HAL
open science

L'Analyse économique des architectures de marché électrique. L'application au market design du temps réel.

Marcelo Saguan

▶ To cite this version:

Marcelo Saguan. L'Analyse économique des architectures de marché électrique. L'application au market design du temps réel.. Economies et finances. Université Paris Sud - Paris XI, 2007. Français. NNT : . tel-00281131

HAL Id: tel-00281131
https://theses.hal.science/tel-00281131

Submitted on 21 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Supérieure d'Électricité
Département Énergie

Université Paris-Sud 11
Faculté Jean Monnet

L'ANALYSE ECONOMIQUE DES ARCHITECTURES DE MARCHE ELECTRIQUE. APPLICATION AU *MARKET DESIGN* DU « TEMPS REEL »

Thèse pour le Doctorat en Sciences Economiques

Présentée et soutenue par

Marcelo SAGUAN

Avril 2007

Directeur de Recherche :
M. Jean-Michel GLACHANT, Professeur, Université Paris-Sud 11

Rapporteurs :
M. Ronnie BELMANS, Professeur, Université Catholique de Louvain, Belgique
M. Christian von HIRSCHHAUSEN, Professeur, Université Technique de Dresde, Allemagne

Suffragants :
M. Stéphane SAUSSIER, Professeur, Université Paris-Sud 11
M. Michel MASSONI, Directeur des accès aux réseaux, Commission de Régulation de l'Energie
M. Philippe DESSANTE, Professeur, Ecole Supérieure d'Electricité (SUPELEC)

L'Université Paris-Sud 11
n'entend donner aucune approbation
ni improbation aux opinions
émises dans les thèses.
Ces opinions doivent être considérées
comme propres à leurs auteurs.

Remerciements

Un professeur m'a dit une fois : « Une bonne thèse est une thèse finie » (je me réserve le nom de l'auteur de cette phrase). En tout cas, ma thèse est finie. Or, elle n'aurait pu être finie sans l'aide et les encouragements que j'ai reçus au cours de ces dernières années. C'est le moment de remercier toutes les personnes qui m'ont entouré pendant la réalisation de ce travail. Malheureusement, les remerciements exprimés dans cette page, si courte et limitée par ma maîtrise maladroite de la langue, ne représentent qu'une partie minimale de mes sincères remerciements.

Je remercie énormément Jean-Michel Glachant, mon directeur de thèse. Mes remerciements envers lui concernent en premier lieu l'attention et le temps qu'il a consacrés à mon travail. Je tiens aussi à le remercier pour ses conseils et ses réflexions intelligentes, son soutien et son encouragement. Sincèrement, grâce à lui j'ai pu apprendre beaucoup des choses dont certaines fort utiles pour mes travaux académiques bien sûr, mais aussi des choses importantes pour mon développement personnel.

Cette thèse a été réalisée dans le cadre du Projet Fédérateur à Supélec. Je remercie Patrick Bastard qui m'a encadré pendant les premières années. Je le remercie de m'avoir fait confiance au début de mes travaux. Je remercie aussi Philippe Dessante qui a pris généreusement la relève et m'a encadré durant les dernières années de cette thèse. Tous les deux, je les remercie pour leurs réflexions, leurs conseils et leurs explications qui m'ont permis de progresser et d'améliorer mon travail. Mes remerciements s'adressent aussi à Jean-Claude Vannier et Sophie Plumel pour leur soutien et encouragement.

Je tiens à remercier Michel Massoni et Christophe Gence-Creux de la Commission de Régulation de l'Energie pour avoir suivi mes travaux pendant la durée de ma thèse. Leurs conseils et remarques, toujours pertinents, m'ont permis de confronter (et ajuster) mes travaux à des problématiques pratiques réelles.

Je remercie spécialement mes relecteurs et correcteurs (et amis) : Virginie Pignon, Céline Hiroux, Yannick Perez, Ute Dubois, Vincent Rious et Florent Maupas. La justesse de leurs critiques ont été très constructives et utiles. Mises à part leurs contributions de fond, ils ont eu la pénible tâche de corriger mon français maladroit ! Un grand merci !

Pendant les années de ma thèse, j'ai eu la chance de travailler dans des groupes de recherche fantastiques. Je remercie toutes les personnes du GRJM et de l'ADIS d'un côté, et du département Energie de Supélec d'autre côté, pour leur support, leurs enseignements et leur amitié. Je remercie spécialement Gino Hadjee avec qui j'ai partagé le bureau (et beaucoup d'autres choses) pendant plus de 3 ans. Je me permets de remercier aussi l'équipe de football « des jeudis » à Supélec ainsi que l'équipe de baby-foot à Fontenay qui ont fait plus conviviale la durée de ma thèse. J'ai eu la chance aussi de travailler avec de groupes de recherche étrangers. En représentation de toutes les personnes intéressantes avec lesquelles j'ai eu des échanges enrichissants, je remercie en particulier deux chercheurs qui sont devenu des amis à force de travailler ensemble : Leonardo Meeus et Nenad Keseric.

Cette dernière partie des remerciements devrait s'appeler « remerciements et excuses » ou remerciements pour la patience. La recherche est très preneuse de temps ! Et le temps consacré à la recherche implique moins de temps pour la famille et les amis… Je remercie énormément à toute ma famille : Abuela, Abuelo, Mamá, Papá, Eugenia, Graciela, Leticia, etc. Depuis très loin, ils m'ont soutenu et encouragé constamment. Je les remercie de leur patience, car à force de travailler dur, j'avoue ne pas leur avoir consacré le temps qu'ils méritent réellement. Je remercie aussi tous mes amis, d'ici et de là-bas. Merci pour m'avoir accompagné et aidé pendant ces dernières années.

Gracias… totales !

Fontenay aux Roses, le 18 mars 2007.

Sommaire

Introduction Générale 1

Partie I : La modularité des architectures de marché électrique

Chapitre 1
Système électrique et exploitation optimale par une entreprise intégrée verticalement
Introduction..... 8
Section 1: Description des éléments d'un système électrique 9
Section 2: Exploitation optimale par une entreprise intégrée 31
Section 3: Coordination d'entreprises intégrées sur un même réseau 45
Conclusions du chapitre 53

Chapitre 2
Architectures de marché électrique
Introduction..... 56
Section 1: Contrôles-commandes *vs.* Marchés 60
Section 2: Architectures de marché sur une seule zone de contrôle..... 72
Section 3: Architectures de marché sur plusieurs zones de contrôle..... 102
Conclusions du chapitre 121

Partie II : Le *Market Design* du « temps réel »

Chapitre 3
Séquence des marchés d'énergie et design du module du temps réel
Introduction..... 125
Section 1: Séquence des marchés d'énergie à court terme 127
Section 2: Design du module du temps réel..... 143
Conclusions du chapitre 188

Chapitre 4
Modélisation de la séquence des marchés d'énergie et du module du temps réel
Introduction..... 192
Section 1: Modélisation du module du temps réel (une zone de contrôle)..... 194
Section 2: Modélisation de l'intégration d'architectures (deux zones de contrôle).... 235
Conclusions du chapitre 259

Conclusion Générale 262

Annexe.....268

Bibliographie 276

Table des matières 293

Introduction Générale

L'introduction de la concurrence dans le secteur électrique s'inscrit dans la lignée des mouvements de réformes des secteurs économiques relevant traditionnellement de monopoles intégrés verticalement (Staropoli [2001]). Les caractéristiques technico-économiques de ces secteurs, notamment l'existence d'une infrastructure en monopole naturel, ont longtemps justifié l'existence de l'intégration verticale. De puissants facteurs de changement sont venus remettre en cause les formes traditionnelles d'organisation de ces industries. Dès les années soixante, des travaux de recherche ont conduit à remettre en cause les schémas d'organisation des industries de réseaux, ainsi que l'efficacité des monopoles dans l'allocation de ressources (Averch-Jonhnson [1962], Demsetz [1968]). Un consensus s'est progressivement établi, dans les dernières décennies du XXe siècle, pour préconiser l'introduction de certaines formes de concurrence. Ce consensus a touché successivement les industries des télécommunications, des transports (aériens et ferroviaires), et finalement le secteur électrique.

En une à deux décennies, des réformes électriques ont été réalisées dans la plupart des pays (Glachant-Finon [2003], Newbery [2005b], Joskow [2006a], Sioshansi [2006]). L'introduction de la concurrence sur un secteur longtemps organisé autour d'un monopole intégré fait apparaître des problèmes de coordination qu'il convient de traiter par la mise en place de dispositifs organisationnels et institutionnels adéquats (Staropoli [2001]). Or la base de ces dispositifs n'émerge pas spontanément du jeu des acteurs, mais résulte des choix d'autorités publiques en charge de l'introduction de la concurrence. Pour ce qui concerne la création d'un marché de gros, la conception d'une architecture du marché (« *market design* ») doit réunir certains dispositifs organisationnels et institutionnels afin d'assurer une nouvelle forme de coordination (Wilson [2002], Stoft [2002]).

Bien que, depuis le début de l'introduction de la concurrence dans le secteur électrique, l'état de l'art des théories économiques sur les architectures de marché ait progressé considérablement, les connaissances actuelles ne permettent pas de trancher définitivement sur le choix d'un seul design optimal qui déboucherait sur des solutions pratiques robustes d'architectures. Ces développements théoriques se heurtent très certainement aux spécificités du bien électricité et, en fait, la question de l'architecture de marché est trop complexe pour être traitée sans être découpée en plusieurs modules (Wilson [1998b], Wilson [2002]).

Le manque de cadre théorique complet sur les architectures de marché rend très intéressantes les expériences des architectures de marché mises en place dans différentes régions du monde. En effet, même sans un cadre théorique complet, on pourrait utiliser l'expérience acquise depuis près de 20 ans pour réaliser des comparaisons et pour développer un classement d'après les performances. Cependant, cette méthodologie, bien que très utile, présente aussi une difficulté de mise en œuvre. Il existe en fait une très grande

diversité d'architectures de marché. La comparaison directe des performances des différentes architectures de marché devient donc très difficile (CE [2004], Green *et al.* [2005]).

La difficulté devient encore plus grande quand on considère l'intégration entre plusieurs architectures de marché. En fait, une architecture de marché est normalement définie pour une zone géographique déterminée (une région, ou un état). Mais l'introduction de la concurrence peut aussi passer par la stimulation des échanges entre plusieurs zones (c'est par exemple, le cas du marché intérieur européen) (Pignon [2003], CE [1996], CE [2003]). Dans ce cas, le choix du design pour intégrer les architectures de marchés des zones adjacentes est donc déterminant pour un bon fonctionnement de nouveaux marchés électriques plus vastes. Cependant, l'analyse économique appliquée à l'étude de l'intégration de marchés ne permet pas non plus de trancher définitivement sur la meilleure manière de mettre en place un marché électrique régional.

Le manque de maîtrise rationnelle dans le choix des options optimales de design, que ce soit au niveau national ou régional, montre d'abord l'importance d'une recherche plus approfondie sur les différents designs possibles. Mais, pour approfondir cette étude des architectures, il faut pouvoir la découper en plusieurs thèmes. Il est d'abord nécessaire de développer un cadre d'analyse modulaire des architectures de marché, afin de pouvoir identifier et caractériser en quoi consiste les designs de marché en électricité. Ensuite seulement on pourra se concentrer sur l'étude approfondie des règles de certains modules spécifiques. La modularité de l'architecture rend possible l'analyse plus approfondie d'une seule sous-partie du système car la modularité permet d'étudier certains modules séparément des autres (Baldwin-Clark [2000]).

Le premier des objectifs de cette thèse est donc de développer un cadre d'analyse modulaire des architectures de marché et de l'intégration des architectures de marché.

La construction d'une architecture de marché consiste à remplacer la séquence de contrôles-commandes (*unit commitment*, *optimal dispatch*, etc.) utilisé dans une entreprise intégrée verticalement par une séquence de marchés ou de modules (Wilson [2002]). Ces marchés ou modules doivent assurer les fonctions essentielles de coordination qui étaient gérée internement par les outils de contrôle-commande. Une architecture de marché électrique sera ainsi composée de 4 modules : 1° le module d'énergie *forward*, 2° le module de transport *forward* et 3° le module de réserves *forward* et 4° le module du temps réel. Puisque l'architecture de marché est « modulaire », il existe différentes options de design ou formes d'organisation possibles pour chacun des modules. L'analyse des outils de contrôle-commande permettra de dériver ces options ou variantes possibles. Nous verrons que ces options sont contraintes par des limitations techniques et économiques. L'architecture de marché sera donc à la fois un choix d'option particulier pour chaque module, et un choix d'articulation entre les modules. L'analyse modulaire des architectures permet aussi de caractériser différentes formes de design pour réaliser l'intégration des

architectures de marché entre des multiples zones de contrôle liées par un réseau de transport commun.

L'identification et l'isolement de quelques modules essentiels d'une architecture de marché permet d'approfondir la recherche sur le design de ces éléments. Quel module particulier doit-on choisir ? Si l'on regarde l'architecture de marché du point de vue de l'effort des autorités pour la construire, on peut identifier une architecture minimale comme le noyau qui est indispensable au fonctionnement de toute industrie électrique sur une base concurrentielle. Cette architecture minimale correspond au design du module du temps réel. Ceci découle du fait que le bien électricité n'est pas stockable. Tous les échanges physiques ne prennent donc effectivement place qu'en temps réel. Le module du temps réel est ainsi au cœur de toute architecture de marché, et il nécessite un traitement attentif de son design. Cependant, c'est le module comparativement le moins étudié depuis le début des réformes.

Le deuxième objectif de cette thèse est alors d'étudier le design du module du temps réel.

L'étude du module du temps réel et son inscription sur une séquence de marchés présente un intérêt académique et un intérêt empirique.

L'importance académique de l'étude du module du temps réel vient de ce que ce module n'a pas été suffisamment étudié depuis le début des réformes du point de vue de la théorie économique. En effet, depuis le début de l'introduction de la concurrence dans l'industrie électrique il y a eu une séparation tacite des tâches entre les problèmes « économiques » de l'électricité et les problèmes d'« ingénierie » de l'électricité.

Ainsi, d'un côté, les économistes se sont concentrés sur les modules de l'architecture de marché qui, par leurs caractéristiques, étaient plus cohérents avec l'analyse économique typique de transactions sur un pur marché. De l'autre côté, toute la problématique « technique » de l'opération du système électrique en temps réel a été laissée aux ingénieurs de réseau, de fait les seuls compétents pour l'exploitation opérationnelle du système en temps réel. Les études d'analyse économique formalisée se sont centrées donc sur le design de « bourses d'électricité » (*day-ahead market*) car les mécanismes d'enchère étaient le « cœur » de la concurrence en production. Ainsi, les premières études des économistes sur les architectures de marché étaient basées sur la théorie des enchères (Von der Fehr-Harbord [1993]). Parallèlement, les économistes ont privilégié l'étude de la concurrence imparfaite notamment les structures de marché, d'abord en utilisant des modèles classiques (concurrence à la Cournot ou à la Bertrand) et ensuite avec des modèles plus complexes du type de « fonction d'offre » (« *supply function equilibria* ») (Green-Newbery [1992], Bolle [1992], Green [1994]). Cette analyse économique appliquée aux marchés électriques s'est orientée ensuite sur les caractéristiques du réseau de transport et l'influence des contraintes de transport dans un environnement de concurrence imparfaite (Cardell *et al.* [1997], Borenstein *et al.* [2000], Neuhoff [2003], Hobbs-Rijkers [2004],

Barquin *et al.* [2004]). Bien qu'importantes pour l'avancé d'une connaissance plus approfondie des marchés d'électricité, ces études ont été limitées car elles ne considéraient qu'une seule des étapes temporelles d'un marché électrique, ignorant ainsi la séquence de marchés. Quelques études ont pris en compte les caractéristiques intrinsèques du marché électrique, organisé comme une séquence de marchés (Allaz-Vila [1993], Green-McDaniel [1999], Bessembinder-Lemmon [2000], Siddiqui [2002], Kamat-Oren [2004], Yao *et al.* [2004]). Mais le but principal de ces études n'était pas orienté sur les caractéristiques de designs du seul marché physique de l'électricité. L'étude formalisée du design du module du temps réel inscrit sur une séquence de marché nécessite la combinaison de connaissances économiques avancées et de connaissances « ingénieur ». D'un côté, considérer une séquence de marchés, implique l'utilisation de modèles d'équilibre à plusieurs étapes et la prise en compte des comportements d'agents face au risque. D'un autre côté, l'étude détaillée du fonctionnement du module du temps réel, gouverné fortement par des règles conçues généralement par des ingénieurs réseau, est nécessaire car c'est ce module qui ferme la séquence de marché et déterminera les effets sur toute la séquence de marchés.

L'intérêt empirique de l'étude du module du temps réel a deux volets principaux. D'un part, la comparaison des alternatives de design du module du temps réel est d'un réel intérêt empirique dans l'Union européenne, puisque plusieurs pays cherchent à améliorer leurs architectures de marché « nationales » (Meeus *et al.* [2005], Meeus *et al.* [2006]). Dans ce contexte, il est important d'avoir un cadre d'analyse formalisé afin de pouvoir comparer les différentes variantes de design pour une architecture de marché correspondant à un seul pays ou une seule zone de contrôle. Concrètement, il existe en Europe différents designs du module du temps réel. Ils forment toute une gamme qui se répartit entre deux designs extrêmes : d'un côté, le design du type « marché » et, de l'autre côté, le design du type « mécanisme ». Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel. Par contre, un mécanisme d'ajustement recourt à des systèmes de prix plus complexes incluant normalement l'utilisation de pénalités. D'autre part, l'étude des différents designs du module du temps réel devient encore plus importante lors de la conception de l'intégration d'architectures de marché entre plusieurs zones de contrôle. Dans un contexte où l'intégration se réalise entre architectures qui ne sont pas parfaitement harmonisées, différents designs peuvent coexister pour les modules du temps réel. C'est notamment le cas pour l'intégration de trois pays de l'Europe Continentale : la France et la Belgique présentant un design du type « mécanisme » et les Pays-Bas présentant un design proche d'un « marché ».

Ainsi, cette thèse apportera-t-elle les éléments formalisés nécessaires pour la compréhension, l'évaluation et la comparaison des principales variantes de design du module du temps réel sur une séquence de marchés d'énergie.

Organisation de la thèse

Cette thèse comprend deux parties.

La **première partie** présente un cadre d'analyse modulaire et met en évidence les principaux éléments d'une architecture de marché électrique.

Le **premier chapitre** est consacré aux caractéristiques économiques et techniques spécifiques du système électrique composé de trois parties : la consommation, les moyens de production et le réseau de transport. Nous étudions ensuite la forme d'organisation et de coordination la plus simple pour exploiter un système électrique de manière optimale. Cette forme est une entreprise verticalement intégrée, bienveillante ou parfaitement régulée. Nous en déduisons la séquence de contrôles-commandes étant le dispositif de coordination nécessaire pour une exploitation optimale.

Le **deuxième chapitre** caractérise les éléments essentiels d'une architecture de marchés ainsi que leur diversité. Notre démarche va du plus simple au plus complexe. Nous caractérisons tout d'abord l'architecture de marché pour une zone de contrôle isolée, avec un réseau de transport unique. Puis nous caractérisons l'intégration d'architectures de marché entre plusieurs zones de contrôle.

A ce titre, nous analyserons les éléments essentiels d'une architecture de marché électrique en concevant l'introduction de la concurrence comme le remplacement des outils de contrôle-commande d'une entreprise intégrée par une combinaison de marchés et de mécanismes. Nous nous intéressons à deux horizons temporels, le court terme (au jour J-1) et le très court terme (le temps réel). Et nous insisterons sur la partie du design qui relie les opérations en J-1 aux opérations en temps réel. Grâce à cette analyse, nous définirons quelles composantes du bien électricité et quels modules (ou sous-marchés potentiels) constituent le cœur d'une architecture de marché (Wilson [1998b], Wilson [2002]). Du point de vue de l'analyse économique, cette modularité des architectures ainsi mise en évidence permettra d'étudier certains modules séparément des autres.

L'analyse modulaire d'architectures de marché montrera aussi, que parmi la grande diversité existante d'architectures de marché, il existe un noyau commun à toutes les architectures, c'est-à-dire une architecture minimale. Cette architecture minimale porte sur le design du module du « temps réel ». Toutes les architectures de marché qui ont été mises en place dans le monde ont formellement défini des règles a minima pour ce module. Ceci souligne l'importance du module du temps réel pour un bien difficilement stockable comme l'électricité, dont les échanges physiques ne prennent effectivement place qu'en temps réel. En conséquence, nous centrerons les développements de cette thèse sur le module du temps réel.

La **deuxième partie** analyse le design du module du temps réel et sur son influence sur la séquence des marchés d'énergie.

Dans le **chapitre 3**, nous présentons le fonctionnement de la séquence des marchés d'énergie et nous réalisons une typologie des designs du module du temps réel. Nous insistons particulièrement sur les signaux de prix envoyés par le module du temps réel, car ces signaux gouverneront les comportements des agents économiques sur la séquence des marchés. Ce module est le dernier module de la séquence des marchés d'énergie et il gère la seule place où l'énergie physique est échangée entre acteurs de marché. Tous les autres marchés, qui fonctionnent avant le temps réel, sont des marchés *forward*, où on négocie bien des prix et des volumes, mais sans échange physique d'énergie. Pour cette raison, les caractéristiques de design du module du temps réel auront des conséquences importantes sur le reste de la séquence de marché.

Nous analyserons les principales variantes de design pour le module du temps réel. Nous distinguerons deux types de design : le « marché du temps réel » et le « mécanisme d'ajustement ». Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel. Par contre, un mécanisme d'ajustement recourt à des systèmes de prix plus complexes pour le règlement, incluant généralement des prix différenciés pour les différents types de transactions et l'utilisation de pénalités. Grâce à l'étude de plusieurs cas nous verrons que les différents designs existants forment toute une gamme, allant du design « marché », sans pénalités, aux designs du type « mécanisme » avec différentes intensités de pénalité. Le choix du *market design* en temps réel aura d'importantes conséquences car il fixera les incitations données aux participants du marché et influencera ainsi leur comportement. Nous verrons alors que les caractéristiques propres de chaque design produisent différentes conséquences économiques qui s'inscrivent sur la séquence de marchés d'énergie.

Le **chapitre 4** est consacré à la modélisation de la séquence des marchés d'énergie en prenant en compte les différents designs pour le module du temps réel. Nous utiliserons un modèle d'équilibre à deux étapes. Ces modèles d'équilibre en présence d'incertitudes se présentent comme l'alternative la plus pertinente pour les marchés d'électricité (Bessembinder-Lemmon [2000], Siddiqui [2002]). Cette modélisation permet de traiter des caractéristiques intrinsèques de l'électricité comme la non-stockabilité, les incertitudes et la valorisation des risques par les participants qui opèrent sur les différents marchés de la séquence. Nous développons une approche originale par rapport aux travaux antérieurs (Bessembinder-Lemmon [2000]) en ajoutant des caractéristiques spécifiques du design des modules du temps réel. Néanmoins l'interprétation économique directe de cette formalisation n'est pas aisée en raison de la complexité analytique du modèle. C'est pourquoi nous avons recours à des simulations numériques comparatives pour évaluer deux problématiques différentes concernant le *market design* du temps réel. Nous évaluons d'abord les conséquences économiques des différents designs dans une seule zone de contrôle. Puis nous évaluons celles de différents designs après l'intégration de deux zones de contrôle adjacentes.

Chapitre 1

Système électrique et exploitation optimale par une entreprise intégrée verticalement

Introduction 8

Section 1: Description des éléments d'un système électrique 9
1.1 Consommation 10
1.2 Moyens de production 15
1.3 Réseau de transport 21
1.4 Conclusions de la section 29

Section 2: Exploitation optimale par une entreprise intégrée 31
2.1 Réseau de transport et zone de contrôle 32
2.2 La séquence de contrôles-commandes (une zone de contrôle) 34
2.3 Conclusions de la section 43

Section 3: Coordination d'entreprises intégrées sur un même réseau 45
3.1 La coordination parfaite et bénéfices associés 45
3.2 Coordination de séquences de contrôles-commandes 47
3.3 Conclusions de la section 52

Conclusions du chapitre 53

Introduction

Les systèmes électriques présentent des caractéristiques spécifiques de fonctionnement, différentes des autres types d'industries. Ces caractéristiques sont propres à la technologie actuelle de l'industrie électrique, et indépendantes des formes institutionnelles d'organisation (que cette organisation soit basée sur des monopoles verticalement intégrés et régulés, ou sur l'introduction de la concurrence dans certains maillons de la chaine). Or, le choix et la conception des formes organisationnelles, et les performances inhérentes, dépendront de la manière dont ces caractéristiques spécifiques sont prises en compte.

Ce chapitre présente d'abord une description de ces caractéristiques spécifiques en considérant chaque maillon de la chaine du système électrique. Ainsi, dans la section 1 nous abordons les caractéristiques de variabilité de la consommation d'électricité, les caractéristiques économiques et les contraintes physiques des différents moyens de production et, finalement, la complexité du fonctionnement d'un réseau de transport.

Ensuite, dans la section 2, nous considérons la forme d'organisation la plus simple pour cette industrie : le monopole verticalement intégré bienveillant ou parfaitement régulé. Cette forme d'organisation est logiquement la plus simple, car toutes les décisions sont concentrées dans une seule entité prenant, par hypothèse, les meilleures décisions possibles. Le monopole verticalement intégré et parfaitement régulé utilise une séquence d'outils d'optimisation afin de prendre les décisions optimales pour la gestion de tous les éléments de ce système complexe. Les deux principaux outils d'optimisation utilisés par le monopole sont connus comme l'outil d'*unit commitment* et l'outil d'*optimal dispatch*. L'outil d'*unit commitment* est utilisé la veille du jour de livraison ; il sert à déterminer quelles unités de production démarrer ou arrêter ainsi que les niveaux de production prévisionnels pour chaque sous période de l'horizon de programmation (par exemple une journée). L'outil d'*optimal dispatch* est utilisé très près du temps réel et sert à déterminer le niveau de production définitif des unités disponibles. La description de ces deux outils est importante car, comme nous verrons dans le chapitre 2, ils sont aussi la base pour la mise en place d'une nouvelle forme d'organisation pour le système électrique : la concurrence.

Finalement, dans la section 3, nous examinons comment plusieurs monopoles verticalement intégrés, et fonctionnant sur un même réseau interconnecté, doivent se coordonner afin d'optimiser leurs performances globales. Cette étape sert également de base aux formes ultérieures de coordination entre les nouvelles formes d'organisation concurrentielles qui seraient mises en place dans les différentes régions correspondant à ces anciens monopoles.

Section 1: Description des éléments d'un système électrique

L'objectif de cette section est de fournir une vision générale des systèmes électriques et de leurs caractéristiques propres, les différenciant des autres systèmes industriels.

Nous nous concentrerons sur les facteurs technico-économiques de chaque composante du système électrique, mais en évitant de privilégier le point de vue organisationnel. En effet, nous cherchons une description des fondamentaux techniques qui soit valable pour toute forme d'organisation : monopole verticalement intégré ou organisation avec l'introduction de la concurrence. De cette description, nous dégagerons ainsi les problèmes de coordination propres à l'industrie électrique.

Un système électrique est composé par quatre maillons : les moyens de production, réseau de transport, réseau de distribution et la consommation (figure 1-1).

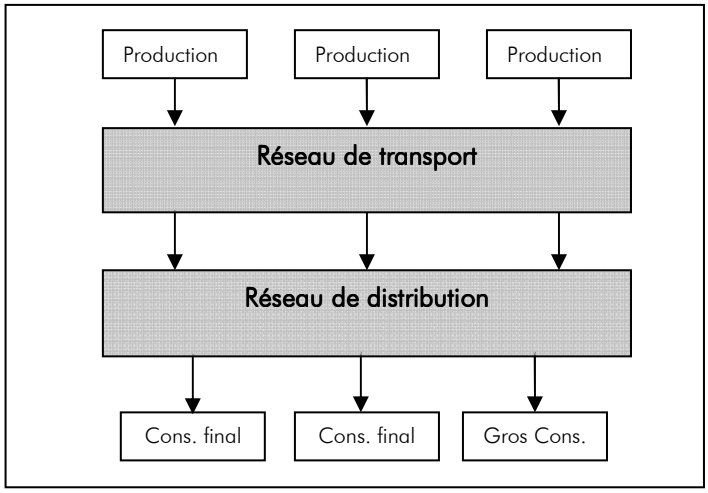

Figure 1-1 : Représentation schématique du système électrique

Notre description simplifiée se limitera aux points nécessaires pour la poursuite de la thèse. Ainsi nous présenterons les trois parties du système électrique qui nous concernent : la consommation (sous-section 1.1), les moyens de production (sous-section 1.2) et le réseau de transport (sous-section 1.3). Le réseau de distribution, quant à lui, ne sera pas traité ici et peut être considéré comme une sous-partie du réseau de transport.[1] Dans la sous-section 1.4, nous dégagerons les éléments conclusifs.

[1] Pour une description plus détaillée du réseau de distribution, voir par exemple Gonen [1986], Raineri-Rudnick [1996] ou Ajodhia-Hakvoort [2005].

1.1 Consommation

La consommation d'électricité correspond à un appel de puissance active[2] sur le réseau pour une période de temps déterminée. Cela correspond à une consommation d'énergie. La puissance est un flux d'énergie (une quantité d'énergie par unité du temps). Elle est mesurée en watts (W) ou dans un de ses multiples[3]. La quantité d'énergie consommée pendant une période de temps donnée se calcule comme l'intégrale de la puissance active pendant la période de temps considérée (l'énergie peut être mesurée, par exemple, en kilowattheures – kWh, unité de puissance par unité du temps). Il est important de remarquer que, ayant fixé la durée de la période de temps à considérer, les valeurs moyennes de puissance active peuvent être converties directement en énergie consommée (en multipliant la puissance moyenne par la durée de temps de la période). Quand l'unité de temps est sous-entendue, les termes « puissance active consommée » et « énergie » sont dès lors utilisés indistinctement.

La consommation d'énergie électrique appelée sur un réseau de transport est la somme des consommations des appareils électriques branchés à chaque instant (aux pertes près). L'électricité est consommée par des différents types de consommateurs (résidentiels, commerciaux et industriels) et, étant donné la pluralité de leurs utilisations individuelles de l'énergie électrique, elle varie à chaque moment.

La consommation d'électricité se caractérise par (1) ses fortes fluctuations, (2) la difficulté de la prévoir de manière exacte et, (3) sa faible élasticité-prix (voire nulle).

1.1.1 Fluctuations de la consommation

La consommation d'électricité varie en permanence en fonction des multiples variations de consommations qui interviennent « sans préavis » sur l'ensemble du système. Les différentes utilisations individuelles de l'énergie électrique, à chaque moment, se traduisent par de fortes fluctuations de la consommation dans le temps. Cependant, ces fluctuations ont un certain caractère cyclique au cours de la journée, de la semaine, et de l'année.

La figure 1-2 met en exergue les fluctuations de la consommation en représentant la consommation d'électricité française sur certains jours de l'année 2005. Le tableau 1-1 résume quelques valeurs importantes concernant la consommation. Les valeurs représentées ici correspondent à des moyennes de consommations en puissance active pour chacune des 48 périodes d'une demi-heure d'une journée déterminée.

---

[2] La puissance active est la valeur moyenne de la puissance sinusoïdale échangée entre un appareil électrique et un réseau électrique à courant alternatif. La puissance active est souvent qualifiée de puissance utile, par opposition à la puissance réactive, puisqu'elle correspond à la partie de la puissance qui transfert durablement l'énergie.

[3] Par exemple en kilowatts [kW] (1 kW=10^3 W), mégawatts [MW] (1MW=10^6 W) ou en gigawatts [GW] (1GW = 10^9 W).

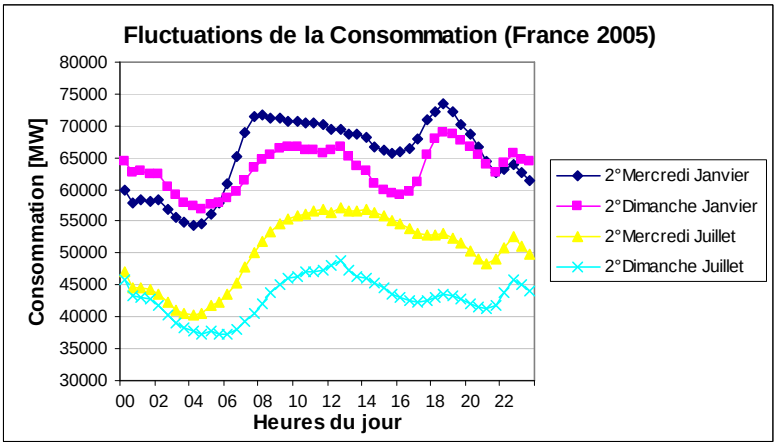

Figure 1-2 : Fluctuations de la consommation en France, 2005 (source : www.rte-france.net)

Tableau 1-1 : Fluctuations de la consommation en France, 2005 (source : www.rte-france.net)

		Minimum [GW]	Maximum [GW]	Delta Min-Max [GW]
Janvier	Semaine (2°Merc)	54	73	19
	Week end (2°Dim)	57	69	12
Juillet	Semaine (2°Merc)	40	57	17
	Week end (2°Dim)	37	49	12
Toute l'année		31	86	55

Une simple analyse de ces chiffres nous montre que la consommation évolue d'une façon cyclique au cours de la journée, de la semaine, et de l'année en créant une saisonnalité. Cette saisonnalité est constatée à trois niveaux :

- **Fluctuations infra journalières** : durant une journée, la consommation d'électricité peut fluctuer d'une façon importante (il peut avoir un écart entre la consommation minimale et maximale de 19 GW… soit 22 % de la consommation maximale de l'année). Ces fluctuations infra journalières sont récurrentes le long de l'année.
- **Fluctuations hebdomadaires :** jour ouvrable / week-end : les différences entre les consommations maximales entre un jour ouvrable et un jour du week-end du même mois peuvent aller jusqu'à 10 % de la demande maximale de l'année.
- **Fluctuations saisonnières :** jour été / jour hiver : les différences entre les consommations (maximales et minimales) entre un jour ouvrable en été et un jour ouvrable en hiver peuvent aller jusqu'à plus de 20 % de la consommation maximale.

Un autre point important à noter est la vitesse de fluctuation de la consommation. En définissant la vitesse de fluctuation comme la différence de la consommation entre deux périodes du temps (consommation en T1 moins consommation en T0) divisée par l'intervalle du temps (T1-T0) et en regardant les chiffres correspondant à la consommation en France de l'année 2005 (tableau 1-2), nous remarquons que la consommation d'électricité peut fluctuer très rapidement : elle peut changer de plus de 10% de la consommation maximale en seulement 1 heure.

Tableau 1-2 : Fluctuations de la consommation entre deux heures consécutives (France 2005, www.rte-france.net)

		Minimum [GW/h]	Maximum [GW/h]
Janvier	Semaine (2°Merc)	-4,87	8,23
	Week end (2°Dim)	-4,04	8,45
Juillet	Semaine (2°Merc)	-4,79	5,14
	Week end (2°Dim)	-4,77	4,14
Toute l'année		-7,27	11,15

Les fluctuations mentionnées précédemment correspondent à des valeurs moyennes pour un intervalle de temps d'une demi-heure. Il faut noter qu'il existe des fluctuations pour des échelles de temps inférieures. Ces fluctuations ont un caractère aléatoire. On ne peut pas assigner une quelconque périodicité à ces fluctuations. La figure 1-3 permet de constater que les fluctuations de la consommation pour chaque minute peuvent avoir des variations entre +/- 60 MW, soit plus de 3% de la valeur moyenne (ces valeurs correspondent à une région des Etats-Unis) (Kirby-Milligan [2005]).

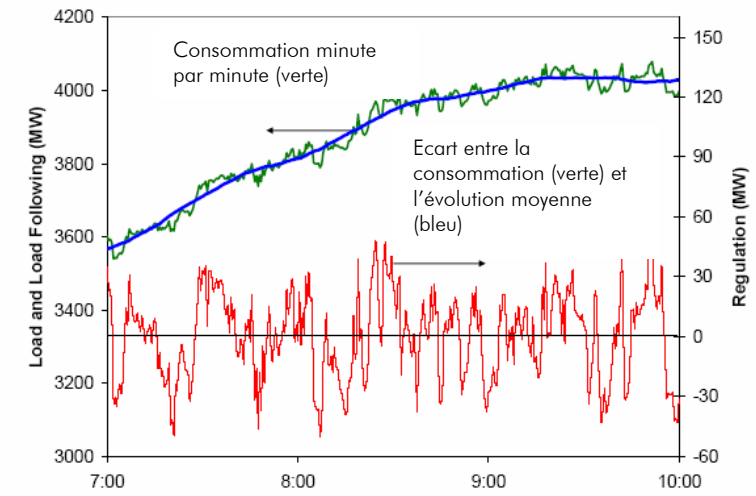

Figure 1-3 : Exemple des fluctuations de la consommation minute par minute

1.1.2 Difficultés de prévision de la consommation

Connaître la consommation de l'électricité d'une période future est important pour l'exploitation du système électrique. Pour ce faire, une multitude de variables sont traditionnellement utilisées pour expliquer et prédire le niveau de consommation d'électricité : la température, l'heure de la journée, le jour de la semaine (jour ouvrable, week-end), le prix, etc. L'impact de la plupart de ces variables est lié aux conditions climatiques, aux habitudes de consommation, aux rythmes de vie et au pays considéré. Naturellement, plus la prévision est réalisée en avance par rapport au moment de la consommation, moins elle est précise. En effet, les valeurs de ces variables, notamment celles liées aux conditions météorologiques, peuvent se modifier dans ce laps de temps. Une prévision éloignée du temps réel génère des erreurs de prévision, plus ou moins conséquentes. Les prévisions de consommation effectuées plusieurs jours à l'avance se basent principalement sur la combinaison des consommations réelles des jours précédents et

la prévision des conditions climatiques. Par exemple, une baisse de la température moyenne de 1° C sur l'ensemble de la France peut entrainer, en hiver, une augmentation de la consommation de plus de 1000 MW (approximativement la taille d'une tranche nucléaire). Une autre variable climatique importante pour la prévision est le caractère nuageux du ciel ou nébulosité (RTE [2003]).

Bien que la prévision de la consommation s'affine lorsque l'on s'approche de la période prévue, il existe encore des écarts entre les prévisions faites la veille et la consommation réelle. Ces écarts, ou erreurs de prévision, peuvent provenir des erreurs de prévision des variables explicatives (Température, nébulosité) ou/et des simplifications du modèle de prévision. Les figures 1-4 et 1-5 montrent des exemples d'écarts entre les valeurs de la consommation prévue et de la consommation réelle pour le cas de la France. Les différences entre les prévisions et la consommation peuvent être assez importantes, jusqu'à 3000 MW (3,5 % de la consommation maximale de l'année).[4]

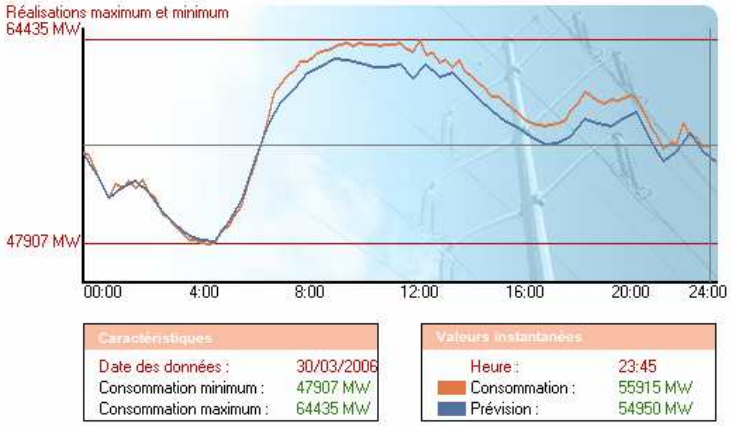

Figure 1-4: Diff. entre prév. et consommation (prév.<conso.;max écart = - 2485 MW à 12h45)[5]

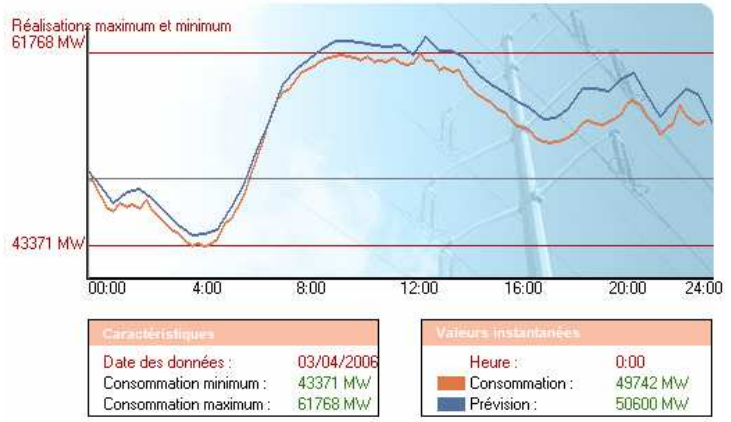

Figure 1-5: Diff. entre prévision et consommation (prév. > conso. ; max écart = 2758 MW à 14h)[6]

[4] Il est important de remarquer que ces prévisions considèrent des valeurs moyennes pour des pas de temps d'une demi-heure. Etant donné le caractère aléatoire de la consommation, des prévisions pour des pas de temps horaires plus fins qu'une demi-heure s'avéreraient moins fiables.
[5] Données correspondant au 30/03/06. Source : www.rte-france.net.
[6] Données correspondant au 03/04/06. Source : www.rte-france.net.

1.1.3 Elasticité-prix de la demande

La consommation d'électricité est la conséquence physique de la demande économique d'électricité. L'élasticité-prix de la demande d'électricité est très faible, voire nulle selon l'horizon temporel considéré (Joskow-Schmalensee [1983]). Sur le très court terme (réaction instantanée), la demande d'électricité est considérée comme inélastique au prix.

Cette inélasticité est due principalement à trois facteurs. Le premier facteur est de caractère historique. L'inélasticité-prix de la demande s'explique en grande partie par le fait que, par le passé, les consommateurs n'ont pas reçus des signaux tarifaires sur les coûts et sur la rareté aux moments de pénurie (coûts/prix élevés) (Rassenti-Smith [1998]). Le deuxième facteur, lié au premier, est d'origine technologique et propre à l'information disponible (Staropoli [2001]). Pour pouvoir réagir, les consommateurs doivent être informés instantanément de l'évolution du prix. De plus, les consommations doivent être mesurées avec la précision temporelle adéquate pour déterminer qui a consommé quoi aux différentes périodes. A l'heure actuelle, même si des innovations récentes (Internet, compteurs intelligents...) permettent d'envisager plusieurs solutions dans le futur, les technologies nécessaires au comptage en continu entraînent des coûts de mesure et de transaction qui rendent difficile la mise en œuvre de ces nouveaux systèmes (Staropoli [2001]). Le troisième facteur vient de la difficulté à trouver des substituts au bien électricité à court terme, et de son utilité très élevée pour les consommateurs à très court terme. Le choix de ne pas consommer immédiatement ou de se tourner vers d'autres produits énergétiques est très coûteux ou techniquement impossible pour les consommateurs dans la plupart des usages courants.[7]

Sur un horizon temporel plus large, par exemple en ce qui concerne le degré de réactivité des consommateurs au prix anticipé de quelques jours après, l'élasticité-prix de la demande pourrait être plus importante, au vu du temps supplémentaire disponible pour la gestion de la consommation ou la recherche de substituts. Néanmoins, les valeurs d'élasticité-prix de la demande d'électricité restent très faibles sur le court terme. Il est retenu, dans la suite de la thèse, que, sur le court terme, l'élasticité-prix de la demande est très faible, voire nulle, sauf indications contraires.

Nous avons présenté les caractéristiques de la consommation d'électricité. Nous avons montré qu'elle se caractérise par ses fortes fluctuations, par la difficulté de la prévoir de manière exacte et par sa faible voire inexistante élasticité-prix. La sous-section suivante sera consacrée à la description des caractéristiques des moyens de production d'électricité.

[7] Dans certains cas les consommateurs peuvent choisir de combiner plusieurs produits énergétiques. Par exemple, une maison équipée avec chauffage électrique et chauffage au bois, un gros consommateur doté de groupes électrogènes, etc.

1.2 Moyens de production

Nous définissons un moyen de production d'énergie électrique comme toute installation capable de convertir une source d'énergie primaire en énergie électrique pouvant être injectée sur un réseau.

Il existe une grande diversité de types de moyens de production d'énergie électrique. Chaque type de moyen de production a des caractéristiques techniques et économiques propres. Ces caractéristiques expliquent comment ces moyens peuvent être utilisés pour satisfaire la consommation d'électricité. Le but ici n'est pas la description détaillée de chacun des moyens de production existants, mais la présentation sommaire de certaines de leurs caractéristiques les plus pertinentes.

Nous nous intéressons principalement aux caractéristiques techniques et économiques ayant une influence sur l'exploitation des moyens de production à court terme (entre J-1 et le temps réel). Nous ne considérons pas, par exemple, les coûts fixes d'investissement. A long terme, l'investissement ou l'augmentation en capacité de production est lui aussi variable. Les moyens de production se différencient par leur rapport coût variable/coût fixe. Cette différenciation justifie un premier critère de choix d'investissement sur le mix de technologies, notamment lorsque l'on prend en compte la durée d'utilisation effective pendant chaque période de temps (Stoft [2002], Green [2006]). A court terme, ces coûts n'interviennent pas dans les décisions de production. C'est pourquoi les coûts fixes des installations après investissement ne sont plus considérés dans ce travail de recherche.

La centrale de production est la composante élémentaire de l'organisation des moyens de production d'électricité. Une centrale peut regrouper sur un même site plusieurs unités de production, ou tranches, souvent de même technologie et de même puissance. Chaque unité de production est composée d'une turbine (ou en général un moteur) et d'un alternateur[8] pour produire de l'énergie électrique à courant alternatif[9]. Les moteurs ou turbines peuvent être entrainés par différentes sources d'énergie primaires. La source d'énergie primaire est un élément technologique déterminant ; elle sera considérée par la suite pour une typologie des moyens de production.

[8] Un alternateur est une machine électrique rotative qui convertit l'énergie mécanique fournie par un moteur (turbine, diesel, éolienne...) en énergie électrique à courant alternatif.

[9] Le courant alternatif est un courant électrique qui change de sens plusieurs fois par seconde (fréquence 50 Hz ou cycles par seconde). Un circuit alimenté en courant alternatif présente en effet une tension et une intensité sinusoïdales, corrélées par un angle de déphasage. Les électrons ont alors un mouvement alternatif, sortant puis rentrant dans les appareils au même rythme. La plupart des réseaux de transport et de distribution fonctionnent en courant alternatif (Bastard [1998], Stoft [2002]).

1.2.1 Types de centrales électriques

Nous classons les centrales de production selon leur technologie, c'est-à-dire selon leurs sources d'énergie primaire (Menager [2002]). Nous distinguons les centrales thermiques à flamme, les centrales nucléaires, les centrales hydrauliques, et les centrales éoliennes.

Centrales thermiques à flamme

Les *centrales thermiques à vapeur* brûlent du charbon, du pétrole ou du gaz pour vaporiser de l'eau. La vapeur ainsi produite se détend dans une turbine à vapeur, qui entraîne un alternateur produisant l'électricité. Ces centrales sont caractérisées par un rendement énergétique moyen. Leurs coûts variables dépendent notamment du prix du combustible utilisé. Elles sont réputées pour être souples à l'exploitation, une fois en fonctionnement, mais lentes à mettre en fonctionnement (une centrale à fioul ou au charbon prend environ 16h pour démarrer à froid, et 8h pour démarrer à chaud).

Les *turbines à combustion,* ou les *moteurs Diesel* à combustion interne, utilisent la détente des gaz produits par la combustion de pétrole ou de gaz. Ces centrales sont caractérisées par un rendement énergétique bas. En revanche, elles sont réputées pour être très flexibles et rapides au démarrage (entre 15 minutes et 1 h).

Les *cycles combinés* sont le fruit d'une technologie hybride. Une turbine à combustion (ou plusieurs) fonctionne en parallèle d'une turbine à vapeur. Le gaz d'échappement de la turbine à combustion est utilisé pour produire de la vapeur dans une chaudière classique. Cette chaudière alimente la turbine à vapeur. Ces centrales sont caractérisées par un très bon rendement, et donc un coût variable relativement faible (mais dépendant du prix du combustible). En terme de souplesse, ces centrales sont un peu plus souples qu'une centrale thermique classique en cycle simple. Le temps de démarrage est, également, plus court qu'une centrale classique.

Centrales nucléaires

On y utilise l'énergie émise lors de la fission d'un atome à base d'uranium. La chaleur qui en résulte est transmise à un fluide caloporteur qui est utilisé pour produire de la vapeur. La vapeur, comme dans le cas des centrales thermiques classiques, est utilisée dans un turboalternateur pour produire l'électricité. Ces centrales sont caractérisées par un coût variable très faible. Leur souplesse dépend d'une programmation conçue avec anticipation. De plus, les centrales de ce type ne sont pas flexibles quand les variations sont demandées à des délais très courts. Le temps de démarrage à froid est assez long (~40 h) (Menager [2002]).

Centrales hydrauliques

L'énergie potentielle de gravité de l'eau est utilisée pour produire de l'électricité. Deux types principaux de centrales hydrauliques sont distingués : les *centrales hydrauliques au fil de l'eau* et les *centrales à réservoir*. Dans le premier type, l'eau est turbinée au fil du courant des cours d'eau. On dit que cette production est « fatale » ; si cette énergie ne sert pas à produire de l'électricité, elle est perdue. Les centrales installées au fil de l'eau

fonctionnent en continu au rythme des affluents. Dans le deuxième type, l'eau est stockée dans des réservoirs (barrages), et turbinée au rythme des besoins. Certaines turbines sont conçues aussi pour fonctionner comme pompes, afin de stocker de l'eau dans les moments de la journée où il y a un excès d'électricité. La caractéristique de stockabilité de cette technologie amène à considérer non seulement le coût variable de production pour leur exploitation, qui est réputé être très faible, mais aussi le coût d'opportunité à produire à un moment où la production sera mieux valorisée. En outre, bien que les centrales à réservoir soient caractérisées par un coût variable très faible, étant donné qu'elles ont aussi des caractéristiques de souplesse extraordinaires, elles sont utilisées en grande partie pour suivre les fluctuations brusques de la consommation.

Centrales de production éolienne

Les centrales éoliennes utilisent l'énergie mécanique contenue dans le vent pour produire de l'électricité. Elles sont caractérisées par un coût variable très faible. La production électrique des éoliennes est particulièrement intermittente, et est considérée comme fatale. Les difficultés de prévision de cette énergie (données météorologiques) renforcent la nature aléatoire de cette technologie (Hiroux-Maupas [2006]).

1.2.2 Caractéristiques des centrales électriques

Les caractéristiques technico-économiques des centrales électriques sont déterminantes pour leur exploitation. Trois types de caractéristiques ont une influence pour l'exploitation d'une centrales électriques à court terme : son coût de production ; ses contraintes techniques et sa fiabilité.

La caractéristique principale pour l'exploitation d'une centrale électrique est son coût variable de production. Pour les centrales thermiques, c'est-à-dire pour la technologie la plus répandue, le coût variable de production reflète principalement le coût du combustible utilisé et les autres coûts d'exploitation et de maintenance de la centrale (en anglais : *Operation & Management Costs*). Le coût du combustible est évalué en utilisant des valeurs (ou graphiques) de consommation spécifique de chaleur (en anglais : *heat rate*[10]) de la centrale et le prix du combustible. Le *heat rate* est proportionnel à l'inverse du rendement énergétique : plus le *heat rate* est grand, moins la centrale est performante.

La figure 1-6 montre un exemple de coût d'exploitation d'une centrale thermique (Hirst [2001]). Ce coût est normalement représenté par une équation quadratique du type $Coût = a + b\ q + c\ q^2$ où q est la quantité produite. La constante a est normalement appelée coût de marche à vide (*No-load cost* en anglais) et représente le coût pour tenir démarrée une unité de production à production nulle. Le coût incrémental (ou marginal) de

[10] La consommation spécifique de chaleur (*ou heat rate*) est la mesure de la conversion, exprimée en kJ/kWh ou en BTU/MWh, d'une quantité d'énergie thermique nécessaire pour produire de l'électricité. Les graphiques de consommation spécifique de chaleur (*heat rate*) représentent le flux d'énergie entrant dans la centrale (énergie contenue dans le combustible) et l'énergie en sortant (énergie électrique) (ECE [2006]).

production est le coût pour produire une unité supplémentaire d'énergie. Ce coût est d'important pour prendre les décisions d'exploitation à court terme ($Coût\ \ marginal = d\,Coût/dq = b + 2c\ q$), figure 1-7.

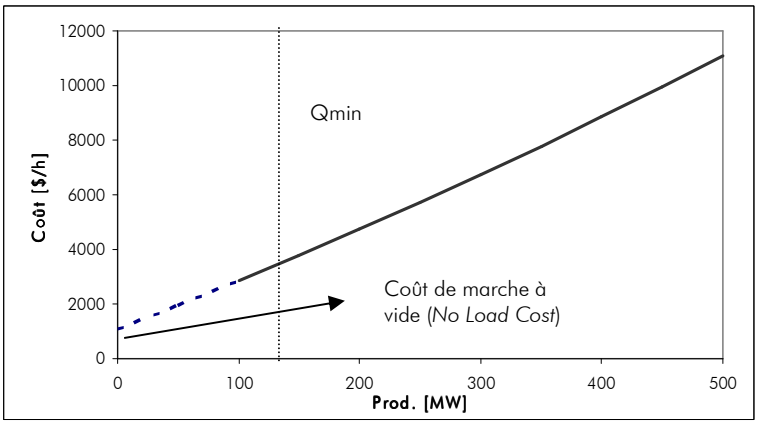

Figure 1-6 : Exemple de coût total d'exploitation d'une centrale thermique

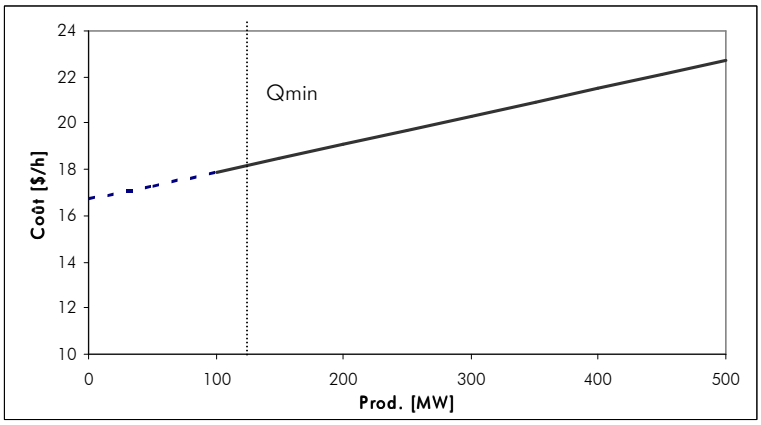

Figure 1-7 : Exemple de coût marginal d'une centrale thermique

Outre le coût variable à court terme, d'autres caractéristiques spécifiques sont importantes à mentionner pour la production d'électricité. C'est le cas notamment du coût spécifique pour démarrer ou arrêter l'unité de production (coût de démarrage et d'arrêt, *start-up cost* et *shutdown cost* en anglais). Par exemple, le coût de démarrage correspond au coût de l'énergie nécessaire pour mettre en fonctionnement toutes les installations permettant la production d'électricité (chaudières, pompes, etc.). Ce coût dépend normalement de l'état de l'unité de production au moment de l'appel à démarrer (démarrage à froid ou à chaud). Certaines contraintes techniques sont aussi importantes pour l'exploitation. Généralement, l'unité de production ne peut fonctionner de manière stable qu'à partir d'un niveau de production minimal (capacité minimale de production) et jusqu'à un niveau maximal de production (capacité maximale de production). L'inertie propre des moyens de production limite la vitesse à laquelle les unités de production peuvent changer leur niveau de production. La vitesse maximale de changement du niveau de production pour une période de temps donné est appelée contrainte de rampe (*ramp*

rate). Il existe aussi un temps minimal pour le démarrage (temps de démarrage, *start-up time*).

Enfin, les unités de production présentent différents degrés de fiabilité et d'incertitude. Ce degré de fiabilité peut être interprété comme le degré de précision dans la prévision de la capacité de production d'une centrale. Les erreurs de prévision de capacité peuvent venir du manque de prévision sur la force motrice (par exemple, courant d'eau ou vitesse du vent). L'exemple le plus typique est ici la production éolienne, dont le niveau de production dépend de la vitesse du vent. Cette vitesse est un phénomène climatique qui dépend de plusieurs variables, et qui est très difficile à prévoir avec exactitude.[11] Les erreurs de prévision peuvent venir aussi de la défaillance forcée d'une unité de production ou d'autres facteurs qui l'empêchent d'atteindre leur niveau normal de production. Le cas le plus extrême est quand l'unité n'arrive pas à démarrer comme prévu, ou qu'elle doit être arrêtée complètement pour des problèmes techniques (Hirt-Kirby [1998]).[12]

Dans le tableau 1-3 nous présentons quelques valeurs indicatives concernant les coûts et les contraintes d'exploitation des différents moyens de production de l'électricité.

Tableau 1-3 : Coûts et contraintes d'exploitation indicatifs pour chaque type de centrale[13]

	Centrales nucléaires	Centrales thermiques à flamme			Centrales hydrauliques	Centrales éoliennes
		Cent. Therm. à vapeur (C. Fossiles)	Cent. Therm. à combustion	Cent. Therm. à cycle combiné		
Taille Moyenne	~de 400 à 1300 MW par tranche	~200 à 800 MW par tranche	~ 1 à 200 MW	~400 à 100 MW	~de 50 à 1300 MW	~ de 1 à 100 MW
Coût variable / Rendement (*Heat rate*)	CV Bas Rend : ~haut (HR~10 MBTU/MWh)	CV Moyen / R: ~moyen (HR~9-13 MBTU/MWh)	CV Haut / R : ~bas (HR~13 MBTU/MWh)	CV Moyen / R : ~haut (HR~6-9 MBTU/MWh)	CV Bas / R : ~haut	CV Bas/ R : ~haut
Coût de démarrage	-	De 20 à 40 U$S/MW			Négligeable	Négligeable
Capacité minimale de production	~20 % (180 MW pour une tranche de 900 MW)	~50 %	~ 50 % à 80 %	~50 %	~ε	~ε
Contraintes de rampe *(ramping rate limits)*	~ 1 %/min (temps réel) ~ 5 %/min (prog. veille)	~ de 0,5%/min à 5 %/min	~10%/min	~7%/min	-	-
Temps de démarrage	~40 h (à froid) ~12 à 18h (à chaud)	~de 11h à 20 (à froid) ~ 5 h (à chaud)	~10 min à 1 h.	~1h à 4h	~5 min	~ε

[11] Par exemple, l'erreur moyenne de prévision pour la production éolienne peut aller entre 5 et 10% selon le moment de prévision (4h ou 24h respectivement) (Klobasta [2006]).

[12] Il y a certaines occasions où les unités de production ne peuvent pas atteindre leur niveau stipulé de production. Par exemple, une centrale à charbon peut avoir des problèmes pour atteindre le niveau maximal de production si les stocks de charbon sont humides, ou si la température de l'eau de refroidissement est élevée, ou si l'un des équipements ne fonctionne pas correctement.

[13] Source : Menager [2002], Stoft [2002] et ECE [2006].

De ce tableau nous pouvons remarquer la grande diversité des caractéristiques technologiques des moyens de production. L'opération de chaque type de moyen de production devra donc prendre en compte ces caractéristiques (cf. section 2). Nous pouvons aussi en tirer trois conséquences de ces caractéristiques sur l'exploitation du système électrique. Premièrement, la diversité des coûts variables de production et de la « flexibilité » des moyens de productions déterminera l'utilisation ou l'appel des différents moyens de production. Deuxièmement, les caractéristiques de la fonction de coût de production introduit des indivisibilités ou des non-convexités. Et, troisièmement, les erreurs de prévision de la production pour certaines technologies introduisent des incertitudes supplémentaires sur le système électrique.

Les coûts variables (et marginaux) de production sont fondamentaux pour la prise de décision de production à court terme. Parmi toutes les technologies de production disponibles, il existe une grande diversité de ces coûts marginaux de production. Cette diversité explique que des technologies différentes entreront en fonctionnement pour différentes situations de consommation du système. Pour exploiter le système de manière optimale, on devrait utiliser d'abord toutes les capacités disponibles des centrales à coût marginal faible, avant d'utiliser d'autres technologies à coût marginal plus important (cf. section 2). La figure 1-8 représente les coûts marginaux (ordonnés de manière croissante) et les capacités des centrales thermiques correspondant à l'état de Californie en 1999. Sur cette figure on remarque d'abord la grande diversité de technologies à coût marginal différent. En outre, on remarque que, quand le système arrive à ses limites de capacité de production (toutes les centrales disponibles sont utilisées), les coûts marginaux augmentent de manière considérable[14].

