual frequency restoration reserves. This means a rather static merit order list of balancing energy bids to be activated. Another major difference is that settlement of balancing energy is based on pay-as-bid, rather than pay marginal as in the Netherlands.
Not knowing the actual bid curves it can only be said that the curve for Germany in Figure 7-10 contains a.o. the effects of the 'guess the strike price' effect around the origin in the discontinuity of the moving average around 0 [MW], as envisaged in Section 4.2. Other mark-ups due to gate closure times prior to clearing of the day ahead market may also be present. Negative Δ's presumably are attributable to wrongly guessed day ahead market prices. Since the average balancing energy price is calculated as balancing energy cost divided by net balancing energy volume, large Δ's result around the origin in Germany referred to as the "crossing the 0" issue, the consequence of counter-activation. Not using a marginal price is manifest towards the edges, where only little progressive relation of average price to balancing energy volume is observed, nor is there much evidence of asymmetry in this relation. This distinct pattern has of course been noted by others like Just & Weber (2012) who refer directly to average balancing energy prices rather than to balancing energy price Δ's. They note the persistency of this pattern from at least since 2005. Furthermore they note some overall relation to commodity value, especially an increase in average balancing energy prices during periods like the year 2008 of "steeply rising oil and overall energy prices" (Just & Weber, 2012). But their main conclusion is that the [average] "balancing prices are highly dependent on the direction of the imbalance of the control zone, but largely independent from the actual size of the imbalance". This feature has serious impact on the strategy and behaviour of the German balance responsible parties, as the average balancing energy price serves as the imbalance price, assuming that $|\int S_{FRR,RR}(t)dt| + \sum Q \approx 0$ [MWh/ISP] per imbalance settlement period.
7.3 Imbalance margin $\Delta$ value
That imbalance prices are related to the day ahead market price as proxy for the commodity price is evident from designs with dual imbalance prices, where the price for imbalances reducing system imbalance is the day ahead commodity price. Such an implicit $\Delta = 0$ [€/MWh] figures Norway, Sweden, Finland, Denmark, France, Spain, Italy, Slovenia, Lithuania as of 2014 (ENTSO-E, 2014, AS Enquiry), and Belgium up to 2012. To what extent the day ahead price reflects its real-time commodity price is increasingly becoming doubtful, given increasing volumes of intermittent generation entering markets, including intraday markets. Croatia uses an intraday commodity price to this purpose (ENTSO-E, 2014, AS Enquiry) in a dual pricing system.

7.3.1 The Netherlands
The simple and sensible anti gaming condition that for balance responsible parties imbalance should be less attractive than the commodity, at least on average, has been an explicit objective in the imbalance pricing in the Netherlands from its inception (TenneT, 2000). Comparing average imbalance prices to average commodity prices was performed in the Netherlands quickly after its introduction (Beune & Nobel, 2001). The results showed that average prices for positive respectively negative imbalances straddle average day ahead market prices from APX as proxy for $P_{OP}$, thus hinting at positive margin $\Delta$'s for imbalance (Beune & Nobel, 2001). A similar test on this property has been performed more extensively by Boogert & Dupont (2005), who evaluated potential gaming strategies by balance responsible parties, and concluded: “Our results show that profits generated by these strategies are rarely positive on average and always characterized by very large potential losses, which dwarf the mean profit when the latter is positive”.

In the Netherlands it is the users themselves that establish the relation between imbalance prices and market prices through their balance responsible parties or as balancing service providers.

- Balancing service providers are allowed to adjust their balancing energy bids and bid prices close to delivery. This allows incorporation of up to date information on availability and price in their balancing energy bids. Publication of aggregate merit order list information allows the balancing service providers to place their bids at their preferences on the merit order list. Marginal pricing however invites to bidding at marginal cost.

- Balance responsible parties are allowed freedom of dispatch to manage their imbalance volume risk. Through the aggregate merit order list information they are enabled to assess imbalance price risks. Through close to real-time on relevant system information on actual balancing energy prices they are enabled to manage their own imbalance, even without knowing their own actual position.

Data from the Netherlands for 2012 show in Figure 7-11 on the horizontal MW axis $\int ACE(t) dt$ in average [MW] per imbalance settlement period. Net system shortage is to the right of the origin, net surplus to the left. At any imbalance settlement period there may, or may not be settlement in imbalance between system operator and balance responsible parties in either direction, shortage or surplus. Per imbalance settlement period all transactions are per direction settled at the same price, resulting in different margin $\Delta$'s over the corresponding transaction against day ahead whole sale market prices.
Figure 7-11 Imbalance margin $\Delta$ versus $\int\text{ACE}_{\text{OL}}(t)\,dt$, NL 2012

The vertical axis shows the volume weighted average $\Delta$ in [€/MWh] for each imbalance settlement period:

$$\Delta = \frac{(|Q_{\text{shortage}}| \cdot (P_{\text{shortage}} - P_{\text{OP}}) - |Q_{\text{surplus}}| \cdot (P_{\text{surplus}} - P_{\text{OP}}))}{(|Q_{\text{shortage}}| + |Q_{\text{surplus}}|)} \ [€/\text{MWh}]$$

For $(|Q_{\text{up}}| + |Q_{\text{down}}|) = 0 \ [\text{MWh}]$, the $\Delta$ has been forced to 0 [€/MWh].

The rationale for choosing these parameters is that the data is publicly available, at least for the Netherlands; ACER (2012, FWGL EB) will make publication obligatory. The figure is cut off at $|600 \text{ MW}|$, to allow detail to become visible, and shows a 96 point moving average. These represent a day's worth albeit obviously not a calendar day. Although the individual data points show considerable dispersion the trend of these data, shown as the 96-point moving average line (white) conforms to the margin $\Delta$ model:

- margin $\Delta$'s are close to 0 [€/MWh] when $\int\text{ACE}_{\text{OL}}(t)\,dt$ is close to 0 [MWh/ISP],
- increasing positive margin $\Delta$'s with increasing system imbalance in both directions,
- an asymmetry in the gradients; the gradient in case of a market shortage that has to be recovered by upward balancing energy, is larger than the gradient in case of a market surplus,
- an increase in volatility of the margin $\Delta$’s with increasing system imbalance in both directions.

The distribution of $\Delta$'s around the trend-line is highly skewed: positive deviations are larger than negative deviations. It is not only the increase in the margin $\Delta$ [€/MWh] way from equilibrium, but also the increase in volatility that may act as an incentive to balance responsible parties to avoid aggravating $\int\text{ACE}_{\text{OL}}(t)\,dt$ [MWh/ISP]. Both these trends support the conclusions of Boogert & Dupont (2005) that the imbalance pricing system in the Netherlands is robust, or gaming-proof.

The volume weighted average margin $\Delta$ over 2012 was 19.6 [€/MWh] i.e. to a balance responsible party the average MWh settled in imbalance with the system operator lost him 19.6 [€/MWh] over the same
transaction at commodity price $P_{\text{OP}}$. For 2013 a similar pattern as for 2012 emerges as shown in Figure 7-12. This is no surprise as this pattern has established itself quite soon after 2001.