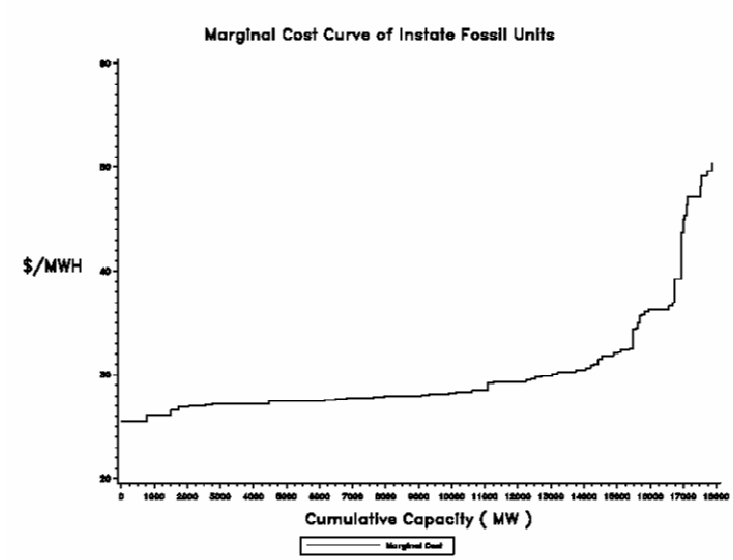

Figure 1-8 : Courbe de coût marginal des centrales thermiques, Californie (Borenstein *et al.* [1999])

[14] Cette forme de courbe, très convexe par rapport aux abscisses (aux niveaux de production), est parfois appelée « bâton de hockey » (*hockey stick*).

Le caractère de flexibilité ou de souplesse de moyens de production à court terme représente la vitesse à laquelle chaque moyen de production peut changer le niveau de sa production après un signal donné. Nous trouvons des moyens de production plus flexibles, comme les centrales hydrauliques (avec réservoir) et les centrales à combustion ou les moteurs diesel (avec des temps de démarrage faibles et des contraintes faibles de *ramping*). Par opposition, les centrales nucléaires et les centrales thermiques sont des moyens de production peu flexibles. Il est important de remarquer que cette flexibilité doit être obtenue rapidement après un ordre. Certains moyens de production peuvent avoir un caractère flexible, mais nécessiter plus de temps pour préparer cette vitesse de changement. Par exemple, certaines centrales nucléaires peuvent être programmées la veille pour réaliser des variations assez grandes de production, mais, à une échelle de temps plus proche du temps réel, les variations de production possibles pour ces centrales sont beaucoup moins élevées.

Certaines contraintes et certaines caractéristiques des moyens de production introduisent des non-convexités dans la fonction de coût. Le coût de démarrage et d'arrêt, le coût de marche à vide et la capacité minimale de production donnent un caractère de non-convexité à la fonction de coût (O'Neill *et al.* [2005]). La convexité de la fonction de coût nécessite la propriété suivante : pour produire le double, le coût total doit être au moins le double (rendement d'échelle constant ou décroissant). En présence de coûts de démarrage, cela n'est pas vérifié, c'est-à-dire qu'avant de prendre la décision de démarrer une centrale, le coût moyen de produire la première unité est plus grand que le coût moyen de la deuxième et ainsi de suite. La prise en compte des non-convexités oblige à utiliser des variables binaires pour représenter les décisions de démarrage ou d'arrêt d'unités de production (cf. section 2).

Finalement, la dernière caractéristique importante pour les moyens de production est la difficulté de prévision exacte de la capacité de production de certaines technologies. Les erreurs de prévision pour la production s'ajouteront aux erreurs de prévision de la demande, et augmenteront l'incertitude présente sur le système.

Nous avons présenté les caractéristiques de la consommation de l'électricité ainsi que celles des moyens de production. La sous-section 1.3 porte sur le réseau de transport qui relie les centres de production et les centres de consommation.

1.3 Réseau de transport

Le rôle principal du réseau de transport est la liaison entre les grands centres de consommation (gros consommateurs et distributeurs) et les moyens de productions. Ce rôle est particulièrement important car le stockage d'énergie électrique à grande échelle (tant au niveau de la production, qu'au niveau de la consommation) n'est pas économiquement viable à l'heure actuelle (Abou Chacra [2005]).

La plupart des réseaux de transport fonctionnent en courant alternatif.[15] Le transport d'électricité se réalise à haute et à très haute tension[16] (>200 kV), en raison de la diminution des pertes obtenue. En effet, le transport d'électricité est source de pertes par l'effet Joule : la puissance électrique se dissipe par échauffement lors de l'écoulement des électrons dans un corps conducteur. L'élévation du niveau de tension permet de réduire les pertes d'énergie. En fait, pour une même puissance à transmettre, en doublant la tension d'un réseau de transport, les pertes sont diminuées par un rapport de 4. Puisque la production de l'électricité et la consommation se réalisent à des niveaux de tension plus faibles[17], des transformateurs[18] sont nécessaires pour la mise à niveau de tension.

Les réseaux de transport d'électricité sont composés de lignes[19] (ou de couloirs de lignes) reliant différentes « barres » ou « nœuds » et ils ont généralement une topologie « maillée ». Ceci renvoie à l'existence de plusieurs chemins pour aller d'un nœud à un autre du réseau (figure 1-9 pour le réseau de transport français). Bien entendu, des topologies radiales (un seul « chemin » électrique reliant deux nœuds du réseau) sont aussi possibles (comme par exemple le réseau de transport en Argentine ou au Chili – figure 1-10 – Sioshansi [2006]).

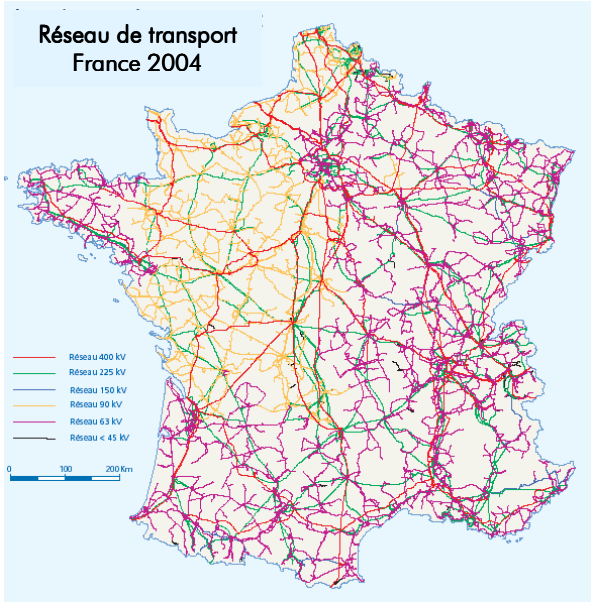

Figure 1-9 : Exemple de réseau maillé : Réseau de transport de la France (source RTE [2005])

[15] Certains réseaux de transport ont des parties qui fonctionnent en courant continu (par exemple, le système électrique nordique, cf. www.nordel.org). De même, les interconnexions entre réseaux de transport non synchronisés se réalisent aussi par le biais d'interconnexions à courant continu (par exemple la liaison France-Angleterre).

[16] La tension reflète l'énergie potentielle qui est susceptible d'être libérée par une charge électrique (Bastard 1998]. Elle est mesurée en Volts [V] ou dans un de ses multiples (1 kV = 1000 V).

[17] Par exemple, un alternateur de centrale thermique fonctionne sous 20 à 30 kV, les moteurs industriels fonctionnent souvent sous quelques kV et les foyers domestiques sont alimentés en 220 V (Menager [2002]).

[18] Un transformateur est une machine électrique statique qui a comme fonction d'augmenter ou de diminuer le niveau de tension.

[19] Les transformateurs peuvent être représentés comme des lignes avec des caractéristiques spéciales.

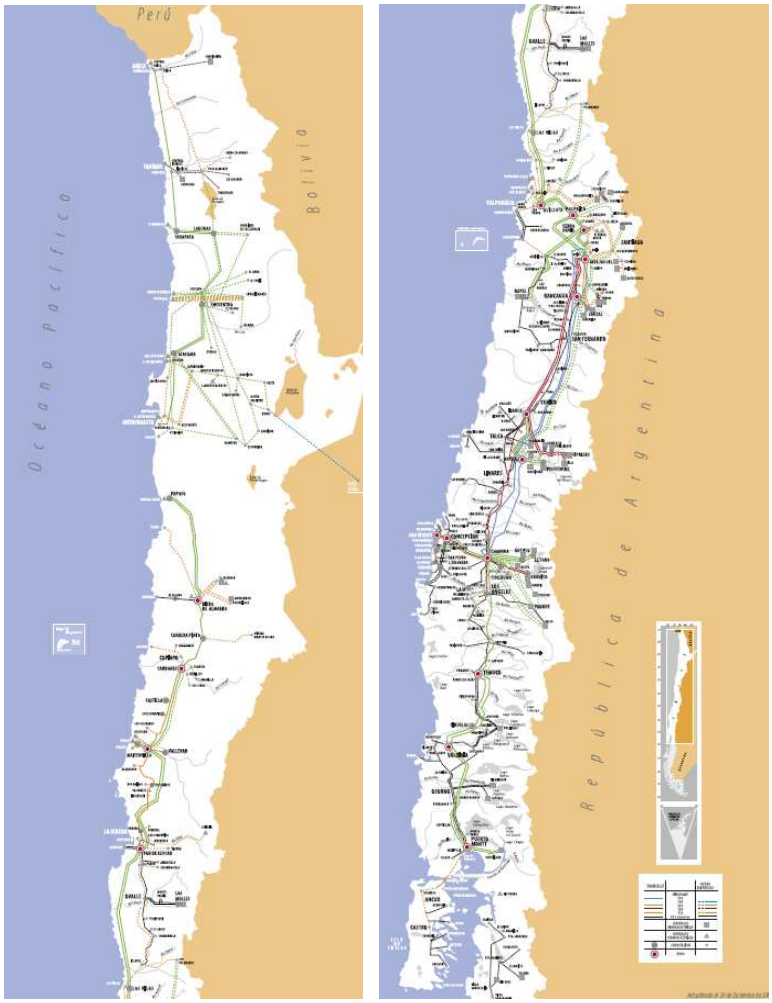

Figure 1-10 : Exemple de réseau radial : Réseau de transport de Chili (source Transelec [2005])

Pour profiter des bénéfices d'un réseau de transport, celui-ci doit être exploité d'une manière particulière : il doit être exploité dans les limites de fonctionnement autorisées. Ces limites ou contraintes du réseau sont exprimées par des valeurs maximales ou minimales sur certaines variables du réseau (fréquence, flux de puissance sur les lignes ou transformateurs, niveau de tension, etc.). Si ces limites sont dépassées, le réseau risque de devenir instable ou de nuire à la « vie » de ses éléments ; tous les bénéfices apportés par le réseau peuvent disparaître, générant dès lors un coût très élevé.

Pour simplifier, nous ne considérons que deux types de contraintes de fonctionnement : l'équilibre du système et le respect des contraintes de capacité de transport. Comme nous le verrons par la suite, la contrainte d'équilibre production-consommation est directement liée à la fréquence du système. Les contraintes de capacité de transport, quant à elles, sont liées principalement aux flux maximaux de puissance qui peuvent circuler sur chacun des éléments du réseau.[20] Ces contraintes de capacité ont une importance particulière dans les réseaux électriques car les flux d'électricité sont difficiles à

[20] Par souhait de simplicité, les autres types de limites, comme celles sur le niveau de tension aux nœuds ou celles de stabilité, ne seront pas traitées.

contrôler et suivent des chemins gouvernés par des lois physiques complexes, dites lois de Kirchhoff. Nous décrivons par la suite ces contraintes et les lois de Kirchhoff afin de comprendre le fonctionnement du réseau et ses limites.

1.3.1 Equilibre du système électrique

L'équilibre d'un réseau de transport exige qu'à tout moment, la puissance injectée (= la production) soit égale à la puissance soutirée (= la consommation) plus les pertes générées sur le réseau.[21] Ainsi, les systèmes électriques subissent-ils une forte contrainte d'équilibre en temps réel entre les injections et les soutirages.

Cet équilibre production-consommation est nécessaire tout d'abord car les systèmes électriques à courant alternatif fonctionnent comme une « grande » machine synchronisée. Le fonctionnement de cette machine est très complexe et particulièrement vulnérable aux instabilités. Ces instabilités se produisent principalement quand l'équilibre production-consommation (P=C) n'est pas respecté.[22] Assurer l'équilibrage continuel, même lors des incidents, est un moyen de maintenir la stabilité du système. En fait, même de faibles écarts hors de la position d'équilibre modifient la « vitesse » de fonctionnement de l'ensemble du système, qui est la fréquence, exprimée en Hertz [Hz ou cycles/seconde].

En fonctionnement normal, on peut considérer que la fréquence est uniforme à un instant donné sur l'ensemble du réseau : les alternateurs[23], étant reliés entre eux par le jeu des forces électromagnétiques, tournent tous à la même vitesse électrique ou fréquence.[24] La fréquence reste avant tout un excellent indicateur de déséquilibre entre production et consommation. Ainsi, pour maintenir cet équilibre, la fréquence est asservie à une valeur nominale. En effet, de nombreux éléments techniques du système électrique (moyens de production, certains appareils électriques comme les moteurs) ont été conçus pour fonctionner à une fréquence de référence ; en Europe, cette valeur nominale est de 50 Hz. De trop grandes variations de fréquence sont, en outre, inadmissibles pour certains composants, dont les groupes de production, qui se déconnectent du réseau pour des écarts de fréquence de 2 à 4 Hz (RTE [2003]). Ainsi, des écarts, même mineurs, de la fréquence de référence peuvent déstabiliser ou endommager des éléments du système de transport et entraîner des conséquences dommageables comme une panne totale du système ou *black out*.[25]

---

[21] Approximativement 2 % de la puissance produite (Quiquempoix *et al.* [2004]).
[22] Les instabilités peuvent être causées par des défaillances apparues à n'importe quel endroit ou sur n'importe quelle composante du système. Le défaut d'une ligne ou d'une unité de production peut provoquer des problèmes de stabilité et ainsi générer la déconnexion des autres unités de production ou machines.
[23] Chaque unité de production est composée d'une turbine (ou en général un moteur) et d'un alternateur pour transformer l'énergie mécanique en énergie électrique (à courant alternatif).
[24] C'est valable pour des systèmes à courant alternatif. Les systèmes interconnectés avec des liaisons à courant continu ne sont pas synchronisés, et chaque zone a, donc, sa propre fréquence.
[25] Le défaut d'une ligne ou d'une unité de production peut provoquer des phénomènes en cascade pouvant se développer rapidement. Si lors d'un incident, les actions correctives nécessaires ne sont pas effectuées, le système peut défaillir complètement (*black out*) (CRE-AEEG [2004]).

Par ailleurs, l'équilibrage permanent du système électrique est rendu difficile par le fait que l'électricité est très coûteuse à stocker (cf. le coût des piles électriques). A cette absence de stockage bon marché s'ajoutent de nombreuses incertitudes, notamment sur la consommation qui change presque constamment mais sans préavis ni engagement (cf. sous-section 1.1). Il en résulte que les systèmes électriques sont en permanence en cours d'ajustement de leur production pour maintenir leur équilibre et la fréquence de référence.[26]

Outre l'équilibrage à tout instant de la production avec la consommation, l'exploitation d'un réseau de transport doit respecter les contraintes de capacité de transport ; en d'autres termes il faut veiller à ce que les flux de puissance transitant par les différents éléments d'un réseau soit inférieurs au seuil permis (flux maximaux de puissance). Ces contraintes de capacité ont une importance particulière dans les réseaux électriques car les flux d'électricité sont difficiles à contrôler et suivent des chemins gouvernés par des lois physiques complexes, dites lois de Kirchhoff. Nous décrivons par la suite les lois physiques qui gouvernent les flux électriques sur un réseau de transport.

1.3.2 Les équations du réseau, les lois de Kirchhoff et les PTDFs

Les lois de Kirchhoff sont fondamentales pour comprendre comment les éléments du réseau sont utilisés et ainsi pour s'assurer que l'utilisation restera dans les limites autorisées (contraintes). Les flux de puissance sur les différents éléments d'un réseau de transport à courant alternatif ne peuvent pas être contrôlés directement. Lorsqu'on injecte de l'électricité dans un nœud et que l'on soutire dans un autre nœud du réseau, on ne peut pas choisir le chemin suivi par cette électricité. Les flux électriques suivent tous les chemins possibles entre le nœud d'injection et le nœud de soutirage et la répartition de ces flux est conforme aux équations de réseau développées sur la base des lois de Kirchhoff (Bastard [1998]).[27] Il est important de noter qu'elles ne posent pas de problème *per se* mais elles entraînent une difficulté supplémentaire à l'heure de la gestion des contraintes du réseau. Les lois de Kirchhoff sont une spécificité des réseaux électriques.

A partir des deux lois de Kirchhoff (loi des mailles et loi des nœuds), il est possible d'écrire les équations dites « de réseau » (Bastard [1998], Stoft [2002]). Ces équations expriment les flux de puissance transitant sur les lignes en fonction des caractéristiques électriques du réseau de transport (topologie, impédances des lignes et transformateurs)[28] et de l'ensemble des injections et soutirages sur le réseau. Ces lois de Kirchhoff permettent

[26] L'ajustement significatif du côté de la consommation est aussi possible mais moins répandu en conditions normales de fonctionnement. Le seul ajustement significatif du côté de la consommation est involontaire et brutal : c'est la coupure de courant (effacement ou délestage).

[27] La répartition des flux conformément aux lois de Kirchhoff est à l'origine de flux « parallèles » entre deux nœuds, souvent appelés « *loop flows* ».

[28] Les propriétés électriques des composantes d'un réseau de transport (lignes, transformateurs, etc.) sont représentées par un ensemble de paramètres, dont l'impédance est l'un des plus importants. L'impédance électrique mesure l'opposition d'une composante électrique au passage du courant alternatif sinusoïdal. Le concept d'impédance est une généralisation de la « résistance », des circuits en courant continu (loi d'Ohm), dans l'étude des circuits en courant alternatif.

donc de modéliser la répartition des flux sur le réseau sous un ensemble d'hypothèses concernant l'état du système électrique (ensemble des lignes effectivement connectées et ensemble des éléments du réseau en maintenance) et les utilisations simultanées du réseau par les consommateurs et les producteurs d'électricité.

Les équations de réseau (tirées des lois de Kirchhoff pour un réseau en régime de courant alternatif) sont non-linéaires et considèrent la puissance active [29] comme la puissance réactive.[30] Cependant, les principales conséquences directes des lois de Kirchhoff peuvent être vues dans une version simplifiée (linéaire) des équations de réseau sous l'approximation du courant continu (*DC load flow model*).[31] Cette approximation permet de connaître la répartition de flux sur les lignes du réseau par le biais des coefficients d'influence (ou PTDF, *Power Transfer Distribution Factor* en Anglais) et l'ensemble des injections et soutirages. Un coefficient d'influence (PTDF) permet d'exprimer la variation relative de flux de puissance, sur une ligne particulière, due au changement de l'injection et du soutirage associés à une paire de nœuds (Wood-Wollenberg [1996], Baldick [2003b]). Sous l'approximation du courant continu, les PTDFs peuvent être obtenus à l'aide d'inversions matricielles basées sur les caractéristiques du réseau (les impédances et la topologie).[32]

Considérons un réseau simple de trois nœuds et trois lignes (figure 1-11). Les trois lignes possèdent la même impédance[33]. En utilisant l'approximation du courant continu et en choisissant un nœud de référence (nœud 3 pour cet exemple) nous pouvons calculer les PTDFs (tableau 1-4). Les valeurs représentées sur le tableau correspondent à la répartition de puissance sur les différentes lignes électriques (lignes du tableau) après une injection sur un nœud quelconque (colonnes du tableau) et un soutirage au nœud de référence. A partir de ces PTDFs, il est possible d'évaluer la contribution au flux sur chaque ligne de n'importe quelle paire injection-soutirage. Pour notre exemple, une injection d'une unité de puissance (1 MW) sur le nœud 1 et un soutirage sur le nœud 3 crée un flux de 1/3 MW sur les lignes 1-2 et 2-3, et un flux de 2/3 MW sur la ligne 1-3. Ceci exprime simplement que la répartition de flux sur les lignes d'un réseau répond à l'idée du chemin de moindre « résistance ».

---

[29] La puissance électrique active est la valeur moyenne de la puissance sinusoïdale (alternative) échangée entre un alternateur ou un appareil électrique et le réseau.

[30] La puissance réactive, « collatérale » à la puissance active, est de valeur moyenne nulle. Elle fournit de l'énergie puis la reprend immédiatement. Elle permet néanmoins de stabiliser le fonctionnement du réseau en contrôlant le niveau de tension du réseau.

[31] Notez bien que, contrairement à son appellation, la méthode d'« approximation du courant continu » est utilisée pour les réseaux fonctionnant à courant alternatif. Cette méthode d'approximation réalise une série de simplifications : elle considère seulement la puissance active (ignorant la puissance réactive et considérant le niveau de tension comme fixe) ; elle ignore les pertes d'énergie sur le réseau ; et enfin, les différences de phases (angle du vecteur de tension) entre nœuds sont supposées très faibles. Ces hypothèses permettent de linéariser et de simplifier les équations définissant les flux de puissance active sur chaque ligne (Baldick [2003b]).

[32] En théorie, les PTDFs dépendent de la topologie du réseau, du point de fonctionnement et du comportement des éléments composant le système électrique (Baldick [2003b]).

[33] Cf. note 28.

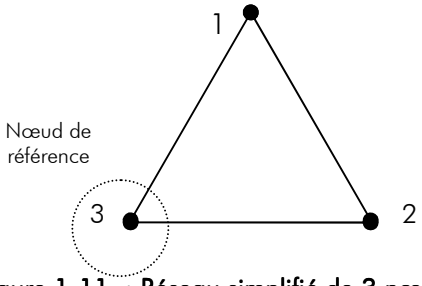

Figure 1-11 : Réseau simplifié de 3 nœuds

Tableau 1-4 : PTDF

	Nœud 1	Nœud 2	Nœud 3
Ligne 1-2	1/3	-1/3	0
Ligne 2-3	1/3	2/3	0
Ligne 1-3	2/3	1/3	0

En conclusion, l'utilisation de coefficients d'influence (PTDF) calculés sous l'approximation de « courant continu » permet de connaître la répartition de flux sur les lignes du réseau d'après un ensemble des injections et soutirages. Malgré les simplifications de cette méthode d'approximation (absence de considération des pertes d'énergie et des problèmes de tension ou de puissance réactive), elle suffit pour comprendre la grande majorité des spécificités technico-économiques des réseaux de transport. Nous retiendrons pour la suite cette version simplifiée et linéaire des équations de réseau.

Une fois décrites les équations de réseau et la fonction de répartition des flux électriques sur les différents éléments du réseau, nous pouvons décrire les différentes limites de capacité de transport d'un réseau électrique.

1.3.3 Limites de capacité de transport d'un réseau électrique

Les éléments (lignes, transformateurs, etc.) composant le système de transport d'électricité ont des limites de fonctionnement. Si l'on dépasse ces limites de fonctionnement (les flux de puissance autorisés), les éléments peuvent être gravement affectés ou se déconnecter intempestivement.[34] Cela impose une certaine attention à l'utilisation de ces éléments. Quand un élément du réseau atteint sa capacité d'utilisation maximale, on dit qu'il y a une « congestion ». Puisque la répartition des flux sur les différents éléments du réseau suivent les équations de réseau (dérivées des lois de Kirchhoff), et qu'on ne peut pas contrôler directement comment ces éléments sont utilisés, les actions possibles afin d'éviter les congestions sont limitées. En effet, le contrôle des flux sur les éléments du réseau est réalisé principalement en modifiant les injections et soutirages.[35]

Les congestions non-contrôlées font courir le risque d'une rupture d'approvisionnement extrêmement coûteuse pour la société. En effet, lorsqu'une ligne se déconnecte suite à une surcharge, le flux supporté initialement par cette ligne va se reporter

[34] Par exemple la surcharge d'une ligne peut provoquer une élongation par un effet de dilatation ; cette élongation entraine la diminution de la distance entre la ligne et le sol et peut être à l'origine d'un court circuit.
[35] La modification de la topologie du réseau (par connexion et/ou déconnexion de certaines lignes, ou par modifications de commande des éléments « flexibles » du réseau à propriétés variables transformateurs déphaseurs/FACTS) peut être utilisée pour contrôler la répartition des flux. Cependant, la modification des injections et des soutirages est la méthode la plus couramment utilisée.

sur les lignes voisines, ce phénomène est appelé « report de charge » (RTE [2003]). En outre, ce report automatique des flux risque à son tour de surcharger d'autres lignes interconnectées du réseau, et d'entraîner un écroulement en chaîne du réseau. Pour cela, les systèmes de transport ne sont pas exploités à leurs vraies limites physiques mais avec des marges de sécurité en utilisant des critères ou des règles de sécurité. Pour ce faire, deux types de limites sont considérés : les limites physiques en fonctionnement normal (limites thermiques) et les limites provenant des règles de sécurité.

Limites physiques (thermiques) en fonctionnement normal

Les lignes électriques et les transformateurs ont des capacités de transport physiquement limitées. Ces limites sont principalement thermiques.[36] Les limites thermiques sont liées à l'échauffement des lignes lors du passage du courant électrique. L'effet Joule entraîne une transformation de puissance électrique en puissance thermique. Cette énergie thermique provoque une augmentation de la température de la ligne. Ce changement de température de la ligne modifie les caractéristiques mécaniques de la ligne et provoque une dilatation des conducteurs. L'énergie dégagée par l'effet Joule augmente avec le courant électrique circulant sur la ligne. Or, à tout instant, on doit garantir que le courant de transit dans les éléments du réseau de transport se situe au dessous du seuil fixé : intensité maximale du courant admissible en régime permanent (IMAP). En cas de dépassement, on dispose alors d'un temps limité, variable selon l'ampleur du dépassement, pour ramener le flux électrique à une valeur acceptable. Si les actions nécessaires ne sont pas effectuées dans ce temps limité, le mécanisme de protection de surcharge déconnectera l'élément du réseau. En général, les limites physiques (thermiques), peuvent être présentées comme une limite maximale de puissance active (par exemple en [MW]) qui peut transiter sur un élément déterminé du réseau.[37] L'exploitation du réseau de transport doit assurer que les flux de puissance transitant par les différents éléments respectent toujours ces limites physiques maximales.

Limites provenant des règles de sécurité

Les règles de sécurité sont utilisées pour l'exploitation du réseau de transport afin de diminuer le risque d'un écroulement en chaîne du réseau suite à la déconnexion fortuite d'un ou de plusieurs éléments du réseau. La règle de sécurité la plus répandue est celle connue sous le nom de règle « N-1 ».[38] La règle du « N-1 » impose que le réseau puisse admettre un fonctionnement acceptable (respect de contraintes physiques) en cas de

[36] Les autres limites, non traitées ici, sont les limites de stabilité (Stoft [2002]). Les limites de stabilité en régime permanent sont dues aux différences de phase entre les tensions de deux nœuds du réseau. En théorie, le flux de puissance active dans une ligne est susceptible de croître jusqu'à ce que la différence de phase entre les deux tensions aux extrémités de cette ligne atteigne 90°. Ceci met une limite théorique à la puissance qui peut circuler sur la ligne. En pratique, les limites de stabilité sont fixées à un niveau bien inférieur (30° par exemple) car il faut aussi tenir compte des instabilités transitoires. Comme pour les limites thermiques, une limite de stabilité peut être spécifiée comme un maximum de flux qui peut passer par la ligne.

[37] Ceci représente néanmoins une approximation puisque c'est physiquement la valeur du courant (en Ampères) et non celle de la puissance active (en MW) qui limite le transit dans une liaison.

[38] Il existe cependant des règles de sécurité « N-k » plus exigeantes. Par exemple, la règle sur les lignes d'interconnexion entre la France et l'Italie est celle du « N-2 » qui considère les cas où deux éléments du système deviendraient défaillants (Glachant-Pignon [2005]).

défaillance de n'importe quel équipement du système.[39] Pour ce faire, on définit une série de scenarios qui considèrent l'état du réseau suite à la défaillance d'un élément (par exemple une ligne) à la fois : un « incident hypothétique » (ou en Anglais « *contingence* »). En effet, pour un ensemble d'injections et de soutirage donné (*load flow patern*), on évalue la répartition des flux de puissance pour chacun des scenarios. Les flux de puissance sur chaque élément évalués « après incident » doivent respecter les contraintes physiques.

En conclusion, un réseau de transport doit être exploité en respectant deux types de contraintes de fonctionnement : l'équilibre du système et le respect des contraintes de capacité de transport. Les contraintes de capacité de transport sont liées principalement aux flux maximaux de puissance qui peuvent circuler sur chacun des éléments du réseau. Les flux circulant sur les différents éléments du réseau doivent respecter deux types de limites : les limites thermiques et les limites venant de règles de sécurité (N-1).

1.4 Conclusions de la section

Cette section porte sur une vision générale des systèmes électriques et de ses caractéristiques propres. Ces caractéristiques propres d'un système électrique sont indépendantes des formes institutionnelles d'organisation. Or, l'étude de ces caractéristiques technico-économiques est une étape indispensable parce qu'elles conditionneront les différences de nature du choix et de la conception des différents dispositifs de coordination pour gérer ces caractéristiques.

Les trois parties du système électrique ont été analysées : la consommation, les moyens de production et le réseau de transport. Ainsi, la consommation est caractérisée premièrement par ses fluctuations importantes. Ces fluctuations ont un caractère cyclique pendant la journée, la semaine et l'année, mais elles ont aussi un caractère aléatoire minute par minute. Deuxièmement, la consommation est caractérisée par la difficulté de la prévoir. Les outils actuels de prévision arrivent à s'approcher des valeurs réelles en utilisant différentes variables explicatives (les conditions climatiques, etc.) mais il y a encore des erreurs de quelques pourcents pour la prévision de la consommation. Ces erreurs de prévision sont la principale cause de la présence d'incertitudes sur le réseau. Troisièmement, la consommation d'électricité est la conséquence d'une demande inélastique au prix à court terme.

Ensuite, les moyens de production se caractérisent par leur grande diversité et leurs caractéristiques. Il y a trois caractéristiques importantes à prendre en compte pour la suite de la thèse : la diversité et la forme des fonctions de coûts pour les différentes technologies ; la flexibilité ou souplesse des moyens de productions ; et les erreurs de prévision de la production pour certaines technologies. Les coûts variables de production sont fondamentaux pour les décisions de production à court terme. Les différents coûts

[39] En pratique, une liste d'incidents « critiques » est considérée. Ces incidents sont définis par la défaillance d'un ou plusieurs éléments du réseau.

marginaux de chaque technologie guideront l'utilisation optimale des moyens de production. Certaines contraintes et caractéristiques des moyens de production introduisent des non-convexités dans la fonction de coût. Cette caractéristique des non-convexités a des conséquences sur la manière de gérer l'utilisation des moyens de production. Le caractère de flexibilité ou de souplesse de moyens de production à court terme, représente la vitesse à laquelle chaque moyen de production peut changer le niveau de sa production après un signal donné. Par ailleurs, la difficulté liée à la prévision de la production de certains moyens de production (par exemples les éoliennes) accroît l'incertitude (initialement liée aux erreurs de prévision de la demande).

Enfin, le rôle principal du réseau de transport est la liaison entre les grands centres de consommation (gros consommateurs et distributeurs) et les moyens de productions. Ce rôle est particulièrement important car le stockage d'énergie électrique à grande échelle (tant au niveau de la production, qu'au niveau de la consommation) n'est pas économiquement viable à l'heure actuelle. Le réseau de transport doit être exploité conformément à ses limites de fonctionnement. Pour qu'un réseau fonctionne correctement, deux types de contraintes doivent être respectés absolument : l'équilibre du système et le respect des contraintes de capacité des éléments du réseau (lignes et transformateurs). D'une part, un réseau de transport doit, à tout moment, être en équilibre, c'est à dire que la puissance injectée doit être égale à la puissance soutirée (en tenant compte des pertes générées sur le réseau). Les perturbations de cet équilibre entrainent des variations importantes de fréquence ainsi que des problèmes de stabilité ; elles peuvent même provoquer une panne totale du réseau et du système. D'autre part, les éléments du réseau de transport (lignes, transformateurs, etc.) ont des limites de fonctionnement. Si l'on dépasse ces limites de fonctionnement, les éléments peuvent être gravement affectés ou se déconnecter intempestivement. Les limites ou contraintes sur un réseau électrique sont exprimées par des valeurs maximales ou minimales sur certaines variables. Deux types de limites sont reconnus comme étant très importants : les limites physiques en fonctionnement normal (thermiques) et les limites venants des règles de sécurité (par exemple : règle «N-1»). La plupart des limites peut être exprimée comme le flux maximal de puissance transitant par l'élément concerné. Le contrôle des flux sur les éléments du réseau n'est pas une tâche simple, car la répartition des flux sur le réseau ne peut pas être commandée directement. La répartition de flux sur les différents éléments du réseau suit les lois de Kirchhoff. Ces lois peuvent être approximées par l'utilisation des coefficients d'influence (PTDF) qui nous indiquent les flux de puissance engendrés sur chaque ligne pour une paire injection-soutirage donnée.

Une fois présentées les caractéristiques propres d'un système électrique, nous identifions les véritables problèmes de coordination présents dans cette industrie. Dans la section suivante nous nous concentrerons sur l'étude de la forme d'organisation la plus simple, d'un point de vue opérationnel, pour gérer de manière optimale ce système complexe. Cette forme est celle d'une entreprise verticalement intégrée bienveillante ou avec régulation parfaite.

Section 2: Exploitation optimale par une entreprise intégrée

Nous avons vu les caractéristiques particulières des différents éléments d'un système électrique : la consommation d'électricité, les moyens de productions et le réseau de transport. Un résumé de ces caractéristiques peut être trouvé dans le tableau 1-5. Exploiter un système avec ces caractéristiques nécessite, dès lors, la mise en place de dispositifs de coordination adéquats.

Tableau 1-5 : Résumé des caractéristiques propres d'un système électrique

Consommation	• grandes et rapides fluctuations (jour, saison) • prévision imparfaite (incertitudes) • inélasticité au prix
Moyens de production	• diverses technologies de production à caractéristiques propres o coût variable, coût de démarrage/arrêt, coût à marche vide. o contraintes min/max de production, *ramping*, temps démarrage • prévision imparfaite : défaillances, éolien (incertitudes)
Réseau de transport	• l'électricité n'est pas stockable • contraintes du réseau o production égale à la consommation à tout moment o limites de capacités des différents éléments du réseau.

L'exploitation du système électrique peut être organisée selon différentes formes, allant d'une version centralisée (cas de l'entreprise intégrée régulée) à des versions plus concurrentielles, basées sur le marché. Cette section considère la forme d'organisation correspondant à une entreprise verticalement et horizontalement intégrée, bienveillante ou parfaitement régulée. Cette forme d'organisation est la plus simple, du point de vue du fonctionnement, car toutes les décisions sont concentrées dans une seule entité prenant, par hypothèse, les meilleures décisions possibles. Notre but est d'avoir une référence d'exploitation optimale ; nous utilisons donc l'idée d'une entreprise verticalement intégrée « idéale », parfaitement régulée. L'hypothèse supplémentaire d'une régulation parfaite suppose que l'autorité de régulation (Ministère, régulateur, etc.) pourra veiller à ce que l'entreprise intégrée suive l'objectif de maximisation du bien être social. L'entreprise intégrée idéale fonctionne donc comme un planificateur bienveillant.[40]

Cette entreprise verticalement intégrée idéale utilise une séquence de contrôles-commandes afin d'atteindre son objectif. Les deux principaux maillons de cette séquence de contrôles sont deux outils d'optimisation sous contraintes : un programme d'optimisation

[40] Sans cette hypothèse supplémentaire, cette forme d'organisation ressemblerait en quelque sorte au fonctionnement typique des systèmes électriques avant les années 90, c'est-à-dire avant l'introduction de la concurrence dans le secteur électrique (Hunt [2002])). Cette hypothèse est loin d'être vérifiée en pratique étant donné le degré de spécificité des entreprises et les ressources limitées du régulateur. En effet, l'asymétrie d'information rend difficile la régulation parfaite des entreprises et les difficultés trouvées (ou imperfections) dans la mise en place de différents systèmes de régulation se révèlent comme l'une des principales causes du changement des formes organisationnelles privilégiées dans l'industrie électrique, notamment avec l'introduction de la concurrence (Stoft [2002], Lévêque [2004]).

appelé « *unit commitment* » et un autre appelé « *optimal dispatch* ». Une troisième étape correspond aux contrôles ou asservissement automatiques du système en temps réel. L'utilisation de cette séquence de contrôles-commandes permet à l'entreprise de sélectionner la meilleure combinaison des moyens de production et leur utilisation afin de minimiser le coût total de production tout en respectant les contraintes et caractéristiques propres du système électrique.

Bien que l'expérience internationale indique que les formes d'organisations présentées dans cette section (entreprises intégrées et régulées) soient en quelque sorte « démodées » et qu'elles sont de plus en plus abandonnées, il est important d'avoir une vision d'entreprise intégrée idéale pour fixer les bases nécessaires (techniques et économiques) pour comprendre des formes d'organisation plus complexes comme celle introduite par la concurrence. Autrement dit, la forme organisationnelle d'un planificateur bienveillant représenté par un monopole verticalement intégrée idéal en tant que référence permettra par la suite de comprendre les mécanismes concurrentiels qui sont intégrés lorsque des réformes électriques sont mises en œuvre.

Selon un ordre logique du plus simple au plus complexe, cette deuxième section va considérer la forme d'organisation correspondant à une seule entreprise qui gère un système électrique « isolé ». Cette entreprise contrôle donc tous les moyens de production connectés au réseau de transport sur une seule zone de contrôle. Dans la troisième section de ce chapitre, nous considérerons le cas où plusieurs entreprises verticalement intégrées fonctionnent sur un même réseau de transport avec plusieurs zones de contrôle. Pour y parvenir, il est important de développer le concept d'étendue d'un réseau de transport et le concept de zone de contrôle.

Cette section va s'organiser comme suit. Dans la sous-section 2.1, nous présentons la notion de « zone de contrôle » sur un réseau de transport. Dans la sous-section 2.2, nous présentons la séquence des outils de contrôle-commande (*unit commitment*, *optimal dispatch*, asservissements automatiques) pour exploiter de manière optimale un système électrique sur une zone de contrôle. Dans la sous-section 2.3, nous présentons les conclusions.

2.1 Réseau de transport et zone de contrôle

L'étendue géographique d'un réseau de transport et l'étendue géographique d'une zone de contrôle sont deux concepts différents. L'étendue d'un réseau de transport fait référence aux caractéristiques technologiques et de fonctionnement physique du réseau. L'étendue de la zone de contrôle, en revanche, fait référence à la forme d'organisation pour le contrôle du réseau ou d'une partie du réseau. Par exemple, l'étendue géographique du réseau de transport interconnecté UCTE correspond à une grande partie de l'Europe. Sur ce réseau de transport, qui fonctionne physiquement comme une seule machine synchronisée, on trouve plusieurs zones de contrôle, en général une zone de contrôle par pays.

2.1.1 Étendue du réseau de transport

Un réseau de transport à courant alternatif fonctionne comme une grande machine synchronisée. Des actions dans n'importe quel nœud du réseau (injection ou soutirage) peuvent modifier sa fréquence et les flux de puissance sur toutes les lignes composant le réseau. Pour définir « physiquement » l'étendue d'un réseau de transport, il est nécessaire de prendre en compte tous les éléments constituant cette grande machine. Ainsi, nous considérons la définition de l'étendue du réseau de transport dans son sens large, correspondant à l'ensemble du système de transport synchronisé et « isolé », au moins partiellement, d'autres systèmes électriques.

Cette définition considère un réseau de transport composé par plusieurs sous-réseaux de transport et interconnectés par des liaisons à courant alternatif. Car ces sous-réseaux, une fois interconnectés, se comportent physiquement comme un seul réseau. Un exemple typique d'un réseau de transport composé de plusieurs sous-réseaux est le réseau UCTE. Ici, les sous-réseaux correspondant à chaque pays européen sont interconnectés et synchronisés fonctionnant comme un réseau de transport unique (figure 1-12). Bien entendu, cette définition de l'étendue du réseau de transport correspond aussi aux réseaux de transport d'une région isolée physiquement et du coup électriquement ou n'ayant que des liaisons à courant continu avec d'autres systèmes électriques (par exemple, le Texas aux Etats-Unis ou l'île de la Grande Bretagne en Europe). En effet, les liaisons à courant continu permettent un très grand contrôle et les réseaux peuvent être considérés comme indépendants du point de vue du comportement physique. [41]

Figure 1-12 : Réseau de transport UCTE (source : www.ucte.org)

[41] La technologie utilisée pour les éléments de courant continu permet de contrôler le flux d'énergie circulant par les différentes lignes par opposition aux éléments en courant alternatif où les flux sont généralement difficiles à contrôler.

2.1.2 Étendue de la zone de contrôle

Le concept de zone de contrôle fait référence au mode d'organisation. Une zone de contrôle est définie comme la sous-partie d'un réseau de transport assignée à une entreprise intégrée particulière (figure 1-13). Cette entreprise est responsable de la gestion et du contrôle de l'ensemble des éléments du réseau de transport (nœuds, lignes, transformateurs et les moyens de production) inclus à l'intérieur de la zone de contrôle. Cela dit, il existe des cas où la définition de la zone de contrôle coïncide exactement avec celle du réseau de transport (figure 1-13a). Ce cas est le plus simple à étudier et correspond à celui considéré dans la suite de cette section. Le cas de plusieurs zones de contrôle pour un même réseau de transport déterminé (figure 1-13b) est aussi intéressant et fait l'objet de l'analyse dans la section 3.

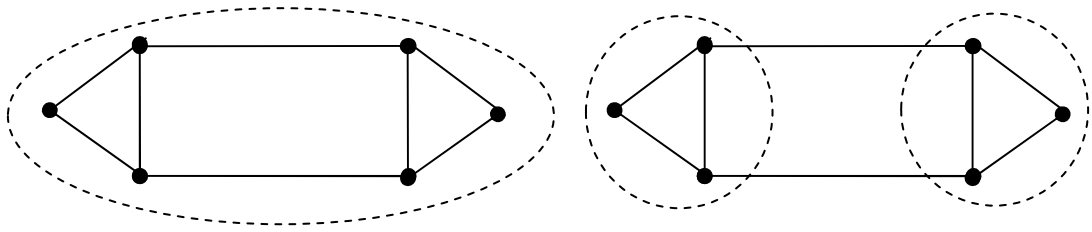

Figure 1-13a : une zone de contrôle Figure 1-13b : deux zones de contrôle

Figure 1-13 : Réseau de transport et zone de contrôle

Dans cette deuxième section, on suppose donc la forme d'organisation suivante : une seule entreprise gère un système électrique « isolé », c'est-à-dire une seule zone de contrôle sur tout le réseau de transport. Nous verrons par la suite quels sont les outils nécessaires pour exploiter le système électrique de manière optimale. Ces outils conformeront une séquence de contrôles-commandes.

2.2 La séquence de contrôles-commandes (une zone de contrôle)

Par hypothèse, l'entreprise verticalement intégrée idéale est parfaitement régulée pour exploiter le système (réseau de transport et moyens de production) à court terme de manière optimale. Son objectif peut être représenté comme la minimisation du coût total de la production nécessaire pour fournir la quantité d'électricité demandée.[42] L'entreprise verticalement intégrée idéale prend les décisions de production adéquates afin d'atteindre l'objectif en respectant les contraintes propres du système électrique.

Afin d'atteindre son objectif, l'entreprise verticalement intégrée utilise une séquence de contrôles-commandes. Les deux principaux maillons de cette séquence de contrôles sont

[42] La demande électrique, étant inélastique au prix à court terme, ne nécessite pas être représentée dans la fonction objectif (cf. section 1.1). La maximisation du bien être social devient donc la minimisation du coût de production.

deux outils d'optimisation sous contraintes : un programme d'optimisation appelé « *unit commitment* » (UC) et un autre appelé « *optimal dispatch* » (OD) (Wood Wollenberg [1996]). Une troisième étape correspond aux « contrôles automatiques » du système. Ces contrôles automatiques ne réalisent pas d'optimisation au sens économique : ils ne minimisent pas le coût de production. Ils se basent principalement sur des caractéristiques techniques. L'utilisation de cette séquence de contrôles-commandes permet à l'entreprise de sélectionner la meilleure combinaison des moyens de production et leur utilisation afin de minimiser le coût total de production tout en respectant les contraintes propre du système électrique. Les figures 1-14 et 1-15 représentent la séquence de contrôles-commande sur différents aspects : l'étendue de contrôle et l'axe temporel.

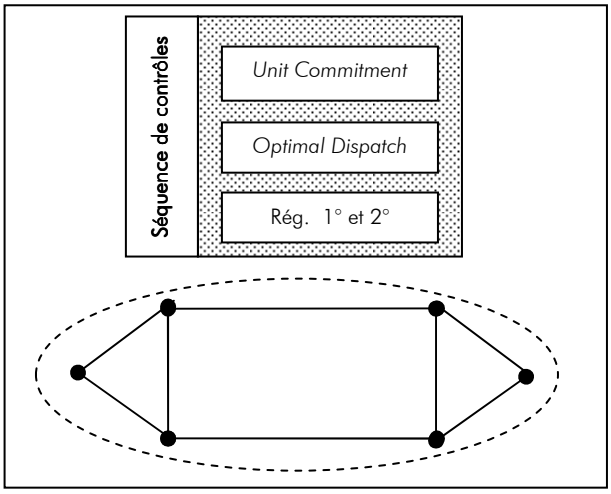

Figure 1-14 : Séquence de contrôles-commandes d'une entreprise intégrée (une seule zone de contrôle)

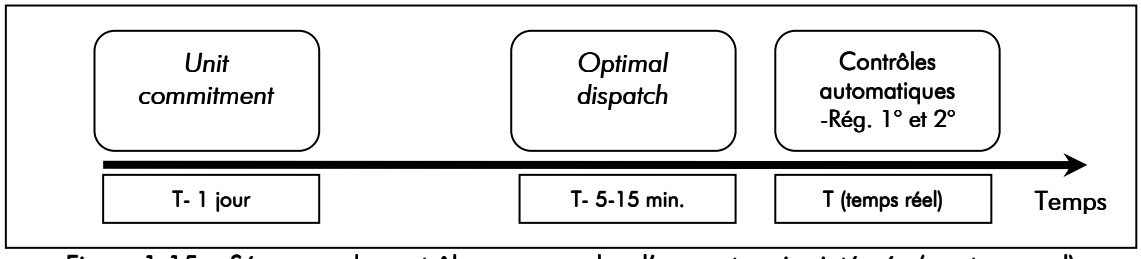

Figure 1-15 : Séquence de contrôles-commandes d'une entreprise intégrée (axe temporel)

L'utilisation de ces optimisations successives est du principalement à trois éléments : l'existence de fortes incertitudes, les caractéristiques des moyens de production et les contraintes du fonctionnement du réseau. Les incertitudes obligent à réaliser les contrôles au plus près du temps réel, lorsque l'environnement est moins incertain. Mais les contraintes techniques des centrales de production ne permettant pas d'y répondre instantanément. Par ailleurs, certaines décisions comme le démarrage d'une centrale sont discrètes et concernent plusieurs périodes. Ces caractéristiques des moyens de production obligent à l'optimisation sur une période de temps assez longue afin de pouvoir prendre en considération les délais

nécessaires et les effets de cycle.[43] Cet horizon de temps est celui de l'horizon de programmation. Enfin, les contraintes du fonctionnement du réseau de transport (l'équilibre et les contraintes de capacité de transport) doivent être respectées en temps réel.

Le premier outil utilisé est le *unit commitment* (UC). L'outil UC est utilisé la veille du jour considéré comme horizon de programmation. Il sert à déterminer les unités de production à démarrer ou arrêter ainsi que les niveaux de production prévisionnels pour chaque sous-période de l'horizon de programmation. En fait, la programmation de la production se réalisera en discrétisant cet horizon de programmation en T sous-périodes ou intervalles (t) (figure 1-16).[44] Programmer les unités pour un horizon de temps dans le futur sous la présence d'incertitudes signifie que des prévisions devront être faites sur certains paramètres qui ne sont pas connues de manière certaine au moment de la programmation. Par exemple, la consommation et la production éolienne sont des sources d'incertitude. Les principales variables de décisions sont les niveaux de production prévisionnels de chaque unité i pour chaque sous période t, $q_{i,t}^{UC}$. L'entreprise doit aussi décider à chaque période de temps quelle unité est démarrée et quelle unité est arrêtée au vu des délais de mise en œuvre et de la considération des coûts issus du changement d'état de l'unité (par exemple, le coût de démarrage). Ainsi, l'état de chaque unité i pour chaque sous-période t, $u_{i,t}^{UC}$ (démarrée, arrêtée) correspond à une autre variable de décision.

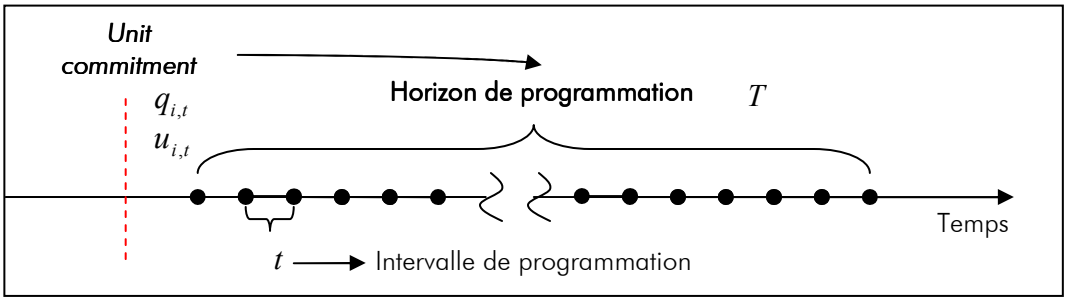

Figure 1-16 : *Unit commitment* et horizon de programmation

L'outil d'*optimal dispatch* (OD), quant à lui, est utilisé très près du temps réel (figure 1-17). Le modèle d'optimisation d'OD sert à déterminer le niveau de production des unités qui sont disponibles à ce moment (q_i^{OD}). C'est, en fait, un sous-problème du problème d'UC. L'outil d'OD est nécessaire car les donnés introduites au programme d'UC ne sont pas des valeurs déterministes, et l'UC peut utiliser une modélisation simplifiée de certains éléments du système électrique (par exemple, l'outil d'UC peut représenter le réseau de transport d'une manière simplifiée). L'outil d'OD fonctionne près du temps réel où les prévisions deviennent de plus en plus précises.

[43] Par exemple, la consommation électrique durant une journée (24h) passe par différentes phases décrivant un cycle : heures creuses, semi-pointe, pointe, etc. (voir section 1.1.2). Ce cycle se répète chaque journée. Différents moyens de production, selon leurs caractéristiques, doivent être démarrés et mis à produire afin que la production puisse suivre ce cycle de consommation.

[44] Par exemple, un horizon de 24h peut être découpé en 48 périodes d'une ½ heure.

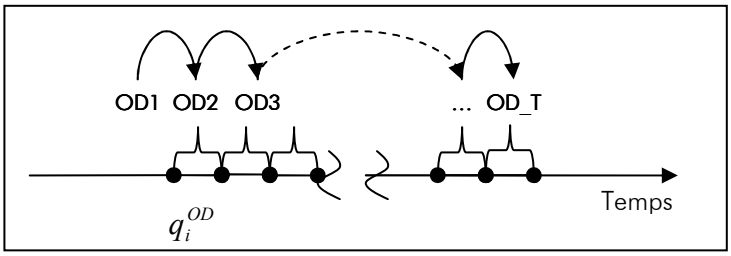

Figure 1-17 : *Optimal dispatch*

Finalement, des outils supplémentaires de contrôle automatique et de surveillance en temps réel sont aussi nécessaires pour le fonctionnement du réseau (contrôle de la fréquence, système de réglage de la tension, *dispatch* manuel de réserves, etc.). Ces contrôles automatiques assurent une vitesse de réaction du système adéquate au fonctionnement du réseau.

Nous décrivons maintenant le fonctionnement de chacun des outils de contrôle-commande mentionnés : l'*unit commitment*, l'*optimal dispatch* et les contrôles automatiques en temps réel. Notre système électrique est composé d'un réseau de transport avec L lignes et I nœuds. Des coefficients d'influence ($PTDF_{l,i,c}$) sont définis par rapport à un nœud de référence.[45] Ils représentent la contribution au flux électrique sur une ligne l suite à une injection au nœud i et un soutirage au nœud de référence pour un état du réseau c. La consommation d'électricité est concentrée dans le nœud de référence. Dans chacun des nœuds i, il existe une unité de production (thermique) nommée aussi avec l'index i. Chaque unité de production est simplement caractérisée par une fonction de coût composée de deux termes : le coût de production proprement dit et le coût de démarrage.[46] Les unités de production ne peuvent produire qu'à un niveau compris entre sa capacité minimale ($Q_i^{\min}$) et sa capacité maximale ($Q_i^{\max}$) de production.

2.2.1 *Unit commitment* (UC)

L'*Unit commitment (UC)*, ou engagement des unités, est un algorithme d'optimisation. Il couvre la décision de démarrer ou arrêter les unités de production ($u_{i,t}^{UC}$) et leur niveau de fonctionnement prévisionnel ($q_{i,t}^{UC}$). Il cherche à trouver la solution présentant le minimum de coût de production total en considérant les caractéristiques et les contraintes du système pour un horizon de programmation donné. Le problème d'optimisation ci-dessous représente l'*unit commitment* simplifié.[47]

---

[45] Cette version linéarisée des équations du réseau limitent notre étude à la puissance active (la puissance réactive et les niveaux de tension ne sont pas considérés). De plus les pertes sur le réseau ne sont pas prises en compte (cf. sous-section 1.3.2).
[46] Les autres contraintes techniques des unités de production ne sont pas représentées, comme par exemple les contraintes de *ramping*, les temps minimum de démarrage ou d'arrêt, etc.
[47] Voir Tseng *et al.* [1999] pour un développement similaire du problème d'optimisation de l'UC.

$$\min_{\underline{u}^{UC},\underline{q}^{UC}} \sum_{t=1}^{T}\sum_{i=1}^{I} C_i(q_{i,t}^{UC})u_{i,t}^{UC} + S_i(u_{i,t}^{UC},u_{i,t-1}^{UC}) \quad \text{(UC1)}$$

Sous contraintes :

Contraintes d'équilibre :

$$\sum_{i=1}^{I} q_{i,t}^{UC} u_{i,t}^{UC} = D_t^{UC} \quad \forall t \quad \text{(UC2)}$$

Contraintes de transport (limites sur le réseau)

$$-F_l \le \sum_{i=1}^{I}\left(PTDF_{l,i,c}\, q_{i,t}^{UC}\, u_{i,t}^{UC}\right) \le F_l \qquad \forall l,t,c \quad \text{(UC3)}$$

Contraintes de réserves

$$\sum_{i=1}^{I}\left(Q_i^{\max} - q_{i,t}^{UC}\right)u_{i,t}^{UC} \ge R^{UC} \quad \forall t \quad \text{(UC4)}$$

Contraintes par unité de production

$$Q_i^{\min} \le q_{i,t}^{UC} \le Q_i^{\max} \quad \forall i,t$$

$$u_{i,t}^{UC} = \{0,1\} \quad \text{(UC5)}$$

$$S_i(u_{i,t}^{UC},u_{i,t-1}^{UC}) = \begin{cases} SC_i & si\ u_{i,t}^{UC} > u_{i,t-1}^{UC} \\ 0 & si\ u_{i,t}^{UC} \le u_{i,t-1}^{UC} \end{cases}$$

La fonction objectif (UC1) est de minimiser le coût total de production (représenté par les coûts de fonctionnement et les coûts de démarrage des centrales). $C_i(q_{i,t}^{UC})$ est la fonction de coût de production de l'unité de production i en fonction de son niveau de production ($q_{i,t}$). $S_i(u_{i,t}^{UC},u_{i,t-1}^{UC})$ est la fonction de coût de démarrage (SC_i entre les périodes $t-1$ et t) de l'unité de production i.

Les contraintes d'équilibre (UC2) indiquent simplement que la quantité d'énergie injectée sur le réseau doit être égale à la quantité soutirée pour toutes les sous périodes. D_t^{UC} est la consommation prévisionnelle pour la période t. La consommation est représentée par une demande inélastique.

Les contraintes de transport (UC3) signalent que le flux sur chaque ligne ne doit pas dépasser les limites thermiques. Le réseau de transport est composé par un ensemble de lignes l chacune avec une limite thermique F_l. La contribution d'une injection sur un nœud i est calculée comme le produit de l'injection $q_{i,t}^{UC}\, u_{i,t}^{UC}$ et le coefficient d'influence $PTDF_{l,i,c}$. Puisque les coefficients d'influence ($PTDF_{l,i,c}$) correspondent aux différents états du réseau (c, état normal ou une des contingences – pertes d'une ligne), les limites de sécurités sont également considérées (règle de sécurité « N-1 »). De cette manière, les contraintes de transport incluent l'aptitude du système à supporter la perte de n'importe quel élément du réseau (ligne) et à maintenir les flux dans les limites thermiques.

Les contraintes de réserve (UC4) sont nécessaires : du fait des nombreuses incertitudes, le système doit être préparé à respecter, à chaque instant, la contrainte d'équilibre. Pour cela, le système garde en réserves de la puissance qui puisse être activée rapidement. Un certain montant de capacité de production doit donc être disponible sur le système.[48] Ce « minimum » de capacité requis R^{UC} est déterminé sur la base de la probabilité de défaillance du système (impossibilité de respecter la contrainte d'équilibre). Cette probabilité de défaillance du système dépend de la combinaison de la probabilité de défaillance des moyens de productions programmés pour chaque période et de l'erreur de prévision de la consommation. R^{UC} doit être assez grand pour que la probabilité de défaillance soit inférieure à une valeur déterminée.[49] Puisque R^{UC} dépend de la « somme » de toutes les incertitudes (généralement non-corrélées), il décroit avec la taille du système.

Finalement, les contraintes des unités de production (UC5) prennent en compte les capacités minimales ($Q_i^{\min}$) et maximales ($Q_i^{\max}$) de production, ainsi que la définition des variables d'état de l'unité ($u_{i,t}^{UC}$) comme variables binaires (1 démarrée et 0 arrêtée). Notre représentation des unités de production est simplifiée car nous ne considérons pas d'autres contraintes comme les contraintes de rampe, les temps de démarrage et d'arrêt, etc.

Le modèle d'optimisation UC fait partie d'une famille de modèles d'optimisation appelée « programmation mixte » (« *mixed-integer-programming* ») car il a deux types de variables : continues ($q_{i,t}^{UC}$) et discrètes ($u_{i,t}^{UC}$). Les variables discrètes génèrent le problème de la non-convexité du programme d'optimisation. Les techniques utilisées pour l'optimisation sont, par ailleurs, assez particulières et difficiles à résoudre par grands systèmes (Hobbs *et al.* [2001]). Nous supposons ici que l'entreprise verticalement intégrée idéale est capable de résoudre exactement ce problème avec l'aide d'outil UC.[50]

Il est à noter que la vérification de toutes les contraintes revêt un caractère prévisionnel, puisque les contraintes dépendent des niveaux de production prévisionnels. Etant donné les incertitudes présentes, il est nécessaire de refaire une optimisation plus proche du temps réel, dès que les incertitudes ont diminué considérablement. C'est la tâche de l'outil d'*optimal dispatch*.

---

[48] Il faut remarquer que nous considérons seulement un type de réserve. En réalité, plusieurs types de réserves sont considérés dans les contraintes de l'*unit commitment*. Ces types de réserves sont généralement différenciés selon la flexibilité des moyens de production (contraintes de *ramping*) comme par exemple : les réserves secondaires, les réserves rapides/5 min., les réserves tournantes/10 min., etc.

[49] Par exemple en France le risque de défaillance ne doit pas dépasser 1% (RTE [2003]). Puisque la détermination exacte et optimale du coefficient R^{UC} peut s'avérer assez compliquée, d'autres méthodes de dimensionnement des réserves peuvent être trouvées empiriquement. Par exemple, certaines se basent sur la taille de la plus grande centrale en fonctionnement (Hirst-Kirby [2003]).

[50] Les algorithmes utilisés pour résoudre ce problème ont été principalement développés en utilisant des techniques de relaxation du lagrangien, de branch&bound, etc. Actuellement, des logiciels d'optimisation permettent de trouver des solutions pour des grands systèmes assez rapidement. Par exemple à PJM le temps de calcul est de l'ordre de 200 secondes pour un système de 50 GW avec plus de 10000 nœuds (PJM [2004]).

2.2.2 *Optimal dispatch* (OD)

L'*optimal dispatch* (OD) ou *optimal power flow* est le deuxième outil d'optimisation utilisé par l'entreprise verticalement intégrée. Cet outil est utilisé entre 5 min et 15 min avant le temps réel (quasi temps réel) et il est lancé à des intervalles réguliers. L'outil d'OD est utile et nécessaire puisqu'il fonctionne au moment où la plupart des incertitudes sur le système sont levées. L'OD fonctionne avec des données en quasi temps réel (niveau de production actuel de chaque unité démarrée, conditions du réseau, *etc.*).

L'outil d'OD sert à déterminer le niveau de production « définitif » (q_i^{OD}) des centrales disponibles (ou démarrées, $u_i^{UC^*} = 1$) pour chaque période de temps afin optimiser une fonction objectif de minimisation du coût total en respectant une série de contraintes. En fait, le processus de l'OD identifie les modifications des niveaux de production des unités démarrées par rapport aux niveaux de production prévisionnels sortis de l'outil *unit commitment* ($q_{i,t}^{UC}$).

Le modèle d'optimisation d'OD est un sous problème du problème d'UC et il plus simple à résoudre car il se base sur un seul type de variables (continues)[51]. Ainsi, généralement, le modèle d'optimisation d'OD est « convexe » et il est résolu par des techniques de programmation linéaire ou quadratique (*linear programming* ou *quadratic-programming*).

Une version simplifiée de cet outil d'optimisation est donnée ci-dessous.

$$\min_{\underline{q}^{OD}} \sum_{i=1}^{I} C_i(q_i^{OD}) \quad \text{(OD1)}$$

Sous contraintes :

Contraintes d'équilibre :

$$\sum_{i=1}^{I} q_i^{OD} = D^{OD} \quad \text{(OD2)}$$

Contraintes de transport (limites sur le réseau)

$$-F_l \le \sum_{i=1}^{I} PTDF_{l,i,c}\, q_i^{OD} \le F_l \qquad \forall l, c \quad \text{(OD3)}$$

Contraintes de réserves

$$\sum_{i \in R1°2°} Q_i^{\max} - q_i^{OD} \ge R^{OD} \quad \text{(OD4)}$$

Contraintes par unité de production

$$Q_i^{\min} \le q_i^{OD} \le Q_i^{\max} \quad \forall i \quad \text{(OD5)}$$

[51] En effet, cela n'est pas exactement vrai. Dans certains modèles d'OD, il y a aussi des décisions binaires comme la décision de démarrer des unités à gaz (GT) très rapides (5-10 min) ou d'arrêter des unités fonctionnant au niveau minimum. Par souci de simplicité, nous supposons que le modèle d'OD considère seulement des variables continues et qu'il n'y a pas de non-convexités en temps réel.

La fonction objectif (OD1) est de minimiser le coût total de production représentant les coûts de fonctionnement des centrales disponibles ou démarrées.

Les contraintes d'équilibre (OD2) signalent simplement que la quantité d'énergie injectée sur le réseau ($\sum_{i=1}^{I} q_i^{OD}$) doit être égale à la quantité soutirée. D^{OD} est la demande (ou une prévision à très court terme, 15 minutes avant la réalisation).

Les contraintes de transport (OD3) indiquent que le flux de puissance sur chaque ligne ne doit pas dépasser son limite thermique maximale (F_l). Le flux total sur une ligne est calculé comme la somme des contributions des injections sur chaque nœud. La contribution d'une injection sur un nœud est calculée comme le produit de l'injection q_i^{OD} et le coefficient d'influence $PTDF_{l,i,c}$. Puisque les coefficients d'influence ($PTDF_{l,i,c}$) correspondent aux différents états du réseau, c , les limites de sécurités sont aussi considérées. De cette manière, les contraintes de transport incluent l'aptitude du système à supporter la perte de n'importe quel élément du réseau (ligne) et de maintenir les flux dans les limites thermiques (règle de sécurité N-1).

Les contraintes de réserve (OD4) expriment d'une manière générale le niveau minimum R^{OD} des réserves de capacité des unités de production qui participent aux réglages primaire et secondaire (voir ci-dessous).[52]

Finalement, les contraintes des unités de production (OD5) prennent en compte les capacités minimales ($Q_i^{\min}$) et maximales ($Q_i^{\max}$) de production.

L'outil d'OD permet ainsi d'ajuster le système pour trouver un fonctionnement à coût minimal et sous contraintes, prés du temps réel au moment où la plupart des incertitudes sont levées. Au moment précis du temps réel, le système passe en contrôle automatique. Ce contrôle automatique sert à commander le système et à le maintenir en fonctionnement entre deux actions de l'outil d'OD. Nous présentons maintenant les contrôles automatiques utilisés en temps réel.

2.2.3 Contrôles automatiques supplémentaires (en temps réel)

Après les ordres envoyés par l'outil d'*optimal dispatch,* le système passe en contrôle automatique. Ces contrôles ne réalisent pas d'optimisation au sens économique ; ils se basent principalement sur des caractéristiques techniques. Les deux contrôles automatiques les plus importants (réglages primaire et secondaire) sont ceux dont l'objectif est d'assurer

[52] Notez qui si cette contrainte ne peut pas être respectée en modifiant les niveaux de production q_i^{OD}, le seul moyen pour la respecter est la réduction de la demande par effacement ou délestage.

un réglage fin de la fréquence et de la rapprocher de la valeur nominale pour assurer l'équilibre production-consommation. Ces contrôles automatiques sont nécessaires car, bien que l'outil d'OD considère comme contrainte l'équilibre du système, il présente aussi des limites. Par exemple, le temps de réponse nécessaire au système d'ordres de l'OD est plus long que la fréquence d'apparition de nouveaux événements, et ce notamment pour les erreurs de prévision et la variation stochastique de la consommation.

Ces contrôles automatiques pour asservir la fréquence à sa valeur de référence sont le réglage primaire et le réglage secondaire. Quand un déséquilibre survient, les réglages primaire et secondaire se mettent tout de suite en marche pour rétablir l'équilibre, la fréquence, et assurer ainsi la sécurité du système[53].

Le réglage primaire (ou « réserves primaires ») est assuré par les boucles de régulation (« régulateurs de vitesse ») situées sur les unités de production. Il assure une correction rapide (en quelques secondes) et décentralisée des écarts production-consommation. Ce réglage permet de retourner à l'équilibre après une perturbation. Mais la fréquence stable de ce nouvel équilibre ne correspond plus à la valeur de consigne. Pour une unité donnée, le régulateur de vitesse agit sur les organes d'admission du fluide moteur (eau, vapeur, gaz, etc.) de la turbine et cherche à imposer, à l'équilibre, une relation linéaire entre la vitesse (image directe de la fréquence) et la puissance. Ces actions sont automatiques et décentralisées.

Le réglage secondaire (ou « réserves secondaires ») sert à ramener rapidement la fréquence à sa valeur nominale après une perturbation. L'adaptation rapide de la production à la consommation, qu'effectue le réglage primaire, laisse, en fin d'action, un écart de fréquence. Pour une zone isolée, l'objectif du réglage secondaire est de ramener la fréquence à sa valeur de consigne.[54]

Sur la figure 1-18, nous pouvons voir un schéma de fonctionnement de ces contrôles. Une variation soudaine de la consommation est couverte dans un premier temps par les réserves primaires décentralisées (en utilisant toutes, ou la plupart, des unités de production). Une fois la fréquence stabilisée, le réglage secondaire commande des unités de production prédéterminées (capables de réagir avec une certaine vitesse) pour pouvoir ramener la fréquence à la valeur de référence. Toute cette procédure est menée en quelques minutes. L'action combinée de ces deux types de réserves automatiques laisse au modèle

[53] Quand tous ces réglages ne suffisent pas à rétablir l'équilibre et la fréquence s'écart trop de sa valeur nominale des délestages automatiques (consommation ou production) aident à rétablir un équilibre et évitent ainsi la panne totale du système.

[54] Le réglage primaire après une perturbation provoque également des variations de transits entre les zones de contrôle (s'il en existe plus d'une). En fait, toutes les machines des différentes zones réagissent à la variation de la fréquence commune du système interconnecté. Nous verrons ensuite d'autres versions du réglage secondaire où l'objectif n'est pas seulement de ramener la fréquence à la valeur de consigne, mais aussi de ramener les transits inter-zone aux valeurs préétablies.

d'*optimal dispatch* un temps d'approximativement 10-15 minutes pour ramener le système à l'équilibre au moindre coût.

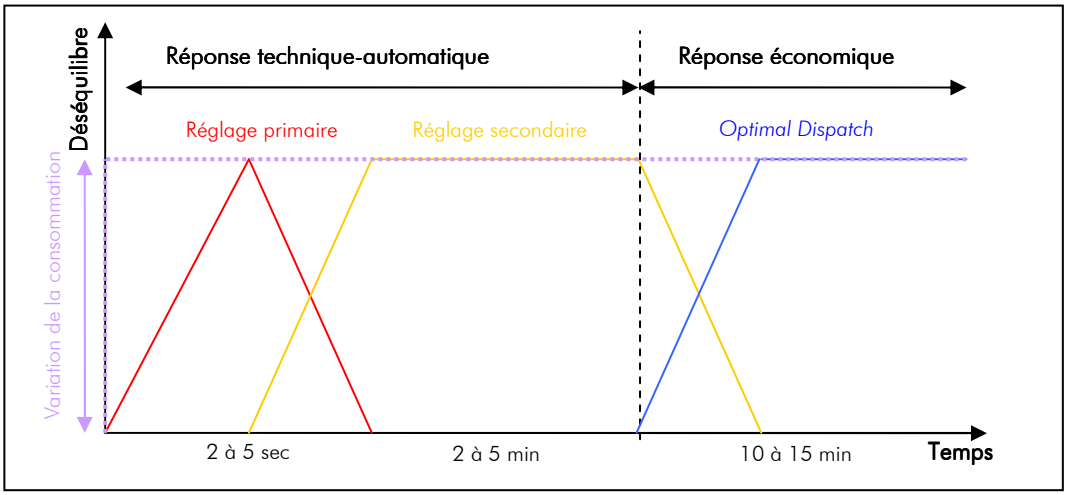

Figure 1-18 : Schéma des réponses temporelles de différents contrôles d'un système électrique

2.3 Conclusions de la section

Cette section a considéré la forme d'organisation la plus simple, d'un point de vue du fonctionnement, pour l'exploitation optimale d'un système électrique. Cette forme d'organisation correspond à une entreprise verticalement intégrée idéale (parfaitement régulée) qui gère tous les éléments d'un système électrique « isolé », c''est-à-dire quand la zone de contrôle assignée à l'entreprise colle parfaitement avec l'étendue du réseau de transport.

Cette entreprise idéale a pour objectif la minimisation du coût de production et doit s'assurer de respecter toutes les contraintes liées aux caractéristiques du système. Pour accomplir son objectif, l'entreprise a recours à une séquence de contrôles-commandes. Ces contrôles sont des outils d'optimisation, dans certains cas, ou juste des asservissements automatiques, dans d'autres.

Deux outils d'optimisation sous contraintes sont d'un intérêt particulier : l'outil d'*unit commitment* et l'outil d'*optimal dispatch*. L'outil d'*unit commitment* est utilisé la veille du jour considéré comme l'horizon de programmation. Il sert à déterminer quelles unités de production démarrer ou arrêter ainsi que les niveaux de production prévisionnels pour chaque sous période de l'horizon de programmation. Cet outil considère l'équilibrage prévisionnel du système (contraintes d'équilibre) ainsi que les limites de capacité de transport des différents éléments du réseau (contraintes de capacité de transport), un niveau déterminé de réserves (contraintes de réserves) et les caractéristiques technico-économiques des unités de production. L'outil d'*optimal dispatch* est utilisé très près du temps réel et sert à déterminer le niveau de production définitif des unités qui sont disponibles en temps réel. Une ré-optimisation est nécessaire (*optimal dispatch* après *unit commitment*), car les donnés introduites dans l'*unit commitment* ne sont pas des valeurs déterministes. L'*optimal dispatch*

fonctionne proche du temps réel où les prévisions sur les conditions du temps réel sont très précises. Cet outil assurera l'équilibrage du système et que les limites de capacité de transport ne soient pas violées avec les nouvelles valeurs des injections et des soutirages.

Dans la section 3, nous considérerons le cas où, sur un même réseau de transport, fonctionnent plusieurs entreprises verticalement intégrées (plusieurs zones de contrôle définies sur un même réseau de transport). Dans ce cas, chaque entreprise utilisera une séquence de contrôles-commandes pour exploiter sa partie du système électrique de manière optimale. Puisque les entreprises partagent le même réseau de transport, il est important d'étudier comment elles peuvent se coordonner (et coordonner leurs séquences de contrôle) pour viser une exploitation optimale de tout le système électrique.

Section 3: Coordination d'entreprises intégrées sur un même réseau

Nous n'avons jusqu'ici pris en compte qu'une seule entreprise intégrée verticalement et horizontalement, étendue sur tout un réseau de transport. En fait, nous avons considéré le cas d'une zone de contrôle unique. Un cas plus général consiste maintenant à imaginer plusieurs entreprises verticalement intégrées partageant le même réseau de transport, c'est-à-dire que plusieurs zones de contrôle, une par entreprise intégrée, coexistent dans le même réseau. Notez que nous considérons le « réseau de transport » comme étant le réseau interconnecté (et synchronisé) complet (cf. sous-section 2.1).

Dans cette section, nous étudions comment plusieurs entreprises verticalement intégrée, et fonctionnant sur un même réseau de transport interconnecté, doivent se coordonner afin d'optimiser leurs performances globales. A cet effet, nous verrons que les entreprises intégrées peuvent coordonner de différentes manières leurs séquences de contrôles-commandes. Les formes de coordination entre entreprises intégrées développées ici serviront de base pour comprendre les formes ultérieures de coordination entre les nouvelles formes d'organisation concurrentielles.

Cette section est organisée comme suit. Dans la sous-section 3.1, nous étudierons d'abord en quoi consiste la notion d'une coordination parfaite des entreprises par fusion des séquences de contrôles-commandes puis quels en sont les bénéfices. Dans la sous-section 3.2, nous considérerons la coordination entre séquences de contrôle ainsi que les différents degrés de coordination possibles. Dans la sous-section 3.3, nous présenterons les conclusions.

3.1 La coordination parfaite et bénéfices associés

La coordination entre plusieurs entreprises intégrées verticalement partageant le même réseau a des bénéfices largement reconnus par les économistes et les ingénieurs (Joskow-Schmalensee [1983]). Pour recueillir ces bénéfices, la coordination doit s'organiser sur la base des outils de contrôle utilisés par chacune des entreprises, c'est-à-dire, en coordonnant les séquences de contrôles-commandes. Ces bénéfices sont maximaux pour un fonctionnement coordonné parfait des entreprises et un tel fonctionnement se matérialise par une « fusion » des outils de contrôle. Nous présentons d'abord le concept de coordination parfaite et ensuite les bénéfices qu'une telle coordination peut apporter.

3.1.1 Coordination parfaite entre entreprises intégrées par fusion des séquences de contrôle

Dans notre environnement idéal, une coordination parfaite entre entreprises intégrées sur un même réseau de transport pourrait être comparé à la forme d'organisation avec une seule entreprise intégrée. En effet, les entreprises se coordonnent de sorte que le

système fonctionne comme si une seule séquence de contrôles commandait tout le système. Tous les outils de contrôle sont parfaitement coordonnés : *unit commitment* commun ; *optimal dispatch* commun ; et dimensionnement et partage de réserves. La coordination parfaite entre plusieurs entreprises intégrées sur un même réseau suppose la fusion des outils de contrôle sur la réunion de deux zones de contrôle (figure 1-19).

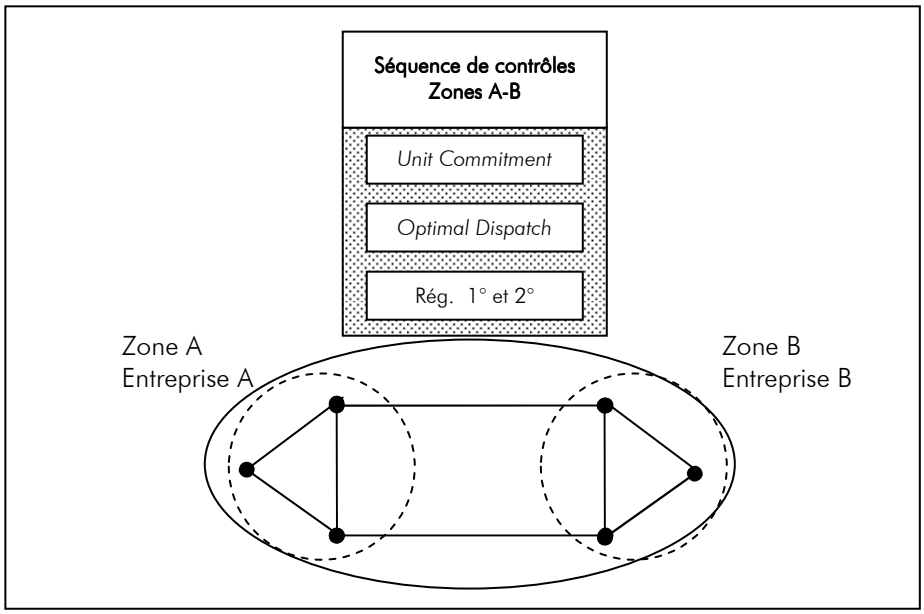

Figure 1-19 : Fusion des séquences de contrôle de deux entreprises intégrées (deux zones de contrôle)

3.1.2 Bénéfices de la coordination

La coordination parfaite entre entreprises intégrées, par le biais d'une séquence de contrôle fusionnée, permet une utilisation plus efficace des moyens de production et des capacités du réseau. Les bénéfices de la coordination justifient (et ont justifié) l'interconnexion des réseaux adjacents et ont donné naissance aux systèmes interconnectés de grande envergure en diverses parties du monde (Menager [2002]).

A court terme, un fonctionnement coordonné est bénéfique pour diverses raisons :

- **Utilisation efficace des moyens de production**. Une forte coordination entre entreprises permet de minimiser le coût total de production grâce à une combinaison optimale des unités de production utilisées à chaque moment. Par exemple, une coordination permettrait d'utiliser une centrale de grande capacité et de rendement énergétique haut plutôt que d'utiliser plusieurs centrales de petite taille et de rendement énergétique bas pour fournir de l'électricité nécessaire à plusieurs zones.

- **Utilisation efficace du réseau de transport**. La coordination entre entreprises au sein d'un même réseau est fondamentale pour profiter des capacités techniques maximales d'un réseau de transport. Ceci est d'autant plus vrai pour les réseaux maillés où une

utilisation efficace des éléments du réseau de transport nécessite la considération d'injections et de soutirages sur tous les nœuds du réseau et la connaissance de la topologie complète du réseau (cf. sous-section 1.3). Par exemple la coordination permet de profiter de contre-flux provoqués sur certains éléments du réseau ; ces contre-flux libèrent de la capacité de transport (supplémentaire) sur ces éléments.

- Economies grâce à la diversité des demandes. Le comportement de la demande dans les différentes régions peut être assez différent. La demande de l'électricité dépend des conditions climatiques, des habitudes de consommation, etc. Ces caractéristiques varient d'une région (pays) à l'autre. De fait, les caractéristiques de consommation d'électricité diffèrent d'une région à l'autre. Par exemple, les pointes de consommation coïncident rarement entre deux régions adjacentes. Une coordination dans l'utilisation des moyens de production permet d'agréger les demandes et d'utiliser la diversité des demandes pour réaliser des économies.

- Economies des aléas. Les incertitudes présentes dans des zones isolées peuvent être réduites en les gérant d'une manière coordonnée. Ces incertitudes sont rarement corrélées et en les agrégeant elles deviennent relativement moins importantes. De cette manière, la coordination permet d'abord une réduction du niveau de réserves nécessaire pour un même niveau de fiabilité du système, c'est-à-dire le partage de réserves entre les différentes régions.

- Collaboration en cas d'urgence. La coordination entre plusieurs entreprises permet d'accroître la capacité à faire face aux situations d'urgence. En effet, lors d'une situation d'urgence, par exemple provoquée par des défauts en cascade suite à la défaillance d'un grand groupe de production ou déconnection forcée d'un élément du réseau, une entreprise intégrée peut s'appuyer sur les autres pour sortir rapidement de cette situation d'urgence. Une telle coordination permet de diminuer les délestages de charges ou de diminuer leur durée.

La coordination parfaite par fusion des séquences de contrôle pour des grandes zones de contrôle, est difficilement atteignable en pratique. En conséquence, l'organisation d'un fonctionnement coordonné devrait passer par la mise en place de mécanismes de coordination explicites entre les séquences de contrôle pour simuler un fonctionnement coordonné du système. Ainsi une coordination partielle, ou concernant seulement certains outils de contrôle est envisageable et les bénéfices de ces formes de coordination dépendront de leur écart à la coordination parfaite. Dans la sous-section suivante nous analysons les manières et les degrés dont cette coordination peut être réalisée.

3.2 Coordination de séquences de contrôles-commandes

L'utilisation d'une seule séquence de contrôle pour des grandes zones de contrôle, peut présenter des obstacles ou des difficultés. D'abord, les contraintes institutionnelles

nécessaires peuvent empêcher la mise en place d'une séquence de contrôle unique. De plus, pour de grandes zones de contrôle, l'utilisation d'une séquence de contrôle unique peut présenter des problèmes de faisabilité technique et des surcoûts importants.[55] Les ressources nécessaires (capacité de gestion de l'information et de calcul, etc.) pour une gestion centralisée de la séquence de contrôle peuvent être importantes. Une gestion centralisée peut présenter des problèmes de sécurité si la zone considérée est trop grande. Par exemple, un contrôle centralisé unique peut avoir un niveau inférieur de fiabilité que de multiples contrôles décentralisés (il est moins probable que tout le système de contrôle souffre d'une défaillance avec de multiples contrôles décentralisés qu'avec un contrôle centralisé unique) ; lors d'une panne totale, la remise en service du système (« *blackstart service* ») peut être réalisée plus facilement par une gestion décentralisée (Stavros [2000]).

Dans un environnent où les contraintes institutionnelles et les problèmes de faisabilité technique ne permettent pas la mise en œuvre d'une séquence de contrôle unique, chacune des entreprises intégrées utilisera une séquence de contrôle pour exploiter sa partie du système électrique de manière localement « optimale ». Mais, étant donné le fait que les entreprises partagent le même réseau, ces séquences de contrôle n'aboutiraient pas, *a priori*, à une exploitation optimale globale (qui considèrerait tout le système) résultat de la coordination parfaite. En conséquence, il est nécessaire de mettre en place des mécanismes de coordination intermédiaires entre les séquences de contrôle afin de simuler un fonctionnement coordonné du système. Ces mécanismes de coordination prennent forme par le biais de l'échange d'information et des protocoles de fonctionnement coordonnés entre les entreprises (figure 1-20).

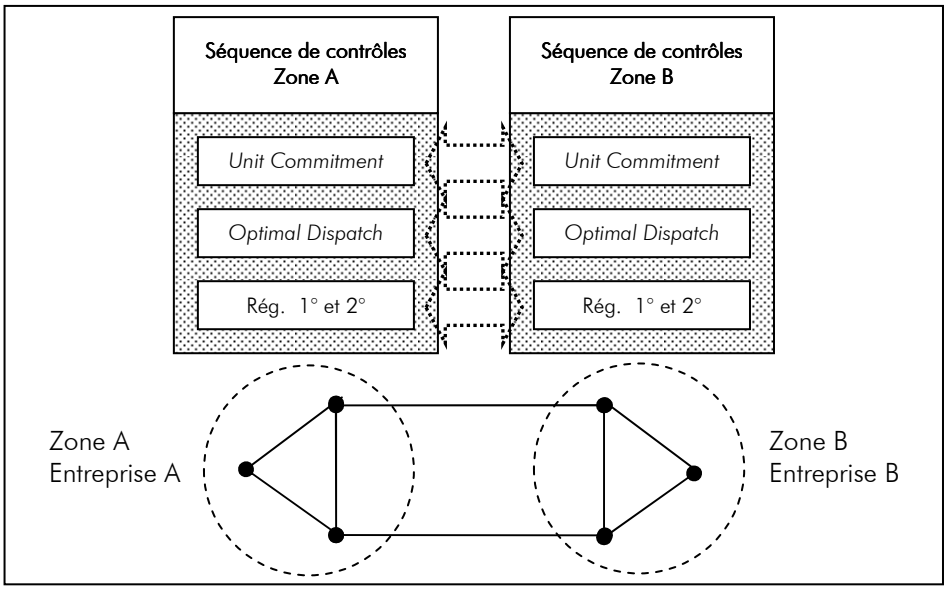

Figure 1-20 : Coordination des séquences de contrôle de 2 entreprises intégrées

[55] Si l'on prend en compte ces nouveaux facteurs, on peut considérer l'existence d'une taille optimale de zone de contrôle. Au delà de cette taille, l'entreprise ne bénéficie plus des économies d'échelle.

La coordination entre les outils de contrôle nécessite un échange important d'information et un mécanisme qui assure la coordination entre les différentes séquences. Il ne faut pas oublier que par définition chaque entreprise n'a que l'information du réseau et des moyens de production concernant sa zone (figure 1-21). Alors dans leurs modèles d'optimisation (*unit commitment* et *optimal dispatch*) elles doivent représenter d'une certaine manière la partie du réseau manquante. Les informations nécessaires pour un fonctionnement coordonné devraient porter sur les coûts de moyens de production, les disponibilités, les capacités de transport, la topologie et PTDF, etc.

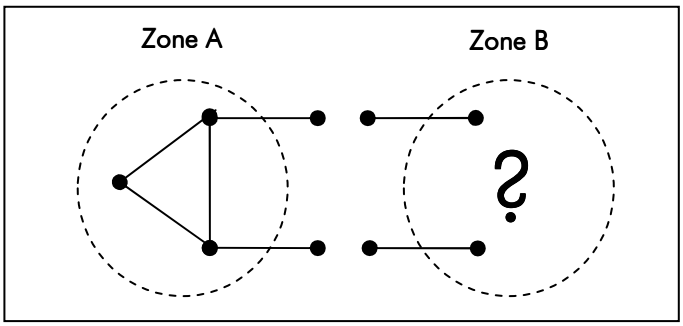

Figure 1-21 : Nécessité d'échange d'information entre entreprises intégrées

Les échanges d'information entre les entreprises et les mécanismes de coordination aident les entreprises de chaque zone à considérer les effets des leurs actions sur le reste du réseau. Les mécanismes de coordination devraient considérer le processus d'échange d'information itératif entre les séquences. En fait, les modifications réalisées dans une sous partie du réseau peut provoquer des modifications dans les autres sous parties. Des itérations devraient être réalisées jusqu'à ce qu'elles convergent et si possible, vers un fonctionnement coordonné et correspondant à un optimum global.

L'échange d'information et les mécanismes de coordination pour les différents outils des séquences peuvent être mis en place avec différents degrés de profondeur. Plus la coordination est importante, plus on peut recueillir les bénéfices d'une coordination parfaite. Nous présentons les différents degrés de coordination dans la sous-section suivante.

La coordination entre entreprises verticalement intégrées peut présenter différents degrés. Les degrés de coordination possibles vont d'un idéal de coordination où tous les outils d'optimisation de la séquence sont coordonnés à des degrés de coordination inférieurs, où seulement certains outils sont coordonnés.[56] Ces différents degrés de coordination sont similaires aux cas réels présentés dans Joskow-Schmalensee [1983] où des entreprises verticalement intégrées (*utilities*) se coordonnent avec une intensité différente par le biais d'arrangements informels ou d'arrangements formels plus structurés (contrats bilatéraux, « *loose pool* » et « *tight pool* »).

[56] Certes, les coûts de mise en œuvre dépendent du degré de coordination. En conséquence des études du type coût/bénéfice sont nécessaires pour déterminer le niveau de coordination optimal pour chaque système en particulier.

Nous présentons les différents degrés de coordination en analysant quels outils de contrôle sont coordonnés. Ainsi, ces différents niveaux de coordination peuvent être représentés de la manière suivante : le degré minimum de coordination (partage de réserves 1° et 2°) ; degré moyen de coordination (*optimal dispatch* et partage de réserves 1° et 2°) et degré maximum (*unit commitment*, *optimal dispatch* et partage de réserves 1° et 2°). En fait des degrés intermédiaires de coordination, qui sortent de cette représentation schématique, peuvent aussi exister.

Le degré minimum de coordination consiste juste en un partage de réserves primaires, et une gestion du réglage secondaire dans chacune des zones qui prenne en compte les échanges entre zones après perturbation (figure 1-22). En fait, si le système interconnecté est composé de plusieurs zones de contrôle, l'objectif du réglage secondaire dans chacune des zones doit être double (cf. sous-section 2.2.3). D'une part, il doit ramener la fréquence à la valeur de consigne mais, d'autre part, il doit aussi ramener les échanges inter-zones aux valeurs programmées. Dans le degré de coordination le plus bas, nous considérons que la valeur des échanges programmés entre zones est nulle. Une mesure qui combine ces deux facteurs est connue sous l'appellation *Area Control Error (ACE)*. Les contrôles automatiques du réglage secondaire (*Automatique Generator Control* – AGC) répondent proportionnellement à l'*ACE* en agissant sur les puissances de consigne des unités de production. Finalement, ce réglage secondaire ramène les échanges inter-zones aux valeurs de consigne (zéro dans notre cas) le plus vite possible, ainsi que la fréquence à la valeur nominale. Il est important de remarquer que dans le cas d'interconnections entre zones de contrôles par plusieurs lignes, si les échanges (agrégés) entre les zones sont bien ramenés à zéro, les flux des différentes lignes d'interconnexion peuvent quand même avoir des valeurs non nulles.[57]

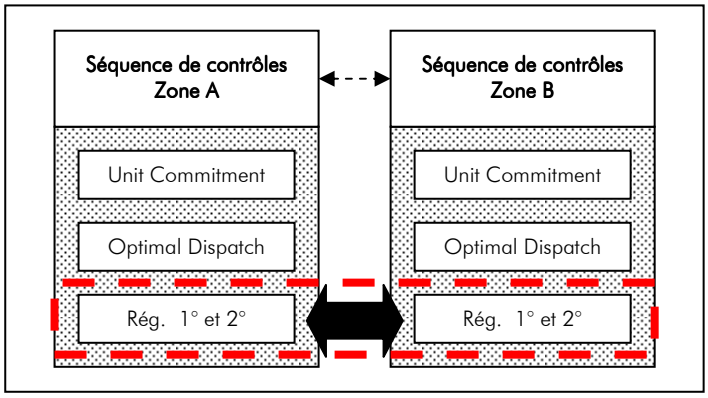

Figure 1-22 : degré de coordination minimum

[57] La coordination entre les anciennes entreprises intégrées des pays européens du réseau interconnecté UCTE (France, Espagne, Allemagne, Belgique, etc.) est un cas similaire pour ce degré de coordination. A la différence de notre description, cependant, certaines de ces entreprises intégrées avaient des accords d'échanges programmés (non nuls) assurés en temps réel (par exemple, les accords d'échange entre EDF-France- et ENEL-Italie- avant l'introduction de la concurrence (Glachant-Finon [2003])).

Le degré moyen de coordination combine l'outil *optimal dispatch* coordonné et le partage de réserves primaires et secondaires (figure 1-23). Ce degré de coordination peut être mis en place de deux manières : en centralisant l'information et l'utilisation d'un outil d'optimisation commun ; ou par le biais des protocoles qui coordonnent plusieurs outils séparés comme s'ils fonctionnaient ensemble (Cadwalader *et al.* [1998], Cadwalader *et al.* [1999]).

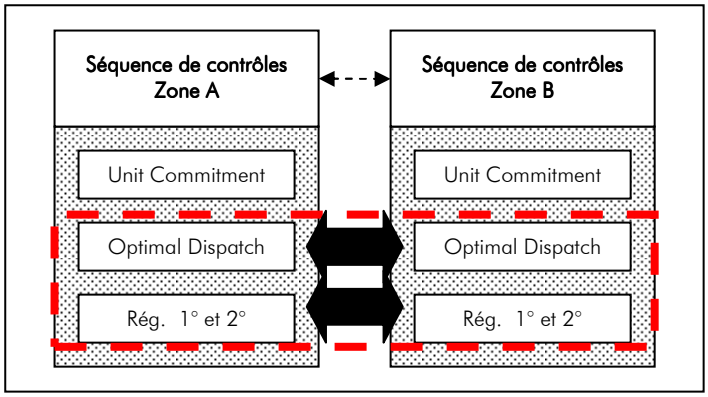

Figure 1-23 : degré de coordination moyen

Dans le degré maximal de coordination, tous les outils de contrôle sont coordonnés : *unit commitment* coordonné (avec partage des réserves) ; *optimal dispatch* coordonné ; et partage de réserves primaires et secondaires (figure 1-24).

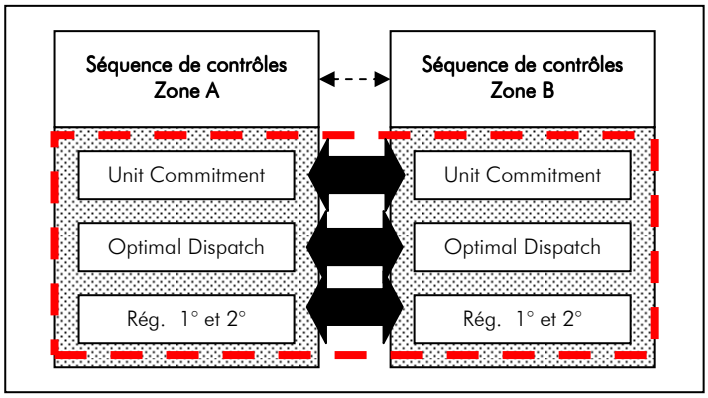

Figure 1-24 : degré de coordination maximum

Ce degré de coordination suppose aussi que les niveaux des réserves (R^{UC} dans notre modèle simplifié d'*unit commitment*, cf. section 2.2.1) sont calculés en considérant tout le système. Cela permet des réductions dans la quantité nécessaire de réserves car, pour des systèmes plus grands, les incertitudes deviennent relativement moins importantes. Ce degré de coordination peut être aussi mis en place de deux manières : en centralisant l'information et l'utilisation d'un outil d'optimisation commun ; ou par le biais des protocoles qui coordonnent plusieurs outils séparés comme s'ils fonctionnaient ensemble (Harvey-Pope [2001]). Un exemple de coordination par centralisation peut être trouvé avec la naissance du (*tight*) « Pool » PJM aux Etats-Unis, où l'information des unités de production disponibles était mise en commun et utilisée par un *unit commitment* et un

optimal dispatch centralisés qui gouvernaient les unités de production de toutes les entreprises intégrées (Joskow-Schmalensee [1983]).

3.3 Conclusions de la section

Dans cette section nous avons étudié un cas plus général où plusieurs entreprises verticalement intégrées partagent le même réseau de transport (plusieurs zones de contrôle).

Un fonctionnement coordonné entre les entreprises partageant le même réseau de transport est bénéfique pour diverses raisons : utilisation plus efficace des moyens de production et du réseau de transport, économies grâce à la diversité des demandes, économies des aléas, et la collaboration pour les urgences. L'exploitation optimale de tout le système électrique serait obtenue par fusion des séquences de contrôles-commandes de toutes les entreprises (réglages primaires et secondaires communs, *optimal dispatch* commun, *unit commitment* commun) pour aboutir à une coordination parfaite.

Néanmoins, ce niveau de coordination avec une séquence de contrôle unique peut être difficilement atteignable en réalité (contraintes institutionnelles, faisabilité technique et coûts de mise en œuvre). Pour cela nous avons considéré la coordination entre différentes séquences de contrôle par le biais de mécanismes de coordination intermédiaires. En analysant quels outils de contrôles sont coordonnés pour chaque cas, nous avons caractérisé ces différents degrés de coordination : degré minimum de coordination (partage de réserves 1° et 2°) ; degré moyen de coordination (*optimal dispatch* et partage de réserves 1° et 2°) et degré maximum (*unit commitment*, *optimal dispatch* et partage de réserves 1° et 2°).

Les différents degrés de coordination entre entreprises intégrées qui ont été développés ici, c'est-à-dire la coordination complète ou partielle des séquences de contrôles, serviront de base pour comprendre les formes ultérieures de coordination entre les nouvelles formes d'organisation concurrentielles (cf. Chapitre 2, section 3).

Conclusions du chapitre

Ce chapitre a donné une vision générale des systèmes électriques et de leurs caractéristiques propres. Trois parties du système électrique ont été analysées : la consommation, les moyens de production et le réseau de transport. La consommation électrique est caractérisée par ses fortes fluctuations cycliques et de caractère aléatoire, par la difficulté de la prévoir de manière exacte, et par son inélasticité au prix à court terme. Les différentes technologies de moyens de production se caractérisent par la diversité des coûts de production et des formes de fonction de coût non-convexes, par la flexibilité temporelle limitée et par les erreurs de prévision de la production. Le stockage d'énergie électrique à grande échelle n'est pas économiquement viable à l'heure actuelle. Le réseau de transport sert de liaison entre les moyens de productions et la consommation. Il doit être exploité dans ses propres contraintes de fonctionnement. En tout moment, il doit être en équilibre (la puissance produite étant égale à la puissance consommée), et les flux électriques circulant dans les différents éléments (lignes, transformateurs, etc.) doivent être inférieurs aux limites de capacité (limites physiques et règles de sécurité). Or le contrôle des flux sur les éléments du réseau n'est pas simple. Car les flux sur le réseau ne sont pas dirigeables et suivent les lois de Kirchhoff.

L'exploitation d'un système avec telles caractéristiques impose un niveau de coordination très élevé. Pour gérer ce système complexe, la forme d'organisation la plus simple, d'un point de vue opérationnel, est une entreprise verticalement intégrée bienveillante ou parfaitement régulée (zone de contrôle unique). Car cette entreprise centralisera alors toutes les décisions et internalisera les problèmes de coordination. Nous avons examiné les outils nécessaires à une entreprise intégrée idéale, parfaitement régulée et possédant toutes les informations nécessaires, pour atteindre un objectif de minimisation du coût de production dans le respect des contraintes du système. Une telle entreprise a besoin d'une séquence de contrôles-commandes. Ces contrôles sont des outils d'optimisation, dans certains cas seulement des asservissements automatiques. Deux outils d'optimisation sous contraintes sont d'un intérêt particulier : l'outil d'*unit commitment* et l'outil d'*optimal dispatch*. La description de ces deux outils est importante car ce sont aussi les briques de base de l'architecture d'un marché électrique.

Enfin, nous avons étudié un cas plus général où plusieurs entreprises verticalement intégrées partagent le même réseau de transport sur plusieurs zones de contrôle. Chacune de ces entreprises utilise sa séquence de contrôle (*unit commitment*, *optimal dispatch* et contrôles automatiques) pour gérer sa zone de contrôle. Or, comme elles partagent le même réseau de transport et toutes les interactions correspondantes, une exploitation optimale collective nécessitera une forte coordination entre entreprises toutes ces intégrées. Les bénéfices maximaux d'une coordination entre entreprises intégrées seraient atteints si l'on fusionnait complètement leurs séquences de contrôles. Néanmoins, la coordination partielle de certains outils d'optimisation entre les entreprises intégrées est aussi possible.

Cette notion de coordination, complète ou partielle, des séquences de contrôle nous servira de base pour l'étude de l'introduction de la concurrence dans différentes régions géographiques partageant le même réseau de transport, c'est-à-dire pour l'intégration d'architectures et de marchés électriques régionaux. En effet, bien que les formes d'organisation soient différentes, la problématique technico-économique est ici invariable, indépendante des formes d'organisation.

Chapitre 2

Architectures de marché électrique

Introduction 56

Section 1: Contrôles-commandes vs. Marchés 60
1.1 Contrôles-commandes *vs.* marchés pour un bien simple 60
1.2 Contrôles-commandes *vs.* marchés pour le bien électricité 63
1.3 Conclusions de la section 70

Section 2: Architectures de marché sur une seule zone de contrôle 72
2.1 Le module d'énergie *forward* 74
2.2 Le module de transport *forward* 78
2.3 Le module des réserves *forward* 84
2.4 Le module du temps réel 86
2.5 Articulation entre modules et arbitrage 90
2.6 Analyse empirique d'architectures de marché 93
2.7 Architecture minimale : le module du temps réel 98
2.8 Conclusions de la section 100

Section 3: Architectures de marché sur plusieurs zones de contrôle 102
3.1 L'intégration d'architectures de marché de plusieurs zones de contrôle 103
3.2 L'intégration d'architectures par coordination des modules 110
3.3 Conclusions de la section 118

Conclusions du chapitre 121

Introduction

Depuis une ou deux décennies, des réformes électriques ont été réalisées dans la plupart des pays (Glachant-Finon [2003], Newbery [2005b], Joskow [2006a], Sioshansi [2006]). Ces réformes ont eu comme objectif principal l'introduction de la concurrence dans un secteur électrique longtemps organisé autour d'un monopole intégré en production-transport.

Dans la chaine de l'industrie électrique (production, transport, distribution et vente/consommation) l'introduction de la concurrence n'est possible que dans certains maillons. Elle est introduite, dans la plupart des cas, au niveau de la production (marché de gros) et de la vente d'électricité (marché de détail) (Hunt [2002]). D'autres maillons de la chaine, comme le transport et la distribution, ont gardé une organisation de monopole régulé à cause de leurs caractéristiques de monopole naturel. Bien évidemment la mise en place de dispositifs organisationnels adéquats a été nécessaire pour l'articulation entre maillons en concurrence et maillons régulés (Curien [2000], Glachant [2002], Lévêque [2004]).

Les réformes électriques se sont centrées d'abord sur les structures industrielles, principalement par le démantèlement des entreprises verticalement intégrées. Ainsi, des nouveaux acteurs de l'industrie ont émergé (figure 2-1). Dans les activités en concurrence - marché du gros et marché du détail - les participants sont : les producteurs (entreprises qui possèdent les moyens de production et vendent de l'électricité sur le marché de gros à des revendeurs ou à des gros consommateurs) ; les revendeurs ou *retailers* (entreprises qui achètent de l'électricité sur le marché de gros et la revendent sur le marché de détail) ; les traders (ne possédant pas des moyens de production, ils achètent et revendent de l'électricité sur le marché de gros), et les consommateurs (qui achètent sur le marché de gros ou sur le marché du détail l'électricité qu'ils consomment).[58]

Le transport et la distribution restent organisés comme monopoles régulés avec deux fonctions principales. La première est l'opération du réseau. La seconde concerne la maintenance et le développement des infrastructures du réseau. Ces fonctions peuvent être assurées par une seule entreprise régulée, ou bien gérées séparément. Pour le cas du transport,[59] en Europe, les deux fonctions sont assurées par une seule entreprise, appelée le Gestionnaire du Réseau de Transport (GRT ou TSO)[60]. En revanche, aux Etats-Unis, des entreprises différentes assurent ces fonctions de manière séparée. L'opération du réseau est

[58] Cependant, des participants intégrés, par exemple un producteur-revendeur ou un producteur-trader, sont aussi envisageables.
[59] Pour le cas de la distribution, l'entreprise responsable de l'opération et du maintien du réseau est appelée en Europe « gestionnaire du réseau de distribution » (GRD).
[60] TSO: *Transmission System Operator*.

toujours assurée par l'opérateur du système (ISO ou RTO)[61] mais une autre entreprise, propriétaire du réseau de transport, est responsable de la maintenance et du développement des infrastructures. Enfin, le régulateur est l'acteur chargé de mettre en place des procédures de régulation pour les activités régulées, et il peut être aussi chargé de contrôler les activités en concurrence.[62]

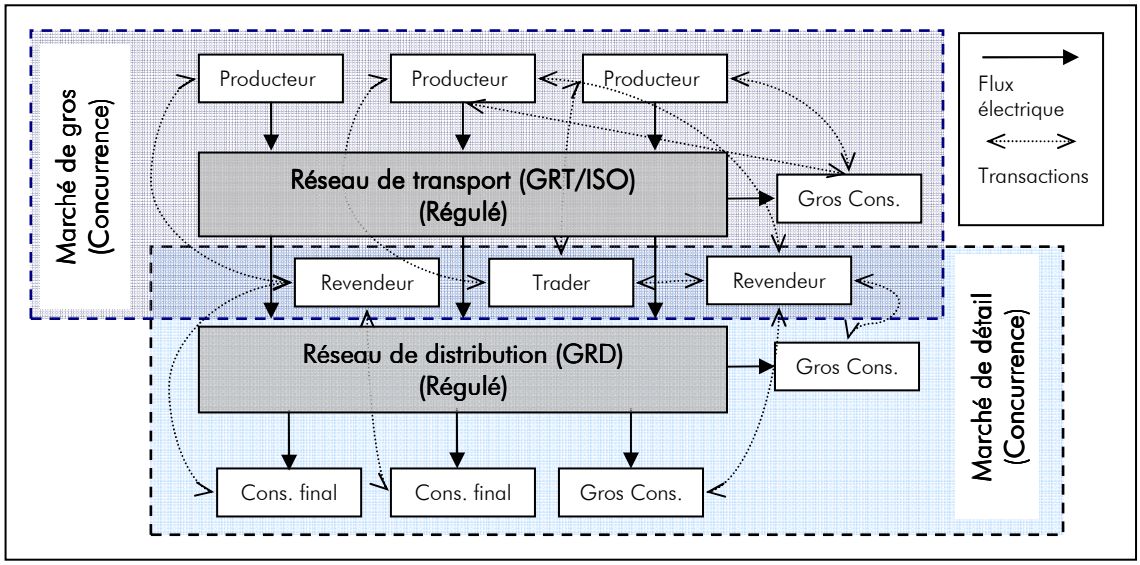

Figure 2-1 : Schéma de l'organisation de l'industrie électrique après la reforme.

Nous nous intéressons à l'introduction de la concurrence dans la production avec la création d'un marché de gros. L'introduction de la concurrence dans un secteur longtemps organisé autour d'un monopole intégré fait apparaître des problèmes de coordination qu'il convient de traiter par la mise en place de dispositifs organisationnels et institutionnels adéquats (Staropoli [2001]). Mais ces dispositifs n'émergent pas spontanément du jeu des acteurs. Ils résultent des choix d'autorités, privées ou publiques. La conception d'une architecture du marché (« *market design* ») doit intégrer ces dispositifs organisationnels et institutionnels afin d'assurer une nouvelle forme de coordination.

L'électricité est un bien complexe qui a des caractéristiques spécifiques (non-stockabilité, contrainte d'équilibre P=C, etc. cf. Chapitre 1). Ces caractéristiques doivent être prises en compte dans la construction d'une architecture du marché. D'autres biens plus « simples » (e.g. les commodités) n'ont pas besoin d'efforts particuliers des autorités pour mettre en place un marché. Créer une architecture de marché c'est chercher, en pratique, à faire de l'électricité un bien plus standard qu'elle l'est réellement. C'est chercher à en faire une « commodité » (Glachant-Saguan [2006]). Concevoir une architecture de marché efficace, qui prenne en compte toutes ces caractéristiques de l'électricité, n'est donc pas une tâche simple et peut aboutir à un design complexe. Même si d'autres marchés efficaces nécessitent normalement un design complexe de règles (Coase [1988]). Le cas particulier de

[61] ISO: *Independent System Operator* ; RTO: *Regional Transmission Organization*.
[62] Dans certains cas, le contrôle des activités en concurrence peut être partagé ; par exemple, avec l'autorité de la concurrence.

l'électricité permet de comprendre pourquoi, après plus de 15 ans d'expérience sur les choix des architectures de marché, il n'y a pas de solution unique qui s'impose à tous.

Une architecture de marché prenant en compte les caractéristiques spécifiques de l'électricité sera donc complexe. Mais l'on peut la découper en sous-parties (ou modules) pour l'étudier. L'objectif général de ce chapitre est ainsi de présenter un cadre d'analyse modulaire des architectures de marché électrique. Ce cadre d'analyse modulaire nous servira ensuite pour étudier d'une manière plus approfondie le design de certains modules et leurs effets sur les autres modules de l'architecture.

Nous proposons d'étudier les éléments essentiels d'une architecture de marché électrique en considérant l'introduction de la concurrence comme le remplacement des outils de contrôle-commande, utilisés dans une entreprise intégrée idéale, par une combinaison de mécanismes et de marchés. Les outils de contrôle-commande sont ceux de la séquence de contrôles développée dans le premier chapitre : l'*unit commitment*, l'*optimal dispatch* et les contrôles automatiques du temps réel. Nous nous intéressons ici au court terme et au très court terme, c'est-à-dire à la partie du design qui relie les opérations en J-1 aux opérations en temps réel.[63] De plus, nous nous limitons aux éléments les plus importants d'une architecture de marché, c'est-à-dire l'énergie, la capacité de transport et les réserves.[64] Ces trois éléments contribuent à construire le bien utile au consommateur. D'ailleurs, dans la plupart des architectures observées, ces trois éléments sont généralement bien pris en compte.

Notre démarche ira du plus simple au plus complexe. Dans la première section nous dériverons les modules essentiels d'une architecture de marché selon les éléments qui remplacent les outils de contrôle-commande. Grâce à l'analyse des outils de contrôle-commande et de leurs implications économiques, nous dériverons les composantes du bien électricité et les modules ou sous-marchés potentiels nécessaires pour construire une architecture de marché (Wilson [1998b], Wilson [2002]).

Dans la deuxième section, nous analysons tout d'abord les différentes options de design pour chacun des modules dans une zone de contrôle[65] unique et isolée. Nous

[63] Nous ne traitons pas les mécanismes qui ont des effets de long terme comme les mécanismes de capacités (marchés et/ou obligation de capacité, *scarcity pricing*, etc.) ou les marchés de droits financiers de transport. Pour plus d'information, voir par exemple Joskow [2006b], Joskow-Tirole [2006], Cramton-Stoft [2006a], Hogan [2002a], Hogan [2006a], Kristiansen [2004] et Sun [2005].

[64] Nous ne considérons pas, par exemple, ni les effets de pertes sur le réseau ni la production de puissance réactive (réglage de la tension) absents par hypothèse des outils de contrôle développés dans le premier chapitre. Nous ne considérons pas non plus d'autres « services auxiliaires » (*Ancillary Services*) : « *blackstart service*», etc. Voir par exemple Hennebel [2007] ou Moreno [2005] pour un traitement technico-économique de la puissance réactive et Pignon [2003], Phulpin *et al.* [2005] ou Stoft [2002] pour les pertes.

[65] Dans ce nouvel environnement, le concept de zone de contrôle développé dans le chapitre 1 doit être adapté (cf. chapitre 1, section 2.1). La zone de contrôle est définie désormais par rapport à la responsabilité des gestionnaires du réseau de transport. Le GRT assure la gestion et le contrôle de l'ensemble des éléments du réseau de transport (sous-partie du réseau) inclus à l'intérieur de la zone de contrôle dans le respect de leurs contraintes techniques.

développerons un cadre d'analyse modulaire, où concevoir une architecture du marché électrique consiste à choisir une option pour chacun des modules et pour leurs articulations (Baldwin-Clark [2000]). Nous montrerons ensuite qu'il existe une grande diversité et complexité de ces architectures de marché, ce qui justifiera une étude approfondie de certains modules en particulier. Le cadre d'analyse modulaire nous permettra d'étudier ces modules séparément les uns des autres. Enfin, nous montrerons que dans toute cette diversité d'architectures, il existe un noyau commun à tous les designs pour remplir des fonctions « électriques » essentielles. Nous appellerons ce noyau commun l'« architecture minimale », et elle guidera la suite de notre réflexion dans les chapitres 3 et 4.

Dans la troisième section, l'approche modulaire sera utilisée pour traiter de l'intégration d'architectures de marché entre des zones de contrôle multiples sur un réseau de transport commun. En effet, l'architecture de marché est normalement définie pour une seule zone géographique déterminée (un pays, ou un état, ou un groupe d'états). Mais l'introduction de la concurrence passe souvent par la stimulation des échanges entre plusieurs zones géographiques. C'est, par exemple, le cas du marché intérieur européen (CE [1996], CE [2003], CE [2004], CE [2005], Pignon [2003]). Dans ce cas, le choix du design pour intégrer les architectures de marchés des zones adjacentes est donc déterminant pour un bon fonctionnement des marchés électriques régionaux. Nous verrons ici qu'il existe différentes formes d'intégration d'architectures. Ces différentes formes d'intégration varient dans le choix des modules qui sont coordonnés entre les zones et dans le degré d'harmonisation entre les designs de ces modules.

Section 1: Contrôles-commandes *vs.* Marchés

L'introduction de la concurrence dans la production d'électricité avec la création d'un marché de gros peut être vue comme le remplacement total ou partiel des contrôles-commandes d'une entreprise intégrée par des mécanismes de marché (Smeers [2004a], Smeers [2004b], Joskow [2006c]).

L'idée du remplacement d'un mécanisme de contrôle-commande par un marché n'est pas nouvelle. Elle est une des bases de la théorie microéconomique standard (Varian [1995], Mas-Collel *et al.* [1995]). En effet, la coordination nécessaire pour une allocation efficace de ressources peut être fournie par un mécanisme de coordination centralisé, basé sur un outil d'optimisation, ou par le biais d'un marché. En théorie, pour un bien simple (e.g. une commodité), un seul marché suffit pour remplacer le mécanisme de coordination centralisé. Par ailleurs, ce marché n'a normalement pas besoin de règles particulières ; en d'autres mots ce marché n'a pas besoin d'une architecture de marché déterminée. En revanche, pour un bien complexe, comme l'électricité, la coordination nécessaire ne peut être fournie que par la construction d'une architecture de marché composée des plusieurs sous-marchés (ou modules) chacun avec des caractéristiques spécifiques.

Cette première section a donc pour objet de dériver les sous-marchés ou modules essentiels d'une architecture de marché électrique. C'est seulement en regardant les outils de contrôle-commande utilisés par une entreprise intégrée (cf. chapitre 1) qu'on peut dériver les caractéristiques intrinsèques des marchés ou des modules nécessaires pour assurer la coordination. Ces outils de contrôle-commande gèrent directement toutes les caractéristiques spécifiques de l'électricité comme par exemple la non-stockabilité, la contrainte d'égalité entre la production et la consommation, l'impossibilité de diriger les flux de puissance, le besoin de réserves, etc. Par contre, la construction d'une architecture de marché devra prévoir tous les éléments nécessaires pour prendre en compte directement ou indirectement les caractéristiques spécifiques ainsi que leurs conséquences économiques, afin de remplacer les contrôles-commandes.

Pour suivre une logique du plus simple au plus complexe, nous introduisons dans la section 1.1 les concepts nécessaires au remplacement des contrôles-commandes par des marchés pour un bien simple. Ensuite, dans la section 1.2, nous présentons le problème de ce remplacement dans le cas d'un bien complexe comme l'électricité.

1.1 Contrôles-commandes *vs.* marchés pour un bien simple

L'équivalence entre un marché concurrentiel et un modèle centralisé d'optimisation, géré par un planificateur bienveillant, est une des bases de la théorie microéconomique standard (Varian [1995], Mas-Collel *et al.* [1995]). Cette équivalence considère d'abord le

cas le plus élémentaire, consistant en un système économique avec un seul bien (de type commodité) en équilibre partiel[66].

Sous certaines hypothèses que nous expliciterons ci-après, on arrive à une conclusion d'efficacité et d'équivalence entre les deux systèmes d'organisation, le planificateur centralisé bienveillant et le marché concurrentiel. C'est, en fait, le paradigme de la concurrence pure et parfaite (Tirole [1991]). Les deux systèmes sont efficaces au sens de Pareto (Varian [1995]) et cette efficacité peut être évaluée avec l'aide du concept du bien-être ou surplus total.[67] En fait, un plan de production et de consommation est efficace s'il maximise le surplus total. Le surplus total est la somme du surplus du consommateur et du surplus du producteur décrits sur la figure 2-2.

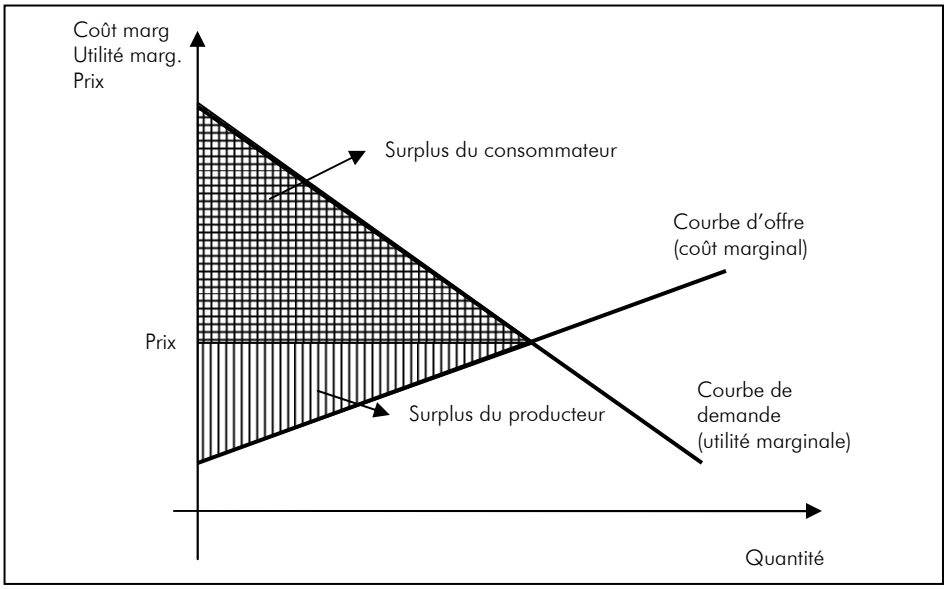

Figure 2-2 : Surplus du consommateur et du producteur

Le résultat d'équivalence en efficacité nécessite une série importante d'hypothèses. D'une part, le planificateur doit être bienveillant, c'est-à-dire qu'il cherche à maximiser le bien-être ou surplus total. Il doit avoir toutes les informations relatives aux coûts de production et aux préférences des consommateurs. Il doit aussi avoir les moyens nécessaires pour trouver et mettre en œuvre une allocation optimale de production et consommation. D'autre part, le marché doit être « parfait », c'est-à-dire fonctionner selon le postulat de la « concurrence pure et parfaite » qui peut se traduire par les hypothèses suivantes : atomicité des acteurs (producteurs et consommateurs), homogénéité des biens, information parfaite, absence d'externalités, absence de caractéristiques de bien public, coûts de transaction nuls, fonctions de coût convexes, etc.

[66] Le terme « partiel » fait référence à l'étude du marché d'un seul bien. Il s'utilise pour le distinguer de l'équilibre général où tous les biens d'une économie sont pris en compte (Mas-Collel *et al.* [1995]).
[67] L'efficacité dans le sens de Pareto signifie qu'il n'existe pas une autre allocation de production et consommation du bien qui fournisse un surplus plus grand (Varian [1995]).

Une représentation mathématique simple permettra de comparer les conditions d'optimalité d'un système d'organisation centralisé, où un planificateur bienveillant prend toutes les décisions de production et de consommation afin de maximiser le surplus total, avec celles d'une organisation de marché.

La version centralisée peut être représentée par un programme d'optimisation (ou de contrôle-commande) dont l'objectif est la maximisation du surplus total[68] (voir tableau 2-1). Le surplus total est représenté comme la différence entre l'utilité du consommateur ($B()$) et le coût de production ($C()$). Le planificateur bienveillant doit alors choisir de manière optimale les quantités à produire, q_p et à consommer, q_c. Les conditions d'optimalité[69] de ce programme déterminent que l'utilité marginale doit être égale au coût marginal, et que ces deux valeurs sont égales au multiplicateur de Lagrange de la contrainte d'équilibre du programme (λ).

Tableau 2-1 : Planificateur bienveillant (organisation centralisée)

Programme d'optimisation du planificateur	Conditions d'optimalité
$\max_{q_c, q_p} \; B(q_c) - C(q_p)$ s.c. $q_c = q_p \quad (\lambda)$ $q_c, q_p \geq 0$	$B'(q_c^*) = C'(q_p^*) = \lambda$ $q_c^* = q_p^*$

Dans une organisation de marché on remplace l'allocation centralisée par contrôle-commande des quantités à produire et à consommer par des actions décentralisées (voir tableau 2-2). Ces actions décentralisées sont coordonnées par un prix (p) qui permet équilibrer le système. Le producteur prend sa décision de production par rapport au prix du marché afin de maximiser son profit étant donné le niveau de prix. Le consommateur prend sa décision de consommation afin de maximiser son surplus (utilité de la consommation moins le coût du consommateur pour se procurer le bien) étant donné le niveau de prix. Ainsi, en développant les conditions d'optimalité pour le producteur et le consommateur on arrive à $p = C'(q_c)$ et à $p = B'(q_c)$, le prix est égal au coût marginal de production et à l'utilité marginale de consommation.

Tableau 2-2 : Marché concurrentiel (organisation décentralisée)

Optimisations des participants du marché	Conditions d'optimalité
Producteur : $\max_{q_p} \; p\,q_p - C(q_p)$ Consommateur : $\max_{q_c} \; B(q_c) - p\,q_c$ Equilibre du marché : $q_c = q_p$	$p = C'(q_c^*) = B'(q_p^*)$ $q_c^* = q_p^*$

[68] Pour simplifier, nous considérons le cas d'un seul consommateur et un seul producteur.