![Image of Figure 7-12 Imbalance margin $\Delta$ versus $\int \text{ACE}_{OL}(t) \, dt$, NL 2013](image)

In 2013 the average $\Delta$ has increased to 22 [€/MWh], again indicating the increasing cost, and hence increasing scarcity of flexibility

### 7.3.2 Germany

Assuming that per imbalance settlement period $\int S_{\text{FRR, RR}}(t) \, dt + \sum Q \approx 0$ [MWh/ISP], the $\Delta$ imbalance margin, valid for all German users, equals the averaging balancing energy price $\Delta$. Plotting against the net balancing energy activated by the German system operators, reveals fundamental differences to the Dutch pattern due to differences in market design (Figure 7-13).
The imbalance margin Δ, like the balancing energy price Δ is largely independent from the actual size of the actual system imbalance. The relation between system imbalance and imbalance margin Δ does not have a stable centre. It may even issue perverse incentives to balance responsible parties to undersupply at times of peak, and oversupply at times of negative commodity prices (Just & Weber, 2012). Therefore the actual imbalance prices, and therefore the margin Δ’s for imbalance are not always equal to the average balancing energy price Δ. They are presently, and in the depicted period, capped or floored at intraday market prices. Imbalance is no longer a profitable alternative to the balance responsible party over intraday trade.

7.4 Energy parameter choice for the net system state ACE\textsubscript{OL}

Data from the Netherlands show in Figure 7-14 and Figure 7-15 per imbalance settlement period the general equality of average block \( \int S_{(F_{rr},rr)}(t)dt \) [average MW] and \( \sum (Q_{\text{surplus}}, Q_{\text{shortage}}) \) [average MW], the joint net imbalance of the market. Deviations are mainly due to ramping over all interconnectors, frequency deviations over the imbalance settlement period, temporal allocation errors in the net imbalance of the market, and imbalances not attributable to a balance responsible party, e.g. the NorNed HVDC link to Norway.
There is no reason to reject the joint net imbalance [MWh] per imbalance settlement period of the balance responsible parties as a sufficiently accurate proxy for $\int ACE_{OL}(t)dt$ over that imbalance settlement period.

7.5 Conclusions and discussion on testing models and design choices
The proposed conceptual commodity price model of a continuous power supply curve intersected by an inelastic, variable load cannot be rejected by empirical data. Even if empirical prices are derived from clearing a dual sided residual day ahead auction, and even without having access to actual power production data. The observed breakdown of this model at the lower and at the upper ends of the inferred supply curve is manifest in price spikes: hi spikes at the upper end, lo spikes at the lower end of the inferred supply curve.
The occurrence of spikes is due to, and increased by the day ahead markets residual character, where the urgency of must buy at the upper end, or must sell, at the lower end prevails over the available opposite trades. 'Jumps' reveal occasional scarcity, or incompressibility of power production. In fact, positive and negative price spikes on day ahead markets may be beneficial in a market based power system. Not only to recover fixed cost, but as well as forewarning of scarcity, enabling users to consider additional resources to cope with or react on such temporary market conditions. In this way potential physical crises are translated into 'financial' crises to be dealt with by all market parties, rather than by only a central dispatch organisation. A testable hypothesis is that fuel prices may drive the entire supply curve up or down on monthly scales in a thermal based power system as the Netherlands. The supply curve then would change its general shape only over longer time scales.

That the power supply model cannot be rejected means that the power supply model for balancing energy prices cannot be rejected either at this stage. It is therefore admissible to use it as basis for subsequent analysis of balancing energy prices according to the analytical model proposed in Section 4.2.

This proposed balancing energy price $\Delta$ model is able to extract relevant information from bidding behaviour, even if the bids themselves or the complete merit order lists per imbalance settlement period are not directly assessed. The proposed model is impervious to the occurrence of negative prices, be it for commodity or for balancing energy. Apparent anomalies resulting from applying the model to data from the Netherlands like negative $\Delta$'s for balancing energy can find a reasonable explanation, without invalidating the conceptual balancing energy price model.

Comparing the results of balancing energy price formation for Germany and the Netherlands supports the conclusion that pay-as-bid in practice does not necessarily contribute to lower cost. Even when considering that the $\Delta$'s are calculated on the net balancing energy, the distribution of $\Delta$'s in Figure 7-10 clearly shows evidence of the guessing the strike price strategy of German balancing service providers.

Compared to the Netherlands the major differences between imbalance margin $\Delta$'s for Germany are:

- an imbalance margin $\Delta$ [€/MWh] peak close to 0 [MW] balance in Germany, whereas in the Netherlands close to system balance the imbalance margin $\Delta$ approaches 0 [€/MWh]. For German balance responsible parties this means that large sums of money are passed on from those balance responsible parties aggravating system imbalance to those reducing system imbalance in a fairly random way,
- a marked discontinuity in imbalance margin $\Delta$'s at 0 imbalance,
- smaller and constant volatility of imbalance margin $\Delta$ in relation to overall imbalance in comparison to the Netherlands, where volatility increases with increasing system imbalance,
- little progression of imbalance margin $\Delta$'s with increasing overall imbalance in Germany in comparison to the Netherlands.
Together with the relatively small and constant volatility this means that the imbalance margin $\Delta$ can be estimated fairly well by German balance responsible parties, and that there is hardly an incentive to reduce individual or system imbalance.

Based on these observations I conclude that the German imbalance settlement mechanism functions as a tariff, and not as a real-time market.

The net imbalance of balancing parties $Q_{\text{surplus}} + Q_{\text{shortage}}$ [MWh] can be used as a good approximation of $\int \text{ACE}_\text{nl}(t) \, dt$ [MWh]. An advantage of using this parameter for most of the analyses shown in this thesis is easy access.

This analytical model is obviously not restricted to a yearly resolution. Anybody can search for seasonal patterns. Likewise it can be applied to other designs, given availability of balancing energy prices and volumes, and commodity prices.
8. Testing design choices

The Electricity Regulation and the Electricity Directive requires "cost-reflective balancing mechanisms", and fulfilling "the need for establishing objective, fair, transparent and non-discriminatory rules for balancing in a cost-reflective way" (ACER, 2012, FWGL EB). There is no standard market design implemented to accomplish such objectives, as can be observed in the variety of balancing market design choices published by ENTSO-E (2014, AS Enquiry). These choices are not fully independent.

The simple model in Figure 8-1 relates the demand of user services required by the system operator to be compliant on balancing on the y-axis to user services supply.

If demand of balancing services by the system operator is larger than the users are prepared to offer, the system operator is non-compliant, and the system is considered inadequate (red triangle). If demand by the system operator equals the users supply, the system operator is compliant. The system is considered adequate, but there is no competition among users. In case of market based balancing service provision, there is absolute market power (red line). The challenge of a market design is to move away (green arrow) from the red field, to become compliant in adequacy and, in case of market based provision, away from the red line to increase efficiency by competition.

The two trivial different ways to accomplish this are:
- reduce balancing services demand from the system operator: move down,
- increase balancing services supply from the users: move right.

One of the easier ways to increase balancing services supply is by removing entry barriers or exclusion mechanisms within the control area of the system operator:
allow for non-procured balancing energy bids (ENTSO-E, 2014 NC EB),
allow all balancing energy bids to be priced by the balancing service provider,
allow for close to real-time gate closure time for balancing energy bids, to enable intermittent and demand resources to participate in balancing energy bidding (ENTSO-E, 2014 NC EB).

More complicated is to remove barriers to external balancing services provision:
- reserve sharing for capability (ENTSO-E, 2013 NC LFC-R),
- common merit order lists for more liquidity on merit order lists (ENTSO-E, 2014 NC EB).

The balancing services demand of the system operator is schematically decomposed in Figure 8-2:
- the results of its users (e.g. size of connection, joint net imbalance),
- the compliancy targets of the system operator (e.g. n-1 security, time to restore frequency),
- the translation of user results into a balancing services demand (grey arrow, e.g. probabilistic or deterministic methods).