[69] Ces conditions d'optimalité correspondent au cas de solution intérieure $q_p^*, q_c^* > 0$

Une simple analyse des conditions d'optimalité venant de deux systèmes d'organisation nous montre qu'ils sont équivalents et efficaces. Ils maximisent le surplus total.

De cet exemple d'application de deux systèmes d'organisation pour la production d'un bien simple et sous les hypothèses de « perfection », nous pouvons en tirer deux conclusions. Premièrement, le remplacement d'un planificateur bienveillant par des marchés pour un bien simple ne nécessite que la mise en place d'un seul marché pour assurer la coordination. Deuxièmement, le remplacement du système d'organisation de contrôle-commande pourrait être a priori direct, sans l'intervention des autorités, car un design particulier de marché n'est pas une condition indispensable pour assurer la coordination.

L'électricité est un bien complexe avec des caractéristiques spécifiques qui l'éloignent des hypothèses de marché parfait. Dans la sous-section suivante, nous verrons que l'introduction de la concurrence dans l'électricité n'est pas directe comme pour le cas d'un bien simple. L'introduction de la concurrence nécessite l'intervention des autorités (Hogan [2002b], Staropoli [2001]). Afin d'assurer la coordination nécessaire au fonctionnement du système électrique, ces autorités doivent construire une architecture de marché comprenant plusieurs marchés.

1.2 Contrôles-commandes *vs.* marchés pour le bien électricité

Dans le chapitre 1 nous avons développé les outils de contrôle-commande d'une entreprise intégrée parfaitement régulée (planificateur bienveillant) pour allouer de manière optimale les quantités de production d'électricité : *unit commitment*, *optimal dispatch* et contrôles automatiques du temps réel (figure 2-3).[70]

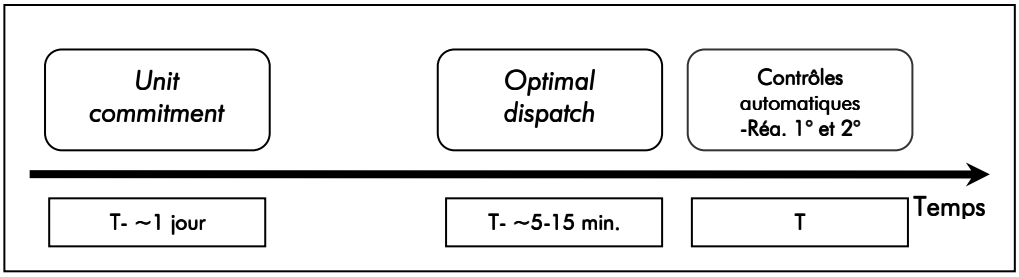

Figure 2-3 : Schéma de la séquence des outils de contrôles-commande pour l'électricité

Cette séquence de contrôles-commandes assure la coordination nécessaire au fonctionnement du système électrique. L'introduction de la concurrence par la suppression complète des outils de contrôle-commande, sans la préalable construction d'une architecture de marché se heurte sur les propriétés économiques qui dérivent de ces caractéristiques spécifiques de l'électricité. Ces caractéristiques spécifiques sont par

[70] Ces outils de contrôle-commande ont comme objectif la minimisation du coût de production et non la maximisation du surplus total car la demande d'électricité est considérée comme inélastique.

exemple la non-stockabilité, la contrainte d'égalité entre la production et la consommation, l'impossibilité de diriger les flux de puissance, le besoin de réserves, etc. (cf. chapitre 1). Ces caractéristiques ne sont pas présentes dans les biens simples et elles créent des effets économiques qui s'opposent à certaines hypothèses de « marché parfait » comme par exemple la présence d'externalités et de caractéristiques de bien public.[71] Ainsi, les effets économiques des caractéristiques spécifiques limitent le remplacement direct des outils de contrôle-commande par des marchés et exigent la construction d'une architecture de marché capable d'assurer la coordination.

Notre but est de dériver les éléments essentiels d'une architecture de marché électrique. Pour ce faire, dans la sous-section 1.2.1, nous décrivons les caractéristiques spécifiques du bien électricité et leurs conséquences économiques. Puis, dans la sous-section 1.2.2, en nous basant sur ces conséquences économiques, nous dérivons les éléments essentiels d'une architecture de marché électrique.

1.2.1 Les caractéristiques spécifiques de l'électricité et leurs conséquences économiques

Les outils de contrôle-commande développés dans le chapitre 1 internalisent les problèmes de coordination créés par les caractéristiques spécifiques de l'électricité. Sans ces outils de contrôle-commande, ces problèmes de coordination s'explicitent via une série de conséquences économiques.

Nous analysons par la suite les conséquences économiques des caractéristiques spécifiques de l'électricité. Pour ce faire, nous regroupons les caractéristiques spécifiques de l'électricité en trois groupes : i) la non-stockabilité qui se combine avec la contrainte d'égalité entre l'production et la consommation ; ii) l'impossibilité de diriger les flux de puissance qui se combine avec les limites de capacité de transport ; et iii) le besoin de réserves.

1.2.1.1 La non-stockabilité et la contrainte d'équilibre production-consommation

L'électricité n'est pas économiquement stockable. Mais ce caractère de non-stockabilité n'est pas la seule caractéristique qui rend l'électricité différente des autres biens ou produits (par exemple, les places d'avion ou les places d'hôtel ne sont pas stockables non plus). La caractéristique de non-stockabilité devient fondamentale à cause d'une autre contrainte de l'électricité : l'équilibre nécessaire et instantané de la production et la consommation en tout moment. Un déséquilibre entre la production et la consommation peut provoquer la panne totale (« *black-out* ») du système et empêcher à tous les

[71] D'autres hypothèses de marché parfait peuvent être écartées pour le cas de l'électricité (par exemple, l'atomicité des acteurs, les coûts de transaction nuls, la convexité des fonctions de coût, etc.). Cependant, nous supposons que ces hypothèses ne seront pas les déterminants principaux pour définir les éléments essentiels d'une architecture de marché.

utilisateurs du réseau de profiter de ses bénéfices. Tout déséquilibre production-consommation est donc interdit, où du moins, les déséquilibres doivent être limités en taille et sur une période temporelle très court[72].

Boucher *et al.* [2006] comparent l'électricité à un autre produit non stockable, les places de transport aérien. Cette comparaison montre bien que la principale différence entre ce produit et l'électricité vient de la deuxième caractéristique, l'équilibre instantané nécessaire entre la production et la consommation. Imaginons, en effet, une situation hypothétique où tous les avions tomberaient en panne s'il y avait un seul des avions où les places n'étaient pas toutes occupées, c'est-à-dire si l'équilibre production-consommation n'était pas assuré.

La non-stockabilité et le respect de la contrainte P=C ont deux conséquences économiques majeures. Premièrement, il faut une forte coordination en temps réel pour trouver l'équilibre très rapidement. Cet équilibre devra donc être assuré de manière centralisé et par l'ajustement de la production car la demande est inélastique à court terme.[73] Deuxièmement, étant donné les contraintes de flexibilité des moyens de production (cf. chapitre 1), il est nécessaire de préparer à l'avance les moyens nécessaires pour l'équilibrage.

Par ailleurs, comme nous le verrons après, ces conséquences économiques sont importantes parce qu'elles interagissent avec les caractéristiques développées par la suite. Nous analysons maintenant une deuxième caractéristique propre à l'électricité : la difficulté de diriger les flux électriques et les capacités limitées de transport.

1.2.1.2 Flux électriques non-dirigeables et limites de capacité de transport

Les flux sur un réseau d'électricité se distribuent selon les lois de Kirchhoff (équations du réseau) et sont difficilement dirigeables (cf. chapitre 1). Une transaction, c'est-à-dire une paire injection-soutirage entre deux nœuds d'un réseau maillé impacte les flux sur toutes les lignes. De plus, deux flux de sens opposés s'annulent mutuellement. La clé de répartition des flux par rapport aux paires injection-soutirage peut être approchée par les coefficients d'influence (PTDFs) que nous avons introduit dans le chapitre 1.[74]

Cette caractéristique de distribution des flux est particulière à l'électricité. En revenant à l'exemple de Boucher *et al.* [2006] sur le transport aérien, imaginons le cas de trois lieux (A, B et C) placés géographiquement comme un triangle et qui sont reliés par des

[72] Un système électrique ne peut supporter des petits déséquilibres que pendant moins d'une seconde (cf. chapitre 1).
[73] Le « coût » d'ajustement côté demande peut être très élevé du fait de l'inélasticité de la demande (délestage forcé).
[74] Les coefficients d'influence (*Power transfert distribution factor* PTDF) représentent linéairement les flux ou transit sur les différents éléments d'un réseau (lignes, transformateurs) engendrés par une paire injection-soutirage unitaire (cf. Chapitre 1, section 1.3).

vols. Pour présenter des caractéristiques similaires à l'électricité, nous devrions imaginer qu'un avion qui veut aller de A à B doit obligatoirement se diviser en deux parties : une partie allant directement de A à B et une autre qui devra d'abord passer par C. Les lois de Kirchhoff imposent un tel lien entre les flux électriques et les injections et soutirages d'électricité sur le réseau.

Mais la non-dirigeabilité des flux ne poserait pas de problèmes s'il n'y avait pas de contraintes de transport. Les différents éléments du réseau (les lignes et les transformateurs) ont des limites de capacité de fonctionnement. L'ensemble des injections et des soutirages sur le réseau crée des flux sur les différentes lignes. Si un de ces flux atteint la limite maximale de transit autorisée sur une des lignes, nous sommes en présence d'une congestion (cf. Chapitre 1, sous-section 1.3.3). Les contraintes de transport ne sont pas une autre chose que l'expression de la rareté des ressources de « capacité de transport » utilisées par les différents participants qui injectent ou soutirent de l'énergie dans les différents nœuds.

Ne pas traiter correctement les caractéristiques des réseaux électriques génère la présence d'externalités. [75] Or, les externalités éloignent le marché électrique des caractéristiques de « marché parfait » et leur présence est considérée comme une défaillance ou incomplétude du marché (Salanié [2000]). Les externalités sur un réseau d'électricité peuvent être mieux comprises si l'on considère la capacité de transport comme une ressource rare. Certains agents économiques peuvent se voir refuser l'accès en cas de congestion, c'est-à-dire si la demande de la capacité de transport est supérieure à l'offre. Si les effets économiques du refus ne sont pas internalisés correctement via le système de prix, des problèmes d'externalités apparaissent. Généralement, on trouve ce type de problèmes lorsqu'aucun mécanisme ou marché ne traduit les contraintes gérées par des outils de contrôle-commande. On dit alors que le marché est « incomplet » (Boucher-Smeers [2002]).

L'électricité n'est pas le seul bien où il est possible de trouver des externalités et la théorie économique standard propose différentes solutions pour les « internaliser » et « réparer » ainsi cette défaillance du marché (Varian [1995], Salanié [2000]). Mais la difficulté pour l'électricité est la manière dont se produisent ces externalités. L'opérateur du système doit gérer l'utilisation du réseau a minima en temps réel. Il doit contrôler que les flux électriques ne dépassent pas les limites du réseau. Son intervention se manifeste en modifiant les injections et les soutirages proposés par les utilisateurs. Comme ces injections et soutirages varient continuellement, les externalités varient continuellement elles aussi. En conséquence, les externalités dans un réseau électrique sont instantanées et changeantes et

[75] En général nous sommes en présence d'une externalité lorsque les actions d'un individu ont une incidence directe sur un autre individu sans que cela passe par le système de prix (Varian [1995]). Nous faisons références à des externalités d'usage du réseau (Pignon [2003]) et non aux « externalités de réseau » (par exemple, une externalité de club, Curien [2000]).

peuvent donc être difficiles à gérer et à anticiper. De telles caractéristiques complexifient les modes d'internalisation standard (Pignon [2003]).

Nous avons analysé les conséquences économiques de la non-stockabilité et de la contrainte P=C puis celles de la difficulté à diriger les flux électriques et du respect des contraintes de capacité de transport. Nous verrons par la suite les conséquences économiques d'une troisième caractéristique de l'électricité : le besoin de réserves pour assurer la sécurité d'approvisionnement.

1.2.1.3 L'utilisation de réserves pour assurer la sécurité d'approvisionnement

La sécurité d'approvisionnement à court terme est une qualité du bien « électricité » qui doit être garantie. Du fait des nombreuses incertitudes, le système doit être préparé à respecter, à chaque instant, la contrainte d'équilibre P=C. Le moindre déséquilibre pourrait entraîner le système à l'instabilité suivie d'une panne totale (*blackout*). Un minimum de capacité de production flexible doit alors être disponible pour parer à un déséquilibre soudain sur le système. Des capacités de production doivent être prêtes à produire en cas de besoin (par exemple, suite à la panne intempestive d'une centrale de production, à la forte augmentation non prévue de la consommation, ou à la déconnexion d'une ligne).[76]

En conséquence, pour assurer la sécurité d'approvisionnement à court terme, un autre bien/service doit être donc produit en parallèle de l'énergie électrique, ce sont les réserves. La production de réserves pose un problème économique du fait qu'elles bénéficient à tous les utilisateurs du réseau, autant à ceux qui contribuent à la maintenir qu'à ceux qui n'y contribuent pas. En effet, on ne peut pas exclure des utilisateurs du réseau du bénéfice des réserves. Cette caractéristique est reconnue comme celle d'un bien public (Varian [1995]). Les caractéristiques de bien public sont considérées aussi comme une défaillance du marché éloignant l'électricité de l'idéal de marché parfait (Salanié [2000]).

Les économistes savent depuis longtemps que la production de biens publics ne peut pas être laissée seulement au marché, car une sous-production par rapport à la quantité nécessaire en résulterait (Varian [1995]). En conséquence, l'une des solutions au problème de la production d'un bien public est de fixer d'abord une demande réglementée puis d'assurer la production par le biais d'obligations, de contrats ou d'enchères. Le caractère non-stockable de l'électricité nécessite que cette production de réserves (ou de capacité de production disponible) soit instantanément adaptée aux conditions changeantes du système pour en assurer la sécurité.

En conclusion, les trois conséquences économiques principales qui émanent de l'étude des caractéristiques spécifiques de l'électricité sont : i) le besoin d'une forte

[76] La dernière option pour équilibrer le système est la diminution forcée de la consommation (délestage). Le niveau de réserves est calculé afin de n'utiliser cette dernière option que très rarement.

coordination en temps réel et d'une pré-coordination avant le temps réel ; ii) la présence d'externalités et, iii) des caractéristiques de bien public pour la production de réserves. Ces conséquences économiques marquent les bases de la construction d'une architecture de marché électrique et guident le choix des éléments essentiels.

1.2.2 Construction d'une architecture de marché électrique

L'art de la conception d'une architecture de marché électrique consiste à choisir quels contrôles-commandes remplacer et par quoi les remplacer, tout en respectant les caractéristiques spécifiques de l'électricité et de leurs conséquences économiques. Une architecture de marché doit remplacer les mécanismes de coordination internes aux outils centralisés de contrôle-commande. En pratique, le résultat final de l'introduction de la concurrence est donc toujours une combinaison entre contrôles-commande et marchés (Boucher-Smeers [2002]). Le rôle exact des marchés et des contrôles-commande ainsi que leur articulation doivent être soigneusement établis (Wilson [2002]).

Une première piste pour comprendre le remplacement des outils de contrôle-commande par des marchés consiste à percevoir le bien électricité comme la somme de trois composantes : l'énergie, le transport et les réserves. Ces trois composantes sont complémentaires et toutes nécessaires pour constituer un bien source d'utilité du consommateur. La composante « énergie » est souvent considérée comme la principale, ou comme la « vraie » commodité (Smeers [2004b]). Or, la composante énergie n'est valorisée par les consommateurs que combinée au transport et aux réserves. Ajouter deux composantes permet de considérer les externalités et les caractéristiques de bien public. La composante « transport » est liée à l'utilisation de la ressource rare nommée « capacité de transport » et sera utilisée lorsqu'on injecte de l'énergie dans un nœud et, qu'on soutire dans un autre nœud du réseau. La composante « réserves » est nécessaire pour assurer le fonctionnement correct du réseau et doit être produite parallèlement aux autres composantes. La nature de ces trois composantes, représentée par les caractéristiques spécifiques de l'électricité, impose un traitement particulier pour établir la limite de remplacement entre contrôles et marchés.

La question de la limite entre contrôles-commandes et marchés a été étudiée par Wilson [2002]. Il conclut que la technologie actuelle ne permet pas d'introduire de purs marchés pour les trois composantes (énergie, transport et réserves) pour les opérations en temps réel. De purs marchés, où l'offre et la demande se trouveraient et s'équilibreraient spontanément par le biais d'un prix, ne suffiraient pas pour combiner et coordonner instantanément (*timeless*) les trois composantes. C'est-à-dire qu'un pur marché ne pourrait pas égaliser l'offre et la demande de chacune des composantes et les coordonner entre elles en moins d'une seconde, avant que le système s'effondre. La réalité physique restreint donc le remplacement des contrôles par des marchés. A cet effet, l'opération du système en temps réel doit être réalisée, contrôlée et gérée par un opérateur centralisé, le gestionnaire du réseau (GRT).

Comme corollaire, des purs marchés pour les trois composantes ne sont réalisables qu'en avance du temps réel donc « à terme », comme des marchés *forward*[77] (Wilson [1998b], Wilson [2002], Boucher *et al.* [2006]). En effet, pour Wilson, l'architecture doit prévoir 3 types de sous-marchés à terme : i) l'énergie *forward*, ii) la capacité de transport *forward* et iii) les réserves *forward*. Cependant, ces marchés *forward*, pour être cohérents, doivent être articulés aux opérations du temps réel, seul moment où l'échange physique se réalise. Pour y parvenir, la solution proposée par Boucher *et al.* [2006] est de créer un pseudo « marché » du temps réel. Pour que ce pseudo marché remplace correctement les purs marchés manquants, il doit assurer la coordination en temps réel des trois composantes et fournir des signaux de prix cohérents avec les conditions économiques du temps réel. C'est pourquoi on parlera de marché « physique ». Ce pseudo-marché physique du temps réel, géré par le gestionnaire du réseau, joue donc un rôle prépondérant dans l'architecture de marché.

Les marchés *forward* ne sont pas des marchés « physiques » (Hirst [2001]). Ceci vient du fait que l'obligation créée entre un producteur et un consommateur sur le marché *forward* est strictement financière. Si, en temps réel, le producteur ne fournit pas ce qui a été spécifié dans le contrat du marché *forward*, il recevra obligatoirement en temps réel le remplacement physique correspondant. C'est donc seulement en temps réel qu'on connait toutes les conditions économiques et physiques de livraison du bien électricité. C'est pourquoi, logiquement, les marchés *forward* devraient être fondés sur des anticipations des signaux des prix donnés par le pseudo-marché du temps réel, le « sous-jacent » en termes financiers. Les marchés *forward* sont déterminants parce qu'ils fournissent aux participants la possibilité de se couvrir des risques. Puisque les signaux de prix du temps réel sont par nature très variables et volatiles, ils provoquent de forts risques pour les participants. De plus, une bonne articulation entre les modules *forward* et le module du temps réel fournit les informations nécessaires à la pré-coordination et à la préparation de la livraison du bien en temps réel.

En conclusion, pour Wilson l'architecture de marché doit prévoir trois marchés *forward*, un pour chaque composante de l'électricité –énergie, transport et réserves– et un pseudo-marché du temps réel géré par le GRT et s'occupant de la coordination en temps réel des trois composantes. Cette conclusion implique que chacun des marchés de l'architecture assure des fonctions essentielles bien définies. Ces fonctions essentielles assurées par les marchés remplacent les fonctions de coordination interne des outils de contrôle-commande. Ainsi, comme nous le verrons dans la section suivante, la création d'un pseudo marché du temps réel apparaît comme une transformation de l'outil de contrôle-commande d'*optimal dispatch*. D'une manière identique, plusieurs caractéristiques des marchés *forward* proviennent du remplacement de l'outil de contrôle d'*unit commitment*.

[77] De la même manière que Wilson [2002] et Boucher *et al.* [2006], nous utiliserons le terme « *forward* » pour designer les marchés financiers ou à terme correspondant au court terme et au très court terme (J-1 et infrajournaliers).

Néanmoins, les contrôles automatiques utilisés en temps réel (réglages primaires et secondaires) ne sont pas remplacés par les marchés dans des environnements concurrentiels. La construction d'une architecture de marché consiste donc à remplacer une séquence de contrôles-commandes par une séquence de marchés couplée à quelques contrôles résiduels (figure 2-4).

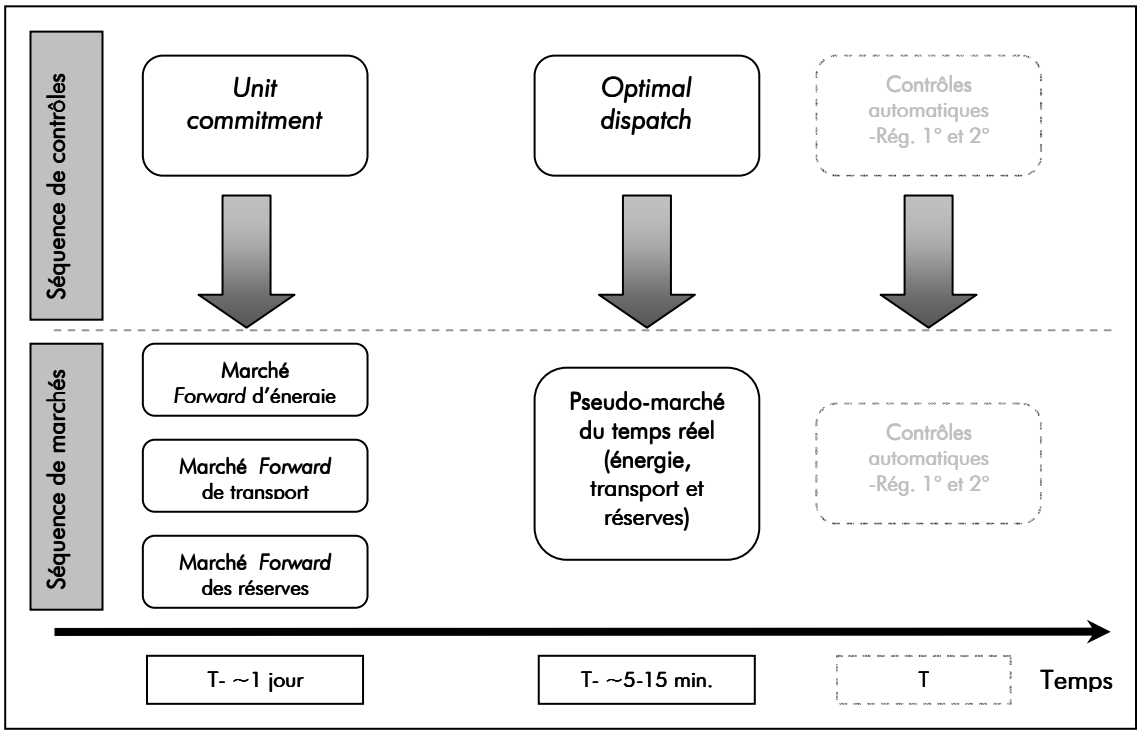

Figure 2-4 : Schéma du remplacement de contrôles par des marchés

D'une manière plus générale, les fonctions essentielles assurées par les marchés de l'architecture décrite ci-dessus peuvent être réalisées par d'autres mécanismes qu'un marché formellement défini. Comme nous le verrons dans la section suivante il existe plusieurs options ou formes organisationnelles différentes (dont la création d'un sous-marché) pour assurer les fonctions essentielles. L'analyse des outils de contrôle-commande permettra de dériver ces options ou variantes possibles. C'est pourquoi nous associerons à chacune de ces fonctions essentielles le terme « module » (Baldwin-Clark [2000]). Une architecture de marché sera donc composée de 4 modules : 1° le module d'énergie *forward*, 2° le module de transport *forward*, 3° le module de réserves *forward*, et 4° le module du temps réel.

1.3 Conclusions de la section

L'introduction de la concurrence, dans un secteur longtemps organisé autour d'un monopole intégré en production-transport, fait apparaître des problèmes de coordination qu'il convient de traiter par la mise en place de dispositifs organisationnels et institutionnels adéquats. Dans cette section, nous avons analysé l'introduction de la concurrence comme un remplacement des outils de contrôles-commande d'une entreprise verticalement intégrée par une combinaison de mécanismes et de marchés.

L'introduction de la concurrence par la suppression complète des outils de contrôle-commande, et sans construction d'une architecture de marché adéquate, est rendue impossible par les caractéristiques spécifiques de l'électricité et leurs effets économiques. Ces caractéristiques spécifiques sont : i) la non-stockabilité et l'équilibrage obligatoire en temps réel, ii) des flux non-dirigeables avec des contraintes de transport et iii) un besoin de réserves. Ces caractéristiques engendrent des conséquences économiques qui ne se retrouvent pas pour les biens simples, à savoir : un besoin de coordination forte et rapide en temps réel, des caractéristiques de bien public ainsi que des externalités volatiles, pouvant changer instantanément.

Pour un bien simple idéal, le remplacement des contrôles-commandes par des marchés, se réaliserait de manière spontanée et « naturelle ». Par contre, le remplacement des contrôles par des marchés électriques nécessite l'intervention d'autorités, du fait des caractéristiques spécifiques de l'électricité (Hogan [2002b]). La construction d'une architecture du marché est donc nécessaire pour l'électricité (Staropoli [2001]).

De ces caractéristiques spécifiques de l'électricité nous avons dérivé une architecture de marché qui considère l'électricité comme un bien combinant trois composantes : l'énergie, le transport et les réserves. Ces trois composantes sont complémentaires car elles sont toutes nécessaires pour fournir un bien utile au consommateur. La logique technico-économique couplée aux contraintes de remplacement des contrôles-commandes débouche sur une architecture composée de 4 modules (ou 4 sous-marchés potentiels) : 1° le module *forward* d'énergie, 2° le module *forward* de transport, 3° le module *forward* des réserves et 4° le module du temps réel. Chacun de ces modules assure une fonction essentielle pour l'architecture de marché. Seuls les modules *forward* peuvent prendre la forme des purs marchés où l'offre rencontre la demande pour définir spontanément un prix. En revanche, le module du temps réel doit être conçu comme un pseudo-marché en temps réel géré de manière centralisé par le GRT. Ce dernier module est très important parce qu'il gère et coordonne en temps réel les trois composantes et doit envoyer des signaux de prix cohérents avec les conditions économiques du temps réel.

Nous verrons par la suite que, pour chacun de ces modules, il existe plusieurs options ou formes organisationnelles différentes (dont la création d'un sous-marché). L'analyse des outils de contrôle-commande, que ces modules cherchent à remplacer, permet de dériver les options ou variantes possibles. Une architecture de marché est construite par le choix d'une option pour chaque module (ou marché potentiel) et pour chaque articulation entre ces modules.

Section 2: Architectures de marché sur une seule zone de contrôle

Dans la section précédente, nous avons vu comment on pouvait introduire la concurrence dans un système électrique en remplaçant par des marchés les outils de contrôle-commande (outils d'optimisation tels que *unit commitment* et *optimal dispatch*). Nous avons montré qu'une architecture de marché électrique doit développer quatre modules pour pouvoir gérer les caractéristiques spécifiques de l'électricité. Ces modules sont : 1° le module d'énergie *forward*, 2° le module de transport *forward*, 3° le module de réserves *forward*, et 4° le module du temps réel.

Le premier objectif de cette nouvelle section est de développer un cadre d'analyse modulaire pour étudier les architectures de marché. Ce cadre d'analyse modulaire consiste à définir les options principales de design pour chaque module (Baldwin-Clark [2000]). En conséquence, concevoir une architecture du marché électrique sur une zone de contrôle consiste à choisir une option pour chacun de ces modules et pour leurs articulations.

Dans une première partie, nous présenterons les principales options de design des quatre modules, en nous basant sur les outils d'optimisation que ces modules sont sensés remplacer. Notre but n'est pas de décrire en profondeur et en détail toutes les caractéristiques techniques ou économiques du *design* de chaque module, mais de montrer deux conséquences essentielles de la modularité d'une architecture de marché. Premièrement, le fait que chaque module ait de multiples options internes suggère une grande diversité possible d'architectures[78]. Cette grande diversité révèle aussi la complexité inhérente des architectures de marché électrique. Deuxièmement, étant donné la complexité des architectures de marché électrique, il serait impossible d'avancer vers une connaissance plus approfondie du *Market Design* sans un cadre d'analyse modulaire qui sépare le problème de l'architecture de marché en sous-problèmes. Ainsi la modularité de l'architecture de marché permet aussi d'étudier les modules séparément les uns des autres.

On observe effectivement une grande diversité d'architectures, et les architectures réelles choisissent des combinaisons très différentes d'options sur la séquence entière ou sur un seul module. Nous le montrerons en analysant trois cas d'architectures de marché qui font référence. Cette diversité d'architectures observée en pratique, déjà suggérée par la notion même de modularité, s'explique principalement par la difficulté à trouver une seule « architecture optimale » valable pour tout type de système électrique. Dans un monde parfait d'un point de vue économique (c'est-à-dire un monde avec atomicité d'acteurs, information complète, coûts de transaction nuls, fonctions de coût convexes, absence d'externalités et de caractéristique de bien public) le choix de l'architecture serait en principe

[78] On pourrait en principe dériver de nouvelles architectures encore inconnues à partir de la seule combinatoire des options pour chaque module.

moins important car plusieurs architectures alternatives pourraient donner le niveau d'efficacité maximum (le « coût minimal de production » si on considère une demande inélastique d'électricité). En réalité, les caractéristiques effectives des systèmes électriques s'éloignent des hypothèses économiques de perfection mentionnées ci-dessus (structures de marchés concentrées, coûts de transactions non nuls, externalités, fonctions de coût non-convexes, caractéristiques de bien public, etc.). Dans ce contexte, les différentes architectures devront être comparées par rapport à leur sensibilité à la prise en compte de ces « défaillances du marché ». Par ailleurs, dans cette optique plus réaliste, le choix d'une architecture de marché devra considérer, outre le « coût minimal de production », d'autres facteurs comme les difficultés de mise en œuvre et leurs coûts, les autres coûts de fonctionnement des institutions de marché, les contraintes institutionnelles, etc. (Coase [1937], Williamson [1985]). Malheureusement, l'état de l'art actuel sur le design d'architectures de marchés électriques ne fournit pas une base théorique « robuste » pour pouvoir trancher définitivement entre toutes les différentes solutions possibles. Le problème semble trop complexe pour trouver des solutions purement théoriques valables pour tous les états des systèmes électriques et leurs environnements.

La diversité et la complexité des architectures de marché révèlent ainsi l'intérêt pour le développement d'un cadre d'analyse modulaire. Ce cadre d'analyse modulaire permet d'étudier les modules séparément les uns des autres. Et, en conséquence, une étude approfondie d'un seul module est rendue possible. Elle est même tout à fait pertinente si le sous-ensemble découpé dans l'architecture de marché y joue un rôle bien identifiable et important.

Le deuxième objectif de cette section est donc de montrer que, dans cette diversité d'architectures, il existe un noyau commun à tous les designs pour remplir des fonctions « électriques » minimum indispensables dans toutes les architectures. Nous l'appellerons l'« architecture minimale », et il s'agit du module « du temps réel ».

Cette section va s'organiser comme suit. Nous présenterons d'abord les options de designs pour chacun des modules d'une architecture de marché (sous-section 2.1 à 2.4). Puis nous présenterons l'articulation entre les modules (sous-section 2.5). Nous développerons ainsi le cadre d'analyse modulaire. Dans la sous-section 2.6, ce cadre d'analyse sera appliqué à l'étude de trois marchés électriques de référence dans le monde (NETA, TEXAS, PJM). Dans la sous-section 2.7 nous aborderons le concept d'architecture minimale. Et dans la sous-section 2.8 nous conclurons.

2.1 Le module d'énergie *forward*

Dans cette sous partie nous considérons le marché *forward* d'énergie qui est placé la veille du jour de livraison (*day-ahead* en anglais).[79] Le marché *day ahead* d'énergie *forward* est celui qui a été le plus étudié. En fait, l'énergie est souvent considérée comme la « vraie » commodité dans le bien électricité (Smeers [2004b]). Les marchés *forward* de l'énergie remplissent deux rôles principaux : un rôle d'information pour faciliter la coordination, donner des bons signaux aux agents pour prendre les bonnes décisions de production et consommation, et un rôle d'outil de couverture face à des signaux de prix du temps réel très volatiles.

Le marché *day ahead* d'énergie qui couvre un horizon temporel de 24h avant le moment de livraison est déterminant pour la pré-coordination de la production. Etant donné les caractéristiques des moyens de production, les prix *day-ahead* sont une des bases pour la programmation de production des différentes centrales (Wilson [2002]). Mais, en fait, le marché *day-ahead* est toujours un marché *forward*, d'un caractère « financier » et non un marché « physique ». Cela vient du fait que l'obligation créée entre un producteur et un consommateur sur le marché *forward* est strictement financière. Par ailleurs, étant donné la présence d'incertitudes sur les conditions du temps réel, ces marchés *forward* sont utilisés pour se couvrir des risques.

Passons maintenant à la description des principales possibilités de design pour ce module en analysant l'outil du *unit commitment* qui est la base, comme nous l'avons déjà mentionné, du remplacement par des marchés. Pour nous centrer sur la composante « énergie *forward* » nous faisons abstraction des contraintes de transport et de réserves et nous considérons qu'il n'y a pas d'incertitudes.[80] Créer un marché d'énergie *forward* consiste donc à remplacer le programme d'optimisation ci-dessous représenté (cf. chapitre 1, section 2.2.1). Ce programme d'optimisation détermine quelles unités de production démarrer ou arrêter et en quel moment ($u_{i,t}^{UC}$) et alloue les quantités de production pour chaque unité ($q_{i,t}^{UC}$) afin de minimiser le coût de production pour un horizon temporel composé de T intervalles de temps t. Le bien électricité, étant non-stockable, est différencié pour chaque intervalle de temps t. En effet, le remplacement de cet outil d'optimisation pourrait se faire directement en le remplaçant par T marchés différents et indépendants (correspondant à chacun des biens différenciés pour un intervalle t) si on pouvait négliger la présence des caractéristiques spécifiques des moyens de production (coût de démarrage, ou les variables binaires -démarrer ou non-, équations UC5 dans notre exemple). Malheureusement ce n'est pas toujours le cas.

[79] Le marché *day-ahead* est un des derniers marchés *forward*. En effet, d'autres marchés *forward* (normalement bilatéraux ou gré à gré) ont des horizons temporels de quelques semaines jusqu'à plusieurs années. Dans le chapitre 3, dédié en partie à la séquence des marchés d'énergie, nous présenterons les marchés infrajournaliers qui se placent entre le marché *day ahead* et le temps réel.

[80] Nous nous concentrons ici sur la coordination. Dans les chapitres 3 et 4 nous analyserons la séquence des marchés d'énergie et le rôle de couverture de risques des marchés *forward*.

$$\min_{\underline{u}^{UC},\underline{q}^{UC}} \sum_{t=1}^{T}\sum_{i=1}^{I}\left[C_i(q_{i,t}^{UC})u_{i,t}^{UC} + S_i(u_{i,t}^{UC},u_{i,t-1}^{UC})\right] \quad \text{.....(UC1)}$$

Sous contraintes :

Contraintes d'équilibre :

$$\sum_{i=1}^{I} q_{i,t}^{UC}u_{i,t}^{UC} = D_t^{UC} \quad \forall t \quad \text{.....(UC2)}$$

Contraintes par unité de production

$$Q_i^{\min} \le q_{i,t}^{UC} \le Q_i^{\max} \quad \forall i,t$$

$$u_{i,t}^{UC} = \{0,1\} \quad \text{.....(UC5)}$$

$$S_i(u_{i,t}^{UC},u_{i,t-1}^{UC}) = \begin{cases} SC_i \ si \ u_{i,t}^{UC} > u_{i,t-1}^{UC} \\ 0 \quad si \ u_{i,t}^{UC} \le u_{i,t-1}^{UC} \end{cases}$$

Les caractéristiques spécifiques des moyens de production ont deux conséquences principales : elles introduisent des non-convexités dans la fonction de coût de production et elles créent des liens intertemporels entre les biens différenciés pour chaque intervalle t. D'une part, les fonctions de coûts non-convexes violent une des hypothèses de « marché parfait » car l'existence d'un équilibre concurrentiel ne peut plus être garantie. Les coûts de démarrage et d'arrêt, les limites minimales de fonctionnement, les coûts de charge à vide sont quelques-unes des caractéristiques qui rendent la fonction de coûts non-convexe (dans notre modèle simple d'*unit commitment*, ces non-convexités sont représentées par les variables binaires de démarrage, pour le coût de démarrage, et par le limites minimales de fonctionnement). Dans des systèmes avec fonction de coûts non-convexes l'existence d'un équilibre concurrentiel (où un ensemble de prix permet d'égaliser l'offre et la demande de tous les biens) n'est pas assurée (Arrow-Debreu [1954], Aliprantis *et al.* [1990], Elmaghraby *et al.* [2004], O'Neill *et al.* [2005]). D'autre part, à cause des liens intertemporels, il est nécessaire de considérer la production de l'ensemble des biens différenciés pour un horizon temporel assez long pour pouvoir prendre en compte les caractéristiques des moyens de production (par exemple, un moyen de production avec des fortes contraintes de flexibilité ne pourrait pas produire une grande quantité d'énergie pour l'intervalle t et rien pour l'intervalle $t+1$). Néanmoins, des problèmes de non-convexités et de liens intertemporels sont présents dans de nombreuses industries et l'utilisation de marchés n'a pas posé des difficultés graves. Ces problèmes peuvent être négligeables si leur taille n'est relativement pas importante pour l'industrie considérée. Pour l'électricité, en règle générale, les problèmes de non-convexités et de liens intertemporels sont moins importants dans des grands systèmes, où la taille unitaire des moyens de production ne représente qu'une fraction marginale du parc (Joskow-Schmalensee [1983], Stoft [2002]).

Les caractéristiques spécifiques des moyens de production (dont les non-convexités et les liens intertemporels) nécessitent un degré de coordination important pour atteindre un fonctionnement à moindre coût. La difficulté à trancher d'une manière définitive cette

question du degré de coordination nécessaire explique la diversité des options de design pour le module « énergie *forward* ». La plupart des solutions mises en œuvre se placent sur un axe où les extrêmes sont d'une part les marchés centralisés (marchés organisés ou enchères) et d'autre part des marchés décentralisés (marchés de gré à gré, contrats bilatéraux, etc.). Un design plus centralisé favorise la coordination. La justification intuitive de ce principe est simple. Un certain degré de centralisation est nécessaire car certaines informations économiques essentielles sont impossibles à obtenir par des transactions seulement bilatérales (par exemple, les non-convexités de la fonction de coût) (Smeers [2004b]). Une organisation décentralisée de marché nécessiterait un nombre important de transactions bilatérales pour arriver à un bon degré de coordination et ceci accroitrait les coûts de transactions.[81] De plus, comme dans ce cas des prix transparents et publics sont plus difficiles à trouver les coûts de transactions peuvent être plus importants (Counsell-Evans [2003], Boisseleau [2004], Baldick-Niu [2005]). Cependant, les designs trop centralisés sont vus comme opaques et gouvernés par des règles complexes, faciles à influencer par la manipulation des règles ou *gaming* (Wilson [2002], Oren [2003b] et Baldick *et al.* [2005]). De plus, une gestion centralisée nécessite une attention particulière pour le design des incitations données à l'entité qui centralise le marché (opérateur du marché ou l'opérateur du système) (Oren [1998], Cameron-Cramton [1999], Stoft [2002], Sioshansi [2006]).

Les options de design pour le module « énergie *forward* » sont cohérentes avec la manière dont on remplace la partie énergie de l'outil d'optimisation de l'*unit commitment* par des organisations de marché. La manière la plus directe de « remplacer » l'outil d'*unit commitment* est de créer un marché centralisé basé sur le même programme d'optimisation. Cette enchère ou marché centralisé considère de manière conjointe tous les biens correspondant aux intervalles t et pour l'horizon T. La logique de remplacement est la suivante. Le *unit commitment* est un modèle d'optimisation ; si on lui fournit des données, il détermine une solution optimale (qui produit et en quelle quantité). De plus, il donne des informations qui peuvent être interprétées comme des prix et des paiements. Cette première approche consiste donc à remplacer les « données » (coûts incrémentaux, coûts de démarrage, etc.) par des offres complexes (prix demandé pour différents nivéaux de production, prix pour démarrer, etc.). L'*unit commitment* est ainsi transformé en marché centralisé (enchère[82]) ou *pool* obligatoire[83]. Les quantités à produire et les prix sont déterminés de manière centralisée. Les non-convexités et les liens intertemporels sont pris en compte en interne de l'enchère. Les designs définis comme des « *pool* » ont utilisé une

[81] Les coûts de transactions comprennent notamment les coûts de découverte du co-contractant, les coûts nécessaires pour l'informer d'une opportunité d'échanger, les coûts de négociation des termes de l'échange, les coûts d'écriture et de mise en œuvre des contrats (Coase [1937], Saussier [1997]).
[82] Ceci n'est pas une enchère commune, c'est en fait une enchère combinatoire avec des offres complexes ou à plusieurs parties (Cramton *et al.* [2005]).
[83] Cette option extrême est connue sous le nom de « *Pool* obligatoire » car toutes les transactions doivent passer par ce marché organisé.

méthode particulière de détermination des prix (Contreras *et al.* [2001], Stoft [2002]).[84] Le prix *day ahead* pour chaque intervalle est fixé égal à l'offre marginale correspondante. Mais comme les prix déterminés de cette façon n'assurent pas que tous les producteurs puissent couvrir tous leurs coûts et que l'offre égalise la demande, un paiement complémentaire (« *side payment* ») est utilisé.[85] Cette option très centralisée de design a été appliquée par exemple en Angleterre avant 2001 (UK Pool).

L'autre extrême de la fourchette d'options réalisables consiste à remplacer l'*unit commitment* par des marchés décentralisés, ou des marchés de gré à gré, basés principalement sur des contrats bilatéraux et sans la création d'un marché organisé. Ce design impose la décentralisation complète de l'*unit commitment* et la mise en place du « *self-commitment* ». Chaque couple de vendeur (e.g. producteur) et acheteur (e.g. consommateur) fixe les conditions auxquelles se font les transactions d'énergie *forward*. Les décisions de production pour chaque unité de production sont prises de manière décentralisée par chaque producteur selon les transactions ou contrats signés. Les résultats en volume des transactions décentralisées (ainsi que les programmes de production et de consommation) doivent être communiqués au gestionnaire du réseau. Cette option de design a été appliquée par exemple au Texas.

Les deux cas présentés précédemment sont des cas extrêmes. Actuellement, les designs utilisés dans la plupart des architectures de marché sont des formes hybrides. D'un coté les designs de marché centralisés permettent aux acteurs de réaliser leur propre *unit commitment* (*self-commitment*). Des designs proches de cette option ont été mis en œuvre dans plusieurs architectures notamment dans certains états du nord-est et du centre-ouest des Etats-Unis (PJM, NE Pool, NY Pool, MidWest Pool).[86] D'un autre coté les designs dits décentralisés fonctionnent en parallèle avec des marchés *day ahead* organisés mais de participation volontaire (bourses d'électricité ou *power exchanges*) (Boisseleau [2004], Meeus [2006]). Ces bourses créent un prix public de référence qui sert à la coordination non pas seulement pour les transactions passant par ces bourses, mais aussi pour le reste du marché non organisé (marché de gré à gré). De plus, dans certaines de ces bourses, le format des offres (offres simples et offres complexes ou blocs[87]) permet aux acteurs de mieux représenter les parties non-convexes et les liens intertemporels de la production d'électricité (Meeus [2006]). Des designs proches de cette option ont été mis en œuvre dans plusieurs

[84] Cette méthode renvoie à la définition de prix non-linéaires. Un prix non-linéaire serait par exemple un prix à deux parties, une partie « fixe » et une partie proportionnelle à la quantité (Meeus [2006]).
[85] Récemment, des développements ont été réalisés sur des méthodes pour dériver des prix non linéaires (prix pour l'énergie, prix pour les décisions binaires, etc.) qui permettraient un équilibre concurrentiel (O'Neil *et al.* [2005], Hogan-Ring [2003], Bjørndal-Jörnsten [2004a] et Bjørndal-Jörnsten [2004b]).
[86] PJM est le marché électrique historiquement centré sur les états de Pennsylvania, Maryland et New Jersey. NE Pool et NY Pool correspondent respectivement aux états de New England et New York. Enfin, MidWest Pool correspond à plusieurs états du centre-ouest américain dont Wisconsin, Michigan et Minnesota.
[87] Un « bloc » correspond normalement à la production d'énergie pour plusieurs intervalles de temps de suite. Les offres complexes correspondant aux blocs incorporent des contraintes spécifiques telles comme l'indivisibilité temporelle et d'acceptation complète (Meeus [2006]).

architectures notamment en Europe (Angleterre, France, Pays-Bas, etc.) et dans certains états des Etats-Unis (Californie, etc.).

Jusque là nous nous sommes concentrés sur le rôle de coordination du module énergie *forward* et nous avons considéré qu'il n'y avait pas d'incertitudes. En réalité, au moment de la réalisation du marché *forward* il y a des incertitudes. En effet, les participants anticipent les conditions du système en temps réel pour participer au marché *forward* d'énergie. Les marchés *forward* sont donc utilisés pour couvrir ces risques. Nous reviendrons sur la couverture des risques dans le chapitre 3.

En conclusion, les principales options de design du module « énergie *forward* » varient selon un axe qui exprime le degré de centralisation de l'option. Par simplicité, dans le tableau 2-3, nous présentons seulement les options extrêmes ainsi que quelques exemples, sachant que la plupart des solutions existantes sont hybrides, mais qu'elles se rapprochent plus d'une de ces solutions extrêmes.

Tableau 2-3 : Options pour le module énergie *forward*

Module d'énergie *forward*	
Degré de centralisation	• Marché centralisé : UK Pool, PJM, NE Pool, NY Pool, MidWest Pool. • Marché décentralisé : Texas, France, Angleterre (après 2001).

Nous avons présenté les principales options de design pour le module d'énergie *forward*. Dans la sous-section suivante, nous présenterons celles du module de transport *forward*.

2.2 Le module de transport *forward*

De la même façon que le marché *forward* de l'énergie, le marché *forward* du transport a deux fonctions principales : 1° faciliter la coordination et 2° donner une couverture contre les risques des signaux de prix du temps réel différenciés géographiquement. Les signaux économiques envoyés par ce module devraient conduire les acteurs du marché à prendre les décisions efficaces étant donné les contraintes de transport. C'est pourquoi, ces signaux devraient prendre en compte les effets des actions individuelles sur celles des autres utilisateurs du réseau, c'est-à-dire, internaliser les externalités. Néanmoins, ce module traite de la pré-coordination ou de la pré-internalisation des externalités car l'utilisation du réseau et les externalités respectives ne se réalisent qu'en temps réel. Les conditions du système peuvent changer entre le module *forward* et le temps réel et les signaux économiques *forward* sont donc complétés par ceux envoyés par le module du temps réel.

En poursuivant avec l'idée de remplacement d'outil d'optimisation par des marchés, le module *forward* du transport naîtrait du remplacement des équations de l'*unit commitment* concernant les limites de capacités de lignes. Deux solutions de design pour ce module

s'imposent pour créer un mécanisme d'internalisation des contraintes de transport. La première option est un module transport *forward* intégré au module énergie *forward* (via des enchères implicites). La deuxième option est un module transport *forward* séparé du module énergie *forward* (via des enchères explicites).

Intégration des modules *forward* énergie et transport (ou « enchères implicites »)

La première option consiste à internaliser les externalités par le biais de prix de l'énergie différenciés par localisation. C'est-à-dire que les composantes « énergie » et « transport » *forward* sont traitées dans un seul marché organisé. Cette option est dérivée du remplacement complet de l'outil d'*unit commitment* par un marché centralisé (ou enchère) qui prend en compte les contraintes des capacités de transport. En effet, de cette enchère on dérive des prix (*forward*) d'énergie différenciés géographiquement. Cette forme d'organisation est connue aussi sous le nom d'« enchères implicites »[88] car l'allocation et la valorisation du bien « capacité de transport » se réalise implicitement sur la base de l'offre et la demande du bien « énergie » dans les différentes localisations ou nœuds du réseau.

Afin de nous concentrer sur l'énergie et sur la capacité de transport, nous considérons une version simplifiée de l'outil d'*unit commitment* sans les contraintes de réserves. Nous ne représentons pas non plus les variables binaires. Ceci nous permettra de faire abstraction des problèmes de non-convexités de la fonction de coût de production déjà traités précédemment.

$$\min_{\underline{u}^{UC},\underline{q}^{UC}} \sum_{t=1}^{T}\sum_{i=1}^{I} C_i(q_{i,t}^{UC}) \quad \text{.....(UC1')}$$

Sous contraintes :

Contraintes d'équilibre :

$$\forall t \quad \sum_{i=1}^{I} q_{i,t}^{UC} = D_t^{UC} \quad (\lambda_t) \quad \text{.....(UC2')}$$

Contraintes de transport (limites sur le réseau)

$$\forall l,c,t \quad -F_l \le \sum_{i=1}^{I} PTDF_{l,i,c}\; q_{i,t}^{UC} \le F_l \quad (\mu_{l,c,t}) \quad \text{.....(UC3')}$$

Contraintes par unité de production

$$\forall i,t \quad Q_i^{\min} \le q_{i,t}^{UC} \le Q_i^{\max} \quad \text{.....(UC5')}$$

L'ensemble d'équations UC3' nous dit que les flux sur chaque ligne du réseau, provoqués par les injections $q_{i,t}^{UC}$, doivent être inférieurs ou égaux aux limites physiques de capacité pour tous les états du réseau (règle N-1)[89] et pour tous les intervalles de temps t. Chaque fois qu'une de ces contraintes est « active », la capacité de l'élément du réseau considéré est utilisée en totalité et devient alors une ressource rare. On dit que cet élément

[88] Et cela sert aussi pour les différencier des « enchères explicites » de la capacité de transport où les modules transport et énergie sont séparés.

[89] Notez que les équations sont représentées aussi pour différents « états » (c) de la topologie du réseau, et qu'alors la règle de sécurité « N-1 » (perte d'une ligne) est incluse.

est congestionné. Une manière économique de donner des signaux de prix pour valoriser l'utilisation de cette ressource rare serait d'utiliser les multiplicateurs de Lagrange des contraintes de transport ($\mu_{l,c,t}$) qui, combinés au multiplicateur de Lagrange de la contrainte d'équilibre (λ_t) permettent de dériver des signaux de prix de la composante « énergie *forward* » localement différenciés. Par exemple, si un élément du réseau n'est pas congestionné, alors le multiplicateur $\mu_{l,c,t}$ est nul.

Les prix « énergie *forward* » peuvent être représentés par la formule suivante :

$$prix_{i,t} = \lambda_t + \sum_{l=1}^{L} \sum_{c=1}^{C} PTDF_{l,n,c}\, \mu_{l,c,t}$$

Ces prix nodaux sont composés de deux parties. Une première partie correspondant au prix de l'énergie au nœud de référence (λ_t). Et une deuxième partie correspondant à la valeur de l'utilisation du réseau de transport pour ramener l'énergie du nœud de référence au nœud i. La valeur de l'utilisation du réseau est dérivée de la valorisation d'utilisation de chaque ligne ($\mu_{l,c,t}$) pondérées par les coefficients d'influence (PTDF). Ces derniers coefficients mesurent l'impact de chaque transaction entre chaque nœud i et le nœud de référence pour chaque ligne l. Un tel système de prix permettrait d'internaliser les externalités en allouent le coût des congestions aux responsables et en rémunérant ceux qui libèrent de la capacité de transport par des contre-flux. Boucher-Smeers [1999] démontrent théoriquement qu'un système de prix représenté par les équations précédentes arriverait à la même allocation de quantités de production qu'un planificateur bienveillant utilisant l'outil d'optimisation similaire. Cette équivalence, et son efficacité, sont seulement valables si des hypothèses fortes sont vérifiées (absence de pouvoir de marché et coûts de transaction, etc.) mais néanmoins cette solution est importante à considérer parce qu'elle résout le problème d'externalités crée par les congestions.

Cette approche, qui se base principalement sur les développements théoriques de Schweppe *et al.* [1988] et Hogan [1992], est connue comme le modèle nodal. Elle spécifie un prix de l'énergie différent pour chaque nœud du réseau et crée une différenciation spatiale du bien électricité. Le modèle nodal a été largement mise en œuvre dans les architectures de marché du nord-est et du centre-ouest des Etats-Unis (PJM, NY Pool, NE Pool et MidWest Pool) et a fait partie du design de référence appelé *Standard Market Design* (SMD) de la FERC (Régulateur de l'énergie aux Etats-Unis).[90] Mise à part les avantages de ce modèle pour internaliser les externalités, la mise en œuvre n'est pas simple. Dans un réseau électrique il y a beaucoup de nœuds. Il faut donc créer beaucoup de biens « énergie électrique » différents, un par localisation.[91] Ceci peut poser plusieurs problèmes étant donné la complexité de gestion de tous ces biens et marchés respectifs intrinsèquement liés,

[90] Actuellement l'architecture du marché au Texas évolue vers ce type de design.
[91] Par exemple, le modèle de réseau utilisé par PJM considère plus de 10000 nœuds et celui de MISO plus de 30000 nœuds (PJM-MISO [2006]).

par exemple : des problèmes de coûts de transaction et de mise en œuvre élevés, des problèmes de manque de liquidité pour certains produits trop spécifiques, des problèmes de pouvoir de marché local, etc. (Harvey-Hogan [2000a], Harvey-Hogan [2000b], Smeers [2004b]). Cette approche nécessite aussi un degré important de centralisation du module énergie *forward* et une participation forte du gestionnaire du réseau.

Séparation des modules *forward* énergie et transport (ou « enchères explicites »)

La deuxième option est basée sur la création de droits de propriété pour l'usage de la capacité de transport et pour que les participants puissent échanger ces droits d'une façon décentralisée, dans un marché *forward* du transport. Les participants du marché voulant injecter l'énergie dans un nœud du réseau et soutirer dans un autre doivent se procurer les droits de propriété nécessaires à l'usage des éléments du réseau. Ces droits de propriété sont normalement mis aux enchères par l'opérateur du système. Pour cela cette approche est connue aussi sous le nom d'« enchères explicites ». Les composantes « énergie » et « transport » *forward* sont gérées séparément. Une version simplifiée de cette option a été mise en œuvre en Californie (Borenstein *et al.* [2000]).

Afin de mieux comprendre cette deuxième méthode, prenons un exemple. Supposons que notre réseau est composé de trois nœuds seulement (figure 2-5). La seule ligne qui a une limite de capacité est celle qui va du nœud 1 au nœud 2 (F_{1-2}). Les PTDFs pour cette ligne sont respectivement 1/3, -1/3 et 0 pour le nœud 1, 2 et 3 (cf. Chapitre 1, sous-section 1.3.2). Les droits de propriétés doivent être créés sur les ressources rares. Dans ce cas, la ressource rare est la capacité de la ligne 1-2 (F_{1-2}). Les PTDFs sont la clé de répartition de l'utilisation de cette ressource rare. Si un participant souhaite réaliser une transaction allant entre le nœud 1 et le nœud 2 il devrait se procurer 2/3 unités de la ressource « capacité de transport de la ligne 1-2 » pour chaque unité de transaction.

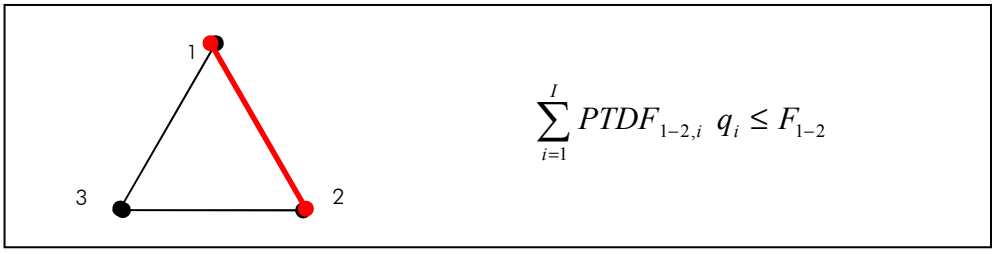

Figure 2-5 : Droits de transport et règles d'échange

Boucher-Smeers [1999] et Chao-Peck [1996] démontrent théoriquement, sous des fortes hypothèses (absence de pouvoir de marché et de coût de transaction, rationalité et anticipation parfaite des acteurs, etc.) que le résultat d'équilibre pour ce modèle, qui repose sur deux marchés (énergie *forward* et transport *forward*), est équivalent au résultat du modèle intégré et que tous les deux sont également efficaces.

Cette approche est notamment plus décentralisée que l'approche « intégrée ». Tout d'abord, parce qu'elle ne nécessite pas une centralisation du module d'énergie *forward* et

parce que l'opérateur du système (le GRT) n'intervient d'une manière centralisée *a priori* que pour définir les droits de transport et les PTDFs. Mais, cette approche présente aussi des difficultés importantes de mise en œuvre. Premièrement parce que le nombre de droits de propriété différents à créer est très grand (Ruff [2001]). En fait, on devrait créer de droits de propriété différents pour chaque ligne du réseau. Le nombre de droits à créer deviendrait encore plus important si l'on considérerait les contraintes de sécurité du réseau (règle N-1). Dans ce cas, on devrait créer des droits de propriété différents pour chaque ligne du réseau et pour chaque état du réseau (états N-1 représenté par c). Deuxièmement, cette approche nécessite une hypothèse supplémentaire par rapport à l'approche intégrée pour assurer l'efficacité théorique (Smeers [2003], Ehrenmann-Smeers [2005], Ehrenmann-Smeers [2006]). Etant donné que les marchés de l'énergie et du transport sont séparés et séquentiels, les acteurs aux marchés doivent avoir une anticipation parfaite pour pouvoir arbitrer correctement entre eux. Finalement, des études théoriques ont démontré que cette approche, qui sépare les modules d'énergie et de transport, est moins robuste au pouvoir de marché (Neuhoff [2003]).

Nous avons vu qu'une des principales difficultés de mise en œuvre des approches présentées (intégrée et séparée) vient principalement du nombre important de biens différents créés (biens « énergie » ou droits de transport différents). Or, les modèles présentés sont des cas extrêmes dans le sens où ils essayent d'internaliser toutes les externalités en représentant tous les éléments du réseau de transport. C'est-à-dire, qu'ils utilisent un « modèle économique de réseau » qui coïncide exactement avec le réseau de transport réel. Il existe néanmoins des solutions simplifiées consistant à utiliser des modèles économiques de réseau simplifiés permettant, en théorie, de choisir le degré d'internalisation d'externalités. Pour le cas du modèle intégré, il est possible de modifier le programme d'optimisation d'enchère pour définir un « modèle économique du réseau » regroupant des nœuds à l'intérieur de zones et en définissant de capacité de transport entre les zones. Ce modèle est dit zonal. La construction des zones est faite pour prendre en compte les externalités principales. Les simplifications réalisées permettent de différencier le prix de l'énergie par zones et ainsi simplifient le fonctionnement du marché. Le cas extrême de simplification consiste à traiter le système, dans un premier temps, comme si les contraintes de transport n'existaient pas. C'est le modèle de zone unique où le bien électricité (énergie) n'est pas différencié géographiquement et les externalités ne sont pas internalisées. Pour le cas du modèle séparé, une des solutions de simplification s'appuie sur la détermination des contraintes principales du réseau (*flowgates*), en agrégeant plusieurs contraintes, et en créant les droits de transport ad hoc sur ces contraintes principales (Chao *et al.* [2000]). Il est important de remarquer que bien que les modèles zonaux simplifient la mise en œuvre, la définition d'un nombre réduit de zones de prix ou de *flowgates* pour ne prendre en compte que les principales externalités n'est pas évidente et peut présenter des difficultés importantes, surtout pour les réseaux maillés[92] (Ott [2000], Hogan [2000], Smeers

[92] Un réseau maillé renvoie à l'existence de plusieurs chemins pour aller d'un nœud à un autre du réseau. Cf. Chapitre 1, sous-section 1.3.

[2001], Ehrenmann-Smeers [2005]). Dans la figure 2-6 nous représentons schématiquement les options de design qui concernent le degré d'internalisation des externalités.

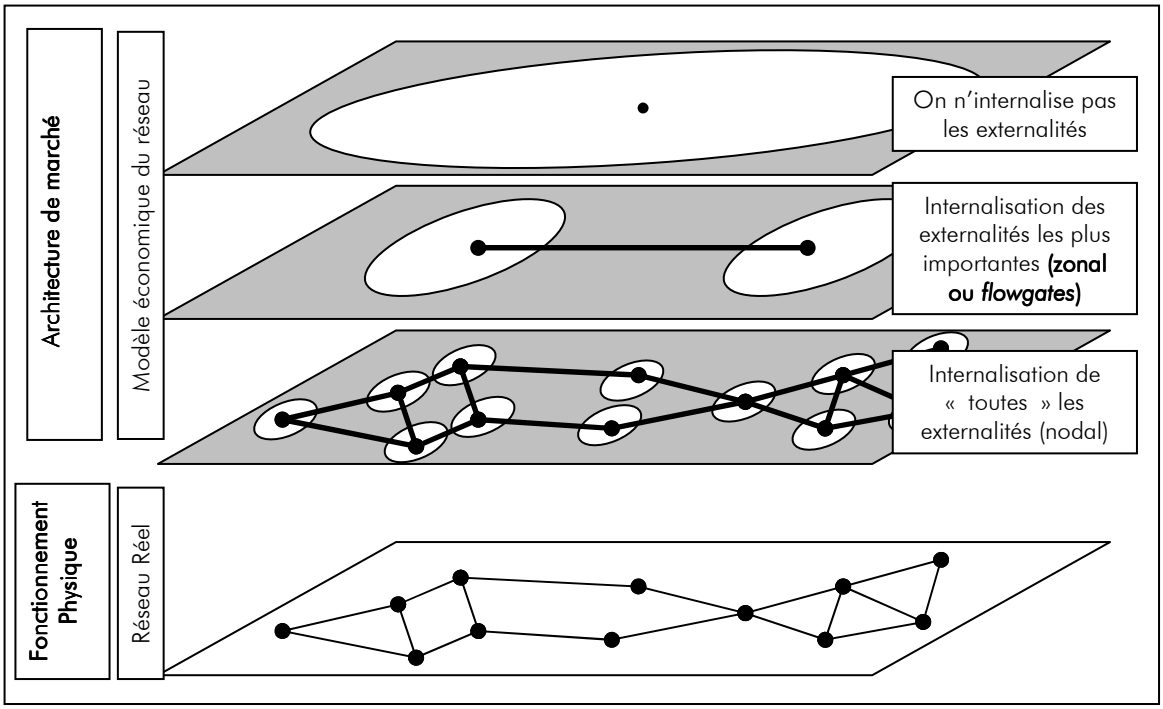

Figure 2-6 : schéma du degré d'internalisation des externalités

Enfin, le module *forward* de transport donne aussi aux acteurs du marché les moyens de se couvrir contre les risques liés au coût du transport en temps réel qui peut être très volatile. Il doit donc y avoir une cohérence entre les définitions du modèle économique du réseau pour ce module *forward* (zone unique, zonal ou nodal) et le modèle économique du réseau du module du temps réel. En général, le degré d'internalisation dans l'étape *forward* peut être aussi détaillé qu'en temps réel.[93]

En conclusion, les principales options de design pour le module de transport *forward* peuvent se différencier selon deux axes : l'intégration (ou séparation) au module d'énergie *forward* et le degré d'internalisation d'externalités du transport. Selon le premier axe, celui de l'intégration au module *forward* de l'énergie, deux solutions de designs s'imposent. La première option est un module transport *forward* intégré au module énergie *forward* (enchères implicites). La deuxième option est un module transport *forward* séparé du module énergie *forward* (enchères explicites). Selon le deuxième axe quant au degré d'internalisation des externalités, les options de design peuvent se repartit entre deux cas extrêmes. D'un côté, un très haut degré d'internalisation d'externalités introduit une grande différenciation

[93] Par exemple dans les architectures du nord-est et du centre-ouest des Etats-Unis (PJM, NY Pool, NE Pool, MidWest Pool) la définition du modèle économique est « nodale », tant dans le module *forward* (*day-ahead*) que dans le module du temps réel. L'architecture du marché au Texas présente une définition zonale, en temps réel mais sans pré internalisation sur le module *forward*. Les architectures où les externalités ne sont pas internalisées en temps réel (par exemple dans la plupart de pays européen) ne nécessitent pas, logiquement, de mécanismes de couverture contre les risques de coûts de transport en temps réel.

spatiale du bien électricité (modèle nodal). D'un autre côté, le cas extrême de simplification consiste à traiter le système comme si les contraintes de transport n'existaient pas (une seule zone). Entre ces deux cas extrêmes la définition d'un nombre réduit de zones de prix ou de *flowgates* permet de ne prendre en compte que les principales externalités (modèle zonal). Néanmoins, les congestions et/ou externalités qui ne sont pas traitées par ce module de transport *forward* devront l'être via le module du temps réel (cf. sous-section 2.4). Nous présentons dans le tableau 2-4 les principales options et quelques exemples de design du module « transport *forward* » selon les deux axes précédemment cités.

Tableau 2-4 : Options pour le module de transport *forward*

Module *forward* de transport	
Intégration	• Intégré au module énergie *forward* (enchères implicites) : PJM, NY Pool, NE Pool, MidWest Pool. • Séparé du module énergie *forward* (enchères explicites) : Californie
Degré d'internalisation d'externalités	• Une seule zone : France, Angleterre, Belgique. • Quelques zones ou « *flowgates* » (zonal) : Californie. • Chacun des nœuds (nodal) : PJM, NY Pool, MidWest Pool, NE Pool.

Nous avons présenté les principales options de design pour le module de transport *forward*. Dans la sous-section suivante, nous présenterons celles du module des réserves *forward*.

2.3 Le module des réserves *forward*

Les réserves sont nécessaires sur le système électrique car la sécurité du réseau à court terme exige d'avoir une certaine capacité de production disponible pour faire faces aux aléas (cf. chapitre 1). Ce besoin de réserves se matérialise en préparant des moyens de production pour produire de l'énergie en cas de besoin. La sécurité du réseau à court terme a des caractéristiques de bien public parce qu'il serait impossible d'exclure des bénéfices des réserves les utilisateurs du réseau qui refusent de payer pour elles (Oren [2003a], Kiesling-Giberson [2004]). La demande des réserves doit donc être réglementée et gérée par le gestionnaire du réseau. Les coûts de l'achat des réserves du GRT sont généralement distribués entre les utilisateurs du réseau sur la base de l'énergie consommée ou produite.

L'introduction de la concurrence pour la production de réserves peut être conçue à partir du remplacement des équations correspondant aux réserves de l'*unit commitment* (UC4) :

$$\sum_{i=1}^{I}\left(Q_i^{\max} - q_i^{UC}\right)u_{i,t} \geq R^{UC} \quad \forall t \quad \text{.....(UC4)}$$

R^{UC} est la demande réglementée des réserves.[94] De cette équation nous remarquons qu'il existe une différence entre la capacité disponible de production ($Q_i^{\max} - q_i^{UC}$) et la production elle-même (q_i^{UC}). La production des réserves consiste à rendre disponible de la capacité de production. C'est-à-dire que dans notre exemple, toute unité de production en fonctionnement pour une période t dont la production effective est inférieure à sa capacité maximale de production pourrait fournir des réserves.[95] De cette équation on voit aussi les liens avec les autres composantes du bien électricité ; par exemple, la quantité d'énergie produite à un moment donné limitera la capacité disponible à offrir comme réserve.

Une fois que la demande de réserves définie par le gestionnaire du réseau, le design du module *forward* des réserves consiste à définir les formes d'organisation pour l'allocation de la production des réserves. Nous distinguons donc deux options principales pour le design de ce module. La première option de design consiste à mettre en œuvre des contrats bilatéraux entre le gestionnaire du réseau et les producteurs de réserves. La plupart des GRT des pays européens achètent leurs réserves *forward* par des contrats bilatéraux avec les producteurs (ETSO [2003]). La deuxième option consiste à mettre en œuvre des marchés journaliers organisés sous la forme d'enchères (marchés *forward* de réserves). Cette dernière option peut avoir deux variantes, la première variante fonctionne d'une manière séparée du marché de l'énergie *forward* et la deuxième variante intègre le marché *forward* de réserves au marché *forward* d'énergie.[96] L'option de design de marchés organisés de réserves *forward* a été appliquée dans plusieurs états des Etats-Unis, comme par exemple le Texas ou la Californie (Wilson [2002], Baldick *et al.* [2005]).

Le design d'un marché organisé de réserves présente des problématiques particulières. Ces problématiques sont aussi valables pour la conception des contrats bilatéraux de réserves. La principale difficulté vient du caractère d'option à produire des réserves.[97] Le bien ou service de « réserves » est composé de deux parties : une partie correspondant à la disponibilité (option d'être appelé) et une partie correspondant à l'appel

---

[94] Notez que la demande de réserves R^{UC} peut ne pas être la même pour une entreprise intégrée verticalement que pour une organisation de marché. R^{UC} devrait être calculé sur la base de la probabilité de défaillance du système (combinaison de la probabilité de défaillance des moyens de productions plus de l'erreur de prévision de la consommation et d'autres incertitudes). Une entreprise intégrée possède toute l'information nécessaire pour calculer R^{UC}. Dans une organisation de marché, le calcul du R^{UC} peut être plus difficile car le GRT ne possède pas le même niveau d'information, et dans certains cas il devra utiliser des niveaux de réserves plus élevés.

[95] Certes, en pratique la capacité en réserves que peuvent offrir les différentes unités de production est aussi limitée par leurs contraintes de *ramping*.

[96] Des designs « hybrides » ont été aussi mis en œuvre, où une partie de la quantité de réserves est allouée entre les utilisateurs du réseau proportionnellement à leur consommation. De cette façon, les utilisateurs du réseau sont libres d'autoprogrammer leurs unités de production comme des réserves, de signer des contrats avec d'autres producteurs, ou d'acheter sur les marchés organisés (Baldick *et al.* [2005]).

[97] Un deuxième problème vient du fait de la substituabilité dans un seul sens des différents types de réserves. Plusieurs types de réserves différenciés par la flexibilité sont nécessaires pour le fonctionnement du réseau (cf. Chapitre 1). Un moyen de production très flexible (e.g. hydraulique) peut offrir sa capacité disponible pour plusieurs types de réserves. En revanche, un moyen de production moins flexible ne peut fournir que des réserves de basse qualité. Ce caractère de substituabilité des différents types de réserves nécessite un design d'enchère plus complexe considérant à la fois tous les types de réserves (Oren [2001a]).

des réserves pour produire de l'énergie. Le design d'enchères sur ce type de bien composé de deux parties a été étudié par Chao-Wilson [1999]. Ils arrivent à la conclusion qu'un design efficace (« *incentive compatible* ») pour ce type d'enchère nécessite la considération des offres composées de deux parties (une offre par disponibilité ou capacité et une offre par énergie). La sélection des offres des réserves (option d'être appelé) se fait en considérant seulement la composante de disponibilité ou de capacité. L'autre composante de l'offre (offre d'énergie) est considérée dans le *merit order* du module du temps réel (cf. sous-section 2.4). Les fournisseurs des réserves sont ainsi toujours rémunérés pour leur disponibilité. En cas de production d'énergie, ils sont rémunérés au prix de l'énergie en temps réel (Chao *et al.* [2005]).

En conclusion, les principales options de design du module *forward* des réserves se différencient selon l'allocation de la production des réserves définie par le GRT. Dans une première option, le GRT met en place des marchés organisés (ou enchères) pour se procurer des réserves. Dans une deuxième option, le GRT met en œuvre des contrats bilatéraux avec les producteurs pour la fourniture de réserves. Nous présentons dans le tableau 2-5, ces options ainsi que quelques exemples des zones de contrôle qui ont opté pour un design particulier.

Tableau 2-5 : Options pour le module *forward* de réserves

Module *forward* des réserves	
Organisation de la production	• Marchés organisés (Texas, PJM, NY Pool, NE Pool, etc.) • Contrats bilatéraux entre les producteurs et le GRT (France, Angleterre, Belgique, etc.)

Jusqu'à présent, nous avons présenté les principales options de design pour les modules *forward* d'énergie, de transport et des réserves. Or, ces modules ne traitent que des « promesses » de fourniture des composantes essentielles du bien électricité. Dans la sous-section suivante, nous présenterons les options du module du temps réel, responsable de la coordination des trois composantes au moment de la livraison.

2.4 Le module du temps réel

Le module du temps réel est fondamental pour l'articulation et le fonctionnement de tous les autres modules. De fait, c'est le module qui nécessite le plus d'attention pour le design (Boucher *et al.* [2006]). D'abord parce que c'est en temps réel que les échanges du bien se réalisent (non-stockabilité) et parce que la coordination entre les composantes de l'électricité (énergie, transport et réserves) est déterminante pour le fonctionnement du réseau.

Avec l'introduction de la concurrence, l'outil d'*optimal dispatch* devrait être remplacé par un module de marché qui assure la coordination entre les trois composantes de l'électricité (énergie, transport et réserves). Ce module ne peut pas fonctionner comme un pur marché car la vitesse nécessaire pour la coordination ne peut pas être assurée par des

procédés ordinaires de transaction entre l'offre et la demande. Le remplacement doit donc être fait par un mécanisme de contrôle qui garde encore quelques propriétés d'un marché ; c'est un « pseudo marché » (Boucher *et al.* [2006]). Ce pseudo marché, centralisé par le GRT, fonctionne comme une enchère dont le résultat est rapidement mis en œuvre par un système d'ordres (contrôle) pour ajuster la production des participants. De cette enchère on peut quand même déduire des signaux de prix. La manière dont ces signaux de prix sont déterminés est fondamentale car elle définit l'articulation du module du temps réel avec les autres modules de marché (*forward*).

Le module du temps réel est le dernier de la séquence, et il sera utilisé pour « corriger » le système (équilibre, gestion de congestions et réserves) par rapport à l'équilibre des échanges financiers (*forward*) venant des modules précédents (marchés *forward*). Bien que les signaux de prix donnés par ce module ne concernent que les « écarts » ou déviation des participants par rapport aux autres modules *forward*, ces signaux influencent tous les autres marchés.

Sans regarder de près tous les détails, la plupart des designs de module du temps réel consiste à remplacer l'outil d'optimisation d'*optimal dispatch* par une enchère de caractéristiques très similaires. En fait, cette enchère utilise pratiquement le même programme d'optimisation mais en remplaçant les entrées de l'outil d'*optimal dispatch* (données de coûts incrémentaux ou marginaux des générateurs, etc.) par des offres de prix des participants (EEI [2001]). Cette enchère est réalisée automatiquement à un intervalle régulier, et elle a comme objectif de ramener le système à un point de fonctionnement au « coût minimal » en respectant les contraintes (équilibre, congestions et réserves). Un point clé de ces enchères est la détermination des signaux de prix. En effet, c'est principalement sur la détermination des signaux de prix que vont se différencier la plupart des options de design pour ce module.

Si l'on considère seulement le problème de l'équilibrage sans les congestions, il existe différentes options pour la détermination de signaux de prix de la composante « énergie ». En regardant le programme d'optimisation de l'*optimal dispatch* on peut déduire que la manière la plus intuitive et économique pour dériver les prix de cette enchère est d'utiliser le multiplicateur de Lagrange de la contrainte d'équilibre (cf. Varian [1995], Newbery [2005a]). Cela correspond au design d'une enchère non-discriminante, car ce prix est appliqué à toutes les transactions du module du temps réel (Baldick *et al.* [2005]). Le prix de l'énergie en temps réel correspond à l'offre marginale, soit le coût marginal de production pour des conditions de concurrence parfaite. Cette version est connue sous l'appellation de « marché du temps réel ». Dans un « marché du temps réel », il existe un seul prix de l'énergie, et ce prix est fixé par enchère. D'autres formes d'enchères et de détermination des signaux de prix ont été mises en œuvre. On désignera toutes les autres formes d'organisation pour ce module sous le terme de « mécanismes d'ajustement ». Les

mécanismes d'ajustement ont généralement choisi des enchères discriminantes ainsi que des signaux de prix pénalisant certaines transactions. [98]

Si l'on considère les congestions, et les externalités engendrées par les contraintes de transport, il existe différentes options possibles pour la détermination de signaux de prix. Les externalités de congestion qui doivent être gérées en temps réel sont celles qui n'ont pas été traitées dans les étapes précédentes de la séquence des marchés ou celles qui varient à conséquence des conditions changeantes du système. Contrairement à la gestion des externalités dans le module *forward*, ici le temps disponible pour trouver un équilibre de marché est faible. Ainsi, le choix du design pour ce module est limité à la différenciation spatiale des prix de l'énergie. Alors, les principales différences du design viennent du degré d'internalisation des externalités et de la détermination de signaux de prix correspondant. Comme pour le cas du module *forward* de transport, les solutions possibles peuvent aller d'une gestion des externalités sans les internaliser à des solutions où on internalise plus ou moins ces externalités. Elles sont dérivées de l'analyse de l'outil d'*optimal dispatch*. On peut remplacer le modèle simplifié d'*optimal dispatch* par une enchère qui prenne en compte aussi les contraintes de capacité de transport pour ensuite dériver des signaux de prix différenciés par la localisation. Cette enchère pourrait être représentée par un programme d'optimisation très similaire à celui de l'*optimal dispatch* (voir ci-dessous)[99].

$$\min_{\underline{q}^{OD}} \sum_{i=1}^{I} C_i(q_i^{OD}) \quad \text{.....(OD1)}$$

Sous contraintes :

Contraintes d'équilibre :

$$\sum_{n=1}^{N} q_i^{OD} = D^{OD} \quad (\lambda) \quad \text{.....(OD2)}$$

Contraintes de transport (limites sur le réseau)

$$-F_l \leq \sum_{n=1}^{N} PTDF_{l,i,c}\ q_i^{OD} \leq F_l \ \forall l,c \quad (\mu_{l,c}) \quad \text{.....(OD3)}$$

Contraintes de moyens de production

$$Q_i^{\min} \leq q_i^{OD} \leq Q_i^{\max} \quad \forall i \quad \text{.....(OD5)}$$

L'ensemble d'équations OD3 nous dit que les flux sur chaque élément doivent être inférieurs ou égaux aux limites physiques de capacité pour tous les états du réseau (règle N-1). Comme pour le cas du module *forward* du transport, des signaux de prix pour valoriser l'utilisation des différents éléments du réseau pourraient d'être dérivées des multiplicateurs de Lagrange des contraintes de transport ($\mu_{l,c}$) combinés au multiplicateur de Lagrange de la contrainte d'équilibre (λ). Ces signaux de prix « énergie » localement différenciés

[98] La détermination des signaux de prix du temps réel pour les différentes options de design du module du temps réel sera traitée plus en détail dans le chapitre suivant (chapitre n° 3).

[99] Par simplicité nous ne représentons pas les contraintes de réserves primaires et secondaires (OD4). Elles peuvent être prises en compte en adaptant les capacités maximales et minimales de production.

devraient être envoyés par le module du temps réel peuvent être représenté par la formule suivante :

$$prix_i = \lambda - \sum_{l=1}^{L} \sum_{c=1}^{C} PTDF_{l,n,c}\, \mu_l$$

Dans le module du temps réel, il est possible aussi d'utiliser différents modèles économiques du réseau afin de choisir le degré d'internalisation des externalités. Le modèle nodal différencie un prix pour chaque nœud du réseau de transport. Cette option a été mise en œuvre au nord-est et au centre-ouest des Etats-Unis, par exemple dans les systèmes connus comme PJM, NY Pool, MidWest Pool et NE Pool.

Dans une conception complètement à l'opposé concernant l'internalisation des externalités, la solution la plus simplifiée consiste à traiter le système, dans un premier temps, comme si les contraintes de transport n'existaient pas. Alors, au moment de l'équilibrage, la définition des signaux de prix se réalise grâce à une enchère qui ne considère pas les contraintes de transport. C'est le modèle de timbre poste ou de zone unique. De cette façon le bien électricité n'est pas différencié géographiquement et les externalités ne sont pas internalisées. En fait, les externalités sont « cachées», gérées hors marché (Pignon [2003], Stoft [2002]). Cette approche a été mise en œuvre dans de nombreuses architectures étant donné sa simplicité, surtout dans des pays où les problèmes internes de congestions ne sont pas forts (c'est souvent le cas des pays européens). Cette option a été mise en œuvre par exemple à l'intérieur de la France, l'Allemagne, l'Angleterre, etc. Bien entendu, s'il y a des congestions, un autre mécanisme doit être utilisé après cet «équilibrage sans contraintes » afin de prendre en compte les contraintes du réseau et réaliser les modifications nécessaires de programmes. Concevoir un système comme s'il fonctionnait « sans » externalités a un coût (Oren [2001b]). Normalement ce coût est distribué entre tous les utilisateurs du réseau sans distinguer sa position géographique (timbre poste ou *uplift*).

Cette méthode simplifie énormément le fonctionnement du marché mais, en revanche, les externalités ne sont pas internalisées et les incitations données aux participants ne sont pas efficaces. Entre ces deux cas extrêmes on trouve les modèles zonaux (par exemple au Texas). De nombreuses solutions intermédiaires sont faisables. Elles consistent à définir des réseaux « économiques » hypothétiques utilisés pour la définition des signaux de prix. Comme nous l'avons annoncé, il existe une vraie difficulté à trancher entre toutes les solutions réalisables car un grand nombre de facteurs a une influence. D'un côté, un très haut degré d'internalisation d'externalités (modèle nodal) introduit une grande différenciation spatiale du bien électricité. Un nombre trop grand de biens différents peut amener des problèmes de manipulation de ces marchés (pouvoir de marché local), de coûts de transaction élevés et aussi de liquidité sur le marché de chaque produit (Smeers [2004b]). L'efficacité réelle de cette architecture dépendra entre d'autres, de la régulation et des règles

spécifiques pour éviter les problèmes de pouvoir de marché local.[100] D'un autre côté, dans les architectures où les externalités sont « cachées » ou qui n'internalisent pas toutes les externalités, les acteurs ne sont pas incités à diminuer le coût de la gestion des externalités et ils peuvent, au contraire, manipuler les règles afin d'augmenter leur revenu (*gaming*[101]) (Hogan [1998], Leuthold *et al.* [2005]). L'efficacité est déterminée aussi par la régulation du GRT, notamment par les incitations données par le régulateur au GRT, acteur clé pour la gestion des externalités (Oren [1998]).

En conclusion, les principales options de design du module du temps réel se distinguent selon deux axes : la détermination des signaux de prix (équilibrage) et le degré d'internalisation d'externalités de transport. Nous présentons ces options, ainsi que quelques exemples dans le tableau 2-6.

Tableau 2-6 : Options pour le module du temps réel.

Module du temps réel	
Détermination des signaux de prix d'« équilibrage »	• Marché du temps réel : PJM, NY-Pool, MidWest-Pool, Texas. • Mécanisme d'ajustement : Angleterre, France, Belgique.
Degré d'internalisation d'externalités	• Une seule zone (modèle timbre poste) : Angleterre, France, Belgique. • Quelques zones (modèle zonal) : Texas. • Chacun des nœuds (modèle nodal) : PJM, NY-Pool, MidWest-Pool, NE-Pool.

Nous avons parcouru sommairement les options de design pour chacun des modules. Mais une architecture de marché est composée à la fois de ces 4 modules et de la manière dont ces modules s'articulent entre eux. Nous présentons donc maintenant l'articulation entre les modules et l'arbitrage.

2.5 Articulation entre modules et arbitrage

Les 4 modules élémentaires d'une architecture de marché (module *forward* d'énergie, de transport et de réserves et le module du temps réel) sont liés car ils remplacent des fonctions essentielles qui étaient toutes gérées par des programmes centralisés d'optimisation. Les articulations nécessaires entre les modules se dérivent donc des outils de contrôles-commandes.

Dans les optimisations centralisées des outils de contrôles-commandes, les capacités des moyens de production (ressources rares) sont allouées à la production des composantes

[100] Toute architecture qui utilise un modèle nodal a incorporé dans son architecture des règles spécifiques pour limiter le pouvoir de marché local (Baldick *et al.* [2005])

[101] Un des problèmes typiques quand on n'internalise pas les externalités d'usage du réseau est le comportement opportuniste des acteurs (*gaming*), plus connu comme « DEC *game* ». Ce « jeu » consiste à provoquer une congestion et à se faire payer après pour la résoudre (cf. Stoft [1998], Oren [2001b], Oren [2001c] et Oren [2003b]).

de l'électricité (énergie, réserves et transport).[102] Ici, l'allocation optimale de la capacité limitée de production est déterminée en terme de coûts totaux pour produire toutes les composantes nécessaires. Le programme d'optimisation compare plusieurs utilisations alternatives d'une ressource capable de produire plusieurs composantes. Cette comparaison se fait en termes de valeur marginale et dépend du coût de production et aussi du coût d'opportunité (ou de substitution).[103] C'est en comparant les valeurs marginales des différentes utilisations des capacités de production qu'un programme d'optimisation trouve une allocation optimale de la production (Ma-Sun [2005]).

Dans un contexte d'ouverture à la concurrence, le design des articulations entre les modules doit permettre également une coordination telle que les ressources soient allouées pour produire d'une manière efficace les différentes composantes. Il existe deux options principales pour les articulations afin d'assurer cette coordination : la coordination dans une enchère centralisée ou par le biais de l'arbitrage direct entre marchés.

D'un côté, la coordination entre les modules peut encore garder la forme d'un programme d'optimisation, c'est-à-dire prendre la forme d'une enchère intégrant les contraintes et les caractéristiques de plusieurs modules et générant des informations économiques qui puissent s'interpréter comme des prix. Les architectures de marché dites « centralisées » ou « intégrées » (Wilson [2002]) utilisent des marchés organisés (enchères) et centralisés et intègrent plusieurs modules dans le programme d'optimisation qui définit l'enchère. Un exemple typique de l'intégration de plusieurs modules sur un même programme d'optimisation est l'intégration des modules *forward* d'énergie et transport (cf. section 2.2 et section 2.4).[104]

D'un autre côté, la coordination peut être assurée par l'arbitrage direct des acteurs entre les modules. L'arbitrage est particulièrement important dans les architectures de marché dites « décentralisées », où chaque module est défini comme un sous-marché (Wilson [2002]). Au sens strict, l'arbitrage est une opération financière apportant un gain certain et il résulte des imperfections momentanées du marché. Si la différence de « prix relatifs » entre modules ne correspond pas aux conditions économiques réelles, un arbitrage est possible. En théorie, à l'équilibre, il ne reste aucune opportunité d'arbitrage entre les marchés (« lois du prix unique », cf. chapitre 3). Le prix de chaque composante correspond à sa valeur marginale. Une comparaison de tels prix d'équilibre assurerait aux acteurs du marché une allocation optimale de leurs ressources.

---

[102] Il est clair que la capacité d'un moyen de production peut être utilisée pour produire de l'énergie ou des réserves. Mais elle peut aussi être utilisée pour créer un contre-flux sur une ligne de transport et produire ainsi de la capacité de transport.

[103] Le coût d'opportunité vient du fait qu'une ressource peut être utilisée pour produire plusieurs composantes.

[104] Un autre exemple est l'intégration entre les modules *forward* énergie et réserves reliés par la co-optimisation (Wilson [1998a], Baldick *et al.* [2005], cf. sous-section 2.3).

En pratique, l'arbitrage n'est pas seulement présent dans les architectures « décentralisés ». Le design réel des modules n'est jamais complètement « centralisé » et les modules sont rarement « intégrés » dans une enchère unique. Ainsi, l'arbitrage est toujours essentiel à la coordination. Trois types d'arbitrages peuvent être présents dans un marché électrique : l'arbitrage temporel, les arbitrages entre marchés d'un même bien, et l'arbitrage entre marchés correspondant à des biens différents (Shahidehpour *et al.* [2002]). L'arbitrage temporel correspond à un même bien qui est proposé dans deux marchés placés sur différents horizons temporels. C'est le cas, par exemple, de l'arbitrage possible entre les marchés d'énergie *day-ahead* et temps réel.[105] L'arbitrage entre marchés d'un même bien correspond, par exemple, à l'arbitrage entre le marché de gré à gré et un marché organisé ou bourse. L'arbitrage entre biens différents et combinés se voit entre le marché d'énergie et le marché de capacité de transport.

Le choix des articulations joue un rôle important pour le fonctionnent d'une architecture de marché. La réussite d'une architecture de marché « décentralisée », avec des modules de marchés peu organisés et centralisés (basés sur des contrats bilatéraux) et peu intégrés (séparation des modules), réside principalement sur la capacité des arbitrageurs à remplir leur tâche. Ici, les sous-marchés fonctionnent normalement en séquence temporelle et la rationalité des arbitrageurs jouera un rôle central sur l'efficacité de l'arbitrage (hypothèse d'anticipation rationnelle) (Smeers [2003], Ehrenmann-Smeers [2005], Ehrenmann-Smeers [2006]). Un arbitrageur qui n'arrive pas à prévoir correctement des différentiels de prix profitables ne réalisera pas la transaction. Les coûts de transactions peuvent aussi limiter les arbitrages. Un degré d'intégration important entre les modules diminue directement les coûts de transaction (les « transactions » sont en fait internes au programme d'optimisation).

En conclusion, il existe plusieurs options de design pour les articulations entre les modules d'une architecture de marché électrique. L'arbitrage est l'option par défaut d'articulation entre les modules et ne nécessite pas de design particulier. L'intégration des plusieurs modules dans une enchère unique (architectures « centralisées ») est la deuxième option et l'arbitrage est ici réalisé en interne dans le programme d'optimisation.

Nous venons d'étudier les options de design pour chacun des modules et de présenter leurs articulations. Le tableau 2-7 synthétise les résultats obtenus. Ces résultats constituent ainsi un cadre d'analyse modulaire qui est maintenant bien développé.

[105] Nous reviendrons sur ce type d'arbitrage dans le chapitre suivant.

Tableau 2-7 : Résumé des options de design des modules

Module *forward* d'énergie	Degré de centralisation	• Marché centralisé • Marché décentralisé
Module *forward* de transport	Intégration	• Intégré au module énergie *forward* • Séparé du module énergie *forward*
	Degré d'internalisation d'externalités	• Une seule zone • Quelques zones ou « *flowgates* » (zonal) • Chacun des nœuds (nodal)
Module *forward* des réserves	Organisation de la production	• Marchés organisés • Contrats bilatéraux producteurs/GRT
Module du temps réel	Détermination des signaux de prix	• Marché du temps réel • Mécanisme d'ajustement
	Degré d'internalisation d'externalités	• Une seule zone (timbre poste) • Quelques zones (zonal) • Chacun des nœuds (nodal)

Nous devons alors vérifier empiriquement, à partir de quelques cas, l'aptitude de notre cadre d'architecture modulaire à rendre compte de la variété existante des principaux designs de référence pour les marchés électriques.

2.6 Analyse empirique d'architectures de marché

Parmi les marchés électriques de référence, au niveau mondial, l'Angleterre, PJM et le Texas sont particulièrement importants comme cas de test de notre grille d'analyse modulaire. PJM, tout d'abord, qui est devenu, après la crise californienne de 2000-2001, le cas mille fois cité du modèle de design centralisé robuste, et aussi la référence empirique du *Standard Market Design* du régulateur fédéral FERC. Le Texas ensuite, qui reste, après la crise californienne, le seul cas de forte décentralisation de l'opération de la séquence des marchés. Au point que le cœur des marchés y demeure bilatéral, à l'exception du module du temps réel. Enfin l'Angleterre, où nous constaterons que l'on peut réussir à bâtir une architecture très décentralisée, fondée elle aussi sur des marchés bilatéraux et un module du temps réel, à partir d'une architecture très centralisée (comme l'était l'*Electricity Pool* de 1990 à 2001).

2.6.1 Le cas de l'Angleterre (Neta)

L'architecture de marché appliquée actuellement en Angleterre a été introduite en 2001 (Staropoli [2001], Newbery [2002], Perez [2002]). Elle est connue sous le nom de NETA (*New Electricity Trading Arrangement*) et elle a remplacé l'ancienne architecture, très centralisée, connu sous le nom d' « UK Pool » (Sweeting [2000], Staropoli [2001], Perez [2002]). A l'inverse, le NETA a une architecture de marché très décentralisée où les participants réalisent la plupart de leurs échanges de façon bilatérale.

Le design du NETA est centré sur le module du temps réel. Ce module fonctionne comme un mécanisme d'ajustement (*Balancing Mechanism*). C'est un mécanisme parce que

ce design s'écarte de notre définition de marché où il y a un seul prix de l'énergie du temps réel pour tous les participants (cf. Chapitre 3). Sur le mécanisme d'ajustement anglais, il y a au moins deux prix en temps réel pour les « écarts » ; c'est-à-dire ceux qui consomment ou produisent une quantité différente à leur position contractuelle *forward* transmise au gestionnaire du réseau (NGC)[106]. Ce système de double prix en temps réel (*dual-cash settlement*) est composé d'un prix pour acheter de l'énergie en temps réel (SBP-*System Buying Price*) quand la production réelle est inférieure à la quantité vendue sur les marchés *forward* (ou consommation réelle > achats sur les marchés *forward*), et d'un prix pour vendre de l'énergie en temps réel (SSP-*System Selling Price*)[107] quand la production réelle est supérieure à la quantité vendue sur les marchés *forward* (ou la consommation réelle < achats sur les marchés *forward*). Dans ce design, il n'y a pas de pénalités explicites. Mais la définition des prix des écarts s'établit sur la base des règles spécifiques qui prennent en compte des coûts additionnels. Les prix des écarts diffèrent du prix unique du marché et pénalisent les écarts (surtout les écarts négatifs) (Cf. Chapitre 3).

Les congestions sont aussi gérées par le biais du mécanisme d'ajustement (*Balancing Mechanism*). Mais les prix du temps réel ne prend pas en compte la localisation géographique de l'injection ou du soutirage (prix zone unique). Il n'y a donc pas d'internalisation des externalités en temps réel. Les coûts de gestion des congestions sont socialisés entre tous les participants par le biais d'un coût supplémentaire (*uplift*).

Pour le module énergie *forward*, nous trouvons une organisation presque complètement décentralisée. La plupart des transactions est réalisée de manière bilatérale. Il existe une bourse organisée (APX Power UK[108]) pour les contrats *forward* de court terme, mais le volume de transactions sur cette bourse est faible[109].

Sur le NETA il n'y a pas de marchés *forward* organisés ni pour le transport, ni pour les réserves. Dans le marché *forward* de l'énergie les participants considèrent que la capacité de transport du système est infinie. NGC se fournit en réserves par le biais de contrats bilatéraux avec les participants. Les coûts de ces contrats sont inclus soit dans le tarif du transport ou soit dans les prix des écarts en temps réel.

Une autre caractéristique spéciale sur le NETA s'exprime par les pouvoirs spéciaux donnés à NGC pour pouvoir agir comme agent de marché, vendre ou acheter de l'énergie pour améliorer le fonctionnement du réseau. Ceci permet de diminuer les coûts de congestions et d'équilibrage en raison des incitations donnés au GRT (Sweeting [2000]).

---

[106] *National Grid Company* (NGC) est le gestionnaire du réseau pour la zone de contrôle correspondant à l'Angleterre.
[107] Sur le NETA il y a aussi une différence entre les écarts qui sont dans le même sens (augmentent l'écart global) que celui du système et ceux qui sont dans le sens contraire (aident à diminuer l'écart global).
[108] Cf. www.apxgroup.com
[109] Au moment de l'introduction du NETA il y avait deux bourses privées mais le volume traité par ces deux marchés est très faible. Actuellement APX a fusionné les deux bourses mais son volume n'est toujours pas important.

En conclusion, l'architecture de marché actuelle de l'Angleterre se caractérise par son module du temps réel fonctionnant comme un mécanisme d'ajustement et sans internalisation des externalités de congestion, par sa forte décentralisation du module d'énergie *forward* et par son module de réserves *forward* fonctionnant sur la base des contrats bilatéraux entre le GRT et les participants.

2.6.2 Le cas du Texas

La deuxième architecture que nous présentons, celle du marché électrique texan, est elle aussi une architecture décentralisée. Mais elle se différencie du cas précédent par un certain degré d'internalisation des externalités et par plus d'organisation sur le module réserves (Yu *et al.* [2005]).

L'architecture du marché texan est elle aussi centrée sur le module du temps réel, géré par le gestionnaire du réseau (ERCOT). En temps réel, le design d'ERCOT fonctionne comme un marché du temps réel. Les écarts entre les programmes et les productions et consommations réelles sont valorisés avec un seul prix de l'énergie en temps réel (Baldick-Niu [2005]). Le module du temps réel considère aussi un certain degré d'internalisation d'externalités. ERCOT utilise un modèle économique de réseau « zonal » dans son algorithme d'optimisation du marché du temps réel. Ces zones sont déterminées sur la base des contraintes de transport les plus importantes du réseau (Baldick [2003a]).[110] Ce modèle économique de réseau simplifié permet de définir des coûts de congestion pour les éléments les plus congestionnées du réseau et faire payer aux utilisateurs la différence des prix entre les zones. Etant donné que ce modèle ne représente qu'une partie des contraintes de transport du réseau, un mécanisme supplémentaire pour résoudre des congestions locales a été conçu pour fonctionner juste après le marché du temps réel. Le coût de gestion de congestions locales est socialisé entre tous les participants de la zone (Potomac [2006]).

Sur le marché texan, il n'y a pas de marchés *forward* organisé, ni pour l'énergie ni pour le transport. Le marché *forward* d'énergie est complètement décentralisé. Il n'existe pas de bourse d'électricité organisée au Texas. En revanche, il y a un marché organisé des réserves *forward*, mais il n'est pas centralisé dans un sens strict (Baldick-Niu [2005]). Les consommateurs ont une obligation de réserves et ils peuvent choisir différentes formes de fourniture (programmer des réserves propres, des contrats bilatéraux, ou participer au marché organisé).

En conclusion, l'architecture de marché actuelle de Texas se caractérise par son module du temps réel fonctionnant comme un marché avec une internalisation zonale des externalités de congestion, par sa forte décentralisation du module d'énergie *forward* et par un module de réserves *forward* fonctionnant sur la base des marchés journaliers organisés ou enchères.

[110] Depuis le début de cette architecture, le nombre de zones a varié entre 4 et 5.

2.6.3 Le cas de PJM

L'architecture de PJM, gérant le système électrique de plusieurs états du nord-est des Etats-Unis, est plus centralisée que les deux autres présentées précédemment. Cette architecture est connue pour avoir une forte intégration entre les modules et un niveau important d'internalisation des externalités. Ce design est aussi très similaire à la définition du *Standard market design* recommandée par la FERC (Régulateur Fédéral Américain) comme le modèle à suivre (FERC [2002]). Ce type d'architecture a été déjà mise en place dans plusieurs régions des Etats-Unis depuis quelques années (par exemple NE Pool, NY Pool et MidWest Pool).

En temps réel, le design sur PJM est du type « marché ». Il existe un seul prix pour valoriser l'énergie en temps réel, pour payer les écarts entre les quantités achetées ou vendues sur les marchés *forward* et celles réellement consommées ou produites. Le marché du temps réel utilise aussi un modèle économique de réseau très complet et représentant le réseau physique afin d'internaliser, avec précision, les externalités du transport (modèle nodal).

Les modules *forward* d'énergie et de transport sont intégrés dans le marché *day-ahead*. Ce marché n'est pas obligatoire (les participants peuvent présenter directement des programmes de production et de consommation). Mais des règles particulières font qu'une part conséquente de la capacité de production passe par ce marché organisé. La majeure partie des participants ayant négocié des contrats bilatéraux présentent des programmes de production dans le marché *day-ahead* organisé pour se couvrir des coûts de transport du temps réel. Le degré d'internalisation d'externalités est maximum, étant donné qu'il y a une différentiation des prix à un niveau nodal, tant dans le marché *day-ahead* que dans le marché du temps réel. Tous les participants réalisant une transaction doivent payer la différence des prix du temps réel entre le nœud de soutirage et le nœud d'injection. Ceux qui participent dans le marché *day-ahead* paieront la différence de prix nodaux *day-ahead* si en temps réel ils respectent exactement leurs engagements pris le marché *day-ahead*.

Le module *forward* de réserves est en partie intégré au marché *forward* d'énergie (réserves opérationnelles ou *operating reserves*) et, en partie, organisé comme un marché *forward* séparé du marché *forward* d'énergie (réserves tournantes et réserves secondaires ou *spinning reserves and regulation*). Pour la partie organisée comme un marché séparé, les consommateurs ont une obligation de réserve et ils peuvent choisir différentes formes de fourniture de ces obligations (programmer des réserves propres, des contrats bilatéraux, ou participer au marché organisé). L'enchère du marché de réserves se réalise près du temps réel, et prend en compte les coûts d'opportunités de ne pas fournir l'énergie mise en réserve (PJM [2006]).

En conclusion, l'architecture de marché actuelle de PJM se caractérise par son module du temps réel fonctionnant comme un marché avec une internalisation nodale des externalités de congestion, par la forte centralisation du module d'énergie *forward*, par un

module de transport *forward* intégré au module d'énergie et basé sur un modèle nodal et par son module de réserves *forward* fonctionnant sur la base des marchés journaliers organisés ou enchères.

Le tableau 2-8 résume les choix des architectures étudiées pour chacun des modules.

Tableau 2-8 : Architectures de marché

Architecture	Module temps réel (énergie, transport et réserves)	Module énergie *forward*	Module transport *forward*	Module réserves *forward*
Angleterre/ NETA	Mécanisme/Zone unique	Décentralisé	-	Bilatéral GRT et producteurs
Texas	Marché/ Zonal	Décentralisé	-	Marché organisé / bilatéral
PJM	Marché/ Nodal	Centralisé	Intégré/Nodal	Marché organisé / bilatéral

L'analyse empirique des architectures de marché de référence a montré leur grande diversité, déjà suggérée par la notion même de modularité. Les marchés ont des architectures dont les combinaisons d'options sur la séquence entière ou sur un seul module peuvent être très différentes. Cette diversité est la conséquence de la difficulté à trouver une seule « architecture optimale » valable pour tout type de système électrique. D'une part, ce qui rend très complexe la détermination d'une architecture optimale est la grande variété des caractéristiques physiques des systèmes électriques. L'architecture de marché s'applique sur un système électrique en fonctionnement, difficilement modifiable de manière radicale à court terme. Les caractéristiques « économiques » réelles des systèmes électriques s'éloignent des hypothèses économiques de perfection (c'est-à-dire atomicité des acteurs, information parfaite, coûts de transaction nuls, absence d'externalités, etc.). D'une autre part, l'état de l'art actuel sur le design d'architectures de marchés électriques ne fournit pas une base théorique « robuste » pour pouvoir trancher définitivement entre les différentes solutions possibles.

Le manque de maîtrise rationnelle dans le choix des options optimales de design montrent l'importance d'une recherche plus approfondie sur les différents designs possibles. Or, étant donné la grande variété et diversité des designs possibles, une étude approfondie oblige à découper l'architecture en plusieurs sous-parties. Grâce au cadre d'analyse modulaire développé, une étude approfondie d'un seul module reste logiquement possible. Elle est même tout à fait pertinente si le sous-ensemble découpé dans la séquence des marchés y joue un rôle bien identifiable. Il faudra donc identifier le noyau commun à tous les designs qui est nécessaire pour remplir des fonctions « électriques » minimum indispensables dans toutes les architectures. Nous l'appellerons l'« architecture minimale ».

2.7 Architecture minimale : le module du temps réel

Une architecture de marché résulte de choix des autorités en charge de l'introduction de la concurrence. Pour toutes les options de design, les autorités ne peuvent pas se dispenser d'un groupe minimal de règles. Ces règles indispensables au fonctionnement de tout système de marché d'électricité constituent un noyau commun à toutes les architectures de marché. Cet ensemble de règles est l'architecture de marché minimale.

Afin de définir cette architecture minimale, nous identifions tout d'abord deux axes qui caractérisent une architecture de marché de manière générale. Ces deux axes sont le degré de centralisation et le degré d'internalisation des externalités de transport. Ensuite, nous constatons que l'architecture de marché minimale se caractérise par un module de temps réel toujours centralisé et devant gérer l'ensemble des externalités.

Le premier axe, le degré de centralisation, combine deux notions : l'organisation d'une part, et la centralisation proprement dite d'autre part. D'une part, l'organisation d'une architecture consiste à définir de façon explicite des règles pour encadrer les transactions dans chaque module. L'exemple typique de la définition de règles pour les transactions est l'utilisation d'enchères pour conduire les échanges. D'autre part, la centralisation correspond ici au fait que toutes les transactions sont obligées d'appliquer les règles d'échange ainsi définies. Le degré le moins élevé d'organisation et de centralisation consiste à laisser les agents faire librement des transactions bilatérales sans leur imposer de règles particulières d'échange. Ce sont des marchés de gré à gré. Pour certains modules cependant, l'organisation et la centralisation sont obligatoires. C'est le cas du module du temps réel. En revanche, pour les modules qui n'ont pas besoin d'être nécessairement organisés et centralisés, le choix d'une architecture s'effectuera entre les deux extrêmes de cet axe.

Le second axe, le degré d'internalisation d'externalités, exprime la prise en compte des externalités d'usage du réseau et de la différenciation spatiale du bien électricité. Plus les externalités sont prises en compte, plus le bien électricité est différencié. On trouve donc sur cet axe la finesse avec laquelle les externalités sont internalisées (définition nodale des prix, zonale, ou absence de différentiation des prix). Or, comme nous l'avons vu, un degré important d'internalisation des externalités de transport implique une définition détaillée de règles permettant la mesure de ces externalités et l'allocation des coûts engendrés aux agents responsables.

Si l'on regarde une architecture de marché du point de vue de l'effort des autorités pour la construire, on peut identifier une « architecture minimale » comme le noyau indispensable au fonctionnement d'une industrie électrique. Ce noyau commun minimal se trouve nécessairement à l'origine des deux axes (figure 2-7).

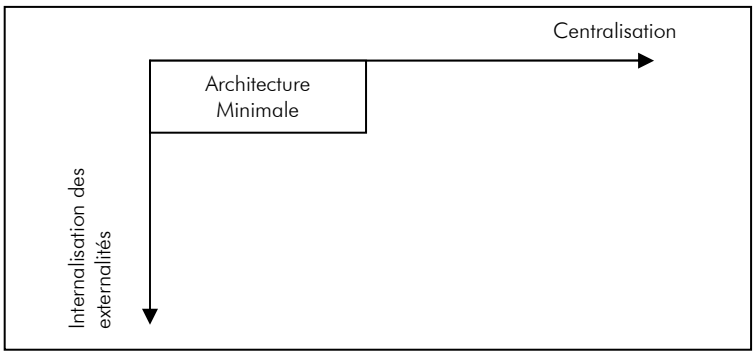

Figure 2-7 : Concept d'architecture minimale

L'architecture minimale ne demande qu'un degré minimal de création des règles pour la centralisation et l'internalisation des externalités. Ceci se voit clairement pour les modules *forward*, car il existe peu de contraintes de design étant donné le caractère « financier » de ces modules engageant peu les acteurs d'un point de vue « physique ». Par exemple, le module énergie *forward* n'a pas un besoin absolu d'être organisé ni centralisé. Les vendeurs (producteurs) et acheteurs (consommateurs) signent des contrats bilatéraux (*forward*) de fourniture de l'énergie. Les seules règles à définir par les autorités concernent la communication au GRT des échanges d'énergie prévus entre les participants. Le module transport *forward* n'a pas un besoin absolu de design « organisé » car il ne traite pas les externalités qui n'apparaissent qu'en temps réel. Ainsi, les acteurs du marché d'énergie *forward* peuvent agir comme si le système avait des capacités de transport assez grandes pour éviter toutes les congestions. Le module réserves *forward* peut être assuré par le GRT. Le GRT s'occupe de signer des contrats avec des fournisseurs du service de réserve.

En revanche, le concepteur d'une architecture de marché électrique n'a pas toutes ces libertés dans le choix des règles pour gérer le système en temps réel. Ainsi l'architecture minimale se centre sur le design du module du temps réel. Le module du temps réel, qui est centralisé par nature, gère d'une manière coordonnée l'énergie, le transport et les réserves. Il amène le système, à partir d'un équilibre *forward* qui n'a pas considéré certaines caractéristiques du réseau, à un équilibre en temps réel qui sera cohérent avec les conditions réelles du réseau physique. Ce module s'occupe obligatoirement de l'équilibre du système en temps réel et des congestions, même si celles-ci n'ont pas été internalisées.

En conclusion, le module du temps réel est la base minimale de toute architecture. Toutes les architectures de marché qui ont été mises en place ont dû fixer des règles concernant ce module. Cela vient de l'importance du module du temps réel pour un bien comme l'électricité qui est difficilement stockable et dont les échanges physiques ne prennent effectivement place qu'en temps réel. En conséquence, dans les chapitres suivants, nous étudierons en profondeur ce module qui est pour nous d'une importance toute particulière.

2.8 Conclusions de la section

Dans cette section, nous avons développé un cadre d'analyse modulaire des architectures de marché électrique sur une seule zone de contrôle. Une architecture de marché électrique est composée de quatre modules fondamentaux : 1° le module énergie *forward*, 2° le module transport *forward*, 3° le module de réserves *forward*, et 4° le module du temps réel. L'architecture de marché est construite par un choix d'options pour chaque module et pour les articulations entre ces modules.

Dans une première partie, nous avons présenté les principales options de design des quatre modules de l'architecture de marché. Les caractéristiques principales des différentes solutions de design des modules sont basées sur les outils d'optimisation des entreprises intégrées (*unit commitment* et *optimal dispatch*).

Nous avons défini des axes représentatifs et simplifiés pour décrire le choix des différentes options de design pour chacun des modules. Premièrement, les options de design pour le module d'énergie *forward* se différencient par leur degré de centralisation. Elles se repartissent entre deux pôles extrêmes : d'une part, un marché *forward* d'énergie complètement décentralisé basé sur des contrats bilatéraux et, d'autre part, un marché d'énergie *forward* basé sur une enchère centralisée qui définit les prix et quantités pour toutes les transactions d'énergie *forward*. Deuxièmement, les options de design du module de transport *forward* se placent sur deux axes. Le premier axe considère la séparation entre le module de transport *forward* et le module d'énergie *forward*, donnant lieu à deux options de design pour la valorisation de la composante « transport » : l'option « modules intégrés » qui valorise implicitement le transport par les biais des prix de l'énergie différenciés géographiquement ; l'option « modules séparés » qui valorise explicitement le transport considérant de droits d'utilisation de la capacité limitée de transport. Le deuxième axe considère le degré d'internalisation des externalités du transport sur l'étape *forward*, dont les options de design concernent la différenciation spatiale du bien électricité (modèle nodal, zonal, zone unique). Troisièmement, le module des réserves *forward* présente différentes options de design pour la forme d'organisation de la production des réserves. La production de réserves peut être organisée sur la base d'enchères ou par le biais des contrats entre le GRT et les fournisseurs. Enfin, les designs du module du temps réel se différencient par deux axes principaux : en premier lieu, celui de la définition des signaux de prix du temps réel, donnant lieu au design du type « marché » et « mécanisme » ; en deuxième lieu, celui du degré d'internalisation des externalités du transport en temps réel (définition nodale, zonale, zone unique). Grâce à cette classification, nous avons pu réaliser une analyse empirique de trois architectures de marché de référence : le cas de l'Angleterre (NETA), le cas du Texas et, le cas de PJM (correspondant à plusieurs états du nord-est des Etats-Unis).

Le développement du cadre d'analyse modulaire ainsi que son application à des architectures de référence, nous a permis de montrer deux points essentiels. Premièrement, il est empiriquement observable et théoriquement compréhensible que diverses combinaisons

sont possibles pour créer une architecture de marché. Nous avons pu vérifier cette grande diversité d'architectures de marché en examinant, module par module, les solutions de design proposées ou mises en œuvre. Les solutions choisies pour chacun de ces modules abordent d'une manière différente les caractéristiques spécifiques du bien électricité, mais aussi de chaque système électrique particulier. Deuxièmement, la complexité des architectures de marché, exprimée par la grande diversité, rend impossible une étude approfondie du *Market Design* sans le découpage du problème en sous-problèmes. Grâce à notre cadre d'analyse modulaire, nous avons vérifié que l'étude approfondie d'un seul module reste logiquement possible.

Nous avons alors choisi d'approfondir la recherche sur le design du module du temps réel. Deux arguments principaux ont guidé notre choix. D'un côté nous avons montré que, dans la diversité d'architectures observées, il existe toujours un noyau commun à tous les designs pour remplir des fonctions « électriques » indispensables. Si l'on regarde ces architectures de marché du point de vue de l'effort des autorités pour les construire, on identifie ce noyau indispensable et commun à toutes les architectures possibles, comme une architecture « minimale » des réformes. Cette « architecture minimale » porte sur le design du module « du temps réel ». D'un autre côté, l'étude analytique du design de chaque module a confirmé l'importance du module du temps réel dans une architecture de marché. Le module du temps réel est fondamental pour l'articulation de tous les autres modules parce que ce module gère la coordination en temps réel des trois composantes de l'électricité (énergie, transport et réserves) et garantit le fonctionnement du réseau. Le design de ce module est donc déterminant pour le fonctionnement d'un marché d'électricité. Les deux arguments qui nous ont guidés à approfondir la recherche sur le module du temps réel ne sont pas surprenants, car c'est en temps réel que les échanges du bien électricité, non-stockable, se réalisent entre vendeurs et acheteurs. Etant donné l'importance centrale du module du temps réel, les chapitres 3 et 4 seront consacrés à l'analyse et à la modélisation de ce module afin d'obtenir une compréhension plus profonde et de préciser l'influence des formes et des paramètres clé de son design.

Dans cette deuxième section, nous avons considéré une architecture de marché isolée, sur une seule zone de contrôle. Dans la section suivante nous étudierons l'intégration d'architectures de marché couvrant plusieurs zones de contrôle.

Section 3: Architectures de marché sur plusieurs zones de contrôle

Jusque là nous avons considéré des architectures de marché correspondant à des systèmes électriques isolés (zone de contrôle unique) gérés par un gestionnaire du réseau unique. Les règles de marché définies pour les différentes transactions et composantes de l'électricité (énergie, transport et réserves) étaient donc valables pour tous les participants du marché placés sur le réseau de transport. Mais, en réalité, les systèmes électriques, et les architectures de marché et institutions respectives, se trouvent rarement isolés. Il est donc important d'analyser les interdépendances physiques et organisationnelles avec les systèmes voisins, surtout lorsque les marchés sont liés économiquement. C'est le cas, par exemple, des pays de l'Europe Continentale ayant des architectures de marché différentes mais fonctionnant sur un réseau de transport interconnecté (Espagne, France, Allemagne, Belgique, Pays-Bas, etc.).

Cette section va donc se focaliser sur un cas plus complexe, mais aussi plus réaliste, celui de plusieurs architectures de marché sur un même réseau de transport. L'architecture de marché et les responsabilités de chaque GRT sont limitées à une zone géographique particulière, appelée zone de contrôle. Une zone de contrôle, normalement définie par les limites géographique d'un pays ou d'un état ou d'un groupe d'états, correspond seulement à une sous-partie du réseau de transport (cf. Chapitre 1, section 2). Dans ce contexte, nous nous intéresserons à la question du design de l'intégration des architectures de marché entre zones adjacentes. Cette question est déterminante pour un bon fonctionnement des marchés électriques régionaux et pour l'intégration géographique des marchés.

L'intégration géographique des marchés, ou l'élargissement de la taille d'un marché, a des bénéfices largement reconnus par les économistes et ingénieurs (Joskow et Schmalensee [1983], Hogan [2001], Harvey-Pope [2001], Hartshorn-Harvey [2001]). Pour le cas de l'électricité, ces bénéfices viennent, d'une part, de l'opération et, d'autre part, de l'amélioration des conditions de concurrence. Tout d'abord, d'un point de vue opérationnel, un fonctionnement coordonné de plusieurs systèmes électriques est bénéfique pour plusieurs raisons (cf. chapitre 1, section 3). Le fonctionnement coordonné permet d'utiliser plus efficacement les moyens de production et le réseau de transport, de profiter des différences de consommation et de production entre les différentes zones, de profiter des effets de foisonnement ou des économies d'aléa, etc. Les bénéfices sont en quelque sorte les mêmes que ceux qui justifiaient une coordination entre entreprises verticalement intégrées fonctionnant sur le même réseau de transport.

Par ailleurs, dans un contexte d'ouverture à la concurrence d'une industrie avec des structures de production encore très concentrées au sein des zones de contrôle, la stimulation de la concurrence par l'augmentation de l'envergure ou de la taille des marchés électriques est a priori très pertinente. En favorisant les échanges entre les zones de contrôle

existantes, la mise en place de marchés régionaux peut assurer les conditions d'une concurrence plus effective entre les producteurs d'électricité et éviter les risques d'abus de pouvoir de marché. Ce point est particulièrement important pour le cas de l'électricité étant donné que les conditions de marché ne sont pas toujours parfaites. L'objectif attendu est que ces marchés électriques régionaux se traduisent par une pression à la baisse des prix de l'électricité (Boucher-Smeers [2002]). C'est la voie poursuivie en Europe avec le marché intérieur européen (Finon [2001], Belmans [2004], Glachant *et al.* [2005], CE [1996], CE [2003], CE [2004], CE [2005]), et aux Etats-Unis avec les organisations régionales de transport (*Regional Transmission Organisation*, RTO) (FERC [1999], Bailey-Eaton [2001], Laughlin [2006]).

Notre objectif est alors de comprendre comment les différentes architectures peuvent s'intégrer et, plus particulièrement, comment leurs modules peuvent s'articuler ou s'intégrer. Pour étudier la question de l'intégration des architectures nous procéderons comme suit. Dans la sous-section 3.1, nous présenterons d'abord l'intégration d'architectures de marché, en utilisant l'analyse modulaire proposée dans la section 2. Puis, nous aborderons le concept d'intégration idéale ou parfaite, et à la suite les différentes modalités d'une intégration qui se rapproche de l'idéal. Nous verrons qu'il existe deux formes principales d'intégration : d'une part, l'intégration par expansion de la zone de contrôle, où tous les modules des architectures sont intégrés et harmonisés, et, d'autre part, l'intégration par coordination de modules où seuls certains modules sont intégrés. Enfin, dans la sous-section 3.2, nous étudierons plus en détail l'intégration par coordination, qui est la plus vraisemblable dans le contexte de l'Europe. Nous verrons que le caractère « modulaire » de cette forme d'intégration d'architectures présente d'un côté l'avantage de permettre une mise en œuvre « graduelle », mais pose alors, d'un autre côté, une question sur les modalités d'harmonisation du design des modules qui seront intégrés.

3.1 L'intégration d'architectures de marché de plusieurs zones de contrôle

L'architecture de marché et les responsabilités du GRT sont déterminées pour une zone géographique délimitée appelée zone de contrôle. Bien qu'un réseau de transport synchronisé corresponde logiquement à une seule énorme machine, sa gestion quotidienne résulte d'un partage des tâches entre plusieurs gestionnaires du réseau (GRT) qui sont autant d'organisations juridiquement indépendantes les unes des autres. La taille de chacune des zones de contrôle, et la taille des GRT qui leur sont associés, est donc généralement liée à la géographie des pays et couvre rarement toute l'étendue physique du réseau de transport.

La figure 2-8 montre un exemple de cette configuration. Sur un réseau de transport, deux zones de contrôle sont définies (zone A et zone B). Il y a une architecture de marché définie pour chacune des zones de contrôle, et la responsabilité sur les sous-réseaux est assignée à différents gestionnaires du réseau (GRT A et GRT B).

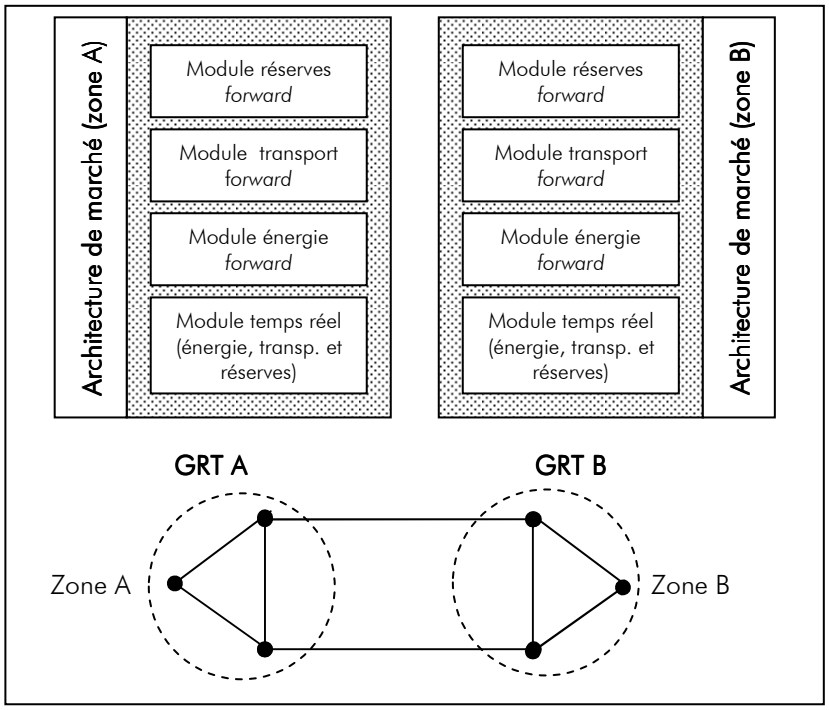

Figure 2-8 : Architectures de marché sur plusieurs zones de contrôle

Dans chacune des zones de contrôle, les transactions entre les agents économiques se réalisent en accord avec l'architecture de marché respective. Quand l'extension des marchés électriques se met en place, elle favorise les échanges entre les zones de contrôle existantes. L'ouverture des transactions entre zones permet un arbitrage entre agents économiques qui valorisent l'électricité et ses composantes de manière différente. Mais, comme nous l'avons vu, les transactions d'électricité doivent obligatoirement parcourir toute la séquence de l'architecture de marché, pour prendre en compte les caractéristiques inhérentes à ce bien complexe. C'est pourquoi la création de marchés régionaux d'électricité se concrétise elle-aussi par l'intégration de leurs architectures de marché.