Each component can be tweaked to reduce the system operator´s demand for user balancing services:
- the user results can be improved/reduced by restricting users,
- the user results can be improved/reduced by incentivizing users,
- compliancy targets can be relaxed,
- the translation from user results to balancing services demand can be optimized.
8.1 Restricting users

A clear example of restrictions imposed on users is the conditionality of system access provision by requiring balance responsibility. Another restriction on users may follow from the minimum compliancy target for system operators or securing frequency restoration reserves: "the largest imbalance that may result from an instantaneous change of active power of a single Power Generating Module, single Demand Facility, and single HVDC interconnector or from a tripping of an AC-Line within the LFC Block" (ENTSO-E, 2013, NC LFC-R). On failure these will cause a power imbalance that the system operator will have to reduce within the time to restore frequency. Therefore, the system operator requires access to frequency restoration reserves from balancing service providers. Securing the required frequency restoration reserves is required to demonstrate commitment to the joint stable frequency responsibility shared with other system operators, whether or not there is actual imbalance. This deterministic dimensioning requirement is a pure security insurance issue. However, resolving capacity for this purpose from the market takes away flexibility otherwise available within the market. Secondly, it comes at a price, which is not linearly related to the volume. Twice as much will more than double the cost. Security insurance costs are presently socialized in most designs through regulated access tariffs.

To reduce some of this cost system operators themselves may want to limit new component size, like HVDC equipment, in case it exceeds current dimensioning requirements. Balance responsibility requires a balance responsible party to settle its imbalance in the event of a large power generating module or demand facility. This is an individual risk that the user has to take into account when considering commissioning an installation larger than anything pre-existing. The system operator on the other hand is obliged to accommodate this individual risk and to socialize associated compliancy costs over all users. Therefore restrictions could be set on maximum size to user connections limiting it to existing dimension requirements. This of course is clearly a limitation on the users’ freedom to connect. Alternatively part of reserve requirements may be delegated to the user, thus reducing the volume of capacity remunerations to be socialized by the system operator.

8.2 Incentivizing users

If imbalance pricing is to reflect scarcity of (remaining) frequency restoration reserves a positive and progressive increase of imbalance margin $\Delta s$ [€/MWh] to $\int \text{ACE}_{OL}(t)dt$ or joint net imbalance $\sum Q$, [MWh] of balance responsible parties (market imbalance) is expected. In the following comparison in Figure 8-3 between Germany and the Netherlands the absolute net market imbalances have been binned in 50 MW intervals. Per interval the average imbalance margin $\Delta s$ [€/MWh] and the number of imbalance settlement periods in that interval have been established. Using absolute values discards any directional asymmetries in the relation between imbalance margin $\Delta$ and the joint net imbalance, but it facilitates direct comparison among different pricing systems.

Different approaches are apparent from the results from the different market designs. Whereas the German design result in rather constant $\Delta s$, the Netherlands design allows for the market to establish (by balancing service providers) and discover (by balance responsible parties) a real-time scarcity related imbalance value.
Therefore the incentive to avoid aggravating imbalances in the Netherlands is much larger than in Germany. The incentive is enhanced by the increase in volatility in imbalance margin Δ's in the Netherlands with increasing net system imbalance, as opposed to the more constant volatility in Germany shown in Figure 7-10. Together with the timely feed-back on system operator actions and prices in the Netherlands, financially attractive opportunities for balance responsible parties to increase system supporting imbalances are much larger in the Netherlands as well. This enables all sorts of flexibility from users to participate in what is effectively a real-time balancing market (Glachant & Saguan, 2007). Consequently, system imbalances in the Netherlands are concentrated at smaller volumes, whereas in Germany they are spread over a much larger interval up to |4 GW|, all included in the rightmost bin. Smaller system imbalances require less balancing energy \( \sum |S_{\text{FRR,RR}}(t)| dt \), thus considerably contributing to the optimization objective of a system operator in a pay marginal balancing energy design. The much larger range of imbalances to be restored by the system operators in Germany requires them to procure huge amounts of frequency restoration reserves.

Above the 200 MW bin, the increasing imbalance margin Δ's between and the Netherlands and Germany are more than offset by a decrease in volumes. Consequently the generally smaller imbalance margin Δ's in Germany conceal the fact that imbalance in Germany was over all more costly to balance responsible parties than it is in the Netherlands. The comparison study mentioned in Section 7.2.2 duly reported (TenneT, 2011):
Just & Weber (2012) also address the issue of perverse incentives to balancing service providers and balance responsible parties in the German market inherent to what they call a balancing energy mechanism: "not really a market, but rather an accounting procedure". They identify the symptoms as basically due to the lack of coherence with previous markets, and the resulting lack of incentives to balance responsible parties to balance themselves or the system in energy. The root cause of this malfunctioning is the German reserve capacity market design. It is essentially disjunct from the long term, day ahead and intraday energy markets. This results in inconsistent prices, and lack of balancing incentives to balance responsible parties as shown already in sections 7.2.2 and 7.3.2. In their proposed solutions Just & Weber (2012) address symptoms rather than fundamental root causes. In this thesis some of the fundamental root causes are indicated, like the temporal resolution of bid prices, the pre day ahead firmness of balancing energy bid prices, the pay-as-bid clearing of balancing energy, and the small divisor issue in the calculation of the imbalance prices.

In 2012, i.e. after TenneT published its comparison study on market design in Germany and the Netherlands, the German regulator explicitly incorporated a condition into the methodology of determining imbalance prices in Germany that imbalance should not be beneficial at least to intraday market prices by capping or flooring imbalance prices by the intraday market price for the same imbalance settlement period.
This is a clear attempt to reduce perverse incentives to balance responsible parties by imposing a cap respectively floor on the imbalance price based on the intraday price for that imbalance settlement period. In addition it attempts to enhance the relation of the imbalance price to scarcity, by imposing a conditional scarcity related component on the imbalance price. Although these adaptations entered into force per December 2012, it took some time to affect the behaviour of the balance responsible parties. In 2013 the average of absolute net activated balancing energy $\sum |S_{(FRR, RR)}(t) dt|$ was 122 [MWh/ISP], its 99th percentile was 543 [MWh/ISP]. Over the first 11 months of 2015 these values had dropped to 92 respectively 408 [MWh/ISP]. Not only has the volume of balancing energy been reduced by about 25%, also the range has decreased. Under stochastic dimensioning this would considerably reduce the volumes reserve capacity to be procured to comply to dimensioning requirements, and thus release capacity on the intraday market. An alternative explanation that this reduction was due to imbalance netting cannot be substantiated as the German potential for imbalance netting had already been largely fulfilled by 2012 (IGCC workshop, 2014).

Obviously this specific approach by the Bundesnetzagentur does not address other flaws in the German balancing markets. But it does show that attempts to reduce volume requirements by internal market reform can be accomplished with significant results.

Recently the British regulator proposed to change the present dual imbalance price, with separate generation and load positions for each balance responsible party, and based on average balancing energy prices (OFGEM, 2014). It proposes a single pricing system, eventually based on the marginal balancing energy price, thus removing the necessity for separate positions or injection and withdrawal. Neither the imbalance settlement period, presently 30 minutes, will be changed, nor will the balancing energy market design be changed.

In 2012 the imbalance settlement in Belgium was changed from dual to single pricing. The result was a steady decrease of $ACE_{OL}$ as shown in Figure 8-4.
8.3 Limiting user incentives

Incentivizing balance responsible parties by imbalance pricing to reduce system energy imbalance may impose risks on system stability and local grid security as reasoned in Section 4.3.1. A possible solution to both drawbacks associated with single imbalance pricing is to remove or reduce the incentive to balance responsible parties to increase the system energy reducing imbalance \( |Q_{\text{reducing}}| \) [MWh].

A trivial option to avoid such an incentive is not to apply single imbalance pricing (Section 4.3.1: option c). All European countries with imbalance settlement periods longer than 15 minutes either apply dual pricing (Section 4.3.1: option b) or distinguish separate imbalance volumes per balance responsible party for generation and for loads, with at least one of these imbalance volumes (generation) subject to dual pricing (Section 4.3.1: option b) (ENTSO-E, 2014, AS Enquiry, 2014). Obviously this requires separate notifications per balance responsible party, and clear criteria which connection is regarded as generation, and which one as load. The Belgian introduction of single pricing described in Section 8.2 is reverse evidence of the initial effectivity of this option to withhold incentives from balance responsible parties.

Imbalance pricing in the Netherlands, with an imbalance settlement period of 15 minutes, is hybrid. In most of the imbalance settlement periods single pricing (Section 4.3.1: option c) is applied. When counter-activation of balancing energy by the system operator has occurred, dual imbalance pricing (Section 4.3.1: option a) can and will be applied according to predetermined public rules and public data (ACM, Systeemcode).
The incentive to reduce $|Q_{\text{reducing}}| [\text{MWh}]$ depends on $\Delta_{\text{reducing}} > 0 [\text{€/MWh}]$. To weaken this incentive from single imbalance pricing (Section 4.3.1: option c) this margin $\Delta_{\text{reducing}}$ should increase only slightly with increasing scarcity of balancing energy bids. This can be achieved by design, e.g. by applying pay-as-bid settlement for balancing energy and using average balancing energy cost as imbalance price as shown in 7.3.2 for Germany. The German introduction of a scarcity related component to the imbalance price mentioned in Section 8.2 is reverse evidence of the initial success of this option to withhold incentives from balance responsible parties.

Another option is not to have the potentially most volatile element of balancing energy, i.e. automatic frequency restoration reserves, determine the balancing energy or the imbalance price. This can be achieved by pro rata activation of balancing energy bids for automatic frequency restoration, and/or by fixed prices for balancing energy bids for automatic frequency restoration (Elia & TenneT, 2014).

The most stringent limitation on user incentives is by delineation of areas. Most continental European countries are a single control block with responsibilities to frequency containment and frequency restoration (ENTSO-E, 2013, NC LFC-R, 2013). This implies that balancing energy settlement and balancing energy prices are per control block. Most continental European countries are a single bidding zone. This implies that within this bidding zone no capacity allocation is required for market participants to exchange energy within this area (Regulation (EC), 2013, 543/2013). This implies that the bidding zone is the geographical area for which the balance responsible party will notify its position, and for which its imbalance will be settled. Imbalance settlement and imbalance prices are per bidding zone. This configuration therefore allow for consistent prices for balancing energy and imbalance within a bidding zone, and having balancing energy scarcity confined per bidding zone. It follows that any real time incentives to balance responsible parties are confined per bidding zone. And since all real time incentives in a bidding zone and balancing energy activation by the system operator aim at reducing system imbalance of the corresponding control block, such incentives will not induce congestion between bidding zones. Balancing responsible parties currently cannot reduce or displace their imbalances over borders between bidding zones. The importance of this observation is underlined in the interconnected Nordic power system, where bidding zones deal with potential capacity limitations on the transmission network on the day ahead market. Potential bottlenecks will entail different balancing energy and imbalance prices. Imbalance settlement per bidding zone, and dual pricing ensure that measures to prevent violation of capacity limitations between bidding zones are nullified by perverse incentives from imbalance prices.

Another consequence of settlement of balancing energy and imbalance per (national) bidding zone is that the financial neutralization of the system operator as explained in Chapter 6, can be performed by country, and will affect only the receivers of the system operator services.
8.4 Imbalance netting as case example of adapting translation scheme

The frequency restoration process per control area in a synchronous zone can be extended to be combined with that of others (ENTSO-E, 2013, NC LFC-R, 2013; BNA, 2009; IGCC workshop, 2014). This may avoid counter-activation of automatic frequency restoration reserves in separate control areas, given available cross zonal capacity. Imbalance netting is based on the triangle inequality theorem that for quantities A, B: |A| + |B| ≥ |A + B|. The imbalance netting extension to the frequency restoration process is shown in Figure 8-5:

\[ |A| + |B| \geq |A + B| \]

Imbalance netting is based on the triangle inequality theorem that for quantities \( A \), \( B \):

\[ |A| + |B| \geq |A + B| \]

The imbalance netting process aims at prevention of simultaneous counter-activation of automatic frequency restoration reserves by different system operators, by combining their individual automatic frequency restoration demands. It is part or all of the automatic frequency restoration demand [MW] of a system operator, that is cancelled out by opposite frequency restoration demands from other system operators, cooperating in this process. It results in an intentional virtual tie-line exchange (interchange). Imbalance netting is real-time and firm, but may be limited by available cross zonal capacity. Compared to activation of frequency restoration reserves by the system operator, that all have delay times and ramp rate limitations and therefore are somewhat inaccurate and imprecise with respect to their impact on the frequency restoration error, the imbalance netting process is an immediate and perfect virtual tie line exchange. In
addition to the variables identified and explained previously in Chapter 3, CIN [MW] is the exchanged power, and \( \int \text{CIN}(t)dt \) [MWh] the energy attributed to the imbalance netting process.

<table>
<thead>
<tr>
<th>Power [MW]</th>
<th>Net Energy/ISP [MWh] ( t = \text{start, end ISP} )</th>
<th>Absolute Energy/ISP [MWh] ( t = \text{start, end ISP} )</th>
<th>( \sum \text{Absolute Energy} [\text{MWh}] ) ( t = \text{start, end ISP} ) ISP = 1, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{ACE}(t) )</td>
<td>(</td>
<td>\text{ACE}(t)dt</td>
<td>)</td>
</tr>
<tr>
<td>( \text{CIN}(t) + \int S_{\text{FRR, RR}}(t)dt )</td>
<td>(</td>
<td>\text{CIN}(t)dt + \int S_{\text{FRR, RR}}(t)dt</td>
<td>)</td>
</tr>
<tr>
<td>( \text{ACE}_{\text{OL}}(t) )</td>
<td>(</td>
<td>\text{ACE}_{\text{OL}}(t)dt</td>
<td>)</td>
</tr>
</tbody>
</table>

Table 2 Power and energy variables in balancing process with imbalance netting

Imbalance netting reduces the activation of automatic frequency restoration reserves by system operators. Therefore it reduces balancing services demand, rather than enlarges supply.

In the Netherlands imbalance netting has been implemented since February 2012. Its impact is illustrated in Figure 8-6 using the methodology developed in Section 3.1. It shows the monthly totals of absolute net energy values [\[MWh]/month\] of the net imbalance of balance responsible parties’ \( \sum |Q| \), as a proxy for \( \sum |\int \text{ACE}(t)dt| \). In a situation of no control this would have resulted in the same value for \( \sum |\int \text{ACE}(t)dt| \). The net absolute volume contribution by the system operators, or through cooperation with other system operators is \( \sum |(\text{CIN} + \int S_{\text{FRR, RR}}(t)dt)| \) is shown as is the difference between \( \sum |Q| \) and \( \sum |(\text{CIN} + \int S_{\text{FRR, RR}}(t)dt)| \), the best attainable (theoretical) result for \( \sum |\int \text{ACE}(t)dt| \). Before February 2012, without imbalance netting CIN was 0 [MW], and therefore \( \sum |(\text{CIN}(t)dt + \int S_{\text{FRR, RR}}(t)dt)| = \sum |S_{\text{FRR, RR}}(t)dt| \) [MWh].