Notre objectif est alors d'étudier comment les architectures de marché pourraient s'intégrer pour étendre l'envergure du marché régional à sa taille maximale. Pour faciliter la compréhension, nous présentons d'abord le concept d'intégration parfaite des architectures qui nous servira de cas idéal. Puis, nous présentons les formes pratiques qui cherchent à s'approcher de cet idéal d'intégration parfaite des architectures.

3.1.1 L'intégration parfaite d'architectures

Le concept d'intégration parfaite correspond au cas hypothétique où, pour un réseau de transport, il existe une seule architecture, valable pour tous les participants, et où la responsabilité du contrôle du réseau est exercée par un seul gestionnaire du réseau. Ceci vaut pour les 4 modules de l'architecture de marché : énergie *forward*, transport *forward*, réserves *forward* et le module du temps réel (voir figure 2-9) et représente le niveau maximal d'intégration : l'architecture de marché et le GRT sont uniques sur tout le réseau de transport.

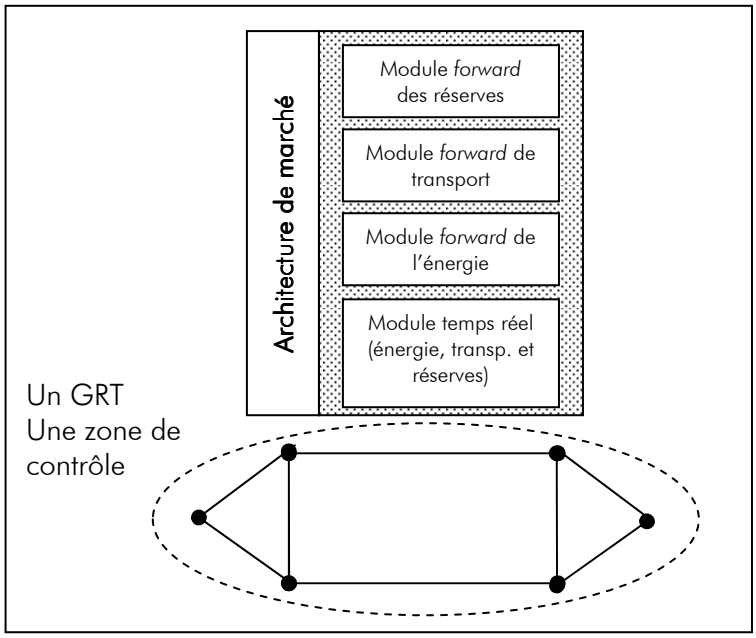

Figure 2-9 : Intégration parfaite d'architectures de marché

Ce concept d'intégration parfaite pourrait être assimilé au concept de coordination parfaite des entreprises intégrées fonctionnant sur un même réseau de transport (Chapitre 1, section 3). Une intégration parfaite d'architectures proviendrait du remplacement des séquences de contrôle (*unit commitment*, *optimal dispatch*, etc.) des entreprises intégrées parfaitement coordonnées. Dans le cas des entreprises intégrées, la coordination est réalisée à un seul niveau (entreprise-entreprise). Dans le cas d'une organisation basée sur la concurrence, la coordination doit se réaliser à deux niveaux : l'intégration des architectures et la coordination des GRTs.

L'intégration parfaite d'architectures est le niveau maximal d'intégration et elle permet de bénéficier au maximum de l'élargissement de la zone d'échange. Ce niveau d'intégration permet de profiter de tous les bénéfices de l'intégration des marchés pour un réseau de transport donné. Grâce à l'intégration, les échanges entre zones sont simplifiés ce qui permet une utilisation plus efficace des moyens de production, une amélioration dans la gestion des congestions, un plus grand partage de réserves, une amélioration de la sécurité du réseau, le bénéfice des effets de foisonnement ou des économies d'aléa, l'amélioration des conditions de concurrence (par élargissement du marché pertinent ou *relevant market*), etc. (Harvey-Pope [2001]). En conséquence, les efforts des autorités devraient être orientés vers la création d'institutions et de structures organisationnelles nécessaires à une intégration parfaite, ou vers une forme d'organisation qui fasse fonctionner le système (architectures de marché et GRTs) comme s'il y avait une intégration parfaite.

Dans un monde idéal, sans contraintes institutionnelles et sans coûts de mise en œuvre, l'intégration parfaite est optimale. Cependant, dans le monde réel, arriver à une intégration parfaite d'architectures, « sans frontières », peut présenter une série d'obstacles ou de barrières institutionnelles. De plus, les coûts de mise en œuvre et de fonctionnement

des mécanismes de coordination peuvent être importants. Ces obstacles, bien réels, obligent à considérer des solutions de deuxième rang qui s'écartent plus ou moins de l'intégration parfaite.

Les obstacles peuvent venir du côté institutionnel ou du côté opérationnel. Du côté institutionnel, les obstacles peuvent venir des résistances aux réformes qui sont nécessaires pour une plus grande intégration de marchés (Finon [2001], Pignon [2003], Glachant *et al.* [2005]). Par ailleurs, l'absence d'institutions « centrales » (régulateur ou autorité public) capables de pousser l'intégration dans des circonstances où tous les participants ne sont pas gagnants ; le manque d'incitations cohérentes des GRTs ; etc. Du côté opérationnel, les obstacles viennent de problèmes de coordination entre zones, entre GRT et entre les architectures de marché respectives, qui sont créés par l'interdépendance des sous-réseaux de transport aux frontières entre zones de contrôles. Par exemple le manque d'harmonisation entre architectures de marché complique davantage l'intégration. Et nous avons déjà vu dans la section précédente (section 2) la grande diversité d'architectures. De plus, l'intégration parfaite d'architectures pour de grands réseaux de transport peut aussi entraîner des coûts et des risques supplémentaires. Les réseaux de transport peuvent avoir des tailles considérables, et alors une conception d'architecture de marché unique et de GRT unique pour tout le réseau de transport peut avoir un coût non-négligeable ou peut être techniquement irréalisable.

On peut donc poser la question du degré optimal et de la forme optimale d'intégration. On devrait comparer les coûts et les bénéfices pour les solutions de deuxième rang possibles d'un point de vue institutionnel. Ce domaine devrait se focaliser sur des études détaillées de tous les bénéfices et coûts pour les différentes alternatives d'organisation pour l'intégration (Costello [2001]).[111] Cependant, bien qu'intéressante et nécessaire, une telle analyse n'est pas l'objet de cette recherche, même si celle-ci nous apporte aussi quelques éléments pour pouvoir réaliser ce type d'analyse coût-bénéfice. Par simplicité, nous considérerons l'intégration parfaite d'architectures comme l'objectif final, et nous cherchons à nous approcher au maximum de cet idéal. Pour ce faire, nous nous concentrons sur la question de l'intégration des 4 modules des architectures et sur la coordination des GRT des différentes zones. Tout écart entre une situation réelle d'intégration partielle et l'idéal d'intégration parfaite donne lieu à des « effets de frontière ». On désigne comme « effets de frontières »[112] les différentes inconsistances ou manque de coordination entre les modules des architectures de marché ou entre les GRT (Costello [2001], Joskow [2006d]).

[111] Mais, quels que soient les résultats futurs de ces recherches, on ne peut pas croire que la taille actuelle des zones de contrôles existantes (déjà définies par pays ou régions) soit déjà optimale, ni que les effets de frontière entre elles soient minimisés (Glachant *et al.* [2005]).
[112] En Anglais : *seams issues*.

3.1.2 Mise en œuvre d'intégration d'architectures

Il existe différentes manières pour avancer vers l'intégration parfaite d'architectures (Pignon [2003]). La première forme, la plus intuitive, correspond à l'expansion de l'architecture et de la responsabilité des institutions (GRT) sur une zone plus grande (ou expansion de la zone de contrôle). C'est l'intégration par expansion. Cette forme d'intégration élimine complètement les effets de frontière. En revanche, elle peut présenter des difficultés de mise en œuvre ou d'acceptation. La deuxième forme, l'intégration par coordination, cherche à mettre en place des mécanismes de coordination entre les architectures afin de simuler un fonctionnement intégré mais sans modifier les institutions préexistantes et les différentes zones de contrôle. Cette forme d'intégration ne garantit pas l'élimination des effets de frontière mais sa mise en œuvre présente moins d'obstacles institutionnels que l'intégration par expansion. Nous présentons maintenant ces deux formes d'intégration.

3.1.2.1 Intégration par expansion de la zone de contrôle

Si l'on cherche à réduire au maximum les effets de frontière, la forme d'intégration la plus adéquate est l'expansion de la zone de contrôle. Cette forme permettrait de s'approcher de l'intégration parfaite car la nouvelle zone agrandie posséderait une architecture complètement harmonisée (sur chacun des 4 modules) et la coordination des GRTs serait assurée par un GRT unique sur cette nouvelle zone. Dans la figure 2-10, nous montrons un exemple d'intégration par expansion de la zone de contrôle. L'architecture de la zone A, avec chacun de ses quatre modules, est étendue à la zone B. De plus un seul opérateur du système (ici le GRT de la zone A) prend le contrôle de la nouvelle zone de contrôle (zone A plus zone B).

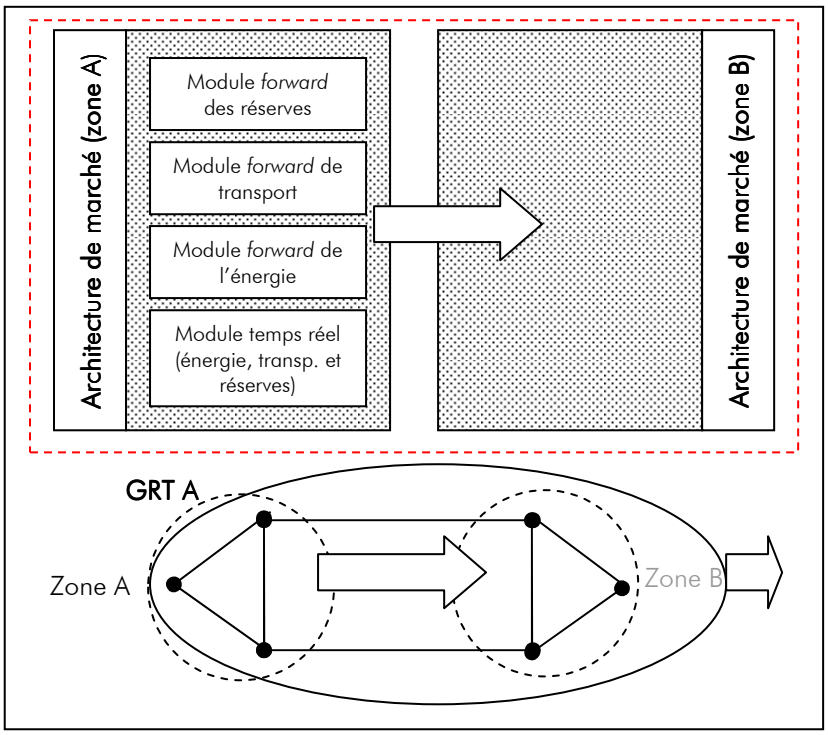

Figure 2-10 : Intégration d'architectures de marché par expansion de la zone de contrôle

L'idéal pour éliminer tous les effets de frontière serait de modifier les zones de contrôle et d'adapter la taille de la zone de contrôle à la taille du réseau de transport. La centralisation de la gestion du réseau de transport (ou d'une sous-partie importante) et de l'architecture de marché respective permettrait de minimiser les problèmes de coordination, de sorte que les effets de frontière soient réduits ou éliminés.

Aux Etats-Unis, ces notions d'expansion de zone de contrôle et d'internalisation des effets de frontières ont incité le régulateur fédéral, la FERC, à pousser la création de « grands » gestionnaires régionaux de réseaux, les RTOs (*Regional Transmission Organisations*) (FERC [1999], Laughlin [2006]). Un exemple d'intégration par expansion de la zone de contrôle est celui de PJM.[113] La zone de contrôle actuelle de PJM résulte d'une expansion géographique importante vers des nouveaux états ou zones (Ott [2002], Krapels-Flemming [2005]). La taille définie par la consommation maximale a triplé en moins de 10 ans. Cette zone correspond à une demande de pointe de 130 GW, soit l'équivalent de la demande de pointe de la France et de l'Allemagne.

L'expansion de la zone de contrôle pour des grands réseaux de transport peut aussi entraîner des coûts supplémentaires et des problèmes de faisabilité technique (Stavros [2000]). Par exemple, un GRT peut présenter des problèmes de sécurité s'il est trop grand (capacité de gestion de l'information et de calcul limitée, perte de fiabilité par contrôle unique centralisé, service de démarrage « *blackstart* ») (MPSC *et al.* [2001]). Or, l'exemple donné précédemment sur l'expansion de la zone de PJM sert d'indice sur la faisabilité technique de zones de grande taille.[114] Outre la faisabilité technique, cette forme d'intégration se voit aussi malheureusement limitée par de nombreuses contraintes organisationnelles et institutionnelles. Il est clair que cette forme d'intégration nécessite des réformes institutionnelles assez profondes. Une autre forme d'intégration, qui présente moins d'obstacles d'un point de vue institutionnel, est basée alors sur la coordination des architectures de marché.

3.1.2.2 L'intégration par coordination des modules

Une autre manière de s'approcher d'une intégration parfaite est de recourir à des mécanismes de coordination entre les architectures et entre les GRTs. C'est une intégration par coordination. Cette coordination devrait se mettre en place au niveau de l'architecture, mais aussi au niveau des GRTs afin d'approcher un fonctionnement coordonné où les frontières entre les zones disparaissent.

---

[113] Il est important de remarquer que l'intégration n'est pas complètement réalisée. Les modules de réserves *forward* sont encore gérés individuellement par zones (PJM [2006]).

[114] Des études plus empiriques de zones de contrôle de taille importante ont été réalisées. Ces études montrent que la faisabilité technique ne pose pas de problèmes pour des zones de contrôle plus grandes que celles existantes. Par exemple Shanker [2001] a montré, qu'un GRT unique est techniquement possible pour une vaste zone, équivalente à une partie importante de l'Europe continentale.

Sur la figure 2-11, le réseau de transport est divisé à la base en deux zones de contrôle. Pour chacune des zones, il existe une architecture de marché différente, et les GRTs fonctionnent aussi d'une manière indépendante. Une intégration par coordination consisterait à mettre en place un mécanisme de coordination pour chacun des quatre modules des architectures plus une coordination étroite entre les GRTs.

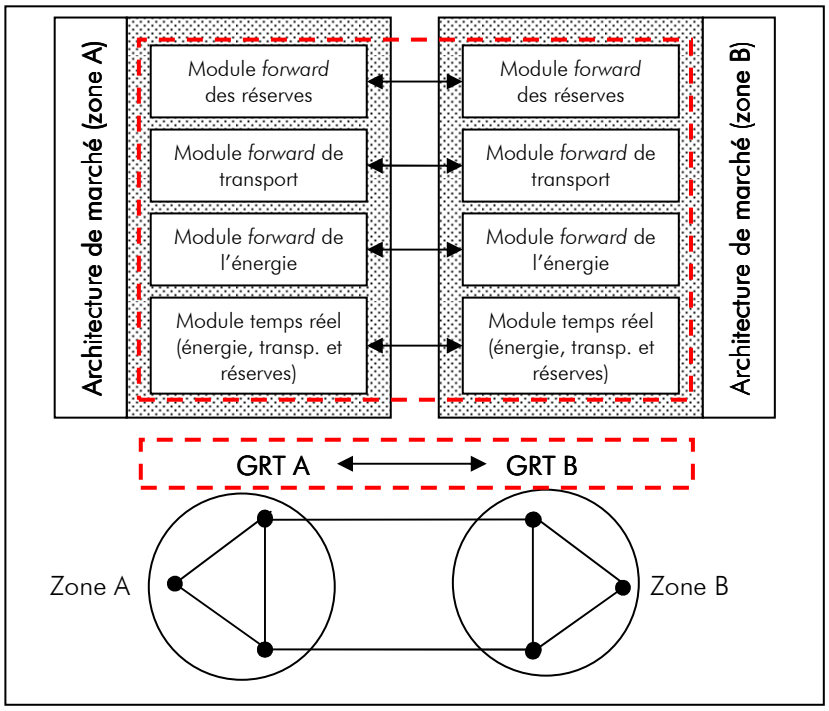

Figure 2-11 : Intégration d'architectures de marché par coordination

Les mécanismes de coordination cherchent alors à simuler un fonctionnement parfaitement intégré des architectures (« intégration parfaite »). Ces mécanismes de coordination pourraient être basés sur la coordination des séquences de contrôles-commandes des entreprises verticalement intégrées fonctionnant sur un seul et même réseau (chapitre n°1, section 3). Car, comme nous avons vu dans la section 1, les modules élémentaires des architectures de marché sont issus de ces outils de contrôle-commande. Dans un environnement du type concurrentiel, l'intégration passerait par la coordination des quatre modules des architectures et des GRT, rattachés d'une manière plus ou moins importante à la commande de ces modules. Ces mécanismes de coordination peuvent être mis en place par le biais d'une combinaison de modules intermédiaires, ou « passerelles », entre les architectures et des protocoles de coordination entre les GRTs (Benintendi-Boccard [2003]). Les modules intermédiaires et les protocoles de coordination devraient fonctionner de manière à simuler une architecture sans frontières et un GRT unique (Harvey [2003]).

Le design et la mise en place de tels mécanismes de coordination, comme le développement des modules intermédiaires et de protocoles de coordination, peuvent se heurter à certaines difficultés. La définition initiale des architectures de marché peut être très différente et incompatible entre les zones (cf. section 2). Ceci pourrait marquer davantage

les « frontières » en compliquant l'intégration des marchés. L'absence d'une institution centrale (par exemple un régulateur central) avec pouvoir sur tout le réseau, qui pousserait l'intégration d'architectures et la coordination entre les GRT, peut compliquer davantage l'intégration (Pignon [2003]). Dès lors, des incompatibilités entre les architectures et le manque de coordination entre les GRTs empêchent une intégration complète des architectures et entretiennent les effets de frontière.

Contrairement à l'intégration par expansion, l'intégration d'architectures par coordination présente la possibilité d'être réalisée module par module ou graduellement. Ceci permet, d'une part, d'avoir une plus grande acceptation à la mise en place de l'intégration, car les reformes sont plus graduelles ; et d'autre part, d'intégrer d'abord les modules dont les bénéfices apportés sont supérieurs aux coûts engendrés ou ceux qui présentent des contraintes institutionnelles moins importantes (Newbery [2005b], Glachant-Lévêque [2005]). Dans la section suivante, nous analyserons l'intégration par coordination.

3.2 L'intégration d'architectures par coordination des modules

L'intégration par coordination des modules entre zones nécessite la mise en place de modules intermédiaires ou de « passerelles » entre les modules afin de les coordonner (figure 2-12).

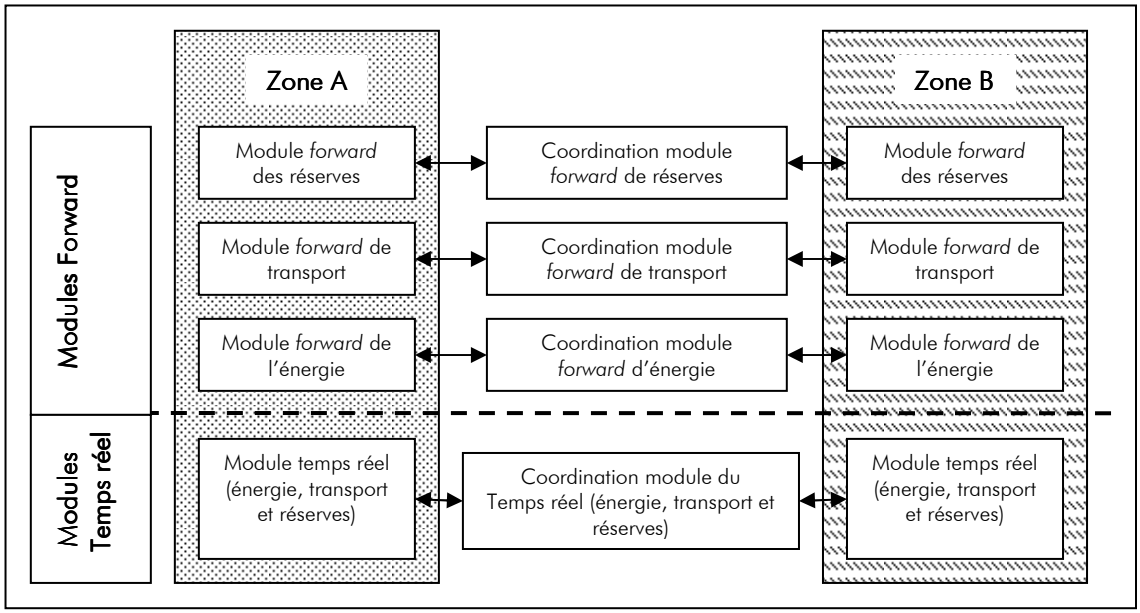

Figure 2-12 : Schéma de l'intégration de modules par coordination

Une intégration complète entre architectures devrait traiter la coordination de tous les modules plus la coordination des GRTs. En pratique, la coordination peut se réaliser de manière graduelle, c'est-à-dire que seulement certains modules peuvent être coordonnés. Le caractère graduel ou modulaire de cette forme d'intégration d'architectures présente l'avantage, par rapport à l'intégration par expansion, d'être moins sensible aux contraintes institutionnelles (Pignon [2003], Meeus *et al.* [2005]). Cependant, la mise en œuvre de

l'intégration d'architectures par coordination présente des enjeux importants du point de vue du design des modules intermédiaires, notamment lorsque les designs des modules à coordonner ne sont pas complètement harmonisés. Les architectures de marché de zones adjacentes peuvent être assez différentes voire incompatibles et, en conséquence, leur intégration par coordination de modules peut laisser la place à des effets de frontière résiduels inévitables.

De la même manière que nous avons étudié la coordination d'entreprises intégrées, en analysant la coordination de leur séquence de contrôle : *unit commitment*, *optimal dispatch*, contrôles automatiques (cf. Chapitre 1, section 3), nous utilisons l'analyse modulaire des architectures de marché par zone pour étudier les différentes formes et degrés pour la coordination des différents modules. En effet, coordonner les modules *forward* des architectures correspondrait à coordonner les outils d'*unit commitment* et coordonner les modules du temps réel correspondrait à coordonner les outils d'*optimal dispatch*. Nous analyserons d'abord la coordination des modules *forward* et, ensuite, celle des modules du temps réel.

3.2.1 Coordination des modules *forward*

Les efforts d'intégration des architectures des marchés régionaux se sont généralement focalisés sur les modules d'énergie et transport *forward*. Nous nous concentrons donc sur la coordination de ces deux modules.[115]

Si on considère d'abord le cas hypothétique d'un réseau sans limites de capacité de transport, le module intermédiaire coordonnant les modules *forward* d'énergie consisterait en des règles et protocoles pour faciliter l'accès des participants de chaque zone au marché *forward* d'énergie de l'autre zone. Sans un module intermédiaire correctement conçu, les difficultés rencontrées par les acteurs du marché pour réaliser une transaction inter-zone pourraient provoquer des effets de frontière et ainsi limiter certaines transactions et arbitrages bénéfiques. Par exemple, pour le cas le plus simple comprenant deux zones, A et B, sans coordination entre les modules, un acteur du marché A voulant réaliser une transaction interzone devra agir dans le module « énergie *forward* » de la zone A puis avec le GRT de cette zone pour ensuite traiter avec le module énergie *forward* et le GRT de la zone B. Les problèmes de frontière proviennent du manque d'harmonisation complète entre les modules à coordonner ; par exemple : le manque d'harmonisation entre les périodes de cotation des bourses et les guichets de programmation des transactions ; les coûts d'accès aux marchés et aux réseaux, etc. (ETSO [2004a], ETSO [2004b], Meeus *et al.* [2005]). La

[115] La coordination des modules de réserves *forward* dans les réseaux interconnectés s'est principalement développée pour certains types de réserves (primaires et secondaires). La réussite de la coordination des modules de réserves dépend fortement de la coordination des modules en temps réel. Dans le cadre de l'intégration des modules du temps réel, une coordination plus approfondie des modules de réserves *forward* est en train d'être évaluée en Europe (ETSO [2007]). Elle sera sans doute bénéfique car le niveau de réserves nécessaire, pour un niveau de sécurité déterminé, diminue lorsque la taille de la zone croît (Joskow-Schmalensee [1983], ETSO [2007]).

conception d'un module intermédiaire entre les modules *forward* d'énergie devrait donc se concentrer sur la simplification de l'accès entre les architectures et sur l'harmonisation des transactions interzones pour résoudre les problèmes de coordination.

Si on considère le cas plus réaliste où il existe des capacités limitées de transport entre les zones, la coordination entre modules *forward* de transport doit être aussi considérée. Alors, la coordination entre les modules *forward* d'énergie et de transport nécessite la mise en place de modules intermédiaires de coordination. Ces modules intermédiaires de coordination entre les modules « *forward* » devrait permettre l'arbitrage spatial entre les modules d'énergie *forward* tout en considérant les capacités de transport limitées du réseau pour passer d'une zone de contrôle à l'autre.

Le module intermédiaire de transport *forward* a un rôle fondamental quand le design des modules *forward* de transport des architectures à intégrer n'est pas « complet » dans le sens d'internalisation des externalités (Hirschhausen-Zachmann [2005], cf. section 2.2).[116] C'est le cas, par exemple, de l'Europe Continentale où les architectures de marché nationales n'internalisent pas les externalités de congestions du réseau[117] dans le cadre de leurs frontières. Dans ces cas, le module intermédiaire de transport nécessite la définition d'un modèle économique du réseau[118] inter-zone. Généralement, dans la littérature, ce module intermédiaire est désigné par les termes « *cross border congestion management* » ou « *international congestion management* » (Smeers [2003]). On cherche alors à internaliser les effets de capacités de transport limitée entre les zones en considérant que les transactions intra-zones n'impactent pas ces capacités de transport.[119] Ainsi, le modèle économique de réseau est généralement simplifié. Il est défini par la capacité de transport entre les zones, comme un bien hypothétique ou « conventionnel », et par la manière dont ces capacités de transport interzones sont utilisées par les différentes transactions (Boucher-Smeers [2002]).

Indépendamment du modèle économique du réseau choisi, l'allocation des capacités d'interconnexion entre les zones (les biens « conventionnels ») peut être mise en place de différentes manières. On peut d'abord distinguer des méthodes sans fondements économiques (mécanismes administrés : règles de priorité ou prorata par exemple) et des

---

[116] Nous utilisons le terme « complet » pour faire référence à une architecture où les externalités d'utilisation du réseau sont internalisées et où l'incomplétude de marché est « réparée ». C'est le cas, par exemple, des architectures avec une définition nodale des prix.

[117] La plupart des architectures de marché de l'Europe utilisent le « timbre poste » comme méthode de tarification des coûts de congestion. Les externalités ne sont pas internalisées. Les congestions sont gérées par divers mécanismes et les coûts sont distribués entre tous les utilisateurs du réseau.

[118] Un modèle économique du réseau est un modèle du réseau hypothétique défini pour internaliser les externalités de congestions par des méthodes économiques (enchères implicites, enchères explicites etc.). Ces modèles peuvent s'écarter de la réalité physique du réseau en choisissant le niveau des externalités qu'on veut internaliser (cf. section 2.2).

[119] Sur les réseaux maillés, les deux types de transactions, intra-zones et inter-zones, peuvent impacter les flux sur les interconnexions. La distinction des effets provoqués par des transactions intra-zones de ceux provoqués par des transactions inter-zone n'est pas fondée sur les réseaux maillés (Boucher-Smeers [2002]).

méthodes à fondements économiques (Benitez-Crampes [2003]). Parmi les méthodes à fondements économiques, nous trouvons d'abord celles où les modules *forward* d'énergie sont séparés du module intermédiaire de transport *forward* : ce sont les enchères explicites non-coordonnées ou « *Contract Path* » et les enchères explicites coordonnées (ETSO [2001]). Nous trouvons aussi les méthodes qui intègrent les modules d'énergie *forward* avec les modules intermédiaires de transport *forward* (enchères implicites[120]) : *Market Coupling* et *Market Splitting*.

La mise en œuvre de chacun de ces mécanismes de coordination implique différents degrés i) d'harmonisation entre les designs des modules à coordonner et ii) de coordination entre les GRTs. Parallèlement, chacun de ces mécanismes utilise différents degrés de simplification de représentation du réseau, et exige une certaine capacité d'arbitrage de la part des participants. Tous ces aspects contribuent à l'efficacité de chaque mécanisme (Ehrenmann-Smeers [2005]). Pour illustrer ces points, nous analysons les formes de coordination entre modules mises en œuvre, particulièrement en Europe. Nous présentons d'abord le cas d'intégration le plus simplifié, les enchères explicites non-coordonnées ou « *Contract Path* », puis deux méthodes où la coordination est plus approfondie : le cas du « *Market Coupling* », qui relie la France, la Belgique et les Pays Bas, et le cas de « *Market Splitting* » appliqué en Scandinavie (NordPool).

L'intégration des modules *forward* par des enchères explicites non-coordonées ou « *Contract Path* » est très appliquée en Europe Continentale (Hirschhausen-Zachmann [2005], ETSO [2006d]).[121] Ici, les modules d'énergie *forward* des zones adjacentes sont coordonnés par un module intermédiaire de transport entre les zones. Des droits de propriété pour l'usage de la capacité de transport entre les zones sont définis et mis aux enchères par le(s) GRT(s).[122]

La méthode de *Contract Path* est un cas particulier où le modèle économique de réseau est très simplifié. Elle consiste à représenter les zones deux par deux comme liées par une interconnexion directe dont les flux pourraient être entièrement contrôlés. Pour utiliser cette méthode, il faut donc calculer *ex ante* (la veille pour le lendemain) des « capacités de transport » conventionnelles entre les zones de contrôle prises deux par deux, puis allouer ces capacités « conventionnelles ». Les simplifications réalisées par cette méthode sont plus facilement justifiées pour certaines topologies du réseau que pour d'autres ; la méthode s'ajuste mieux à des interconnexions dites « radiales » (par exemple l'interconnexion entre

[120] Les enchères implicites intègrent notamment les modules d'énergie *forward* de chaque zone en prenant en compte les capacités d'interconnexion limitées entre les zones.

[121] Certaines interconnections sont toujours gérées par des mécanismes administratifs sans fondement économique (par exemple, l'interconnexion entre la France et la Suisse). Boucher-Smeers [2002] montrent que le recours à ces procédés d'allocation des capacités de transport provoque des incomplétudes qui mettent sérieusement en cause l'efficacité des résultats obtenus.

[122] Cette méthode est similaire à celle appliquée dans une architecture de marché sur une seule zone de contrôle où le module de transport *forward* est séparé du module énergie *forward* (cf. section 2.2). La différence ici est qu'il existe plusieurs GRTs et les designs des modules d'énergie *forward* peuvent être différents.

la France et l'Angleterre). En revanche, la méthode de « *Contract Path* » présente des problèmes quand elle est appliquée à des réseaux maillés (Ehrenmann-Smeers [2005], Ehrenmann-Smeers [2006]).[123] Les problèmes viennent de ce que le modèle économique du réseau ne représente pas exactement la réalité physique.

Le manque de cohérence de l'application de modèle « *Contract Path* » sur des réseaux maillés provient de ce que l'on traite les zones uniquement deux par deux, comme si le reste du système n'impactait pas l'interconnexion entre ces zones. Or, dans un réseau maillé, toutes les zones sont en permanence interdépendantes. Pour arriver à calculer une capacité « conventionnelle » pour une seule paire de deux zones, chaque gestionnaire d'une sous partie du réseau maillé doit quand même estimer *ex ante* ce que seront les injections et les soutirages sur tout le réseau en temps réel. S'il n'existe pas une forte coordination entre les GRTs, deux cas sont envisageables. Le premier consiste à penser qu'aucun GRT ne fera réellement confiance au mécanisme « faible » de coordination existant au moment de préparer les scénarios d'exploitation de sa portion du réseau commun. Chaque GRT prendra ainsi les scénarios les plus pessimistes de son propre point de vue isolé comme norme de préparation de son exploitation. Il se fixe ainsi des limites d'action plus protectrices à l'égard d'incidents et de conjectures qu'il ne peut pas écarter de ses scénarios par manque d'information crédible (Glachant *et al.* [2005]). Un deuxième cas correspond à des GRTs prenant des décisions trop « optimistes » par rapport à l'information dont ils disposent. Au moment du temps réel, des problèmes d'exploitation graves peuvent apparaître.[124]

Cette méthode d'allocation des capacités de transport utilise des enchères séparées des marchés *forward* d'énergie, enchères explicites. Ainsi cette méthode nécessite une capacité importante d'arbitrage des participants (cf. sous-section 2.5). Les participants doivent avoir la capacité d'arbitrer entre trois marchés (marché d'énergie zone A, marché de transport zone A-B et marché d'énergie zone B). Le caractère séquentiel de ces marchés oblige les participants à réaliser des transactions sur la base d'anticipations. Si ces anticipations ne sont pas parfaites, des transactions inefficaces peuvent être opérées ou, les participants peuvent limiter des transactions inter-zones et perdre ainsi des opportunités d'arbitrage (Zachmann [2005], Ehrenmann-Smeers [2005], Ehrenmann-Smeers [2006]).

La méthode des enchères explicites non-coordonnées (*Contract Path*) est relativement simple à mettre en œuvre car une harmonisation complète du design de modules *forward* d'énergie n'est pas une condition indispensable. En revanche, elle a deux points faibles. D'une part, une représentation trop simplifiée du réseau rend difficile l'internalisation des effets de réseau pour les réseaux maillés. D'autre part, les acteurs de marchés doivent avoir

[123] Une interconnexion radiale correspond au cas où il existe un seul « chemin » électrique reliant deux nœuds du réseau. C'est le cas par exemple entre la France et l'Angleterre. Une interconnexion « maillée » renvoie à l'existence de plusieurs chemins pour aller d'un nœud à un autre du réseau. C'est par exemple le cas de l'interconnexion entre la France et l'Allemagne. Cf. Chapitre 1, sous-section 1.3.

[124] Par exemple, ce type de problématique est considéré comme une des causes de la panne partielle du système interconnecté européen du 4 novembre 2006 (cf. UCTE [2007] et Glachant [2007]).

une forte capacité pour réaliser l'arbitrage entre les différents marchés. Les autres méthodes d'enchères dites implicites (*Market Coupling* ou *Market Splitting*) peuvent résoudre les points faibles de la méthode de « *Contract Path* ». En revanche, ces deux méthodes sont plus exigeantes au niveau de l'harmonisation de design des modules.

L'intégration des modules par « *Market Coupling* » a été développée pour relier les modules *forward* d'énergie de la France, de la Belgique et des Pays-Bas.[125] Ce modèle d'intégration prévoit essentiellement un couplage entre les marchés *forward* d'énergie organisés en J-1 (« bourses d'électricité »). Ce couplage est réalisé par un processus décentralisé et itératif d'échanges d'informations économiques (les offres et les demandes de chaque zone) sur un modèle économique zonal de transport (chaque pays ou zone de contrôle est représenté comme un nœud du réseau hypothétique) (Meeus *et al.* [2005]). Cet échange d'informations itératif cherche à utiliser toutes les opportunités d'arbitrage entre les bourses. Les bourses d'électricité gardent leur indépendance et leur propre design. Néanmoins, un niveau minimum d'harmonisation des designs est nécessaire pour réaliser la coordination. En Belgique, par exemple, la mise en place du *Market Coupling* a obligé la création d'une bourse d'électricité (marché organisé - BELPEX) qui a remplacé le design complètement décentralisé du module *forward* d'énergie. Les caractéristiques du design *Market Coupling* sont supérieures à celles de la méthode de *Contract Path* car les marchés d'énergie et les capacités de transport entre les trois zones sont gérés d'une manière coordonnée. De plus, puisqu'il s'agit d'enchères implicites, le problème de la capacité d'arbitrage des participants ne se pose pas. En revanche, plusieurs questions restent en suspens. La première question porte sur la qualité du découpage « zonal » du réseau de transport des trois pays sans considérer l'Allemagne. Un écart trop grand entre le modèle de représentation du fonctionnement du réseau et les caractéristiques physiques du réseau réel laisserait trop de place à des inefficacités de gestion du réseau (ETSO [2004a]). Une deuxième question porte sur la possibilité de gérer efficacement le couplage des bourses d'électricité d'une manière décentralisée en présence d'offres de « blocs » (Meeus [2006]).

L'intégration des modules *forward* par « *Market Splitting* » est une deuxième possibilité de coordination par enchères implicites. Le marché d'électricité des pays scandinaves (Norvège, Suède, Finlande et Danemark), connu sous le nom de NordPool, est un exemple de l'intégration du type « *Market Splitting* ». Le modèle de NordPool est connu pour ses bonnes performances et il est considéré comme un modèle de référence (Von der Fehr *et al.* [2005], Bergman [2005]). Le cœur de son architecture est le module intermédiaire appelé « ELSPOT » qui intègre les modules d'énergie *forward* des différentes zones avec le module *forward* de transport entre zones.[126] Les contraintes de transport entre les zones sont représentées dans l'enchère correspondant au marché *forward* d'énergie

[125] Son design procède d'une combinaison de propositions antérieures d'ETSO [2001] (enchères explicites coordonnées), d'une part, et d'EuroPEX [2003] (enchères implicites) d'autre part (ETSO-EuroPEX [2004]).
[126] En Norvège des prix différents de l'énergie peuvent apparaître pour plusieurs zones internes au pays. On entend par là que les principales externalités sur le réseau norvégien sont internalisées par différenciation des prix sur ELSPOT.

(ELSPOT).[127] Ces contraintes, quand elles sont « actives », permettent de définir des prix différenciés par zone. La figure 2-13 schématise la configuration du marché NordPool.

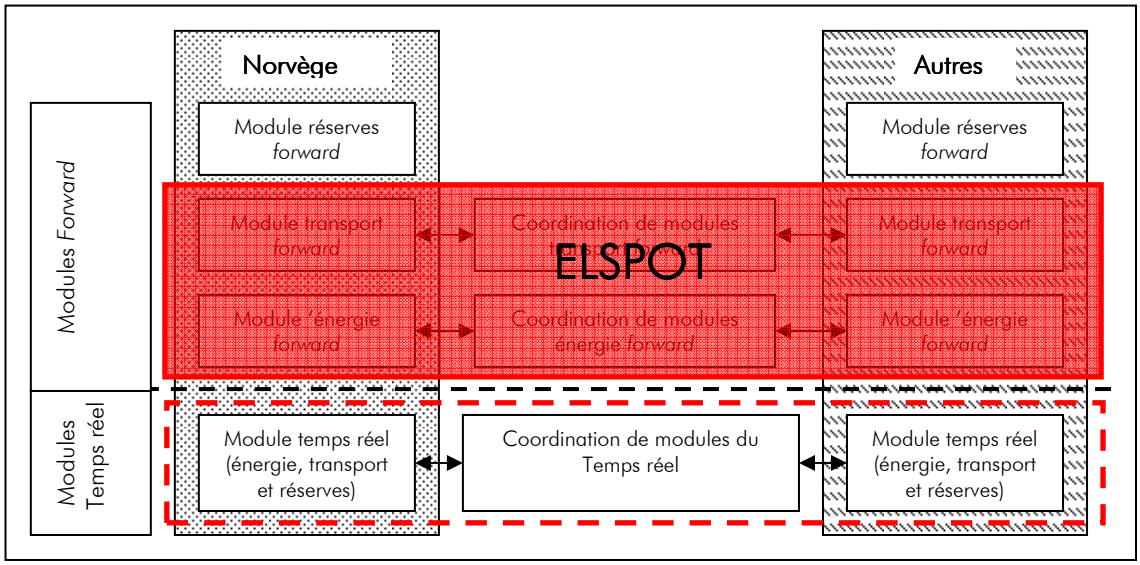

Figure 2-13 : Schéma de NordPool

Les différentes méthodes d'intégration des modules *forward* permettent de profiter des opportunités d'arbitrage et d'utiliser efficacement les capacités limitées du réseau. Cependant, quel que soit le modèle de coordination de modules *forward*, elle trouve son potentiel maximum quand elle est accompagnée par l'intégration des modules du temps réel (Smeers [2001], Boucher-Smeers [2003], Chandley-Hogan [2006]). Logiquement, les marchés *forward* doivent être basés sur les modules du temps réel, les seuls capables théoriquement d'envoyer des signaux économiques cohérents sur les conditions réelles de l'échange.

3.2.2 Coordination des modules du temps réel

L'intégration des modules du temps réel permet de coordonner le fonctionnement en temps réel du système électrique et de créer ainsi les signaux économiques nécessaires à l'articulation des modules *forward*. Il ne faut pas oublier que les transactions sur les modules *forward* sont plus financières que physiques et qu'elles sont basées sur des anticipations des conditions du temps réel. Il se peut qu'une transaction programmée avec anticipation (la veille par exemple) ne soit plus économiquement efficace en temps réel. C'est pourquoi l'intégration de ces modules pour des architectures fonctionnant sur un même réseau interconnecté a une importance centrale.

[127] La forme exacte avec laquelle ces contraintes sont représentées n'est pas connue par les chercheurs (Ehnremann-Smeers [2006]). Cependant, il existe dans la littérature des approximations du traitement des contraintes et de la détermination des prix zonaux dans le modèle de « *Market Splitting* » (Bjørndal-Jörnsten [2001], Ehnremann-Smeers [2005]).

Il est bien connu que cette intégration peut apporter des nombreux avantages. D'abord, elle permettrait de diminuer le volume et le coût du *balancing* nécessaire pour deux raisons. Premièrement, certaines incertitudes peuvent s'annuler mutuellement. Deuxièmement, on peut profiter des opportunités d'arbitrage. Ensuite, l'intégration des modules du temps réel permettrait d'améliorer la gestion des congestions transfrontalières en autorisant la mise en place de *redispatching* transfrontalier (Frontier-Consentec [2005]). Enfin, augmenter le nombre de concurrents dans le module du temps réel permet de diminuer la sensibilité au pouvoir de marché (Smeers [2005a], Hirschhausen *et al.* [2007]).

Actuellement, plusieurs mécanismes de coordination entre modules du temps réel sont utilisés ou sont en train d'être analysés (ETSO [2006a], ETSO [2006b], ETSO [2006c], Glachant *et al.* [2005], PJM-MISO [2006]). Ces modèles d'intégration des modules du temps réel requièrent différents degrés de coordination et d'harmonisation ainsi que la mise en place de procédures entre les GRT et les participants.

Les versions les plus sophistiquées de coordination entre modules du temps réel exigent des designs de modules dotés d'un fort degré d'internalisation des externalités, c'est-à-dire avec une définition nodale des prix du temps réel (cf. section 2.4). Ces versions considèrent les contraintes de transport des zones adjacentes et favorisent une utilisation plus efficace du réseau de transport. Ce type de coordination a été développé théoriquement par Cadwalader *et al.* [1999] sur la base de la coordination de modèles d'*optimal dispatch* pour le cas des entreprises verticalement intégrées (cf. Chapitre 1, section 3). Il consiste à atteindre un optimum global, équivalent à celui d'un module du temps réel coordonné, mais en partant d'un procédé décentralisé qui opère par les biais de protocoles de coordination assez exigeant [128]. Les résultats théoriques montrent qu'une telle coordination est logiquement robuste, et qu'elle est également assez rapide pour pouvoir être mise en pratique (Hogan [2001], Cadwalader *et al.* [1999]). Ce type de coordination forte entre plusieurs GRT indépendants est souvent présenté sous l'expression de « gestionnaire virtuel ». Elle permet notamment d'éviter les principaux « effets de frontières », tout en recourant à une mise en œuvre moins complexe que celle de l'intégration par expansion (Harvey [2003]). Les développements des protocoles de coordination entre GRT, basés sur ce nouveau concept d'intégration, ont déjà débouché sur des applications d'ingénieurs réalistes qui ont été mises récemment en pratique (Ott [2006]). Dans le nord-est et le centre-est des Etats-Unis, plusieurs GRTs/RTOs ont volontairement choisi de mettre en place entre eux des protocoles de coordination pour réduire leurs « effets de frontières » (NEISO-NYISO [2003], MISO-PJM [2004], Ott [2006]). Parmi les actions menées pour la

[128] Les GRTs doivent recourir à un modèle commun de réseau, à la communication entre eux de toutes les informations pertinentes pour l'exploitation du réseau, etc. Voir par exemple, Benintendi-Boccard [2003] qui analysent les problèmes de coordination entre les GRTs dans un environnement d'information asymétrique.

réduction des effets de frontières, PJM et MidWest Pool ont mis en place une solution originale pour l'intégration par coordination des modules de temps réel.[129]

En Europe, l'intégration des modules de temps réel a commencé à être considéré comme indispensable (Glachant-Lévêque [2005], Newbery [2005b], Meeus *et al.* [2005], ETSO [2006a], ETSO [2006b], ETSO [2006c], ERGEG [2006], Hirschhausen *et al.* [2007]). Et, elle est d'une importance centrale pour donner un nouvel élan à l'intégration du marché européen (Glachant-Lévêque [2005]). Les mécanismes d'intégration des modules du temps réel en Europe représentent des versions variées où, par exemple, les GRTs de différentes zones acceptent des offres de producteurs placés en dehors de leur zone de contrôle. Au moment d'accepter une offre, les autres GRT impliqués doivent être mis au courant de la transaction car les échanges programmés entre GRT seront modifiés. Ce modèle a été appliqué en Allemagne et en France (ETSO [2005c]). Dans une autre version du mécanisme de coordination, ce sont les GRT des différentes zones qui s'échangent directement les offres proposées dans leurs zones respectives (liste d'offres commune) et les producteurs ne participent que passivement à la coordination entre les GRTs. Ce modèle est appliqué entre les pays du NordPool (Nordel [2003], ETSO [2005c]).

La coordination de modules du temps réel se réalise principalement par la mise en commun d'offres proposées sur les modules. La mise en commun d'offres et leurs sélections cherchent à approcher un « *dispatch optimal* » coordonné des moyens de production en temps réel pour l'ensemble de zones concernées. Bien qu'un certain degré d'harmonisation entre les designs du module du temps réel soit désirable pour cette forme de coordination, une harmonisation complète n'est pas nécessaire. Par exemple, pour le cas de l'Europe, les designs des modules du temps réel à intégrer peuvent utiliser des règles assez différentes pour le règlement des écarts, c'est-à-dire qu'un design de type « mécanisme d'ajustement » peut être intégré à un design du type « marché du temps réel » (cf. Chapitre 3, section 2).

3.3 Conclusions de la section

Cette section s'est focalisée sur le cas de plusieurs architectures de marché sur plusieurs zones de contrôle composant toutes un seul réseau de transport interconnecté. L'objectif dans cette section était d'analyser la manière dont les architectures de marché peuvent s'intégrer pour étendre l'envergure du marché électrique régional. Pour ce faire, nous avons utilisé le cadre d'analyse modulaire des architectures de marché proposé dans les sections précédentes.

Dans une première partie, nous avons utilisé le concept d'intégration parfaite des architectures considérant une architecture de marché unique, avec les 4 modules

[129] Cette méthode consiste en un échange d'informations sur certains éléments du réseau susceptibles d'être congestionnés et sur lesquels des actions (injections ou soutirages) au sein des zones de contrôle peuvent avoir une influence (Ott [2003], MISO-PJM [2004]). Cf. www.miso-pjm.com.

correspondants, et un seul GRT pour tout le réseau interconnecté. Ce cadre idéal suppose qu'il n'existe pas de « frontières » entre zones, que ce soit dans la définition de l'architecture ou dans la responsabilité du GRT. Ce niveau d'intégration idéal permettrait de capter tous les bénéfices d'un élargissement régional en éliminant complètement les « effets de frontière » issus d'une coordination imparfaite entre zones. Cet idéal nous a servi de référence vers lequel les différentes formes pratiques de mise en œuvre d'intégration doivent s'approcher.

Il existe différentes manières pour avancer vers l'intégration parfaite d'architectures. La première, l'intégration par expansion, correspond à l'expansion de l'architecture de marché et de la responsabilité d'un seul GRT sur une zone de contrôle plus grande. Cette forme d'intégration élimine complètement les effets de frontière. En revanche, elle suscite des difficultés d'acceptation par les autres zones étant donné la radicalité des réformes organisationnelles et institutionnelles qu'elle impose. La deuxième, l'intégration par coordination des modules, met en place des mécanismes de coordination entre les architectures de marché existantes afin de simuler un fonctionnement intégré, mais sans supprimer les institutions préexistantes ni les différentes zones de contrôle. En conséquence, cette forme d'intégration est plus simple à mettre en œuvre d'un point de vue institutionnel. En revanche, cette forme d'intégration ne garantit pas l'élimination complète des effets de frontière. Elle pose aussi une question sur l'harmonisation du design de modules à intégrer.

Dans une deuxième partie, nous avons analysé plus en détail l'intégration d'architectures par coordination des modules car c'est la plus facilement applicable dans un contexte de fortes contraintes institutionnelles (par exemple, le cas de l'Europe). La coordination entre architectures est réalisée par le biais de mécanismes de coordination entre les différents modules des architectures adjacentes.

Nous avons tout d'abord analysé la coordination des modules *forward*. Pour ce faire, nous avons étudié trois principales méthodes de coordination entre les module *forward* d'énergie et de transport appliquées en Europe : les enchères explicites non-coordonnées (e.g. France et Allemagne), les enchères implicites de « *Market Coupling* » (France-Belgique-Pays-Bas) puis les enchères implicites de « *Market Splitting* » (Pays nordiques). Nous avons montré que ces trois méthodes varient selon leur simplicité de mise en œuvre, le besoin d'harmonisation entre modules, leur aptitude à prendre en compte les effets de réseau et la capacité d'arbitrage des agents. D'un côté, la méthode des enchères explicites non-coordonnées (*Contract Path*) est la plus simple à mettre en œuvre ; elle ne nécessite qu'une faible harmonisation du design de modules *forward* d'énergie. En revanche, sa représentation simplifiée du réseau ne prend pas en compte les propriétés de répartition des flux sur les réseaux maillés. Par ailleurs, le fonctionnement de cette méthode implique une forte capacité d'arbitrage des acteurs entre les différents marchés (énergie et transport). D'un autre côté, les enchères implicites (*Market Coupling* ou *Market Splitting*) qui peuvent mieux appréhender les caractéristiques des réseaux maillés et réaliser les arbitrages de manière automatique, nécessitent un degré d'harmonisation des modules plus important.

Quel que soit le modèle de coordination des modules *forward*, cette coordination trouve son potentiel maximum quand elle est accompagnée par l'intégration des modules du temps réel. Nous avons analysé ainsi la coordination des modules du temps réel. Les différentes méthodes pour la coordination de ces modules en temps réel cherchent à s'approcher d'un « *dispatch optimal* » coordonné des moyens de production pour l'ensemble des zones concernées. Divers types de mécanismes de coordination entre modules du temps réel sont envisageables. D'abord, l'application de mécanismes de coordination sophistiqués, considérant l'équilibre et les contraintes de transport des zones adjacentes, nécessite, d'un côté, un degré important d'harmonisation d'architectures et d'un autre côté que les architectures soient « nodales » internalisant fortement les externalités du transport. Ce type de mécanisme a été appliqué pour la coordination de modules du temps réel des architectures du nord-est des Etats-Unis (PJM, MidWest Pool, etc.). Ensuite, d'autres mécanismes de coordination moins sophistiqués, qui ne considèrent pas les contraintes de transport, sont plus simples à mettre œuvre. La coordination des modules du temps réel se réalise ici par la mise en commun d'offres proposées sur chacun des modules des différentes zones. Ces derniers mécanismes de coordination ont été, ou vont l'être, appliqués principalement en Europe. Bien qu'un certain degré d'harmonisation entre les designs du module du temps réel soit désirable, une harmonisation complète n'y est pas nécessaire. On peut, par exemple, combiner un design du type « mécanisme d'ajustement » avec un design du type « marché du temps réel ». Cependant, ce genre d'harmonisation limitée peut provoquer des effets économiques asymétriques sur les participants des différentes zones. Nous reviendrons sur ce point important dans le quatrième chapitre.

Conclusions du chapitre

L'introduction de la concurrence dans le secteur électrique, et plus particulièrement sur le marché de gros, nécessite un design de l'architecture de marché. Dans ce chapitre, nous sommes partis des caractéristiques spécifiques de l'électricité (non-stockabilité, incertitudes, externalités changeantes, caractéristiques de bien public, etc.) et des outils d'optimisation d'une entreprise verticalement intégrée afin de dériver les éléments essentiels de cette architecture de marché.

Dans une première section du chapitre, nous avons étudié le cas le plus simple, celui de l'architecture de marché d'une seule zone de contrôle. Nous avons analysé l'introduction de la concurrence comme le remplacement des contrôles-commandes utilisés dans une entreprise intégrée (l'*unit commitment* et l'*optimal dispatch)* par une combinaison de mécanismes et de marchés devant coordonner les trois composantes du bien « électricité » : l'énergie, le transport et les réserves. Nous avons conclu qu'une architecture de marché est nécessairement composée de quatre modules (ou sous-marchés potentiels) : 1° le module énergie *forward*, 2° le module de transport *forward*, 3° le module des réserves *forward* et 4° le module du temps réel.

Dans une deuxième section, en considérant toujours une seule zone de contrôle, nous avons développé un cadre d'analyse modulaire des architectures de marché. Pour chacun des modules, il existe plusieurs options ou formes organisationnelles différentes. Les caractéristiques principales de ces différentes solutions de design sont logiquement basées sur les outils d'optimisation utilisés dans une entreprise intégrée (l'*unit commitment* et l'*optimal dispatch)*. Une architecture de marché est donc construite en choisissant une option pour chacun de ces modules et pour chacune des articulations entre ces modules.

Nous avons alors défini des axes représentatifs et simplifiés pour décrire le choix des différentes options de design pour chaque module. Premièrement, les options de design pour le module d'énergie *forward* se différencient par leur degré de centralisation. Deuxièmement, les options de design du module de transport *forward* se placent sur deux axes, considérant d'un côté la séparation entre le module de transport *forward* et le module d'énergie *forward* et, d'un autre côté, le degré d'internalisation des externalités du transport sur l'étape *forward*. Troisièmement, le module des réserves *forward* présente différentes options de design pour la forme d'organisation de la production des réserves. Enfin, les designs du module du temps réel se différencient par deux axes principaux, celui de la définition des signaux de prix du temps réel lors de l'équilibrage du système, et celui du degré d'internalisation des externalités du transport en temps réel. Ce cadre d'analyse modulaire nous a aussi permis de caractériser les architectures réelles de trois de marchés de référence : Angleterre, Texas et PJM.

Grâce au développement du cadre d'analyse modulaire, et à son application à des architectures de référence, nous avons montré deux points essentiels. Premièrement, il est empiriquement observable et théoriquement compréhensible que diverses combinaisons sont possibles pour créer une architecture de marché. Nous avons pu vérifier cette grande diversité d'architectures de marché en examinant, module par module, les solutions de design proposées ou mises en œuvre. Le manque de maîtrise rationnelle pour choisir les options optimales de design montre l'importance d'une recherche plus approfondie sur les différents designs et modules. Deuxièmement, grâce à l'analyse modulaire, nous avons vérifié que l'étude approfondie d'un seul module reste logiquement possible.

Nous avons choisi alors d'approfondir la recherche sur le design du module du temps réel. Nous avons montré que, dans la diversité d'architectures observées, il existe toujours un noyau commun à tous les designs pour remplir des fonctions « électriques » indispensables. Si l'on regarde ces architectures de marché du point de vue de l'effort des autorités pour les construire, on identifie ce noyau indispensable et commun comme une architecture « minimale » des réformes. Cette architecture minimale porte sur le design du module du temps réel. Elle sanctionne l'importance opérationnelle du module du temps réel pour un bien qui n'est pas stockable, et dont les échanges physiques ne prennent effectivement place qu'en temps réel. Cette importance centrale du module du temps réel marquera toute la suite de notre recherche. Les deux chapitres suivants (3 et 4) seront consacrés à l'analyse et à la modélisation de ce module afin d'en obtenir une compréhension plus profonde et de préciser l'influence des paramètres clé de son design.

Finalement, dans une troisième section du chapitre, nous avons étudié la question du design de l'intégration des architectures de marchés entre plusieurs zones de contrôle. Cette intégration des architectures de marché est une condition nécessaire du bon fonctionnement des marchés électriques régionaux.

Nous avons étudié les différentes formes pratiques de mise en œuvre qui cherchent à s'approcher d'un idéal d'intégration sans « effets de frontière ». Ceci consiste en une intégration parfaite pour chacun des modules des architectures de marché et en une coordination parfaite des GRTs. Il existe alors différentes manières pour avancer vers l'intégration parfaite d'architectures. La première, l'intégration par expansion, correspond à l'expansion de l'architecture de marché et de la responsabilité d'un seul GRT sur une zone de contrôle plus grande. Cette forme d'intégration élimine complètement les effets de frontière. En revanche, elle peut susciter des difficultés d'acceptation par les autres zones étant donné la radicalité des reformes organisationnelles et institutionnelles qu'elle impose. La deuxième, l'intégration par coordination, met en place des mécanismes de coordination entre les architectures de marché existantes afin de simuler un fonctionnement intégré, mais sans supprimer les institutions préexistantes ni les différentes zones de contrôle. Cette forme d'intégration, en revanche, ne garantit pas l'élimination complète des effets de frontière. Elle pose aussi une question sur l'harmonisation du design de modules à intégrer.

L'intégration d'architectures par coordination est donc la forme d'intégration la plus facilement applicable dans un contexte avec des fortes contraintes institutionnelles, comme celui de l'Europe. Et c'est celle qui nous avons étudié avec le plus de détails. Nous avons tout d'abord analysé la coordination des modules *forward* avec les trois principales méthodes de coordination entre module *forward* d'énergie et module *forward* de transport appliquées en Europe : les enchères explicites non-coordonnées (e.g. France et Allemagne), les enchères implicites de « *Market Coupling* » (France-Belgique-Pays-Bas), puis les enchères implicites de « *Market Splitting* » (Pays nordiques). Nous avons montré que les trois principales méthodes appliquées en Europe pour coordonner les modules *forward* (énergie et transport) varient, d'une part, par rapport à leur simplicité de mise en œuvre et au besoin d'harmonisation entre modules, et, d'autre part, par rapport à leur aptitude à prendre en compte les effets de réseau et la capacité d'arbitrage des agents.

Après la coordination des modules *forward*, la coordination du module du temps réel est elle-aussi nécessaire pour permettre une intégration efficace de ces modules *forward*. Quel que soit le modèle de coordination des modules *forward*, cette coordination trouve son potentiel maximum quand elle est accompagnée par l'intégration des modules du temps réel. En conséquence, l'étude de l'intégration des modules du temps réel est d'une importance cruciale. Logiquement les marchés *forward* doivent être basés sur les modules du temps réel, qui sont théoriquement les seuls capables d'envoyer des signaux économiques cohérents sur les conditions réelles de l'échange.

Nous avons analysé ainsi la coordination des modules du temps réel. Les différentes méthodes pour la coordination de ces modules en temps réel cherchent à s'approcher d'un « *dispatch optimal* » coordonné des moyens de production pour l'ensemble des zones concernées. Des mécanismes de coordination sophistiqués, qui considèrent l'équilibre et les contraintes de transport des zones adjacentes, sont envisageables. Mais ils nécessitent un degré important d'harmonisation des architectures. C'est le cas pour la coordination des modules du temps réel des architectures du nord-est des Etats-Unis (PJM, MidWest Pool, etc.). D'autres mécanismes de coordination moins sophistiqués, qui ne considèrent pas les contraintes de transport, sont plus simples à mettre œuvre. La coordination des modules du temps réel se réalise ici par la mise en commun d'offres proposées sur chacun des différents modules. Ces derniers mécanismes de coordination ont été, ou vont être, appliqués principalement en Europe. Bien qu'un certain degré d'harmonisation entre les designs du module du temps réel soit désirable, une harmonisation complète n'y est pas nécessaire. On peut, par exemple, combiner un design de type « mécanisme d'ajustement » avec un design du type « marché du temps réel » (cf. chapitre 3). Cependant, ce genre d'harmonisation limitée peut provoquer des effets économiques asymétriques sur les participants des différentes zones. Nous reviendrons sur ce point important dans le quatrième chapitre.