Figure 8-6 Impact imbalance netting on \( \sum |\int \text{ACE}(t)dt| \) in the Netherlands
Imbalance netting has resulted in a considerable reduction of $\sum |\int S_{FRR, RR}(t)dt|$ as shown in Figure 8-7, thus considerably contributing to the balancing energy volume reduction objective of a system operator in a pay marginal balancing energy design.

At a first glance it is obvious that with the onset of imbalance netting for the Netherlands at "Start IGCC" $\sum |\int (CIN(t)dt + \int S_{FRR, RR}(t)dt)|$ increased significantly, enabling a better theoretical result. Secondly the actual control quality improved considerably as indicated by the significant decrease shown in actual $\sum |\int \text{ACE}(t)dt|$.

Imbalance netting not just replaced activation of balancing energy from frequency restoration reserves, but $\sum |\int (CIN(t)dt + \int S_{FRR, RR}(t)dt)|$ has actually increased compared to the time that no imbalance netting occurred. This is due to the immediacy of the imbalance netting process, i.e. absence of delay times and ramping limitations which are inherent to the physical nature of the frequency restoration process, compared to the virtual tie line exchange of imbalance netting.
9. Conclusions and discussion on balancing market design

The responsibilities of different actors, users and operators, with respect to the physical components of a power system and to their legal and regulatory limitations and obligations from European legislation are determined according to the conceptual framework developed in this thesis in Chapter 2. This enables to map the contributions of users and system operators in maintaining and restoring the active power and energy balance in a power system in Chapter 3. Since the cost of balancing is the result of volumes and prices, different methods of price derivations are explored in Chapter 4. Different design choices for temporal resolution are discussed in Chapter 5. Different options for price derivations and settlement of balancing energy may have a profound impact on the cost optimization possibilities of the system operator, and vice versa of its users as argued in Chapter 6.

This price decomposition proposed in Chapter 4 facilitates insight in the merit order list for balancing energy bids without requiring access to actual bid data. The $\Delta$ price method allows relating actual prices to actual scarcity, and allows for a direct comparison among different designs, and over longer periods. Moreover reasoning according to this model provides testable hypotheses.

The system operator is responsible for being compliant to power balancing requirements. Consequential necessary conditions then are adequate and secure remunerations for its expenditure, and predetermined, dedicated allotment of revenues. Together, these conditions neutralize the system operator financially and secure its financial disinterest in the balancing processes. This prevents the fundamental flaw of offering potential perverse incentives to system operators by design.

The balance responsible parties, representing the users are financially responsible for settlement of their energy imbalances. Under the necessary condition that settlement of their imbalance is on average, less attractive than the corresponding market value, they become de facto responsible for the energy balance. A fundamental design choice is whether energy imbalance is the real-time product on an energy only market, or whether imbalance serves as a system operator cost recovery mechanism.

Energy imbalance as a real-time product on an energy only user market

If imbalance is the real-time product on an energy only market, it is the termination to the series of long term, day ahead and intraday markets, where energy products and positions are traded among users. In that case the system operator is responsible only for compliance to its power balancing requirements and the system energy balance becomes primarily the responsibility of the users. Provision of balancing energy by the system operator is a result, and not the objective of power balancing. This choice allows, invites, and basically incentivizes active participation and competition between imbalance and balancing energy. This choice requires a consistent set of design features:

- an imbalance settlement period not longer than time to restore frequency,
- imbalance prices equal to balancing energy prices (not more, not less),
- balancing energy price to reveal scarcity,
- real-time system balancing information feedback to all users.
A key element is the pricing and settlement of balancing energy. Bid prices [€/MWh] from balancing service providers can be decomposed in marginal cost and mark-ups. The marginal cost is in principle related to the system marginal cost or commodity value. Efficient market design, being incentive compatible (Stoft, 2002), should induce, or at least allow for the reduction of these mark-ups. The mark-ups are risk premiums due to the balancing service provider’s estimates of:

- the opportunity cost or market value of its resources,
- the availability of its resources
- the difference between the settlement price and its bid price.

The first two mark-ups or risk premiums are related to the balancing energy gate closure time. Late gate closure times enable balancing service providers to assess market prices. Market prices relate to scarcity of the commodity, and therefore form a baseline for bid prices. The mark-up for the estimate of the market value may be negative, in case the balancing service providers judge public market clearing prices as excessive. They should be allowed to do so. Early gate closure times increase risk premiums due to the estimate of the market value, and are effectively exclusion mechanisms for potential sources of flexibility, such as intermittent generation, demand side management and storage. Such resources require the admittance of non-procured bids on the merit order list. The third mark-up is reduced to 0 [€/MWh] in case of pay marginal settlement of balancing energy. Pay marginal settlement of balancing energy, however, necessitates short imbalance settlement periods.

**Imbalance as a system operator cost recovery mechanism**

In the alternative, the imbalance serves as a remuneration source for the system operator. In that case the system operator not only has to be compliant to its power balancing requirements, but also becomes partly responsible for energy balancing and optimization. Typical distinguishing features of such designs are:

- imbalance settlement period > 15 minutes,
- a reserve replacement process performed by the system operator,
- early gate closure times for trading and for bidding for balancing energy.

Such a design requires several implicit or explicit exclusion mechanisms, to avoid competition between imbalance and balancing energy, a.o.:

- pay-as-bid for balancing energy,
- dual imbalance pricing.

These exclusion mechanisms make real-time system balancing information feedback to users superfluous. Exclusion mechanisms may be prejudicial against some users. Moreover, these designs require much more centralized decision making, centralized resources, and therefore centralized risks. In such a design the balance responsible party is more (passive) wallet than active market participant.

The design choices and their consequences have been tested in Chapter 7. Most European countries have established imbalance as a cost recovery mechanism, as can be inferred from inclusion of one or more of the above listed features in their market design. Based on these tests the real-time balancing market design
in the Netherlands indeed resembles an energy only market. The German design contains too much exclusion mechanisms to function as a real-time energy market.

These design choices affect the balance between demand and supply of user balancing services that are explored and tested in Chapter 8. Imbalance as a market product reduces the demand of user balancing services, just like imbalance netting, as shown in Figure 9-1:

**Figure 9-1 Options to reduce user’s balancing services demand**

Late gate closure times (GCT), non-procured bids and common merit order lists (CMO) aim at expanding the supply of user balancing services providers, as shown in Figure 9-2:

**Figure 9-2 Options to increase user’s balancing services supply**
By combining merit order lists from several system operators into common merit order lists indeed does increase supply, but also combines the demands of balancing energy from different system operators, and therefore are less effective than simply reducing such demands.

Single imbalance pricing does not abide well with congestions, unless imbalance is calculated per congested area, and the imbalance price margin $\Delta_{\text{reducing}}$ small, and not highly correlating with scarcity. Incentivizing users to increase $|Q_{\text{reducing}}|$ [MWh] requires bidding zones without internal congestions. This clearly limits applicability of a design with real time scarcity pricing for balancing energy and imbalance in the real world.

Some options are hardly compatible with each other. Common merit order lists work best in a pay-as-bid balancing energy optimization, and presumably are invented to such purpose. But that would be incompatible with a real-time market and market incentives. Late gate closure times and non-procured balancing energy bids result in dynamic merit order lists, which are less suitable for energy optimization.

The choice for a more centralized approach through the system operator or a decentralized approach through market incentives to reach energy balanced imbalance settlement periods is of particular significance in a system that will have to accommodate ever increasing intermittent generation and localized storage. A choice for a decentralized approach i.e. an energy only market depends on the actual grid resembling a copper plate, which was one of the major assumptions in this thesis.

Some other options like imbalance netting and reserve sharing are rather universal and represent no regret options. These kinds of improved cooperation and coordination between system operators most closely resemble the effect of reducing the number of system operators in Europe. But reduction of the number of system operators in Europe would require political determination, just as would a vast reduction of congestions by re-enforcement of the existing transmission system. None of these are foreseen in the near future.

Future research may use concepts developed in this thesis to check the European Guidelines on the presence of implicit and explicit exclusion mechanisms. Future research may use methodologies developed in this thesis on actual data from European balancing markets to assess the presence or absence of inefficient mark-ups in prices on balancing markets.

No countries have changed their initially selected imbalance settlement period yet. Although ACER initially promoted a harmonized imbalance settlement period not exceeding 30 minutes (ACER, 2012, FWGL EB), more recently they appear to prefer a 15 minute imbalance settlement period (ACER, 2015, QR EB). Future research may reveal if a change in imbalance settlement period will lead to other market reforms in the affected countries.
A. Appendices

Business cycles
The business cycles diagrams map process timings in two time dimensions. The vertical axis displays the distance in time from the past to the present of the start or time of a decision. The horizontal axis displays the distance in time from the present into the future of the duration or realization. Time units are year \([Y]\), month \([M]\), day \([D]\), imbalance settlement period \([ISP]\) and the present is indicated \(RT\). LT, DA and ID are respectively long term, day ahead and intra-day. ICE is interconnector capacity entitlement, the right of a market party to access a market over an interconnector. FR and FC stand for the frequency restoration respectively frequency containment process (Section 2.4). These maps are extensions of a one-dimensional graph given by Frunt (2011).

Interconnectors
Controllable generation business cycle

System operator business cycle
USA experience and expertise

Ancillary services as identified previously in US based markets by NERC’s Interconnected Operations Services working group do not yet show a clear directionality, nor do they differ between grid, system and user services (Hirst, 1999).

<table>
<thead>
<tr>
<th>Service</th>
<th>Definition</th>
<th>Time scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System control</strong></td>
<td>The control-area operator functions that schedule generation and transactions before the fact and that control some generation in real-time to maintain generation/load balance</td>
<td>Seconds to hours</td>
</tr>
<tr>
<td><strong>Voltage control from generation</strong></td>
<td>The injection or absorption of reactive power from generators to maintain transmission-system voltages within required ranges</td>
<td>Seconds</td>
</tr>
<tr>
<td><strong>Regulation</strong></td>
<td>The use of generation equipped with automatic-generation control (AGC) to maintain minute-to-minute generation/load balance within the control area to meet NERC control-performance standards</td>
<td>~1 minute</td>
</tr>
<tr>
<td><strong>Contingency reserve spinning</strong></td>
<td>The provision of unloaded generating capacity that is synchronized to the grid that can respond immediately to correct for generation/load imbalances caused by generation and transmission outages and that is fully available within 10 minutes to meet NERC’s disturbance-control standard</td>
<td>Seconds to &lt;10 minutes</td>
</tr>
<tr>
<td><strong>Contingency reserve supplemental</strong></td>
<td>The provision of generation capacity and curtailable load used to correct for generation/load imbalances caused by generation and transmission outages and that is fully available within 10 minutes</td>
<td>&lt;10 minutes</td>
</tr>
<tr>
<td><strong>Energy imbalance</strong></td>
<td>The use of generation to correct for hourly mismatches between actual and scheduled transactions between suppliers and their customers</td>
<td>Hourly</td>
</tr>
<tr>
<td><strong>Load following</strong></td>
<td>The use of generation to meet the hour-to-hour and daily variations in load</td>
<td>10 minutes to hours</td>
</tr>
<tr>
<td><strong>Backup supply</strong></td>
<td>Generating capacity that can be made fully available within 30 to 60 minutes to back up operating reserves and for commercial purposes</td>
<td>30 to 60 minutes</td>
</tr>
<tr>
<td><strong>Real-power-loss replacement</strong></td>
<td>The use of generation to compensate for the transmission-system losses between generators and loads</td>
<td>Hourly</td>
</tr>
<tr>
<td><strong>Dynamic scheduling</strong></td>
<td>Real-time metering, telemetering, and computer software and hardware to electronically transfer some or all of a generator’s output or a customer’s load from one control area to another</td>
<td>Seconds</td>
</tr>
<tr>
<td><strong>System blackstart</strong></td>
<td>The ability of a generating unit to go from a shutdown condition to an operating condition without assistance from the electrical grid and then to energize the grid to help other units start after a blackout occurs</td>
<td>When outages occur</td>
</tr>
<tr>
<td><strong>Network stability services</strong></td>
<td>Maintenance and use of special equipment (e.g., power-system stabilizers and dynamic-braking resistors) to maintain a secure transmission system</td>
<td>Cycles</td>
</tr>
</tbody>
</table>

Source: Eric Hirst, COMMENTS ON ANCILLARY SERVICES, 1999
Hirst (1999) then comments:

"The NERC IOS Implementation Task Force, successor to the Working Group, has further refined these concepts. The Task Force considers real-power-loss replacement, energy imbalance, and dynamic scheduling “related concepts” rather than ancillary services: “Real Power Losses is a component of the generation and demand balance but is not itself an IOS, as the provision of losses occurs through scheduling, regulation, and load following. Energy imbalance is a measure but is not itself an IOS. Dynamic scheduling also is not an IOS, but is a market service to enable the transfer of certain IOS.” The Task Force does not consider backup supply a separate service because it is a purely commercial function.”

Source: Eric Hirst, COMMENTS ON ANCILLARY SERVICES, 1999

It must be remembered that the original NERC report appeared around the same time as Directive 96/92/EC. A peculiarity in the American description of services is the emphasis on generation as the preferred supplier of controllability, which of course did not escape Eric Hirst’s (1999) criticism:

"Most ancillary services are produced by the same pieces of equipment (generating stations) that produce the energy commodity. Therefore, if society wants competitive markets for energy, it must create such markets for ancillary services. As a consequence, embedded-cost pricing cannot last long because it is difficult and inefficient to operate competitive and regulated markets for products from the same machinery, as demonstrated in California in the summer of 1998. Competition for ancillary services is desirable because it will reduce consumer costs, lead to innovations in further unbundling and rebundling these services, charge customers for what they use, and allow for customer provision of some services.”