Chapitre 3

Séquence des marchés d'énergie et design du module du temps réel

Introduction 125

Section 1: Séquence des marchés d'énergie à court terme 127
1.1 Organisation de la séquence des marchés d'énergie 128
1.2 Liens entre marchés et arbitrage temporel 136
1.3 Conclusions de la section 142

Section 2: Design du module du temps réel 143
2.1 Etapes du module du temps réel 144
2.2 Types de design du module du temps réel 155
2.3 Etudes de cas 159
2.4 Le choix du design : incitations et conséquences économiques 177
2.5 Conclusions de la section 186

Conclusions du chapitre 188

Introduction

L'architecture d'un marché d'électricité est composée de plusieurs modules. L'état de l'art actuel de la théorie sur les architectures de marché ne permet pas de trancher d'une façon définitive sur les options optimales pour chaque module. Ce manque de maîtrise rationnelle pour choisir les options optimales de design montre l'importance d'une recherche plus approfondie sur les différents designs et modules. La modularité de l'architecture facilite une analyse approfondie car elle permet d'étudier certains modules séparément des autres.

Dans le chapitre précédent, l'importance du module du temps réel a été mise en valeur. En effet, et en premier lieu, il constitue le dernier module de la séquence de marchés ; son rôle principal est l'équilibrage du système tout en donnant les signaux économiques nécessaires pour une bonne articulation des autres modules. En deuxième lieu, ce module est nécessaire à toutes les variantes de design de toutes les architectures. Si l'on regarde l'architecture de marché du point de vue de l'effort des autorités pour la construire, on identifie une architecture minimale comme étant le noyau indispensable au fonctionnement de toute l'industrie électrique. Cette architecture minimale porte sur le design du module du temps réel. Ceci découle de l'importance du module du temps réel pour un bien qui est difficilement stockable et dont les échanges physiques prennent effectivement place en temps réel.

Le module du temps réel est donc au centre de toute architecture de marché et il nécessite ainsi un traitement attentif de son design. Ce module doit coordonner en temps réel les trois composantes fondamentales de l'électricité : l'énergie, le transport et les réserves. Dans ce chapitre, nous nous concentrerons principalement sur la composante « énergie » qui est la « vraie » commodité dans l'électricité. Les deux autres composantes, transport et réserves, ne seront pas traitées explicitement.

La composante « énergie » est commercialisée sur une séquence de marchés qui s'achève par le module du temps réel. L'objectif de ce chapitre est de montrer quels sont les paramètres de design à prendre en compte sur le module du temps réel pour que la séquence de marchés d'énergie fonctionne correctement.

Pour ce faire, dans la section 1, nous rappellerons l'organisation de la séquence de marchés et les liens entre ses différents modules ou marchés (arbitrage temporel). Nous montrerons ici l'importance des caractéristiques du dernier marché dans une séquence où les transactions peuvent avoir lieu sur des horizons de temps différents mais qui traitent toutes le même bien physique. Ce dernier marché sera pour l'électricité le module du temps réel.

Dans la section 2, nous analyserons les caractéristiques du design du module du temps réel. Grâce à cette analyse, nous distinguerons deux types de design pour ce module : le « marché » et le « mécanisme ». Nous montrerons que la principale différence entre ces deux cas de figure provient de la manière dont certaines transactions du temps réel sont réglées financièrement, notamment avec ou sans utilisation de pénalités. Le choix de la méthode de règlement des transactions du temps réel sera déterminant puisqu'il fixera les incitations données aux participants du marché et influencera ainsi leur comportement en provoquant différentes conséquences économiques.

Section 1: Séquence des marchés d'énergie à court terme

Le chapitre précédent nous a montré que l'électricité pouvait être décomposée en trois éléments (énergie, transport et réserves) et que l'organisation de l'architecture de marché pour ces trois éléments pouvait être définie avec quatre modules (énergie *forward*, transport *forward*, réserves *forward* et module du temps réel). Si l'on considère seulement l'élément « énergie », nous avons la séquence des marchés d'énergie. L'élément énergie est central dans le design de marchés et il a fait l'objet de nombreuses études. En fait, l'énergie est souvent considérée comme la « vraie » commodité dans le bien électricité (Smeers [2004b]).

La séquence de marchés d'énergie est composée du module *forward* d'énergie et du module du temps réel. Le module du temps réel doit assurer l'équilibrage physique en temps réel et émettre des signaux économiques cohérents avec les conditions du temps réel. Le module *forward* d'énergie remplit deux rôles principaux : un rôle d'information pour faciliter la coordination et donner de bons signaux aux agents, ce qui leur permettra de prendre les bonnes décisions de préparation de la production, et un rôle d'outil de couverture des risques du temps réel, qui sont d'une certaine importance étant donné la grande volatilité des signaux économiques délivrés par le module du temps réel (Counsell-Evans [2004]).

Afin de nous focaliser sur la séquence des marchés correspondant à un bien unique, nous faisons l'hypothèse simplificatrice que les biens « énergie » correspondant aux différents intervalles de livraison sont indépendants les uns des autres. Certes, certaines caractéristiques « imparfaites » des moyens de production (coût de démarrage, limites minimales de fonctionnement, contraintes de *ramping*, etc.) peuvent être sources de liens intertemporels et de non-convexités entre les différents biens « énergie » définis au long d'une journée. Deux arguments nous incitent à négliger les problèmes de non-convexités et les liens intertemporels. D'un côté, ces problèmes peuvent être d'une taille négligeable pour les grands systèmes électriques. D'un autre côté, ces problèmes peuvent être traités par le choix d'un degré de centralisation au niveau du module *forward* d'énergie pour fournir le degré de coordination nécessaire (cf. Chapitre 2, sous-section 2.1).

L'objectif de cette première section est ainsi de comprendre la mécanique de l'organisation de la séquence des marchés d'énergie ainsi que les liens ou l'arbitrage entre ces marchés. Ceci nous permettra ensuite d'étudier les caractéristiques de design du module du temps réel et leurs conséquences sur l'articulation avec les autres modules. Cette première section est organisée de la manière suivante. Nous présentons d'abord l'organisation de la séquence des marchés d'énergie (section 1.1). Puis nous présentons l'arbitrage temporel, qui est fondamental pour comprendre l'articulation des modules de la séquence complète des marchés (section 1.2).

1.1 Organisation de la séquence des marchés d'énergie

La séquence des marchés d'énergie est fondée sur un processus organisé en étapes successives (figure 3-1). A court terme, ces étapes considèrent les différents marchés *forward day-ahead* et infrajournalier. Elles se terminent avec le dernier module où les participants peuvent vendre ou acheter de l'énergie. C'est le marché du temps réel[130].

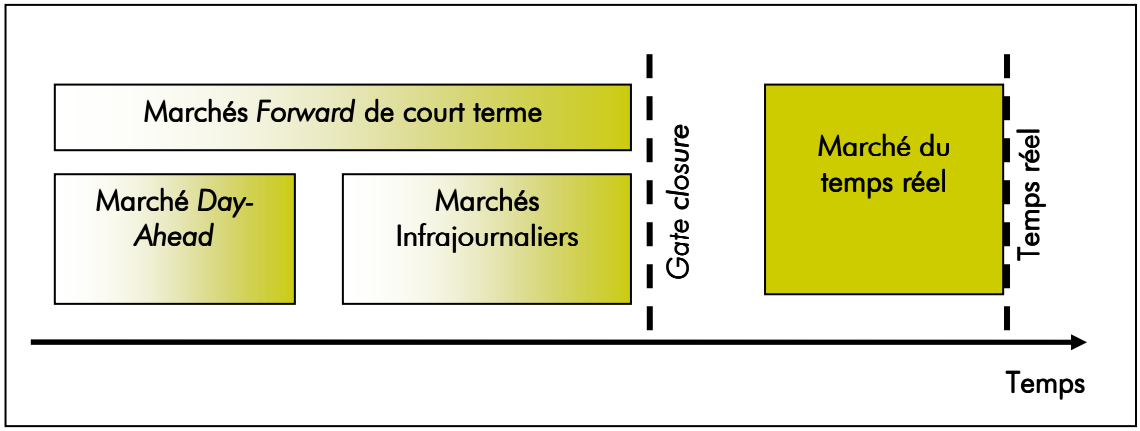

Figure 3-1 : Séquence de marchés d'énergie

L'électricité étant un flux continu nécessite de définir des biens successifs différenciés temporellement. Ces biens sont définis en discrétisant l'axe temporel sur un nombre donné d'intervalles de livraison. En effet, les biens commercialisés sur ces marchés correspondront aux intervalles de temps définis, pendant lesquels, l'énergie injectée ou soutirée sera mesurée et réglée financièrement en temps réel.[131] La définition de l'intervalle de livraison est un paramètre de l'architecture de marché et sert de référence pour comparer les obligations contractuelles avec les injections et soutirages physiques d'énergie mesurées en temps réel. Par exemple, la durée correspondant à une journée peut être divisée en 24 biens horaires différenciés (PJM), en 48 biens différenciés d'une demi-heure (France et Angleterre) ou 96 biens différenciés de 15 minutes (Texas). Pour chacun de ces biens différenciés il existe une séquence de marchés (*day-ahead*, infrajournaliers, temps réel) où l'on peut réaliser des transactions pour diverses quantités d'énergie à livrer ou à soutirer en temps réel pendant l'intervalle de livraison correspondant.

Dans cette section nous utilisons une caractérisation simplifiée des architectures de marché. Cette caractérisation prend en compte le module *forward* d'énergie et le module du temps réel mais se base principalement sur les différences de design du premier module. Le module *forward* d'énergie peut prendre des designs différents sur la dimension de la

[130] Dans cette section nous utilisons le terme « marché du temps réel » d'une manière générale. Dans la section suivante nous verrons que le module du temps réel peut avoir un design qui ne correspond pas exactement à un « vrai » marché (notion de mécanisme d'ajustement).

[131] L'énergie injectée ou soutirée pendant un intervalle de temps est comptabilisée comme l'intégrale de la puissance active (injectée ou soutirée) pendant la durée de l'intervalle. Ainsi, par exemple, une machine de 1 MW de puissance, fonctionnant pendant une heure, consomme une énergie correspondant à 1MWh.

centralisation (Wilson [2002]). [132] Nous considérons deux types d'architectures : « centralisée » et « décentralisée ».[133] L'architecture centralisée est basée principalement sur des marchés organisés (enchères centralisées) gérés par une entité qui fait office à la fois d'opérateur du marché et de gestionnaire du réseau (par exemple aux Etats-Unis : PJM, NY Pool, etc.). Le design décentralisé est fondé principalement sur des transactions bilatérales, ou marché de gré à gré, (par exemple Angleterre ou Texas). Bien que pour le design décentralisé, une bourse d'électricité (*Power Exchange*) avec enchères puisse exister, elle est gérée par une entité extérieure au gestionnaire du réseau. En temps réel, ces deux architectures convergent car c'est toujours le gestionnaire du réseau qui prend en charge l'équilibrage et qui centralise les opérations du module du temps réel.

Cette différenciation du design est importante car elle contraint la manière dont est transférée l'information nécessaire au gestionnaire du réseau. Le gestionnaire du réseau a besoin d'information sur les résultats des marchés d'énergie (volume des positions contractuelles, programmes de production et de consommation estimatifs, etc.) afin de vérifier si les transactions nouées sur les marchés sont réalisables sur le réseau, pour préparer les moyens adéquats pour l'équilibre en temps réel, et pour avoir l'information nécessaire pour comptabiliser les ventes et les achats en temps réel (Stoft [2002]). Dans les architectures « centralisées », une partie importante de cette information est déjà disponible dans les résultats des enchères des marchés organisés. Dans les architectures « décentralisées », des représentants agréés auprès du gestionnaire du réseau, doivent transmettre l'information. Ces représentants agréés peuvent être des responsables d'équilibre (*scheduling coordinator* en anglais) qui regroupent des producteurs, des *retailers* ou revendeurs, des consommateurs et/ou des *traders*, ou bien directement les bourses ou les marchés organisés.

Nous présentons par la suite l'organisation des différentes étapes de la séquence des marchés d'énergie, marchés *forward* et marché du temps réel, pour les différents types d'architecture.

1.1.1 Marchés *forward*

La séquence des marchés *forward* d'énergie à court terme commence avec le marché *day-ahead* et continue avec les marchés infrajournaliers. La différence entre ces divers marchés *forward* correspond à la période de temps qui s'écoule entre la transaction sur le marché et la période de livraison. Le marché *day-ahead* commercialise des contrats d'énergie la veille du jour de la livraison physique, c'est-à-dire 24h en moyenne avant la livraison. Par exemple, on y trouve un contrat correspondant à la livraison de 1 MWh

[132] Cf. chapitre 2. Il faut aussi considérer que dans la définition de l'axe « centralisation » il y a aussi inclus le caractère d'organisation de ces marchés centralisés et la participation du gestionnaire du réseau dans l'organisation de ce marché.

[133] Bien que es architectures existantes correspondent a des formes d'organisation hybride, nous les avons classifiées en architectures « centralisées » et architectures « décentralisées ».

d'énergie à 12h le lendemain. Les marchés infrajournaliers commercialisent des contrats de biens « énergie » livrables le jour même, par exemple quelques heures avant la livraison. Cette différence est plus marquée dans les architectures « centralisées » où existe nécessairement un marché organisé en J-1. En revanche, dans les architectures « décentralisées », sans marché organisé en J-1, les transactions se réalisent continuellement, et le passage entre *day-ahead* et infrajournalier est alors moins marqué.

Nous verrons par la suite comment les options de design se mettent en place en pratique pour les marchés *forward day-ahead* et les marchés infrajournaliers.

1.1.1.1 Marché *day-ahead*

La veille du jour de livraison, les participants au marché (producteurs et acheteurs) doivent équilibrer leurs positions contractuelles. Ils ont déjà des obligations contractuelles venant de leurs contrats *forward* de long terme, et ils doivent ajuster leurs transactions aux conditions actuelles pour chacun des intervalles de livraison. Par exemple, les producteurs doivent considérer l'état de leur parc de production et décider s'ils doivent produire ou acheter sur le marché pour honorer les obligations contractuelles ; les acheteurs doivent comparer leur volume de contrats avec des prévisions de consommation plus précises pour chaque intervalle de livraison ; les traders agissent comme intermédiaires entre producteurs et acheteurs en arbitrant de possibles désajustements, imperfections ou déséquilibres du marché.

Les participants réalisent des transactions en utilisant les différentes modalités d'échange dans l'architecture de marché. Deux types d'informations peuvent être demandés par le gestionnaire du réseau. La première est le volume agrégé contractuel de chacun des participants. La deuxième information est de nature « physique » parce qu'elle couvre, d'un côté, les unités de production et leur programme prévisionnel de production et, d'un autre côté, les points de soutirage et leur programme prévisionnel de consommation (Sweeting [2000]).[134] Ces informations y sont organisées en des positions contractuelles, où on présente pour chaque période de livraison la quantité d'énergie qu'on s'engage contractuellement à « injecter » ou à « soutirer », ainsi que les programmes journaliers prévisionnels de production et de consommation pour chaque unité de production et chaque point de soutirage. En pratique, la première notification des informations, faite en J-1 est seulement indicative. Elle sert principalement à ce que le gestionnaire du réseau puisse commencer à évaluer la sécurité du système.

Les architectures « centralisées » utilisent des enchères centralisées comme base pour le marché *day-ahead*. L'enchère se réalise généralement vers midi de la veille du jour de livraison. Cette enchère commercialise différents contrats *forward* d'énergie

[134] Par exemple sur le marché Anglais (NETA), cette première déclaration est connue comme « *Initial Physical Notification* ».

correspondant à tous les intervalles de livraison du lendemain. Ces différents contrats ne possède pas la même « maturité » (période de temps entre la transaction *forward* et l'intervalle de livraison). L'enchère reçoit des offres d'achat et de vente des unités de production[135], des *retailers* et des *traders*. Les offres qui ne sont pas associées à des moyens de production déterminés sont considérées comme « virtuelles », strictement financières, et sont différenciées par le gestionnaire du réseau au moment d'évaluer la sécurité du réseau.[136] Le résultat de ce marché *forward* se traduit dans un ensemble d'obligations et de positions contractuelles des participants au marché, mais aussi dans des programmes de marche prévisionnelle pour les moyens de production. Etant donné que, pour ces architectures, les fonctions de gestionnaire du réseau et d'operateur du marché (enchère centralisée) sont exercées par la même entité, la transmission de l'information est directe entre le marché et le réseau.

Dans les architectures « décentralisées », les participants réalisent leurs transactions pour les intervalles de livraison du lendemain par le biais des contrats bilatéraux ou en participant volontairement à une bourse d'électricité. Sur les bourses d'électricité on trouve des contrats standardisés (définis strictement par intervalle normalisé de livraison). [137] Ces marchés organisés sont placés normalement à midi en J-1 et le jour de livraison commence à 0h du lendemain (J). A un moment déterminé en J-1, les participants, par le biais des responsables d'équilibre, doivent communiquer au gestionnaire du réseau les informations correspondant aux résultats agrégés des transactions *forward*. Dans ce processus, on agrège l'information dans des programmes journaliers qui combinent les positions des contrats passés antérieurement avec les nouveaux contrats du J-1. L'information sur les positions contractuelles doit être accompagnée par les programmes journaliers prévisionnels de production et de consommation. On parle de « programmes équilibrés » (*balanced schedules*) quand les programmes journaliers de production et de consommation crédibles doivent coïncider avec les positions contractuelles (Baldick-Niu [2004). Normalement le moment où les participants sont appelés à communiquer les informations vient après le déroulement de la bourse d'électricité, si elle existe. Les bourses d'électricité offrent aussi le service de transfert d'information sur les résultats du marché auprès du GRT.

Après le marché *day-ahead,* les participants peuvent continuer à réaliser de transactions, par le biais des marchés infrajournaliers.

---

[135] Ces offres sont accompagnées par d'autres informations qui permettent prendre en compte les caractéristiques techniques des moyens de production.
[136] Ce type d'offres permet à l'opérateur du marché de distinguer les positions contractuelles et les programmes prévisionnels de production et consommation.
[137] Des contrats plus sophistiqués sont aussi commercialisés sur les marchés. Un exemple de contrats plus sophistiqué sur les marchés organisés correspond au « bloc » qui consiste en groupe d'heures continues (intervalles de livraison continus) et qui suppose d'être acceptés complètement ou rejetés (Meeus [2006]).

1.1.1.2 Marchés infrajournaliers

Les marchés infrajournaliers sont les marchés *forward* qui sont placés après le marché *day-ahead*. Ces marchés infrajournaliers permettent d'incorporer les nouvelles informations survenues après le marché *day-ahead* et de réajuster plus finement leurs positions.

De la même manière que pour les marchés *day-ahead*, la forme d'organisation de ces marchés peut varier entre plusieurs niveaux de centralisation. Dans les architectures centralisées, un ou plusieurs marchés infrajournaliers peuvent être organisés avant le temps réel.[138] Mais, dans la plupart des architectures, le design des marchés infrajournaliers n'est pas organisé et il reste très simple. Il consiste seulement à autoriser la modification des positions contractuelles agrégées et des programmes de production et de soutirage. Les participants réalisent de leur côté de nouvelles transactions, et ils communiquent au GRT les changements des volumes agrégés de leurs positions contractuelles ainsi que leurs programmes de production et de consommation.

Des modifications des programmes pour chaque intervalle de livraison peuvent être réalisées jusqu'à un moment très précis, qui correspond à la fermeture par le GRT du dernier marché infra journalier des échanges pour la période concernée. Cette fermeture des marchés *forward* par le GRT s'appelle, pour cette raison, la « *gate closure* ». A ce moment précis, tous les programmes et les positions contractuelles communiqués au GRT deviennent définitifs. Les informations de ces programmes définitifs serviront au GRT pour le calcul des achats et des ventes sur le marché du temps réel, par comparaison avec les mesures réelles d'injection et de soutirage effectuées sur le réseau. Au moment de la *gate closure*, les participants peuvent aussi présenter des offres d'ajustement qui seront utilisées pour équilibrer le système dans le module du temps réel. Ces offres consistent normalement en des propositions en volume et en prix pour modifier volontairement les programmes définitifs de production (ou de consommation).

La fermeture des marchés *forward*, ou « *gate closure* », peut être définie de différentes manières par rapport aux périodes où les participants sont autorisés à modifier leurs déclarations de volume de transaction et leurs positions physiques. Cette autorisation peut être activable à différentes franges horaires (« guichets »), ou peut être continue jusqu'à une limite fixée avant la période de livraison (« *rolling gate closure* »). Dans la première forme d'organisation, des guichets sont ouverts à différents moments de la journée pendant lesquels les participants peuvent modifier leurs programmes initiaux (figure 3-2). C'est pourquoi la durée entre la *gate closure* et le moment du temps réel varie par rapport à la position temporelle des différents intervalles de livraison.

[138] Par exemple, l'architecture de marché OMEL (Espagne) utilise des marchés infrajournaliers organisés (cf. www.omel.es).

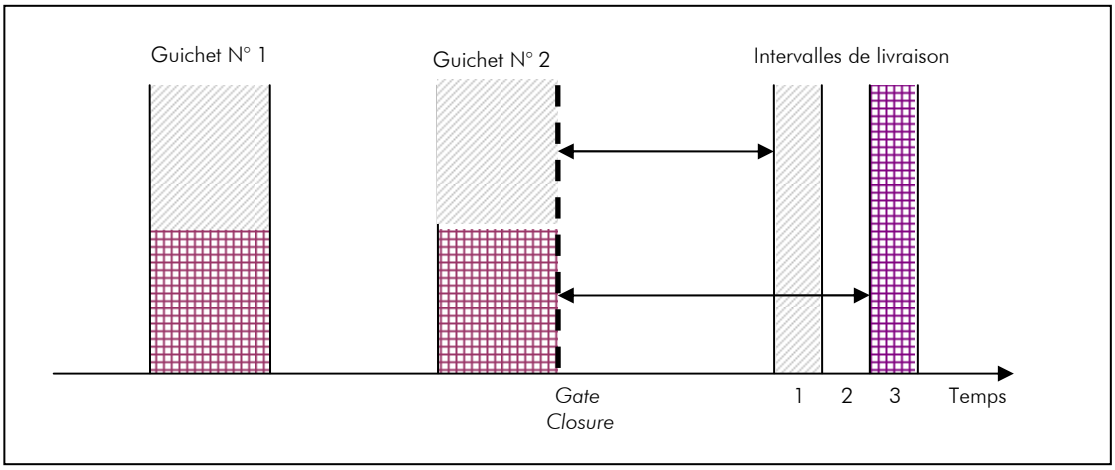

Figure 3-2 : Schéma de *gate closure* par « guichets »

Dans la deuxième forme d'organisation, « *rolling gate closure* », les participants sont autorisés à modifier leur programmes librement jusqu'à la *gate closure* et celle-ci se déplace le long de l'axe temporel pour garder constante la durée du temps qui la sépare du temps réel pour toutes les intervalles de livraison (figure 3-3). Indépendamment du design de la *gate closure*, la durée de temps avant l'intervalle de livraison où elle se place est très importante. En fait, c'est cette période de temps qui définira le niveau d'incertitude présente sur le dernier marché *forward*.

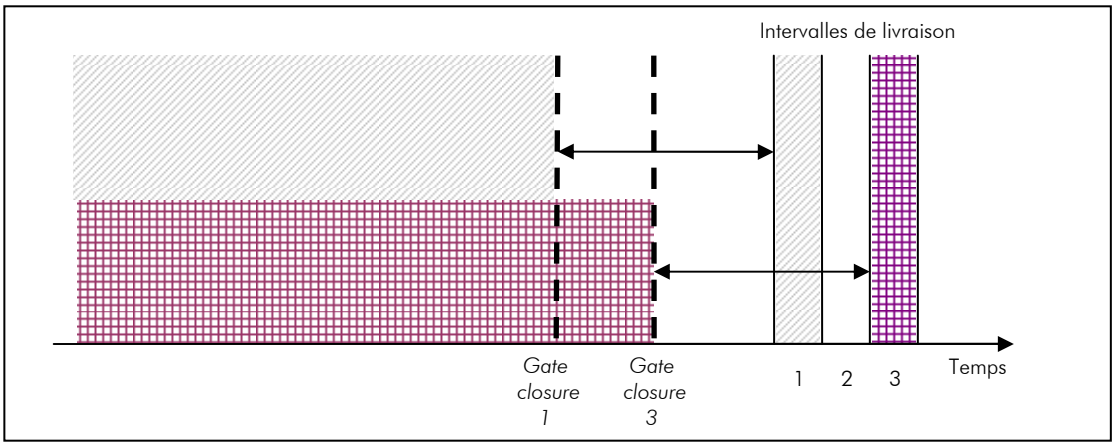

Figure 3-3 : Schéma de *gate closure* type "*rolling*"

Nous avons étudié l'organisation des différents marchés *forward* d'énergie. Indépendamment du type d'architecture de marché, pendant cette étape *forward*, les participants réalisent des transactions par des moyens différents (marchés organisés, contrats bilatéraux, etc.). Parallèlement, les participants doivent, directement ou indirectement, communiquer les informations des programmes prévisionnels de production et consommation et des positions contractuelles nécessaires pour le gestionnaire du réseau. Tout ce processus continue jusqu'à la *gate closure* où les participants doivent déclarer leurs positions contractuelles et programmes de production et consommation définitifs. Etant donné qu'avant la *gate closure* les participants peuvent réaliser des nouvelles transactions et annuler ainsi des obligations engagées précédemment, toutes les transactions *forward* peuvent être agrégées dans une seule étape qui fini à la *gate closure*. Le niveau d'incertitude

présente sur le dernier marché *forward* dépend période de temps qui existe entre la *gate closure* et le temps réel. Après la *gate closure* les participants ne peuvent vendre ou acheter de l'électricité qu'en passant par le marché du temps réel centralisé par le gestionnaire du réseau.

1.1.2 Marché du temps réel

La séquence des marchés d'énergie se termine avec le module du temps réel. C'est le dernier moment où les participants peuvent acheter ou vendre de l'énergie pour chaque intervalle de livraison. Le « marché » proprement dit est une des variantes possibles pour le design du module du temps réel. Par simplicité nous utiliserons ici le terme « marché de temps réel » d'un sens beaucoup plus général.[139]

En temps réel, les producteurs et les consommateurs injecteront ou soutireront une certaine quantité d'énergie sur chaque intervalle de livraison. Cette quantité d'énergie, injectée ou soutirée, est mesurée par le gestionnaire du réseau. Afin de déterminer des volumes de « vente » ou d'« achat » sur le temps réel, le gestionnaire du réseau comptabilise les déviations entre les mesures d'injection et de soutirage et les volumes contractuels qui lui ont été notifiés avant la *gate closure*.

La figure 3-4 montre un exemple de transactions du temps réel pour le cas d'un producteur. Pour un volume de ventes *forward* donné, le producteur achète en temps réel si la quantité injectée sur le réseau est plus petite que le volume des ventes *forward*. En revanche, le producteur vend en temps réel si la quantité injectée sur le réseau est plus grande que le volume des ventes *forward*. La figure 3-5 montre un exemple de transactions du temps réel pour le cas d'un *retailer* ou revendeur. Pour un volume d'achats *forward* donné, le revendeur achète (ou vend) en temps réel si la quantité d'énergie soutirée sur le réseau est plus grande (ou plus petite) que le volume d'achats *forward*.

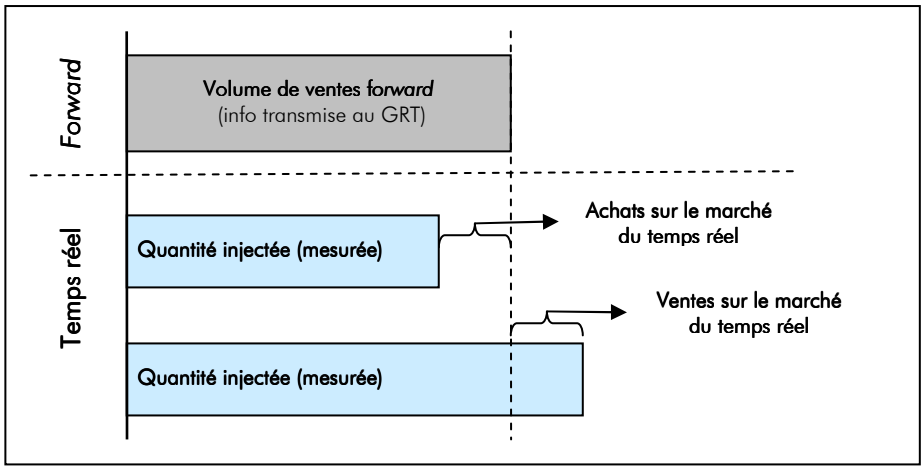

Figure 3-4 : Ventes et achats sur le marché du temps réel pour un producteur

[139] Dans la section suivante nous nous concentrerons sur les différents types de design possibles.

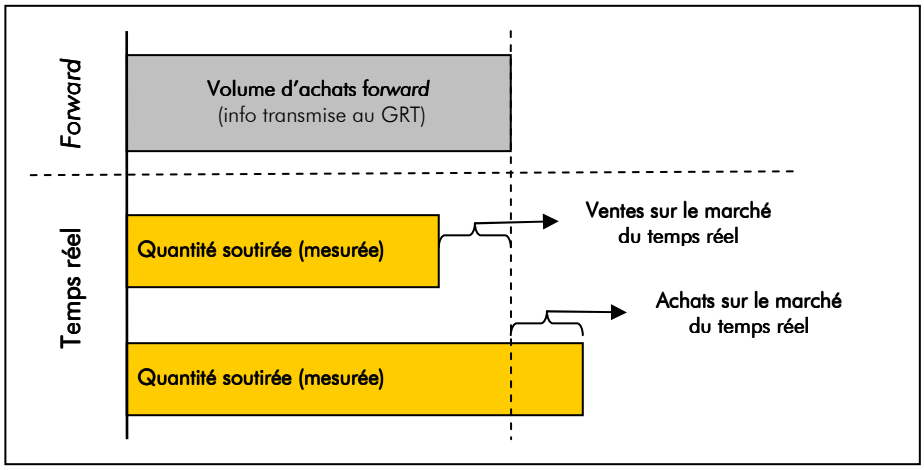

Figure 3-5 : Ventes et achats sur le marché du temps réel pour un consommateur

Pour un *trader*, qui ne possède pas de moyens de production ou auquel aucun point de soutirage n'est associé, la position physique en temps réel est toujours nulle. En conséquence, les positions contractuelles prises par un trader sur les marchés *forward* sont renversées complètement en temps réel (figure 3-6).[140]

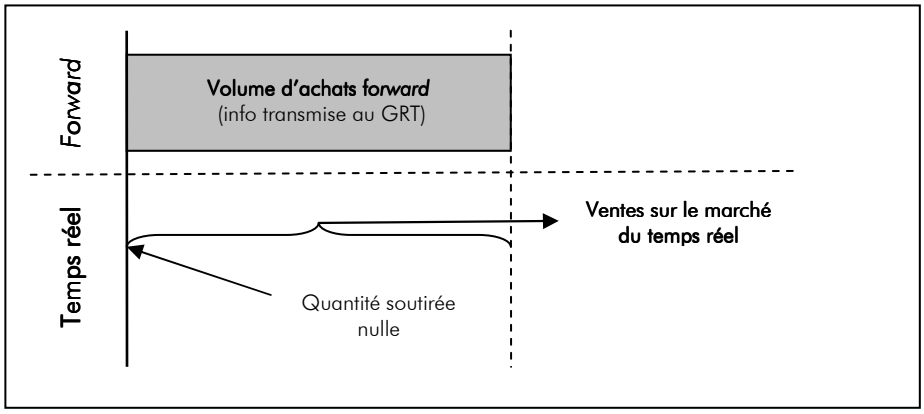

Figure 3-6 : Ventes sur le marché du temps réel pour un trader

Nous verrons dans la section 2 que les ventes et les achats en temps réel peuvent se faire de différentes manières, caractérisées comme « actives » et « passives ». La manière « active » d'acheter ou vendre de l'énergie sur le marché du temps réel se fait par acceptation d'offres d'ajustement. Si, par exemple, un producteur propose une offre d'ajustement à la hausse et si cette offre est acceptée, il « vend » de l'énergie en temps réel. Pour le cas de l'acceptation d'une offre d'ajustement à la baisse, il « achète » en temps réel. La manière « passive » d'acheter ou de vendre de l'énergie en temps réel est d'injecter ou de soutirer une quantité différente de celle déclarée comme position contractuelle. Ce dernier type de transactions du temps réel est connu sous le nom des « écarts » (« écarts positifs » pour les ventes « passives » du temps réel, et « écarts

[140] Ce type de transaction peut être limité par des règles spécifiques. Par exemple, en exigeant des « programmes équilibrés » ou en empêchant la participation d'acteurs qui n'ont pas une présence physique sur le marché (Saravia [2003]).

négatifs » pour les achats « passifs » en temps réel). Cette différenciation est importante car la manière dont ces différents types de transactions sont valorisés est une des principales options de design pour le module du temps réel (cf. section 2).

Nous avons présenté l'organisation de la séquence des marchés d'énergie dans différentes architectures de marché simplifiées (architectures « centralisée » et « décentralisée »). Ces types d'architectures utilisent différentes variantes de design pour chacun des modules de la séquence (marché *day-ahead*, marchés infrajournaliers, et marché du temps réel). Mises à part toutes les caractéristiques particulières d'organisation qui différencient ces architectures, une représentation générale de la séquence peut être dégagée. En effet, la séquence de marchés d'énergie peut être séparée en deux étapes principales, l'étape *forward* et l'étape de temps réel. Les transactions réalisées dans chacune de ces étapes correspondent au même bien physique, livré en temps réel, mais elles sont réglées séparément. Ceci est connu comme le système de « double règlement » (*two-settlement system*) (Stoft [2002])[141]. Dans ce système de règlement, les transactions *forward* sont réglées au prix *forward* et les transactions du temps réel (déviations des injections ou des soutirages par rapport aux transactions *forward*) sont réglées au(x) prix du temps réel. Dans la sous-section suivante (1.2) nous nous concentrerons sur les liens qui existent entre ces deux marchés de l'énergie.

1.2 Liens entre marchés et arbitrage temporel

Les marchés d'énergie composant la séquence sont liés car ils traitent un même bien qui sera délivré en temps réel. En effet, ces marchés sont liés par le fait qu'un arbitrage entre eux est possible. Cependant, les relations liant ces marchés d'énergie seront différentes de celles qui gouvernent les marchés de biens stockables. Sur les marchés de biens stockables, on peut réaliser des transactions d'arbitrage entre biens livrés sur des périodes de temps différentes (par exemple, on peut acheter une tonne de blé aujourd'hui et la revendre après demain). Des transactions de ce type ne sont pas possibles pour l'électricité, par exemple, il est impossible d'acheter un 1 MWh livrable à 4 heures, puis le revendre comme un bien livrable à 20 heures. L'arbitrage pour l'électricité ne s'effectue donc que sur des transactions réalisées à des moments différents mais concernant toutes le même bien physique, délivré exactement au même moment en temps réel. Par exemple, en achetant sur le marché *day-ahead* 1 MWh livrable à 4 heures du lendemain et en revendant sur le marché du temps réel ce 1 MWh à 4 heures du lendemain.

Les marchés *forward* de court terme sont donc des marchés financiers et non des marchés « physiques ». Comme l'électricité est non-stockable, c'est seulement en temps

[141] Le sens normalement donné au « double règlement » dans la littérature est restreint aux architectures centralisées qui prennent en compte le marché *Day-Ahead* et le temps réel pour faire le double règlement. Nous utilisons nous-mêmes un sens plus général en l'élargissant à toutes les architectures prenant en compte deux étapes, une étape *forward* (avant la *gate closure*) et une autre étape en temps réel.

réel qu'on connaîtra les vraies conditions de livraison du bien. Les obligations créées entre un producteur et un consommateur sur le marché *forward*, avant le temps réel, sont strictement financières. Le vrai marché physique, celui où le bien est délivré et où les conditions économiques sont connues, est le marché du temps réel (Hirst [2001]). Il représentera donc le « sous-jacent » en termes financiers et, logiquement, les marchés *forward* devraient être fondés sur des anticipations de ce dernier marché qui est celui du temps réel. La possibilité d'arbitrage créera des liens économiques directs entre ces marchés.

Afin de comprendre la mécanique générale de l'arbitrage temporel, base des liens entre les marchés, nous présentons d'abord l'arbitrage en général, et plus particulièrement l'arbitrage temporel. Puis nous analysons comment l'arbitrage temporel est réalisé dans une séquence de marchés d'énergie.

1.2.1 L'arbitrage entre marchés

L'arbitrage consiste à gagner un profit certain en exploitant des différences relatives de prix entre différents marchés (par exemple, des différences de prix géographiques ou temporelles) (Shahidehpour *et al.* [2002]). Dans des environnements concurrentiels, on peut considérer qu'à l'équilibre il n'existe plus d'opportunités d'arbitrage. Cela peut être expliqué simplement par le fait qu'une seule opportunité restante d'arbitrage conduirait à de nouvelles transactions qui annuleraient la différence relative de prix entre les marchés. L'absence d'opportunité d'arbitrage entre marchés est connue sur le nom du principe de « non-arbitrage » ou « loi du prix unique ».

Pour illustrer ce point, considérons le cas de deux marchés (A et B) pour une même bien que présente la figure 3-7 (Siddiqui [2002]).

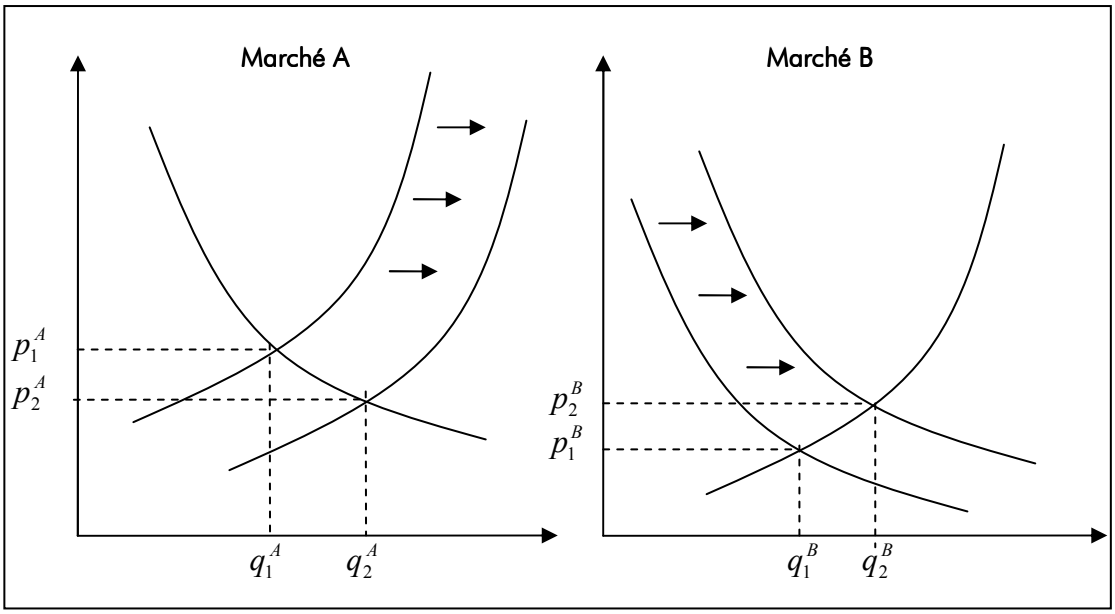

Figure 3-7 : Arbitrage entre marchés

Le prix du bien dans le marché A est supérieur au prix dans le marché B ($p_1^A > p_1^B$). En négligeant les coûts de transaction et le coût de transport, un arbitrageur pourrait acheter le bien dans le marché B à un prix bas (p_1^B) et la revendre dans le marché A à un prix haut (p_1^A) et gagner un profit de p_1^A - p_1^B pour chaque unité du bien commercialisée. Comme d'autres arbitrageurs s'aperçoivent de l'opportunité de réaliser des profits, ils se mettent à acheter du bien sur le marché B et à revendre sur le marché A. Ce processus continue jusqu'à ce que les conditions qui donnaient des opportunités d'arbitrage disparaissent. En termes économiques, les arbitrageurs augmentent la demande sur le marché B et provoquent une augmentation du prix (p_2^B). Simultanément, ils augmentent l'offre sur le marché A et provoquent une diminution du prix (p_2^A). Cette pression sur les prix continue jusqu'à ce que les prix soient égalisés, et alors toutes les opportunités d'arbitrages sont éliminées.

En conséquence, dans les environnements concurrentiels où toutes les opportunités d'arbitrages sont utilisées et où l'hypothèse de non-arbitrage est valide, on peut définir la relation reliant les prix de marchés différents comme étant $p^A = p^B$. Puisque les prix sur les deux marchés sont égaux le principe de non-arbitrage est aussi connu sous le nom de « loi du prix unique ».

Les liens entre marchés concernant un même bien mais placés sur des horizons de temps différents peuvent être basés aussi sur le principe de non-arbitrage. Mais, étant donné la présence des incertitudes, le principe de non-arbitrage a des caractéristiques différentes.

1.2.2 L'arbitrage temporel en présence d'incertitudes[142]

L'arbitrage temporel entre marchés concernant un même bien mais placés sur des horizons de temps différents est particulier quand il existe des incertitudes au moment de réaliser les transactions. A la différence de l'arbitrage présenté précédemment, qui considère les opportunités d'arbitrage présentant un profit certain, il peut exister des risques sur les profits associés aux arbitrages. Quand l'arbitrage se fait en présence d'incertitudes, il faut prendre en compte le comportement des participants face au risque.

Considérons tout d'abord le cas où tous les participants sont neutres au risque. C'est-à-dire qu'ils valorisent de la même manière un profit « certain » et un profit « espéré ».[143] Prenons un bien à livrer à la date t. Ce bien peut être échangé, vendu ou

[142] Le terme « incertitude » dans cette thèse est réservé aux incertitudes probabilisables donnant origine aux problèmes de risques. Nous ne traiterons pas les problèmes des incertitudes radicales (Knight [1921]).

[143] Par exemple, un participant de marché neutre au risque valorise également un gain certain de 10 ou une loterie où le lot est de 100 et la probabilité de gagner est du 0,1. Le gain espéré y est égal à 10 (100 x 0,1 + 0 x 0,9 = 10) et il est équivalent au gain certain (Varian [1995]).

acheté, sur un marché *forward* ouvert avant la date de livraison, $t-T$, ou sur un marché du temps réel (RT)[144] réalisé en t, au moment même de la livraison.

Dans le cas d'un environnement concurrentiel, où il existe une possibilité d'arbitrage entre marchés, la relation entre les prix correspondant à ces marchés s'exprime par l'équation suivante : $p^F = E[p^{RT}]$.[145] p^F est le prix sur le marché *forward*, et $E[p^{RT}]$ est l'espérance mathématique du prix du temps réel (p^{RT}). Cette expression se vérifie grâce au principe de non-arbitrage. Si, par exemple, le prix sur le marché *forward* était supérieur au prix espéré sur le marché du temps réel ($p^F < E[p^{RT}]$), les arbitrageurs en achèterait sur le marché *forward* pour revendre sur le marché du temps réel. Cette transaction d'arbitrage aurait un profit espéré de $E[p^{RT}]$ - p^F par unité de transaction. Au contraire, si $p^F > E[p^{RT}]$, les arbitrageurs changerait en leurs positions sur le marché *forward* afin d'acheter davantage sur le marché du temps réel.

Il est clair que les forces de l'arbitrage soutiennent l'équation $p^F = E[p^{RT}]$. Cette équation permet d'abord d'estimer et de valoriser les contrats *forward*. En plus, les acteurs du marché, participant au marché *forward* pour arbitrager la différence relative des prix, concentrent toute l'information possédée à un moment donné sur les prix du marché du temps réel. Une bonne information agrégée sur les conditions du marché du temps réel est très importante parce que très utile pour la coordination de la production (Counsell-Evans [2004]).

Si on considère le cas plus général où les acteurs sont averses au risque, l'expression qui relie les prix entre les marchés peut être adaptée (modèle du type « *cost-of-carry* »).[146] Cette adaptation nécessite une nouvelle hypothèse relative à la possibilité de stockage du bien. En conséquence, pour les biens non-stockables comme l'électricité, et en présence d'aversion au risque, les modèles du type « *cost-of-carry* » et le principe de non-arbitrage ne sont pas directement applicables (Bessembinder-Lemmon [2000], Vehviläinen [2004]). Cependant, comme nous verrons, l'équation $p^F = E[p^{RT}]$ reste une bonne approximation pour comprendre l'arbitrage entre la séquence de marchés d'électricité, plus particulièrement des marchés d'énergie.

---

[144] En général, dans la littérature, le marché qui a lieu au moment même de la livraison prend le nom de marché *spot*. Nous utilisons ici le terme temps réel pour garder une cohérence avec l'arbitrage temporel pour l'électricité.
[145] Des hypothèses supplémentaires sont nécessaires : absence de barrières à l'arbitrage, absence de coûts de transaction, absence de coûts de stockage ou facteur d'actualisation, absence de pouvoir de marché des arbitrageurs (Borenstein *et al.* [2004]).
[146] Dans ce cas, le prix sur le marché *forward* correspond à la valeur espérée du prix du temps réel calculée sous la base de « probabilités neutres au risque » (*risk neutral probabilities*) (Chance [2002]).

1.2.3 L'arbitrage temporel sur la séquence des marchés d'énergie

Dans le cas de l'électricité, l'arbitrage temporel se réalise d'une manière particulière. Indépendamment du type d'architecture de marché, la séquence de marchés d'énergie peut être séparée en deux grandes étapes, l'étape *forward* et l'étape du temps réel (cf. sous-section 1.1). Chacune de ces étapes possède son propre règlement. Dans ce système de règlement, les transactions *forward* sont réglées au prix *forward*, tandis que les déviations par rapport aux transactions *forward* qui sont les transactions du temps réel seront réglées au prix du temps réel.

En conséquence, les actions d'arbitrage entre les marchés de la séquence consistent à opérer des transactions sur le marché *forward* ou à modifier le volume de transactions par rapport aux quantités injectées ou soutirées en temps réel. Par exemple, le profit d'un producteur sera déterminé par l'équation suivante : $p^F q^F + p^{RT} q^{RT}$ (où q^F, q^{RT}, p^F et p^{RT} sont respectivement les quantités et les prix sur les marchés *forward* et sur le marché du temps réel). Les transactions du temps réel correspondent à la différence entre l'injection d'énergie en temps réel et la position contractuelle sur le marché *forward* ($q^{RT} = q^{RT}_{injectée} - q^F$). S'il existe une grande différence entre le prix *forward* et l'espérance mathématique du prix du temps réel (par exemple $p^F << E[p^{RT}]$) le producteur peut participer au marché *forward* pour avoir une position contractuelle telle que $q^{RT} > 0$ ($q^F < q^{RT}_{injectée}$) et gagner ainsi un profit espéré de q^{RT} x ($E[p^{RT}]$ - p^F).[147]Ces transactions d'arbitrage devraient rapprocher le prix *forward* de l'espérance du prix du temps réel.

Diverses études examinent les relations entre les prix *forward* (*day-ahead*) et les valeurs espérées du prix du temps réel pour le cas de l'électricité (Hirst [2001], Siddiqui [2002], Borenstein *et al.* [2004], Longstaff-Wang [2004], Potomac [2006], Hadsell-Shawky [2006]). Ces études montrent d'abord une forte corrélation entre ces deux valeurs. Ce qui montre qu'il existe des arbitrages entre ces marchés, que les prix *forward* prennent en compte les conditions de prix du temps réel ($E[p^{RT}]$ est une première approximation de p^F). Ceci est important parce que cette information est utile pour la coordination des moyens de production (Counsell-Evans [2004]). Cependant, ces études montrent aussi qu'il existe des différences systématiques et persistantes entre les prix *forward* et les valeurs espérées de prix du temps réel (p^F - $E[p^{RT}]$). Ces différences sont normalement connues comme un « *premium* » dans le jargon des marchés *forward*. Les conclusions générales des études sont que ce *premium* est dû à trois facteurs : l'aversion au risque, les caractéristiques de design de l'architecture de marché et le pouvoir de marché des arbitrageurs.

---

[147] Notez que ce producteur peut présenter, à la *gate closure*, un programme prévisionnel de production correspondant à la quantité injectée en temps réel ($q^{programme}_{injectée} = q^{RT}_{injectée}$) et permettre ainsi au GRT d'évaluer correctement la sécurité du réseau.

L'aversion au risque des acteurs est le facteur principal pour expliquer le *premium* (Bessembinder-Lemmon [2000], Siddiqui [2002], Longstaff-Wang [2004], Karakatsani-Bunn [2005], Hadsell-Shawky [2006]). Cette caractéristique d'aversion au risque sera plus importante dans l'électricité étant donné la volatilité importante des prix en temps réel, provoquée par les incertitudes et l'inélasticité en temps réel de l'offre et de la demande (cf. Chapitre 1). Le fait que l'électricité ne soit pas stockable et la présence d'aversion au risque empêchent l'utilisation de modèles du type « *cost-of-carry* ». Les modèles d'équilibre présentent donc une alternative pour étudier l'arbitrage temporel entre marché pour les biens non-stockables (Bessembinder-Lemmon [2000]). Dans ces modèles plus complexes, une représentation de la valorisation des risques par les acteurs est nécessaire, c'est la fonction d'utilité (Bessembinder-Lemmon [2000], Siddiqui [2002]). Nous reviendrons sur ces modèles dans le chapitre 4.

Les différences systématiques entre les prix *forward* et les prix du temps réel peuvent venir de caractéristiques de design (Saravia [2003], Potomac [2006]). Les différences entre prix peuvent être provoquées par des designs imposant des barrières administratives à l'arbitrage (par exemple, exigeant des « programmes équilibrés » où les positions contractuelles doivent correspondre aux productions ou consommations prévues ou interdisant les offres « virtuelles »), créant des coûts de transaction artificiels ou des distorsions sur les prix du temps réel (par exemple avec des pénalités), etc.

Enfin, Borenstein *et al.* [2004] rendent compte de différences importantes entre les prix *forward* et les prix du temps réel pour le marché californien pendant les années 2000-2001. Ils expliquent ces différences, et les inefficacités d'arbitrage qui leur correspondent, par la manipulation des marchés *day ahead* et du temps réel de la part des arbitrageurs. En fait, ils considèrent que, pour le cas californien, seulement un ou deux arbitrageurs réalisaient des transactions d'arbitrage, et que leur stratégie optimale ne supposait pas l'arbitrage « complet » des marchés où le prix *forward* égale le prix espéré du temps réel.

Nous avons étudié les liens existant entre les différents modules de la séquence de marchés d'énergie. Nous avons montré que les marchés composant cette séquence sont liés car ils traitent le même bien qui sera livré en temps réel. L'arbitrage temporel est donc la base des liens entre les marchés. Les actions d'arbitrage entre les marchés de la séquence pour l'électricité consistent à opérer des transactions sur le marché *forward* ou à modifier le volume de transactions par rapport aux quantités injectées ou soutirées en temps réel. Une première approximation de la relation qui lie les différents marchés placés à différents horizons de temps est $p^F = E[p^{RT}]$, le prix *forward* est équivalent à l'espérance mathématique du prix du temps réel. Bien que la caractéristique de non-stockabilité de l'électricité ne permette pas l'utilisation directe de cette relation, elle rappelle l'importance des prix du temps réel sur toute la séquence de marchés car les prix et les transactions *forward* vont se baser principalement sur les prix des transactions du temps réel.

1.3 Conclusions de la section

Cette section a porté sur l'organisation de la séquence de marchés d'énergie et sur les liens existants entre ces marchés.

A court terme, la séquence de marchés d'énergie commence par les marchés *forward* (marché *day-ahead* et marchés infrajournaliers) et finit par le marché du temps réel. Nous avons d'abord présenté la séquence de marchés d'énergie pour différentes architectures de marché simplifiées (architectures « centralisée » et « décentralisée »). Ces architectures utilisent différentes options de design pour chacun des modules de la séquence (marché *day-ahead*, marchés infrajournaliers et marchés du temps réel). En mettant à part toutes les caractéristiques particulières d'organisation qui différencient ces architectures, une représentation générale de la séquence est dégagée. En effet, la séquence de marchés d'énergie peut être représenté par deux étapes principales, l'étape *forward* et l'étape du temps réel. Ces étapes sont séparées par la *gate closure*. Avant la *gate closure*, les marchés d'énergie sont dits *forward*. Après la *gate closure*, le module du temps réel est en charge de l'équilibrage du système et de la détermination des signaux de prix correspondants. Les transactions réalisées dans chacune de ces étapes correspondent au même bien physique, livré en temps réel, mais elles sont réglées séparément. Dans ce système de règlement, les transactions *forward* sont réglées au prix *forward* et les transactions du temps réel (déviations des injections ou des soutirages par rapport aux transactions *forward*) sont réglées au prix du temps réel. Cette séparation est importante car le règlement entre participants pour les étapes *forward* peut être déterminé librement sur les marchés *forward* (fixation des prix et des quantités). En revanche, les transactions réalisées dans le module du temps réel (offres d'ajustement ou écarts) seront réglées financièrement à partir de règles particulières déterminées dans ce module.

Les marchés composant cette séquence sont liés car ils traitent le même bien qui sera livré en temps réel. Ces liens économiques sont créés par la possibilité d'un arbitrage temporel entre les étapes. Sur une séquence de marchés d'un bien stockable, la relation qui lie les différents marchés placés à différents horizons de temps est $p^F = E[p^{RT}]$ (prix *forward* équivalent à l'espérance mathématique des prix du temps réel). Déterminer la relation exacte entre les prix *forward* et les prix du temps réel pour l'électricité, non-stockable, nécessitera des modèles plus complexes comme nous le verrons dans le chapitre 4. Cependant, cette relation reste une bonne approximation pour comprendre l'arbitrage entre la séquence de marchés d'électricité et la base des liens entre les marchés de la séquence. Remarquons ici l'importance d'étudier les caractéristiques du design du dernier marché de la séquence. Comme les prix et les transactions *forward* vont y être fondés principalement sur les prix des transactions du temps réel, nous nous concentrerons sur les caractéristiques de design influençant la détermination des prix des différentes transactions du temps réel. Ces caractéristiques peuvent avoir des conséquences décisives sur les autres marchés et sur l'arbitrage. La section 2 de ce chapitre portera donc sur le design du module du temps réel.

Section 2: Design du module du temps réel

Le module du temps réel est fondamental dans toute architecture du marché et il nécessite une grande d'attention pour sa conception. C'est en temps réel que les échanges du bien « électricité » se réalisent, et la coordination entre les composantes du bien électricité (énergie, transport et réserves) est déterminante pour le fonctionnement du système électrique. Le design de ce module n'est pas simple car il s'agit d'un « pseudo marché » combinant des caractéristiques d'outil de contrôle et des caractéristiques de marché (cf. Chapitre 2, section 1).

Dans un environnement concurrentiel, le design du module de temps réel devrait remplacer l'outil d'*optimal dispatch* de l'entreprise intégrée (cf. chapitre 1). Cependant, cet outil de contrôle ne peut pas être remplacé par un pur marché (cf. chapitre 2). En effet, le remplacement doit être opéré par un arrangement combinant des éléments de contrôle centralisé et des propriétés d'un marché. C'est donc un « pseudo marché » géré de manière centralisée par le Gestionnaire du Réseau (Boucher *et al.* [2006]). Ce pseudo marché fonctionne d'abord comme une enchère, dont le résultat est mis en œuvre rapidement par un mécanisme d'ordres (contrôle) sur la production des participants. Ensuite, il génère des signaux des prix correspondant aux conditions économiques du temps réel.

Il est important de remarquer que ce module de temps réel est le dernier de la séquence de marchés, et qu'il est chargé de « corriger » le système à partir d'un point d'«équilibre » antérieur donné par les marchés *forward*. Bien que les signaux de prix donnés par le marché du temps réel ne concernent que les déviations par rapport à l' « équilibre *forward* », ces signaux influencent nécessairement tous les autres marchés. Car, du point de vue des marchés, le module du temps réel sera la seule place de marché ayant une nature « physique » puisque c'est en temps réel que l'échange se concrétise. Les marchés qui fonctionnent auparavant ont, en fait, des caractéristiques de marché financier (ou marché *forward*) et seront donc fondés sur les bases des signaux du temps réel (cf. section 1).

Cette deuxième section portera sur les caractéristiques de design d'un module du temps réel. Nous nous concentrerons principalement sur la composante « énergie ». Les deux autres composantes, transport et réserves, ne seront traitées que secondairement.

Pour dégager les paramètres de design capables de différencier les variantes de design, dans la section 2.1, nous présentons d'abord les trois étapes caractéristiques dans tout module du temps réel. Cela nous permettra de distinguer, dans la sous-section 2.2, deux types de designs : d'une part, le « marché du temps réel » - proprement dit - et, d'autre part, le « mécanisme d'ajustement ». Dans la sous-section 2.3, nous réaliserons des études de cas de divers designs de modules du temps réel mis en œuvre (Texas, Angleterre, Belgique, France et Pays Bas). L'analyse des différents designs nous permettra de montrer que la principale différence entre ces designs vient de la manière dont certaines transactions

du temps réel sont réglées financièrement, notamment en pénalisant avec différentes intensités certains types de transaction. Or, le choix entre ces différentes méthodes de règlement des transactions du temps réel sera déterminant puisqu'il fixera les incitations données aux participants du marché et influencera ainsi leur comportement en générant différentes conséquences économiques. En conséquence, chaque type de design aura des propriétés économiques particulières ce qui conduit à une comparaison entre les designs.

Finalement, dans la sous-section 2.4, nous verrons que cette comparaison devrait considérer des conséquences économiques de différent ordre qui seront présentées en trois volets : l'impact du design du module du temps réel sur la séquence des marchés, la contribution à la sécurité du réseau, la résistance au *gaming* et au pouvoir de marché. Nous verrons que, parmi ces différentes conséquences possibles, ce sont celles de l'impact des règles du module du temps réel sur la séquence de marchés qui constituent le cœur de l'analyse économique de l'organisation des marchés. Elles seront donc, pour cette raison, placées au cœur de notre modélisation du quatrième chapitre.

2.1 Etapes du module du temps réel

Pour comprendre les caractéristiques particulières des différentes variantes de design pour le module du temps réel, nous étudions tout d'abord le fonctionnement du module en le découpant en trois étapes.[148] La première étape est celle de la *gate closure* et de l'équilibrage du système (acceptation des offres d'ajustement) ; la deuxième étape couvre le règlement et la détermination des prix pour les transactions « actives » du temps réel ; finalement, la troisième étape correspond au règlement et à la détermination des prix pour les transactions « passives » (ou « écarts ») du temps réel.[149]

2.1.1 *Gate closure* et équilibrage du système

La fermeture des marchés *forward*, ou « *gate closure* », marque le début du module du temps réel. Nous avons vu qu'elle peut être définie de différentes manières, en relation aux périodes où les participants sont autorisés à modifier leurs déclarations de transactions et de positions physiques (cf. section 1). Cette autorisation peut être arrêtée à différentes franges horaires (« guichets »), ou être continue jusqu'à une limite de temps fixée avant la période de livraison (« *rolling gate closure* »). Indépendamment de la forme de la *gate closure*, la durée qui la sépare du temps réel est un paramètre important.[150] Ce paramètre

[148] Il est important de noter que cette séparation en étapes est utilisée pour faciliter la compréhension de l'exposé. En effet, ces trois étapes peuvent se dérouler, au moins en partie, simultanément.
[149] Nous avons différencié les transactions du temps réel entre « actives » et « passives » (cf. chapitre 3, section 1). Les transactions « actives » correspondent aux offres d'ajustement acceptées, alors que les transactions « passives » viennent des différences entre les positions contractuelles et les injections ou soutirages en temps réel (après la prise en compte des offres d'ajustement acceptées).
[150] Le durée entre la *gate closure* et l'intervalle de livraison est aussi connue sous le nom de « délai de neutralisation » (CRE [2005b]).

détermine le niveau d'information des participants sur les marchés *forward* et, en conséquence, le niveau d'incertitude à gérer dans le module du temps réel.

La première étape du module du temps réel commence ainsi à la *gate closure*. Le GRT y reçoit deux types d'informations qui sont, d'une part, les positions contractuelles des participants aux marchés *forward* et, d'autre part, les programmes définitifs de production et de consommation des participants (ou des « responsables d'équilibre »). Ces positions contractuelles et ces programmes de production et consommation sont présentés pour chaque intervalle de livraison ou de règlement.[151] Cet intervalle peut avoir différentes durées entre 10 minutes et 1 heure.[152]

Lors de la présentation de programmes et de positions contractuelles définitifs, les responsables d'équilibre peuvent présenter des offres d'ajustement. Les offres d'ajustement expriment, par le biais d'un prix et un volume, la disposition de chaque participant à modifier son programme de production (ou de consommation). Ces offres peuvent être « à la hausse » (correspondant à une augmentation de la production ou une diminution de la consommation) ou « à la baisse » (correspondant à une diminution de la production ou une augmentation de la consommation). Ces offres sont simples (prix et quantité). Mais elles peuvent aussi être accompagnées de paramètres techniques (par exemples, taux de *ramping* ou d'autres caractéristiques des unités de production). Les offres d'ajustement peuvent être soumises avant la *gate closure* et/ou venir des autres marchés de la séquence (marché *day ahead*).