Source: Eric Hirst, COMMENTS ON ANCILLARY SERVICES, 1999

Another key observation relates to the time scales on which processes (services) occur and the impact this has on the applicability of market concepts. For very fast processes “no attempt is made to implement a least cost dispatch”, for processes associated with longer time a merit order dispatch is generally in place (Hirst, 1999). These statements are prescient to several core aspects of the division of responsibilities between users and system operator as implemented in European markets.
A case example of real-time incentives to users
The current liberalized market design in the Netherlands was preceded by a vertically integrated central planning and dispatch system, according to a power supply configuration as shown in Figure 1-2. Restrictions on market and grid access of users were coupled with obligations on operators regarding public supply. Centrally planned investment and central generation commitment and dispatch optimization aimed at limiting cost to the user. Operator remuneration was based on fixed tariffs, thus safeguarding users from risks, but also denying users opportunities. This left little room for user-innovation, which is not necessarily a bad thing in central planning. This highly centralized economy design with 10-year planning and all was not totalitarian. It still left some room for independent decentralized combined heat and power generation aimed at heat provision. These installations, even if limited in plant-size, were allowed to be planned, built and operated by distribution companies and (industrial) consumers. So effectively, the demand that the centralized system covered was a residual one, albeit with a market share of some 80% to 90% (CBS, Statline).

In 1986 an idiosyncratic mechanism emerged to recover the capital and operational expenditure of the central dispatch generation from its users. Operational expenditure (fuel, maintenance, imports, losses) were recovered yearly through an energy tariff. Capital expenditure (investment in generation, including reserve capacity, and in extra-high voltage grid) was recovered from users according to their individual share in the central system peak demand, in a capacity tariff. So users could reduce their average energy price [€/MWh], the result of (capacity payment [€] + energy payment [€])/[withdrawal [MWh]] by:

- Reducing their share in the central system peak demand,
- Increasing their energy consumption at all other moments.

In fact, users could even attempt to increase central system peak demand, provided their share would be less than the current share. The general idea was by everybody aiming at reduction of its share, central system peak demand would be reduced, thus reducing the required generating capacity and (future) capex. This mechanism would of course not affect capex already spent and yet to be recovered. The actual capacity fee thus resulting provided users a clear reference value for substitutes to central demand, like load shedding or decentralized generation. This mechanism indeed could reduce the actual system demand peak, without lowering capex, thus effectively increasing the capacity tariff. An increased tariff would make substitutes ever more attractive for all expect the most captive users, who eventually would be facing the cost. However, its instability led to its abandonment as of 1997, just prior to the Elektriciteitswet 1998 (Ministry Of Economic Affairs Of The Netherlands, 1989).

I have analysed load data from the central planning and dispatch system from 1972 to 1998. An earlier version of this analysis has been used before [See Electricity Plan 1995-2004]. The analysis uses percentile values of yearly time series of evenly spaced load data. As much as percentile values are loads, in a time series they can be translated into time. The 75th percentile value of a yearly load series is exceeded for about 8760/4 = 2200 hours. The range between 75th percentile and maximum is the upper quartile, which in hours is close to the total of working hours (8) on the about 250 non-weekend days. The upper 5th percentile represents about 440h.
The analysed years show considerable, albeit irregular demand growth in Figure A-1. The years with the incentivizing tariff mechanism are highlighted.

Figure A-1 The Netherlands central dispatch demand 1972-1998

The incentivizing tariff mechanism was accompanied by a (very close to) real-time information service, reporting the actual system power demand per 5 minute moving 15 minute average. At first hierarchically and only to the generation companies, who redistributed it to their distribution companies, later as of 1993 directly to all users, which might have been instrumental in the effect in the later years of its existence. In the absence of internet a television broadcast signal was used to carry the signal. Also in the first years in existence, especially in 1987, an extreme cold spell at the beginning of the year largely rendered any subsequent attempt to change it ineffective. This imperfection could be rectified by using as of 1992 peak values from two separate months per semester, instead of the average two peaks from two different days.

The 95-percentile is extremely well correlated to yearly consumption, regardless of tariff mechanism. Centring the upper quartile on the 95-percentile (as proxy for yearly consumption) will take volume growth into account in the evolution of the upper quartile range in Figure A-2. The years with the incentivizing tariff mechanism are highlighted. Due to the actual growth pattern the data points ordering is not fully chronologically from left to right.
The first 2 years of the tariff mechanism do not show similar effects as shown in later years. Feedback information was not yet fully implemented with easy access, and learning experience had to be accumulated. Furthermore 1986 and especially 1987 had very cold winters with accompanying peak loads.

Without an incentivizing tariff mechanism the upper quartile appears to have been a fairly constant 1500MW, slightly rising with the level of consumption. The tariff mechanism eventually reduced this upper quartile range by some 600 MW, and flattened the monthly peak load pattern considerably. A not desired side effect was that a previous existing summer maintenance window for generators disappeared, and that the system became more vulnerable to cooling water issues in summer. The changes were not permanent however. After abandoning this unique system the more original upper 5\textsuperscript{th} and 75\textsuperscript{th} percentile ranges re-emerged as shown in Figure A-2. This lends credibility to the hypothesis that is indeed was the tariff mechanism that caused the pattern change. The conclusion is that aggregate load is not inflexible, given appropriate incentives and real-time information.

From this example it can be concluded that with the necessary conditions of appropriate incentives and real-time feedback information aggregate load is not inflexible in close to real-time.
Price caps and floors of day ahead price in the Netherlands

Prior to 2001 day ahead bid prices were implicitly capped due to binding price and sales fixes between production and distribution companies (de Lange et al., 2002) At APX day ahead bid prices ([€/MWh], 2 decimals) are capped and floored. Starting in 1999 with a floor of 0.01 [€/MWh] and a cap at 600 [€/MWh], the cap changed successively through 1500 (end of 2001), and 2000 in August 2003 and finally to 3000 [€/MWh] at the onset of implicit market coupling with France and Belgium in November 2006. Only at the start of Central Western European market coupling in November 2010 did the original floor, starting at 0.01 [€/MWh] in 1999, change to -3000 [€/MWh]. But with the onset of the North Western European market coupling on November 26th 2013 the floor went back up to -500 [€/MWh]; the cap remained at 3000 [€/MWh].

The caps are lower than estimates of the value of lost load for the Netherlands of about 8000 [€/MWh] (de Nooij, 2012), and therefore would exclude loads with such value from participating on the day ahead market in the Netherlands. Published estimates on the value of lost load are highly elusive and contested. They span several orders of magnitude, and depend on who has been hit, at what time, and for how long (Reckon, 2012). The largest estimates surpass the price caps at the coupled European markets. Only the smallest estimates fall within the presently applied 3000 [€/MWh] for the Netherlands and its coupled day ahead markets.

Bid prices for balancing energy are only limited by the bidding messages, i.e. bid prices may range from -99999.99 to 99999.99 [€/MWh], and thus exceed even the more extreme estimates of value of lost load.

Intriguingly, apart from bid price caps and floors, the coupled day ahead markets have established a procedure to invoke a "second auction" in case of an "exceptional situation":

* In the event of an Exceptional Situation DAM, APX may, in order to assure the proper functioning of the Power NL DAM, in accordance with the Power Instrument Specifications, apply the "second auction" procedure according to the following process. APX calculates (or receives from an external entity if MC is applied) an indicative MCP for each Instrument eligible for the Auction, based on the situation of the Order Book at the time of calculation, taking MC into account if applicable. If for at least one Instrument the indicative MCP is lower than -150 €/MWh or higher than 500 €/MWh, or if a coupled DAM has a similar situation, a Request for Quotes (hereafter "RIQ") is sent out to all Authorized Users by e-mail and via the Trading System and the MCT DAM is delayed as communicated in the in the RIQ.*

Source: APX, Power NL Market Instrument Specifications, 25-11-2014/version 11.0

Effectively this means that clearing prices outside the range [-150, +500] [€/MWh] have to be confirmed by a second auction, thus annulling the results of the first.
## System operator tariff characteristics in European countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Sharing of network operator charges</th>
<th>Price signal</th>
<th>Are losses included in the tariffs charged by TSO?</th>
<th>Are system services included in the tariffs charged by TSO?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>43%</td>
<td>57%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Belgium</td>
<td>7%</td>
<td>93%</td>
<td>XXX</td>
<td>Not included for grid &gt;=150 kV</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>0%</td>
<td>100%</td>
<td>n/a</td>
<td>Yes</td>
</tr>
<tr>
<td>Croatia</td>
<td>0%</td>
<td>100%</td>
<td>X</td>
<td>No / post stamp</td>
</tr>
<tr>
<td>Cyprus</td>
<td>0%</td>
<td>100%</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Denmark</td>
<td>5%</td>
<td>95%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Estonia</td>
<td>0%</td>
<td>100%</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Finland</td>
<td>18%</td>
<td>82%</td>
<td>X</td>
<td>Yes</td>
</tr>
<tr>
<td>France</td>
<td>2%</td>
<td>98%</td>
<td>/ XXX</td>
<td>No</td>
</tr>
<tr>
<td>Germany</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Great Britain</td>
<td>TNuoS 27% BSuoS 50%</td>
<td>TNuoS 73% BSuoS 50%</td>
<td>XX</td>
<td>No, recovered in the energy market</td>
</tr>
<tr>
<td>Greece</td>
<td>0% (TUOS and Uplift charges)</td>
<td>100% (TUOS and Uplift charges)</td>
<td>X</td>
<td>No, recovered in the energy market</td>
</tr>
<tr>
<td>Hungary</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Iceland</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ireland</td>
<td>25%</td>
<td>75%</td>
<td>No</td>
<td>Generation only, No, recovered in the energy market</td>
</tr>
<tr>
<td>Italy</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Latvia</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>FYROM</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Montenegro</td>
<td>0%</td>
<td>100%</td>
<td>X</td>
<td>No</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>25%</td>
<td>75%</td>
<td>XXX Load</td>
<td>Generation only, No, recovered in the energy market</td>
</tr>
<tr>
<td>Norway</td>
<td>40%</td>
<td>60%</td>
<td>XXX xxx</td>
<td>Yes</td>
</tr>
<tr>
<td>Poland</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Portugal</td>
<td>9%</td>
<td>91%</td>
<td>XX</td>
<td>No, included in energy price</td>
</tr>
<tr>
<td>Romania</td>
<td>19%</td>
<td>81%</td>
<td>No</td>
<td>Generation and Load, Yes</td>
</tr>
<tr>
<td>Serbia</td>
<td>0%</td>
<td>100%</td>
<td>X</td>
<td>Yes</td>
</tr>
<tr>
<td>Slovak Rep.</td>
<td>3%</td>
<td>97%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0%</td>
<td>100%</td>
<td>XX</td>
<td>Yes</td>
</tr>
<tr>
<td>Spain</td>
<td>10%</td>
<td>90%</td>
<td>XXX</td>
<td>No, they are recovered through the energy market</td>
</tr>
<tr>
<td>Sweden</td>
<td>39%</td>
<td>61%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0%</td>
<td>100%</td>
<td>No</td>
<td>No, there is a separate tariff for losses</td>
</tr>
</tbody>
</table>

Source: ENTSOE Overview of Transmission Tariffs in Europe: Synthesis 2015
Financial position gas system operator for balancing

European regulation for gas (balancing) explicitly state principles of neutrality for the transmission system operator, and as means to achieve this objective balancing neutrality cash flows, to be by paid by or to the network user.

* Article 29, Principles of neutrality
1. The transmission system operator shall not gain or lose by the payment and receipt of daily imbalance charges, within day charges, balancing actions charges and other charges related to its balancing activities, which shall be considered as all the activities undertaken by the transmission system operator to fulfill the obligations set out in this Regulation.
2. The transmission system operator shall pass to network users:
   (a) any costs and revenues arising from daily imbalance charges and within day charges;
   (b) any costs and revenues arising from the balancing actions undertaken pursuant to Article 9, unless the national regulatory authority considers those costs and revenues as incurred inefficiently in accordance with the applicable national rules. This consideration shall be based upon an assessment which:
      (i) shall demonstrate to what extent the transmission system operator could have reasonably mitigated the costs incurred when undertaking the balancing action; and
      (ii) shall be made with regard to the information, the time and the tools available to the transmission system operator at the moment it decided to undertake the balancing action;
   (c) any other costs and revenues related to the balancing activities undertaken by the transmission system operator, unless the national regulatory authority considers these costs and revenues as incurred inefficiently in accordance with the applicable national rules.
3. Where an incentive to promote efficient undertaking of balancing actions is implemented, the aggregate financial loss shall be limited to the transmission system operator’s inefficiently incurred costs and revenues."

If the revenues cannot be efficiently used for the purposes set out in points (a) and/or (b) of the first subparagraph, they may be used, subject to approval by the regulatory authorities of the Member States concerned, up to a maximum amount to be decided by those regulatory authorities, as income to be taken into account by the regulatory authorities when approving the methodology for calculating network tariffs and/or fixing network tariffs.

The rest of revenues shall be placed on a separate internal account line until such time as it can be spent on the purposes set out in points (a) and/or (b) of the first subparagraph. The regulatory authority shall inform the Agency of the approval referred to in the second subparagraph."

* Article 30, Balancing neutrality cash flows
1. The neutrality charge for balancing shall be paid by or to the network user concerned.
2. The national regulatory authority shall set or approve and publish the methodology for the calculation of the neutrality charges for balancing, including their apportionment amongst network users and credit risk management rules.
3. The neutrality charge for balancing shall be proportionate to the extent the network user makes use of the relevant entry or exit points concerned or the transmission network.
4. The neutrality charge for balancing shall be identified separately when invoiced to network users and the invoice shall be accompanied by sufficient supporting information defined in the methodology referred to in paragraph 2.
5. Where the information model variant 2 is applied and thus the neutrality charge for balancing may be based on forecasted costs and revenues, the transmission system operator’s methodology for the calculation of neutrality charge for balancing shall provide rules for a separate neutrality charge for balancing in respect of non daily metered off-takes.
6. Where relevant, the transmission system operator’s methodology for the calculation of the neutrality charge for balancing may provide rules for the division of the neutrality charge for balancing components and the subsequent apportionment of the corresponding sums amongst the network users in order to reduce cross subsidies."

Source: EC regulation 312/2014/EC
The main observations on gas balancing neutrality charge from this regulation are:

- the neutrality charge for gas balancing, can be paid by or received by the network user with an entry or exit point, i.e. a connection,
- the neutrality charge is separate from any other charges or tariffs, and specifically,
- this neutrality charge is not included as surcharge on the imbalance charges.
B. Bibliography


ACM. Gebiedsindeling, website ACM, www.acm.nl, 1999-recent


APX website, www.apxgroup.com


ENTSO-E, AHTor, Ad hoc team operational reserves, final report, Brussels, 2012.

ENTSO-E, AS Enquiry, Ancillary services working group, Survey on ancillary services procurement and balancing market design, Brussels, 2014.


M. de Nooij, Social cost benefit analysis & energy policy, Ph.D thesis, Jacobs University, Kiel, 2012.


T. Pynchon: Togetherness, Aerospace Safety December 1960, pp. 6-8.


Reckon, Desktop review and analysis of information on Value of Lost Load for RIIO-ED1 and associated work, UK, 2012.


Regulation (EC) on capacity allocation and congestion management (CACM), 2014.


TenneT website, www.tennet.eu


Twenties project - Final report, 2013.