Peu avant l'intervalle de livraison, en quasi-temps réel, le GRT doit ramener le système à l'état d'équilibre (figure 3-8).[153] Cet équilibrage se réalise près du temps réel car, à ce moment là, il y a peu d'incertitude sur la prévision du déséquilibre du système pour l'intervalle de livraison.[154] Si la durée de l'intervalle de livraison est longue, il se peut que le système doive être équilibré plusieurs fois au long de l'intervalle.[155] L'équilibrage doit être réalisé de la manière la plus efficace possible.[156] Pour choisir les offres d'une manière optimale le GRT se sert d'un algorithme d'optimisation (un *merit order*, dans les cas les plus

---

[151] Dans certains cas les programmes de production peuvent être demandés avec un pas de temps plus court que l'intervalle de livraison (par exemple le NETA britannique et la France). L'intervalle de livraison sera utilisé pour le règlement des transactions du temps réel. Par simplicité nous considérerons les mêmes intervalles pour les deux types d'informations.

[152] Par exemples cet intervalle est 30 minutes en France et en Angleterre, 15 minutes au Texas, 10 minutes en Californie et 1 heure au nord-est des Etats-Unis (PJM, NY Pool, etc.).

[153] Si des déséquilibres importants sont prévus pour le temps réel, par exemple la perte d'une unité de production de grande capacité, des actions d'équilibrage peuvent être entreprises bien avant le temps réel, et même avant la *gate closure*. Car certaines offres d'ajustement ou certaines utilisations de réserves achetées peuvent prendre plus de temps pour être déployées.

[154] Bien que l'état du système puisse changer entre le moment de la prise des décisions de l'équilibrage et l'équilibrage même, ces variations sont minimes et seront amorties par les réglages primaire et secondaire.

[155] Par exemple PJM utilise un intervalle de livraison d'une heure, mais l'équilibrage économique (*dispatch*) du système est effectué toutes les 5 minutes (Stoft [2002]).

[156] Nous considérons que le GRT a les incitations nécessaires pour assurer la minimisation des coûts d'équilibrage (y compris les coûts venants de l'achat de réserves).

simples). Cet algorithme d'optimisation peut être très similaire à l'outil d'*optimal dispatch*. Dans le chapitre précédent, nous avons vu que le modèle d'*optimal dispatch* pouvait être transformé en enchère en remplaçant par les offres d'ajustement des participants les entrées de l'outil de contrôle (cf. Chapitre 2, sous-section 2.4). L'algorithme ainsi modifié a comme objectif d'équilibrer le système à un coût minimal tout en respectant les contraintes de réseau.[157] Le résultat de cette enchère est une série d'offres d'ajustement acceptées qui se traduiront par des ordres de modification de programmes qui seront envoyées aux participants. Noter que le choix et l'acceptation d'offres est seulement une partie de l'enchère. La manière dont les offres acceptées sont réglées est l'autre partie importante de l'enchère, et elle sera présentée dans la sous-section suivante (2.1.2). Cette séparation entre acceptation et valorisation est importante car, dans la plupart des modules du temps réel, les procédés d'acceptation sont généralement de même nature. Ce qui n'est pas le cas pour la valorisation.

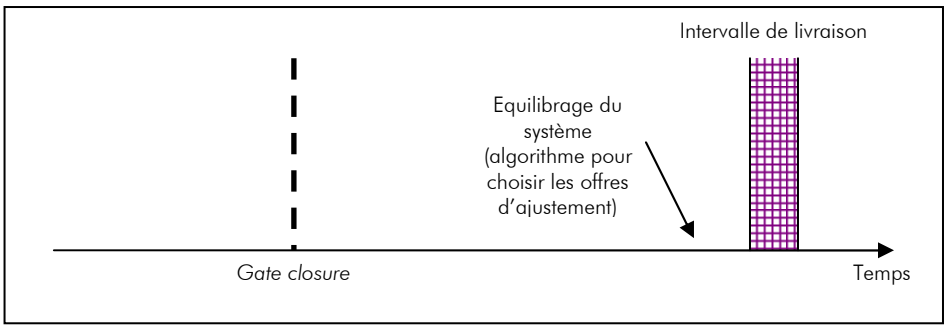

Figure 3-8 : Equilibrage du système

Une version simplifiée de cet algorithme, ne considérant que la contrainte d'équilibre, est équivalent à un *merit order*. Pour l'illustrer, considérons une version modifiée et simplifiée de l'algorithme d'optimisation d'*optimal dispatch* (ci-dessous).

$$\min_{\underline{\Delta q}} \sum PO^H \, \Delta q^H - PO^B \, \Delta q^B$$

Sous contraintes :

Contrainte d'équilibre :

$$\sum \Delta q^H - \sum \Delta q^B = D - \sum q^{programme} \quad (\lambda)$$

Contraintes des offres d'ajustement (unité de production)[158]

$$0 \le \Delta q^B \le q^{programme} - Q^{\min}$$

$$0 \le \Delta q^H \le Q^{\max} - q^{programme}$$

[157] Les contraintes d'équilibre, de transport, de réserves et des moyens de production. Dans ce chapitre nous ne traiterons que la contrainte d'équilibre.

[158] Remarquez que les capacités de production minimale et maximale pour chacune des unités de production ($Q^{\min}$, $Q^{\max}$) sont ajustées en accord avec les contraintes de *ramping*. Ces capacités de productions correspondent aux valeurs minimales et maximales techniquement atteignables pour chaque unité.

La production de chaque unité en temps réel est composée par la production programmée et les modifications venant des offres acceptées ($q^{programme} + \Delta q$). D est la consommation en temps réel[159]. Pour modifier leurs programmes sur cet intervalle, les participants ont proposé des offres à la hausse et à la baisse, respectivement aux prix PO^H et PO^B. Si le GRT détecte un déséquilibre ($D - \sum q^{programme} \neq 0$) l'algorithme d'optimisation doit ramener le système à l'équilibre avec un coût minimal. L'algorithme d'optimisation aide le GRT à choisir les offres à la hausse (Δq^H) et à la baisse (Δq^B) et, en considérant les prix proposés (PO^B, PO^H), à équilibrer le système à un coût minimal. Si la consommation est plus grande que la production programmée, le système est « court » en énergie, et le GRT devrait accepter des offres à la hausse pour l'équilibrer. Si la production programmée est plus grande que la consommation, le système est « long » en énergie, et le GRT devrait accepter des offres à la baisse pour l'équilibrer.

La sélection optimale des offres peut être réalisée par un *merit order* où les offres à la hausse et à la baisse des participants sont ordonnées de manière croissante et décroissante respectivement, en partant de l'équilibre préétabli par les programmes de production présentés à la *gate closure*. La figure 3-9 montre le fonctionnement d'un *merit order* pour le cas d'un système « court » en énergie. Les offres à la hausse (en bleu) ont été ordonnées de manière croissante à partir de la production totale programmée. Pour équilibrer le système, le GRT choisit donc les offres moins chères d'abord jusqu'à couvrir le volume du déséquilibre. Notez qu'après l'équilibrage, le prix de l'offre marginale ($PO^H_{\text{marginale}}$) coïncide avec le multiplicateur de Lagrange de la contrainte d'équilibre en énergie (λ^*) sortant de l'algorithme d'optimisation.

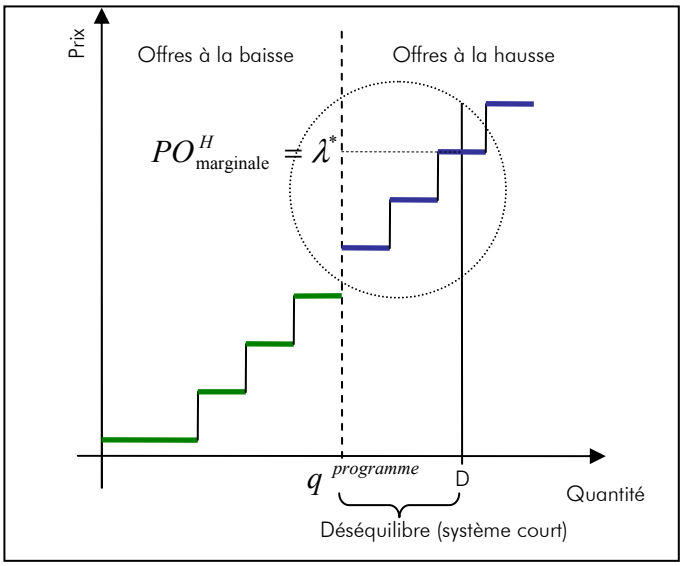

Figure 3-9 : *Merit order* représentant le module du temps réel

[159] D est, en fait, une estimation très précise de la consommation car elle est réalisée presque en temps réel. Elle prend aussi en compte les possibles variations de la production par rapport à ses programmes.

Même s'il n'y a pas de déséquilibre physique, certaines offres peuvent être acceptées. C'est le cas, par exemple, quand une offre a l'hausse à un prix inférieur à une offre à la baisse. Ce type d'acceptation des offres est bénéfique pour l'efficacité de la production car il permet le remplacement d'un moyen de production plus cher pour un autre moins cher (Sweeting [2000]). Ceci est plus important pour les architectures décentralisées puisque des possibilités d'arbitrages peuvent subsister au stade du temps réel à cause de coûts de transaction ou du manque d'information. Les offres acceptées sous ces conditions sont connues normalement comme offres acceptées par « arbitrage », représentant justement le GRT comme l'arbitrageur réalisant des transactions entre les participants (Henney [2002]).

Il est important de remarquer que le GRT ajuste continuellement l'état du système. Il utilise non seulement les offres d'ajustement déposées volontairement, mais peut aussi recourir à des options sur les réserves (qu'il a négociées à l'avance par des contacts bilatéraux ou aux marchés *forward* de réserves - cf. chapitre 2, sous-section 2.3).[160] Dans la plupart des designs, les réserves achetées (sur des marchés *forward* ou par le biais des contrats bilatéraux) sont obligées de rendre disponibles leurs capacités de production sur le module du temps réel (Chao *et al.* [2005]). En pratique, elles doivent soumettre des offres d'ajustement pour le déploiement des réserves en temps réel et, au moment de l'équilibrage, elles devraient être traitées comme les offres volontaires d'ajustement dans le *merit order* (cf. chapitre 2, sous-section 2.3). Dans d'autres designs, le GRT ne peut exercer des options acquises antérieurement pour différentes catégories de réserves que si les offres d'ajustement sont insuffisantes pour équilibrer le système, en quantité ou en qualité. Les réserves peuvent aussi être déployées en temps réel à cause de contraintes spécifiques. Par exemple pour un changement de la consommation très rapide, mais prévisible, à l'intérieur d'un intervalle de livraison (comme une pointe de consommation localisée).

Comme nous l'avons annoncé, la sélection et l'acceptation des offres d'ajustement est seulement une partie de l'enchère réalisée pour équilibrer le système. L'autre partie de l'enchère qui est la manière dont les offres acceptées sont réglées (transactions « actives ») sera traitée dans la sous-section suivante.

2.1.2 Détermination des prix et règlement des transactions « actives »

Il existe différentes variantes de règlement de transactions « actives ». Les offres acceptées par le *merit order,* et les variantes de design définiront en fait le type d'enchère. La première variante consiste à régler toutes les offres acceptées avec un seul prix. Ce qui correspond à une enchère « à prix uniforme » (Staropoli [2001]). Le prix utilisé pour régler toutes les offres acceptées est ici, le prix de l'offre marginale. Une deuxième variante consiste en une enchère discriminante où les transactions actives (c'est-à-dire, les offres

[160] La dernière option qu'a le GRT pour maintenir le réseau en fonctionnement est le délestage forcé des consommations. Cette option ne sera utilisée que quand la probabilité d'une panne totale (black out) est élevée.

acceptées, celles qui « gagnent » l'enchère) seront réglées aux prix qu'elles ont proposé. Ce type d'enchère est donc connu sous le nom de « *pay-as-bid* » (Staropoli [2001]).

La figure 3-10 considère les détails de l'exemple présenté précédemment (voir cercle « zoom » de la figure 3-9). Trois offres d'ajustement à la hausse, offres 1,2 et 3 avec des prix proposés PO_1^H , PO_2^H et PO_3^H , ont été acceptées. Sur une enchère à prix uniforme ces trois offres seront réglées au prix de l'offre marginale (dans notre exemple $PO_{\text{marginale}}^H = PO_3^H$). Sur une enchère discriminante chacune des offres acceptées recevra le prix qu'elle a proposé. L'offre 1 sera réglée au prix PO_1^H , l'offre 2 au prix PO_2^H et l'offre 3 au prix PO_3^H .

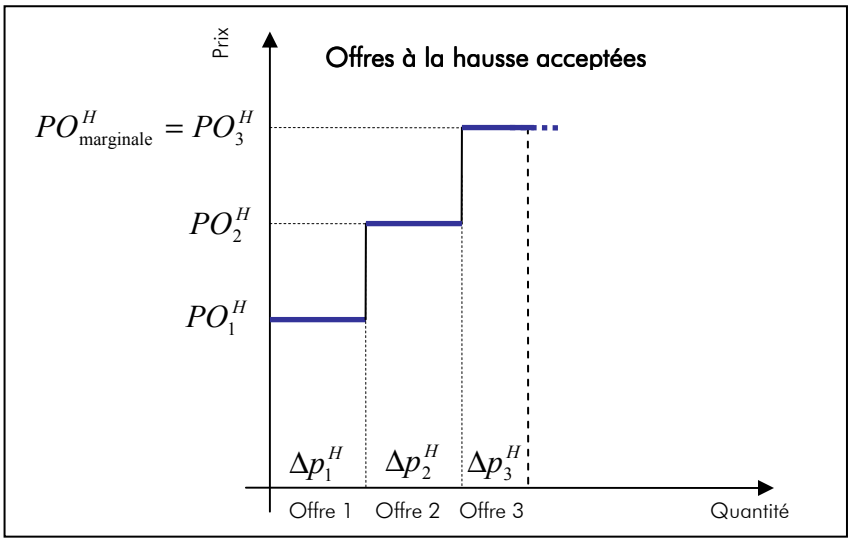

Figure 3-10 : Règlement des transactions actives (enchère discriminante ou à prix uniforme)

Les propriétés économiques de ces deux types d'enchères appliquées à l'électricité ont été largement étudiées (Staropoli [2001], Cramton-Stoft [2006b]). Un point important y est l'adaptation du comportement des participants aux différents types d'enchère. En fait, en fonction du type de valorisation des offres, les participants concevront différemment leurs offres pour pouvoir maximiser leur bénéfice. C'est la conclusion du « théorème d'équivalence du revenu » (Green-McDaniel [1999]). Selon ce théorème, en concurrence parfaite et sans incertitudes, sur une enchère à prix uniforme les participants réaliseront des offres correspondant à leurs coûts marginaux (dans notre exemple $PO_1^H < PO_2^H < PO_3^H$ comme montré dans la figure 3-10). Cependant, sur une enchère discriminante les participants réaliseront des offres correspondant au prix d'équilibre, donc l'offre marginale (dans notre exemple $PO_1^H = PO_2^H = PO_3^H$, figure 3-11). Finalement, dans ce cas idéal, le règlement des offres est équivalent pour les deux types d'enchères.

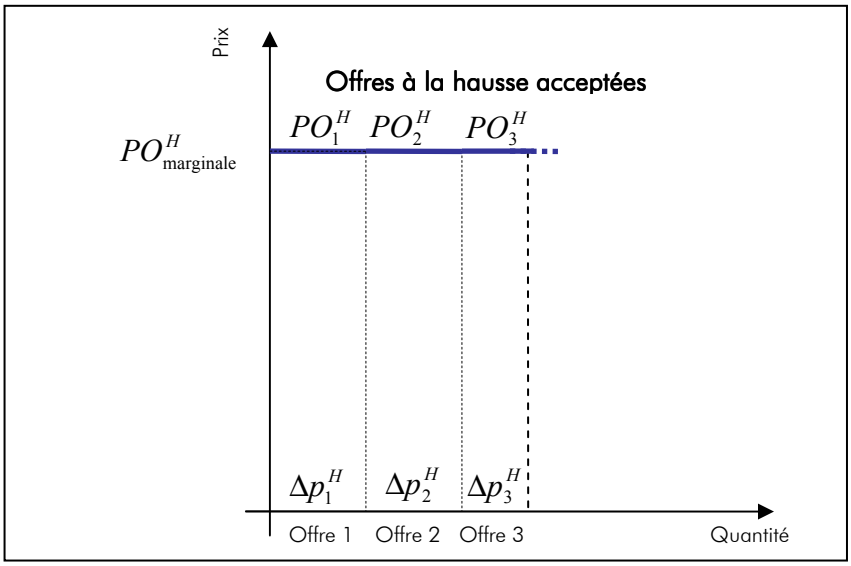

Figure 3-11 : Schéma des offres sur une enchère discriminante

Les résultats d'équivalence entre les deux types d'enchère ne sont plus maintenus si l'on sort des hypothèses de concurrence parfaite. Dans les cas plus réels, les études théoriques qui comparent les deux types d'enchère montrent que les résultats dépendent fortement des configurations et caractéristiques particulières des modèles utilisés (pouvoir de marché, incertitudes, asymétries d'information). En effet, ces caractéristiques conduisent des performances économiques particulières pour chaque variante du design et elles devraient être prises en compte pour le design du module du temps réel.[161] Néanmoins, une étude approfondie sur chacune d'elles irait au delà de notre objectif. Nous nous concentrerons, en revanche, sur les options du design de la troisième étape du module du temps réel, qui est la détermination des prix et le règlement des transactions passives. C'est sur ces options que se trouve la principale différence entre les différents designs du module du temps réel.

2.1.3 Détermination des prix et règlement des transactions « passives »

La troisième étape du module du temps réel porte sur la détermination des prix et le règlement des transactions « passives ». Ce type de transactions du temps réel est connu sous le nom d'« écarts ».

On achète (ou on vend) de l'énergie en temps réel de manière « passive » quand on injecte (ou on soutire) une quantité différente de celle déclarée comme position contractuelle à la *gate closure*.[162] En effet, les « écarts positifs » correspondent à des ventes

[161] Notez, par exemple, que si on ne considère que les problèmes d'information, l'équivalence des deux types d'enchères dépend de la capacité des participants à anticiper le prix d'équilibre au moment de réaliser des offres. L'information disponible et les incertitudes présentes au moment de réaliser les offres est donc un paramètre important (Green-McDaniel [1999]). Plus on s'approche du temps réel pour faire les offres, plus les résultats sont similaires dans les deux types d'enchère.

[162] Notez que, pour les participants réalisant aussi des transactions actives (offres d'ajustement), notamment les producteurs, les positions contractuelles sont ajustées aux volumes des offres acceptées.

« passives » du temps réel et les « écarts négatifs » correspondent à des achats « passifs » en temps réel.

La figure 3-12 montre un exemple pour un participant ayant une position contractuelle donnée (partie supérieure de la figure). Il sera en écart positif si la quantité d'énergie soutirée en temps réel est inférieure à sa position contractuelle. En revanche, il aura un écart négatif si la quantité d'énergie soutirée en temps réel est supérieure à sa position contractuelle.

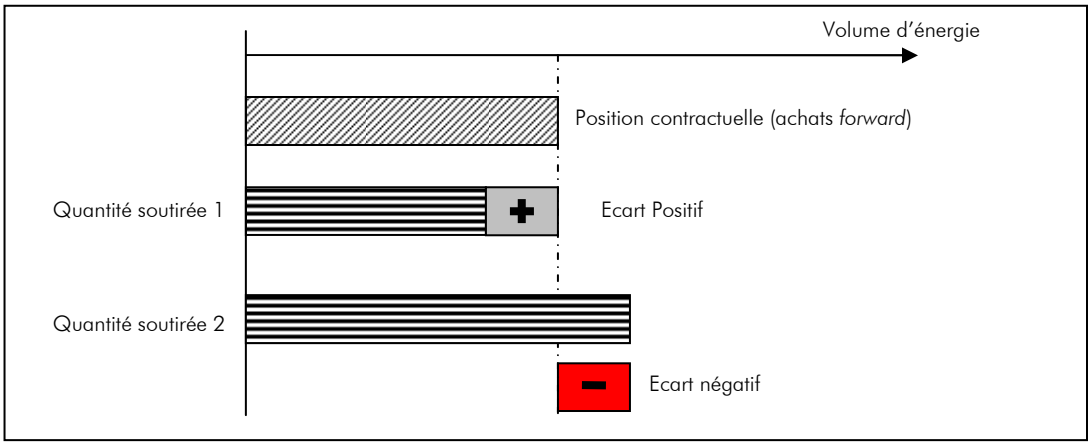

Figure 3-12 : Ecarts positifs et négatifs

Les variantes de design possibles pour cette étape se distinguent par le nombre de « prix » utilisés pour les différentes transactions et par leur forme de détermination. Nous distinguons ainsi deux variantes de design : le règlement des écarts « à prix unique » et le règlement des écarts « à double prix ».

2.1.3.1 Règlement des écarts « à prix unique »

La forme la plus simple pour définir les prix des transactions passives du temps réel est de définir un prix unique pour tous les écarts. Ce prix unique correspond souvent au prix de l'offre marginale acceptée pour équilibrer le système. Ce prix sera ainsi le prix de l'énergie en temps réel.[163]

Des systèmes de règlement à prix unique ont été mis en œuvre dans les architectures du Texas, de Norvège, PJM, NY Pool, etc. Dans ces architectures, tous les écarts sont valorisés au prix de l'énergie en temps réel. Par exemple, les participants ayant un écart négatif devront payer au GRT d'après le volume de l'écart multiplié par le prix du temps réel. Les participants ayant un écart positif seront payés par le GRT le montant correspondant au volume de l'écart multiplié par le prix du temps réel. Nous verrons plus

[163] Généralement, on combine un règlement à prix unique avec des enchères à prix uniforme. En conséquence, dans cette configuration, tant les transactions actives (offres acceptées) et les transactions passives (écarts) sont réglées avec le même prix du temps réel (cf. « marché du temps réel, sous-section 2.2).

tard que le prix unique pour toutes les transactions de temps réel est la caractéristique déterminante pour un design du type « marché » dans le module du temps réel.

2.1.3.2 Règlement des écarts à « double prix »

Une deuxième option pour le règlement des écarts consiste à utiliser des prix différents, deux en général, pour les écarts différents (positifs ou négatifs). Ce type de règlement comporte, tout d'abord, une différenciation marquée entre le règlement des transactions « actives » (offres acceptées) et les écarts. Mais encore, et c'est plus important, cette forme de règlement différencie les types d'écart pour déterminer les prix correspondant et pour réaliser le règlement. Par exemple un écart positif (surplus d'énergie en temps réel) et un écart négatif (déficit d'énergie en temps réel) seront valorisés différemment selon les règles déterminées par le système de règlement.

Les designs des systèmes de règlement mis en place sont très divers, variant surtout sur la forme de fixation des prix des écarts positifs et négatifs dans les différentes conditions du système. Nous donnons par la suite une représentation générale d'un système de règlement « à double prix ». Les détails particuliers des designs concrets seront donnés dans la sous-section 2.3.

Les prix des écarts peuvent être basés différemment : sur les offres acceptées pour l'équilibrage pour chaque intervalle de livraison ; sur un tarif préfixé et indépendant des actions d'équilibrage ; ou sur des prix correspondant à des marchés (*forward) day-ahead* (ETSO [2003]). Ces variantes sont généralement combinées dans les systèmes réels de règlement. Les prix des écarts basés sur les offres d'équilibrage acceptées peuvent être calculés en utilisant le prix de l'offre marginale ou un prix moyen pondéré (en volume et/ou durée) des offres acceptées. Le choix d'une de ces deux variantes est souvent lié au type d'enchère utilisée (règlement des offres ou transactions actives). Généralement, les enchères à prix uniforme sont combinées avec des prix des écarts calculés sur la base de l'offre marginale. Et les enchères discriminantes utilisent souvent une détermination des prix des écarts basés sur des prix moyens.[164]

Indépendamment de la méthode utilisée, prix de l'offre marginale ou prix moyen, le processus de détermination des prix des écarts commence avec la détermination de la « direction » du déséquilibre net du système. Le déséquilibre net du système est le volume d'équilibrage nécessaire et il coïncide avec la somme algébrique des écarts individuels. Ce déséquilibre peut être négatif (système « court ») ou positif (système « long »). La figure 3-13 montre les deux cas de déséquilibre net du système. A gauche, un exemple du système « court », la somme des écarts individuels négatifs est supérieure à la somme des écarts

[164] Si l'on suppose que les participants offrent leur coût marginal pour une enchère à prix uniforme et une estimation du prix de l'offre marginale pour une enchère discriminante, les prix des écarts calculés sur la base de l'offre marginale (combinés à une enchère à prix uniforme) ou sur la base des prix moyens (combinés à une enchère discriminante) seraient équivalents.

individuels positifs, le GRT doit alors appeler des offres à la hausse pour l'équilibrer. A droite, un exemple de système « long », la somme des écarts positifs est supérieure à la somme des écarts négatifs, le GRT doit alors appeler des offres à la baisse pour l'équilibrer.

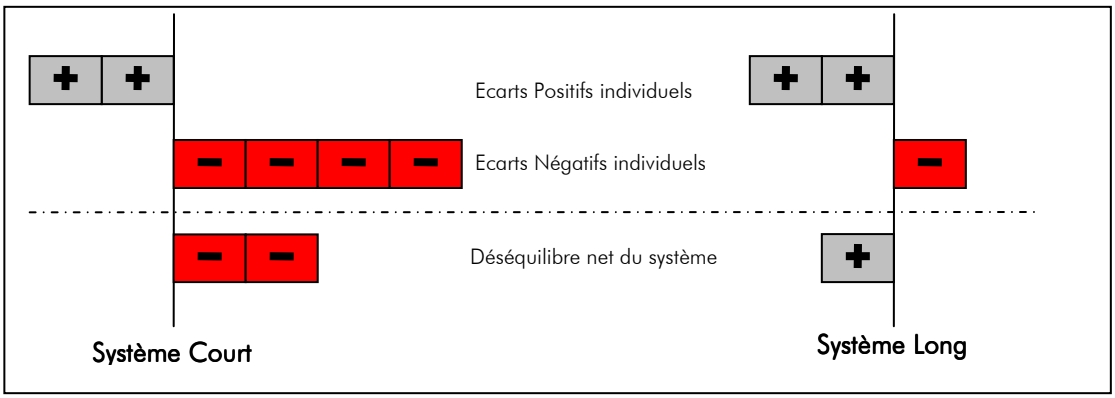

Figure 3-13 : Système « long » ou « court »

Une fois le déséquilibre net du système déterminé, les différents types d'écart peuvent être définis, et les prix des écarts peuvent être déterminés pour chacun des états du système (tableau 3-1).[165] La distinction résulte d'une comparaison de la « direction » (ou signe) de l'écart avec la direction ou tendance du déséquilibre du système.

Tableau 3-1 : Types d'écarts et état du système

	Système long	Système court
Ecart négatif	Ecart négatif/Syst. Long	Ecart négatif/Syst. Court
Ecart positif	Ecart positif/Syst. Long	Ecart positif/Syst. Court

Les prix des écarts ayant la même direction que le système (écarts positifs quand le système est long et écarts négatifs quand le système est court) peuvent être calculés selon deux méthodes :

- La méthode « prix de l'offre marginale » définira le prix des écarts négatifs (positifs) sur la base du prix de l'offre à la hausse (baisse) la plus grande (petite) acceptée.
- La méthode du « prix moyen» utilisera comme base du prix des écarts négatifs (positifs) le prix moyen pondéré par le volume et/où la durée des offres à la hausse (baisse) acceptées.

Les prix des écarts ayant une direction opposée à celle du système (écarts positifs si le système est court et écarts négatifs si le système est long) prennent normalement la valeur du prix *day ahead* ou une autre valeur de référence.[166]

[165] Les écarts peuvent être définis aussi par rapport à leur taille individuelle (avec des seuils d'ampleurs des écarts). Voir après le cas de la Belgique (sous-section 2.3).
[166] Les prix des écarts en direction contraire à celle du système peuvent être calculés aussi sur la base de toutes les offres acceptées pour un intervalle de livraison dans les deux directions, c'est-à-dire les offres acceptées à la hausse et à la baisse (Henney [2002], cf. sous-section 2.3.2).

Les prix des écarts définitifs, utilisant comme base un prix moyen ou le prix de l'offre marginale, peuvent être ajustés par des coefficients de pénalités explicites ou en ajoutant des coûts spécifiques. La pénalité explicite peut prendre la forme d'un facteur multiplicatif appliqué aux prix des écarts (par exemple en France), ou s'ajouter aux prix des écarts comme un facteur aditif ou soustractif (par exemple l'*incentive component* au Pays Bas). Généralement les prix des écarts positifs (excès d'énergie) sont diminués par la pénalité et les prix des écarts négatifs sont augmentés par la pénalité. Certains coûts de réserves sont imputés dans le calcul des prix des écarts négatifs (Henney [2002]). Ces coûts peuvent correspondre à des contrats signés par le GRT avec les fournisseurs de réserves ou des réserves achetées dans des marchés *forward* de réserves pour assurer une certaine disponibilité en temps réel (le coût de l'énergie déployée étant déjà considéré par les offres d'ajustement acceptées).

Les systèmes de règlement réel sont conçus avec un nombre important de règles spécifiques. Ces règles spécifiques sont liées, par exemple, au processus de tri entre toutes les offres acceptées pour calculer les prix des écarts. Les prix des écarts devraient considérer seulement les offres acceptées pour l'équilibrage en énergie du système. Or, les offres dans un module du temps réel sont normalement acceptées pour d'autres motifs que l'équilibrage en énergie (congestions, réglage de la tension, création de réserves ou de marges, etc.). Et, en effet, ces offres devraient être enlevées pour le calcul des prix des écarts. Un GRT, profitant de son expertise, pourrait « marquer » (*tag* –en anglais-) et ainsi distinguer les « types » d'offres acceptées (énergie, congestion, réserves) en réalisant une étude au cas par cas. Cependant, séparer les offres contribuant purement à l'équilibrage en énergie de celles qui ne le font pas peut être d'une extrême difficulté. Car l'acceptation d'une seule offre (et la modification de l'injection d'énergie sur un nœud) peut avoir plusieurs conséquences sur le réseau. Cette difficulté est d'autant plus importante quand cette séparation se réalise d'une manière « systématisée » par des règles ou des algorithmes. Des expériences sur ce type de méthodes ont montré une influence très forte sur la détermination du prix des écarts (Henney [2002]).

Enfin, le fait de différencier la valorisation de plusieurs types d'écart pose une question sur la forme du calcul des écarts en volume. Les responsables d'équilibre peuvent être représentants d'un groupe hétérogène de participants avec des activités différentes (production, consommation, *trading*). Ils présentent auprès du GRT des positions contractuelles par activité. Le calcul des écarts pour chaque responsable d'équilibre peut donc se réaliser ou bien en considérant toutes activités confondues (France, Belgique) ou bien en considérant les activités séparément (NETA, Suède). Comme les écarts viennent principalement des incertitudes, la forme de calcul des écarts permettra plus ou moins l'annulation de quelques incertitudes. Et cela peut influencer le comportement des participants.[167]

[167] Notez que cette question ne se pose pas quand toutes les transactions sont valorisées avec le même prix. Car la forme comment les écarts sont comptabilisés n'affecte pas le résultat financier du règlement. Par exemple un

Une fois abordées les variantes de design pour chaque étape du module du temps réel nous définirons les deux types de design pour ce module : le « marché du temps réel » et le « mécanisme d'ajustement ».

2.2 Types de design du module du temps réel

Depuis le début des réformes électriques, on a constaté l'existence de deux grandes conceptions différentes du design du module du temps réel. D'un côté, surtout aux Etats-Unis, des réformes ont adopté un design de « marché du temps réel » qui recourent à un prix unique de l'énergie en temps réel. D'un autre côté, d'autres réformes, surtout en Europe, préfèrent utiliser des « mécanismes d'ajustement » (*balancing mechanism*) qui recourent à des prix multiples pour régler les transactions de temps réel.

Dans le cadre de l'un ou l'autre de ces designs, le GRT ajuste continuellement l'état du système électrique en utilisant des offres d'ajustement déposées, ou en recourant à des options qu'il a négociées à l'avance (étape n°1). Cela se fait à l'aide d'un algorithme d'optimisation ou un *merit order*.[168] La première étape fonctionne alors d'une manière similaire dans un marché du temps réel et dans un mécanisme. Les deux types de design des modules du temps réel se différencient dans le règlement de transactions « actives » (étape n°2), mais surtout dans le règlement des transactions « passives » ou « écarts » (étape n°3).

2.2.1 Marché du temps réel

Nous désignerons par l'expression « marché du temps réel » le design qui utilise un prix unique de l'énergie en temps réel. Un « marché du temps réel » se sert donc d'une enchère à prix uniforme où toutes les offres acceptées (transactions actives) et tous les écarts sont réglés avec un seul prix, le prix de l'offre marginale. Ce prix sera utilisé pour régler toutes les transactions propres au temps réel. En théorie et dans des conditions de concurrence parfaite, ce prix correspondrait au coût marginal de l'énergie en temps réel.

Ce type de design a principalement été mis en place aux Etats-Unis. C'est le cas par exemple des architectures connues comme « centralisées » dans le nord-est et le centre-ouest des Etats-Unis (PJM, NY Pool, NE Pool, MidWest Pool), mais aussi dans l'architecture

---

écart négatif devrait payer au GRT le montant correspondant à son volume multiplié par le prix du temps réel. Un écart positif (du même volume) recevrait du GRT un montant équivalent. En conséquence, le fait de les comptabiliser sur un même compte ou par comptes séparés par activités ne change pas les incitations du responsable d'équilibre.

[168] Si les offres sont insuffisantes pour équilibrer le système, le GRT peut exercer des options acquises antérieurement pour différentes catégories de réserves. Dans la plupart des marchés ces options obligent les participants à soumettre des offres d'ajustement. En temps réel le GRT utilise les offres d'ajustement venant de deux sources : d'une part des offres d'ajustement et d'autre part l'utilisation d'offres de réserves (venant des marchés *forward* de réserves ou des contrats bilatéraux signés par le GRT).

« décentralisée » au Texas. En Europe, la Norvège utilise aussi un design du type « marché ».

La détermination du prix du temps réel est centrale pour le fonctionnement d'un marché du temps réel. La figure 3-14 montre comment ce prix est déterminé pour les cas d'un système court en énergie (gauche) et d'un système long en énergie (droite).

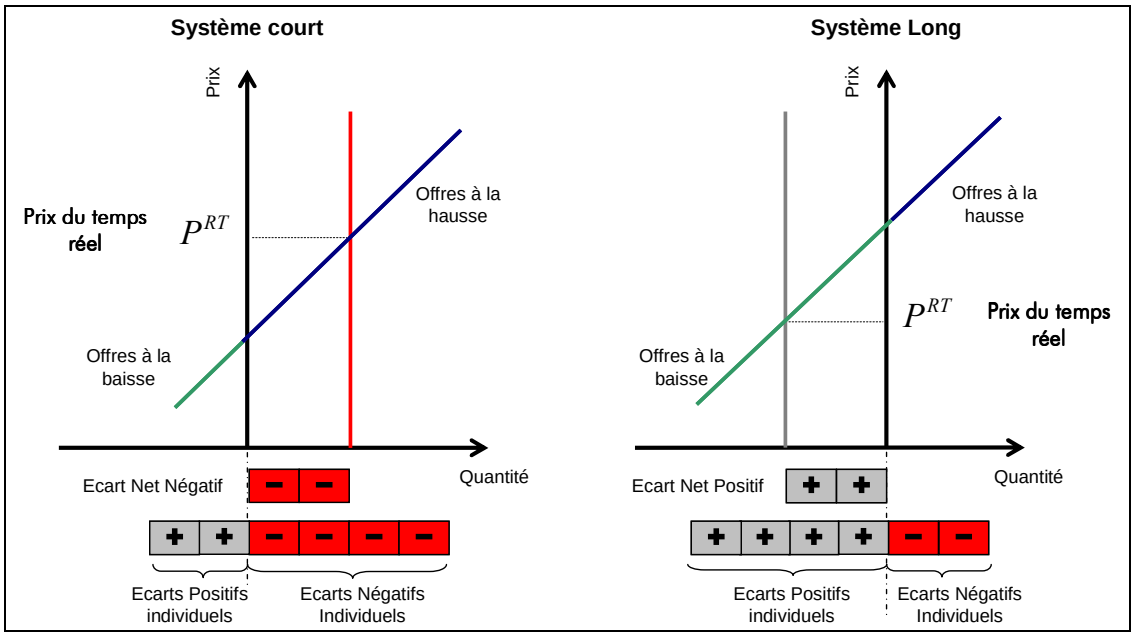

Figure 3-14 : Marché du temps réel – prix du temps réel pour système court et long

Sur cette figure on voit, par exemple, que pour équilibrer un système « court », des offres à la hausse doivent être acceptées pour un volume équivalent à l'écart négatif net (la production programmée est inférieure à la consommation en temps réel). Le prix du temps réel (P^{RT}) est défini comme l'offre marginale pour équilibrer le système. Ce prix sera utilisé tant pour les offres acceptées que pour les écarts individuels (positifs et négatifs). Un participant ayant un écart positif (il produit plus ou consomme moins que sa position contractuelle) est remboursé par le GRT d'un montant qui correspond au prix du temps réel multiplié par le volume de l'écart. Enfin, toutes les transactions passives sont réglées de même manière que les transactions actives.

Il est important de remarquer que le système de règlement des écarts du design du type « marché » implique un bilan économique neutre pour les revenus du GRT. En effet, les recettes du GRT venant des achats des participants en temps réel (écarts négatifs ou offres acceptées à la baisse) seront équivalentes aux coûts du GRT venant de ventes des participants en temps réel (écarts positifs ou offres acceptées à la hausse). Par exemple, si le système est « court » le GRT devra accepter des offres à la hausse pour l'équilibre. D'une part, il devra payer le total du volume des offres acceptées au prix du temps réel. D'autre part, les participants qui ont des écarts négatifs (et qui ont provoqué le déséquilibre du système) devront payer au GRT le volume net des écarts négatifs au prix du temps réel. Etant

donné que le volume des offres acceptées et le volume net des écarts négatifs sont équivalents, le bilan du GRT est neutre.

Le tableau 3-2 résume les caractéristiques (ou variantes de design) d'un marché du temps réel.

Tableau 3-2 : Caractéristiques d'un marché du temps réel

	Variante de design
Etape n°1 *Gate closure* et équilibrage du système	Equilibrage par *merit order* ou algorithme d'optimisation
Etape n°2 Règlement des offres d'ajustement	Enchère à prix uniforme
Etape n°3 Règlement des écarts	**Règlement des écarts à prix unique**

2.2.2 Mécanisme d'ajustement

Nous désignerons par l'expression « mécanisme d'ajustement » (*balancing mechanism*) le design qui recourt à des prix multiples pour régler les transactions du temps réel. Ils visent principalement à décourager les transactions passives en les pénalisant. En effet, l'arrangement du type « mécanisme d'ajustement » fonctionne plus comme un mécanisme de sécurité, géré par le GRT, que comme un marché. Les participants ne peuvent y réaliser des transactions que dans des cas exceptionnels ou comme la dernière option. Les réformes en Europe ont préféré utiliser des « mécanismes d'ajustement » (*balancing mechanism*), C'est le cas par exemple en Angleterre, Belgique, France, Finlande, etc. (ETSO [2003], ETSO [2006b]).

Les « mécanismes d'ajustement » utilisent des enchères pour équilibrer le système (offres d'ajustement et *merit order*). Ils peuvent utiliser soit une enchère à prix uniforme, soit une enchère discriminante (*pay-as-bid*) pour valoriser les transactions actives (offres acceptées). La caractéristique qui définit ces designs comme « mécanisme » est l'utilisation de prix différenciés pour l'énergie en temps réel qui pénalisent les écarts (règlement des écarts à double prix). Un mécanisme d'ajustement différencie clairement les deux types de transactions dans le module du temps réel. D'un côte, les transactions actives qui viennent de l'acceptation des offres d'ajustement et, de l'autre côté, les transactions passives ou écarts qui viennent d'une déviation d'injection ou de soutirage par rapport aux positions contractuelles. Les mécanismes d'ajustement fixent des prix différents pour ces deux types de transactions. Le règlement des écarts est fait ainsi avec des « prix » calculés sur la base des offres acceptées et qui subissent une pénalité, soit au travers de coefficients ou de pénalités explicites, soit que ces pénalités viennent de la forme du calcul des prix. Il y a au moins deux prix des écarts (négatifs et positifs) différents pour chaque situation du système (court et long). Les pénalités sont généralement asymétriques, plus fortes pour les écarts négatifs que pour les écarts positifs. Le tableau 3-3 résume les caractéristiques d'un mécanisme d'ajustement.

Tableau 3-3 : Caractéristiques d'un marché du temps réel

	Variante de design
Etape n° 1 *Gate closure* et équilibrage du système	Equilibrage par *merit order* ou algorithme d'optimisation
Etape n° 2 Règlement des offres d'ajustement	Enchère à prix uniforme ou enchère discriminante
Etape n° 3 Règlement des écarts	**Règlement des écarts à double prix (+pénalité)**

La figure 3-15 représente un schéma de la détermination des prix des écarts dans un mécanisme d'ajustement. A gauche, le cas d'un système « court » en énergie et à droite le cas d'un système « long » en énergie. Sur cette figure, on trouve une des propriétés typique d'un mécanisme d'ajustement : $P^E_{neg} > P^{RT} > P^E_{pos}$. Un prix pour les écarts positifs (P^E_{pos}) auquel les positions en excès d'énergie par rapport aux positions contractuelles sont rémunérées moins que l'offre marginale d'équilibrage en énergie en temps réel (P^{RT}). Un autre prix pour les écarts négatifs (P^E_{neg}) auquel les positions déficitaires en énergie par rapport aux positions contractuelles payent plus cher que l'offre marginale d'équilibrage en énergie en temps réel (P^{RT}). La différence entre les prix des écarts et le prix de l'énergie en temps réel représente la pénalité appliquée aux écarts en temps réel. L'intensité des pénalités appliquées sur un mécanisme d'ajustement en particulier peut être mesurée par la différence entre les prix des écarts négatifs et positifs (P^E_{neg} - P^E_{pos}).

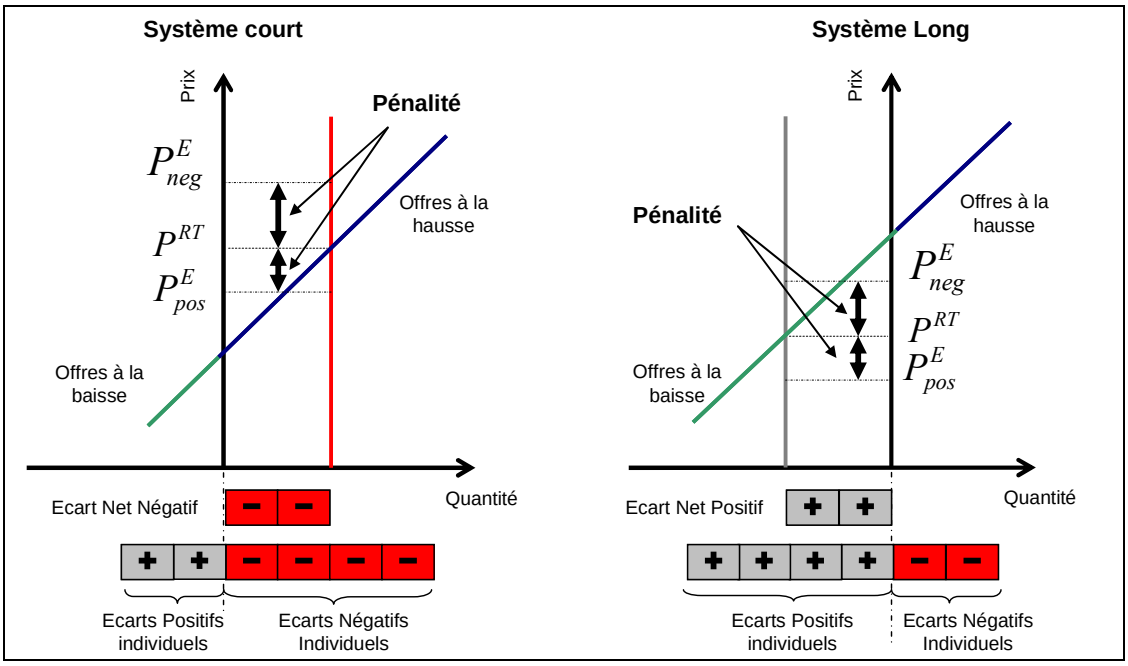

Figure 3-15 : mécanisme d'ajustement – prix des écarts système long et court

Le système de règlements des écarts caractéristique d'un mécanisme d'ajustement implique un bilan économique pour les revenus du GRT souvent positif. Les recettes du GRT venant des achats des participants en le temps réel (les écarts négatifs et les offres acceptées

à la baisse) seront, la plupart du temps, supérieures aux coûts des ventes des participants en temps réel (les écarts positifs et les offres acceptées à la hausse).[169]

Nous avons présenté les caractéristiques des deux types de design du module du temps réel, le « marché » et le « mécanisme ». Les designs du type « marché du temps réel » utilisent généralement des enchères à prix uniforme. Ainsi, toutes les offres acceptées (transactions actives) sont réglées avec un seul prix, le prix de l'offre marginale. Les « mécanismes d'ajustement » utilisent soit une enchère à prix uniforme, soit une enchère discriminante en rémunérant les transactions actives (offres acceptées) au prix proposé (*pay-as-bid*). Bien évidement, les prix des offres déposées par les participants dépendront du type d'enchère (théorème d'équivalence du revenu) et dans certaines conditions idéales, les résultats de ces deux types d'enchère peuvent être considérés comme équivalents (cf. sous-section 2.1.2).[170] La principale différence entre les deux conceptions opposées des arrangements pour le module du temps réel porte sur le règlement des écarts (transactions passives du temps réel) et l'utilisation de pénalités. Le tableau 3-4 reprend les caractéristiques particulières de ces deux types de designs.

Tableau 3-4 : Variantes de design du module du temps réel

	Marché du temps réel	**Mécanisme d'ajustement**
Etape n° 1 *Gate closure* et équilibrage du système	Equilibrage par *merit order* ou algorithme d'optimisation	
Etape n° 2 Règlement des offres d'ajustement	Enchère à prix uniforme	Enchère à prix uniforme ou enchère discriminante
Etape n° 3 Règlement des écarts	**Règlement des écarts à prix unique**	**Règlement des écarts à double prix + Pénalité**

Dans la sous-section suivante nous réaliserons des études de cas pratiques de design des modules du temps réel. L'étude de cas nous montrera que les différentes variantes de design mises en œuvre pratiquement pour le module du temps réel forment une gamme de designs qui se répartissent entre les deux cas extrêmes que nous avons définis selon l'intensité des pénalités appliquées.

2.3 Etudes de cas

Nous étudions ici 5 designs différents de modules du temps réel appliqués respectivement au Texas, en Angleterre, en Belgique, en France et aux Pays-Bas. En général, le fonctionnement de la plupart des marchés de temps réel est similaire car les règles qui les

[169] Le compte qui récupère les différences entre les « recettes » et les « coûts » du GRT est connu dans le langage courant comme « *beer fund* » (Cornwall [2002]).
[170] Bien que ces conditions ne correspondent pas exactement au cas effectif des modules du temps réel, nous considérerons que ces deux designs se comportent de manière similaire afin de pouvoir nous concentrer sur les différences de design de l'étape 3.

gouvernent sont relativement simples. Pour cela nous présentons le cas du Texas comme exemple ou référence de ce type de design. En revanche, le fonctionnement des mécanismes d'ajustement présentent de nombreuses différences, car ils incorporent un nombre important de règles spécifiques. Pour l'illustrer nous avons donc choisi d'analyser quatre mécanismes d'ajustement différents en Europe : Angleterre, Belgique, France et Pays-Bas. Pour chacun des cas, nous étudierons à quel degré ces mécanismes d'ajustement s'éloignent d'un « marché du temps réel » en utilisant une mesure approximée de l'intensité des pénalités qui est qui est la différence entre les prix des écarts négatifs et les prix des écarts positifs (P^E_{neg} - P^E_{pos}).

2.3.1 Le marché du temps réel au Texas

L'architecture de marché au Texas utilise un design du type marché pour son module du temps réel. Ce module est géré par ERCOT, gestionnaire du réseau au Texas, et il est organisé comme une enchère à prix uniforme où l'offre marginale acceptée pour équilibrer le système détermine le prix de l'énergie en temps réel. Ainsi, un seul prix de l'énergie (par zone)[171] est utilisé pour régler toutes les transactions du temps réel. La figure 3-16 montre l'évolution du prix du temps réel entre 2004 et 2006.

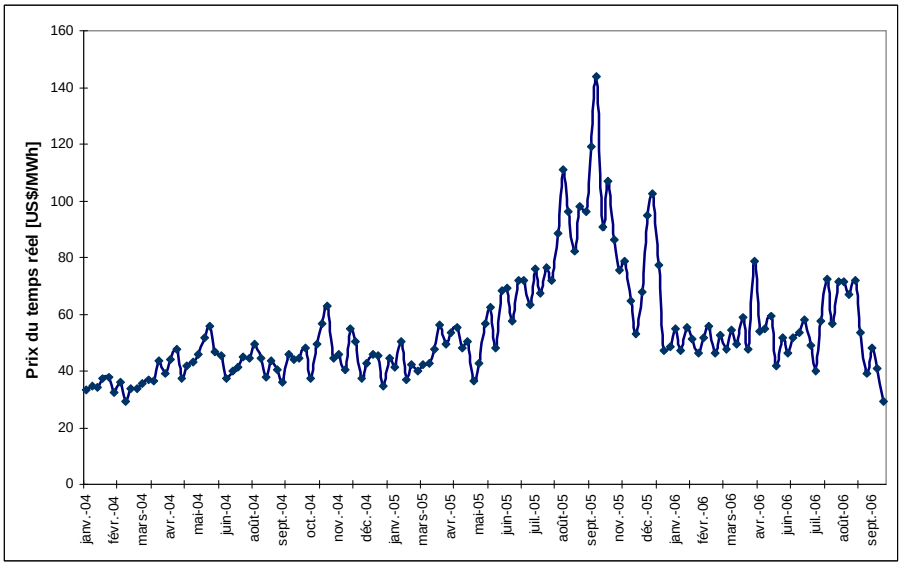

Figure 3-16 : Prix du temps réel au Texas (moyenne par semaine, zone Houston)[172]

Le prix des écarts est unique et coïncide avec le prix de l'énergie sauf sur certaines situations particulières. Par exemple, si, au moment de la présentation de programmes de production, ERCOT détecte une insuffisance de moyens de production programmés par rapport à sa prévision de consommation, il procède à l'achat de réserves de remplacement. Les coûts des réserves de remplacement sont alloués proportionnellement entre ceux qui

[171] Le marché du temps réel au TEXAS différencie les biens « énergie » en 5 zones.
[172] Source de données : www.ercot.com

sont en écart négatif (Potomac [2004]).[173] Cependant, l'expérience montre que ce type de réserves n'est pas appelé couramment et que leur coût n'est pas trop élevé (Potomac [2004], Hogan [2006b]).

En J-1, les responsables d'équilibre doivent présenter une première fois leurs positions contractuelles et leurs programmes de production (« *ressource plan* ») et consommation. Ils peuvent modifier leurs programmes mais leur valeur définitive est arrêtée à la *gate closure* (entre 1 heure et 2 heures avant l'intervalle de livraison). Le GRT (ERCOT) s'assure une bonne information de l'état des unités de production et de leur programme grâce à l'incitation à présenter des programmes de production ou « *ressources plan* » crédibles. [174]

Jusqu'en 2002, une règle obligatoire de « programmes équilibrés » (*balanced schedules*) s'imposait pour les responsables d'équilibre et selon laquelle les positions contractuelles (et les programmes journaliers de production) devaient s'approcher au maximum de la valeur de consommation estimée. Cette règle cherchait à minimiser les échanges réalisés sur le module du temps réel, et ainsi les déséquilibres à gérer par le GRT. Après 2002, cette règle à été supprimée et les responsables d'équilibre (*Qualified Scheduling Entities*) peuvent présenter des positions contractuelles permettant d'acheter et de vendre sur le marché du temps réel. La figure 3-17 montre le volume des transactions actives du temps réel ou « offres d'ajustement acceptées » (à la hausse -*Up Balancing*- et à la baisse –*Down Balancing*-) sur le marché du temps réel géré par ERCOT, entre 2002 et 2005. On remarque que le volume total des transactions en pourcentage de la consommation totale a été de l'ordre du 5 % avec une augmentation après 2002.[175]

Le changement de la règle des « programmes équilibrés » a permis cette augmentation des transactions en temps réel. Le volume net des transactions (déséquilibre net du système) est resté de l'ordre de 2% ou 3% et, bien que sa direction moyenne ait changé après la modification de la règle, sa valeur absolue n'a pas augmenté considérablement. L'acceptation d'offres pour arbitrage a donc été très importante (Potomac [2006]). Il est important de remarquer que bien que les participants sur le marché de gros Texan soient autorisés à acheter ou à vendre « librement » de l'énergie en temps réel, le volume net des transactions gérées par le marché du temps réel reste faible et l'exploitation en temps réel n' donc pas été mise en danger.

---

[173] Ce type de règle spécifique est commun à plusieurs marchés américains (PJM, NY Pool, MidWest Pool), et traite aussi l'allocation des coûts de non-convexités de la fonction de coût (Hogan [2006b]).

[174] ERCOT incite les responsables d'équilibre (*scheduling entitiies*) à révéler des programmes de production crédibles en réalisant périodiquement de classements par ordre de précision et en avertissant à ceux qui mal classés (ou qui donnent couramment des fausses informations).

[175] Les pourcentages moyens pour 2002, 2003, 2004 et 2005 ont été 4,6%, 6,1%, 5,7% et 5,6% respectivement (Potomac [2006]).

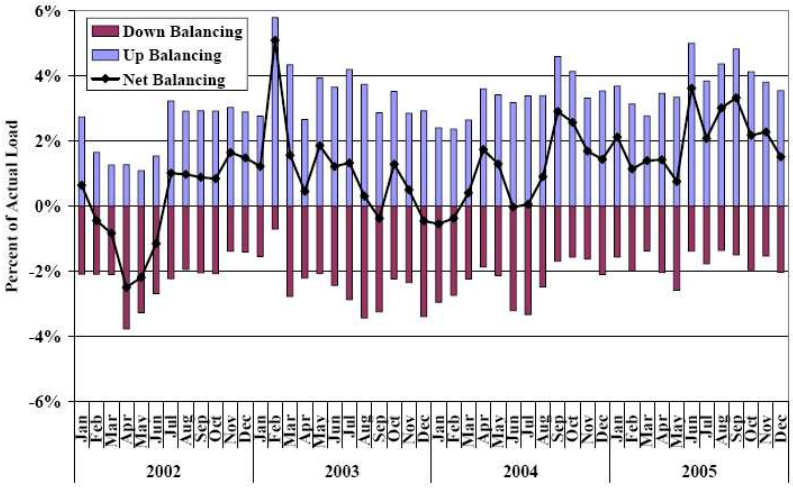

Figure 3-17 : Transactions du temps réel au Texas

En conclusion, le design du module du temps réel au Texas correspond à un design du type « marché ». Ce marché fonctionne sur la base d'une enchère à prix uniforme où un seul prix de l'énergie est utilisé pour toutes les transactions du temps réel (sauf situations très particulières). Il a commencé à fonctionner vraiment comment un marché en 2002 une fois la règle administrative de « programmes équilibrés » qui limitait le volume des transactions du temps réel a été supprimée.

Comme nous l'avons mentionné, la plupart des designs du type « marché » fonctionne avec des règles similaires pour la valorisation des transactions du temps réel. En revanche, les mécanismes d'ajustement utilisent des règles spécifiques pour valoriser les transactions du temps réel. Nous verrons par la suite le cas du mécanisme d'ajustement en Angleterre qui a été l'un des premiers mécanismes d'ajustement en fonctionnement et qui, par ailleurs a été le plus étudié.

2.3.2 Le mécanisme d'ajustement en Angleterre

Le mécanisme d'ajustement anglais (*Balancing Mechanism*) est géré de manière conjointe entre le gestionnaire du réseau (NGC) et Elexon.[176]

Le *Balancing Mechanism* fonctionne sur la base d'une enchère discriminante. Pour le règlement des écarts, ce mécanisme utilise un système de double prix (*dual-cash settlement*) : le prix des écarts négatifs, pour acheter de l'énergie en temps réel (SBP-*System Buying Price*) et, le prix des écarts positifs, pour vendre de l'énergie en temps réel (SSP-*System Selling Price*). Quand le système est « court », le SBP est le prix moyen pondéré en volume des prix des offres acceptées à la hausse pour équilibrer le système en énergie. Quand le système est « long », le SSP est le prix moyen pondéré en volume des prix des

[176] *National Grid Company* (NGC) est le gestionnaire du réseau pour la zone de contrôle correspondant à l'Angleterre et au Pays de Galles. Elexon est l'entreprise responsable du règlement du mécanisme d'ajustement. Elexon est indépendante dans sa gestion mais son capital appartient à NGC.

offres acceptées à la baisse pour équilibrer le système en énergie. Ces prix sont ajustés avec les coûts de contrats signés directement par le GRT pour se procurer des services d'équilibrage (*Balancing service adjustment data*).[177] Ce mécanisme d'ajustement a eu des nombreuses évolutions depuis le début du fonctionnement en 2001.

La règle de définition des prix des écarts en direction contraire à celle du système (écart négatif avec système long, et écart positif avec système court) a été une des principales modifications depuis le début du mécanisme en 2001. Cette règle a été modifiée au début du 2003 grâce à l'amendement P78. Nous expliquerons d'abord le fonctionnement avant P78, puis les modifications réalisées.

Prix des écarts en direction contraire au déséquilibre avant P78

Au début du fonctionnement du mécanisme anglais, les prix des écarts en direction contraire au déséquilibre net du système sont basés sur les offres acceptées en direction inverse au déséquilibre net. Les offres en direction contraire au déséquilibre net peuvent être acceptées pour des déséquilibres momentanés en direction contraire (équilibrage en énergie) ou pour gérer d'autres contraintes du réseau, par exemple gérer une congestion. Selon les règles qui gouvernent le mécanisme, le calcul des prix des écarts doit utiliser seulement les offres acceptées pour l'équilibrage en énergie, et il faut donc les séparer de celles venant des contraintes du réseau.[178] Pour ce faire, le design en Angleterre (NETA) a choisi une méthode automatique de séparation des offres afin de diminuer l'intervention discrétionnaire du GRT. Cette méthode utilisait le concept de « niveau de réserves d'ajustement » (BRL - *balancing reserve level*) (Henney [2002]). Selon cette méthode, les offres acceptées par des « contraintes de réseau » étaient celles acceptées en direction contraire au déséquilibre du système et au delà du niveau de réserves d'ajustement.

La figure 3-18 montre un exemple du fonctionnement de la méthode de BRL pour la séparation des offres et la définition des prix des écarts. On représente la liste d'offres acceptées à la hausse et à la baisse pendant un intervalle de livraison (le système est long : volume d'offres à la baisse > volume d'offre à la hausse). Les offres acceptées à la hausse pour un volume équivalent de BRL sont utilisées pour déterminer le SBP. Les offres acceptées à la hausse au delà de BRL sont considérées comme appelées par des contraintes du réseau. Ainsi, un volume équivalent d'offres à la baisse est neutralisé pour des contraintes de réseau. Le reste des offres acceptées à la baisse sont utilisées pour le calcul du SSP.[179]

---

[177] Ces contrats considèrent le coût d'équilibrage en énergie mais aussi le coût payé pour les options de fonctionnement (réserves). Dans les coûts d'équilibrage en énergie, le GRT introduit ses contrats d'achat ou de vente sur les marchés *forward*. Certains coûts de réserves du GRT sont alloués directement sur les prix des écarts négatifs.

[178] L'allocation des coûts des contraintes du réseau est faite par le biais d'un *uplift*. Pour le cas des contraintes de transport par congestion, il existe d'autres méthodes d'allocation des coûts de congestion aux responsables (tarification nodale ou zonale) (cf. Chapitre 2, sous-section 2.4).

[179] Cette méthode de calcul des prix nécessite aussi des règles auxiliaires de définition quand il n'y a pas d'offres acceptées dans la direction contraire au déséquilibre. Sur le NETA ces « *default prices* » étaient basés sur les offres disponibles mais non-acceptées (Henney [2002]).

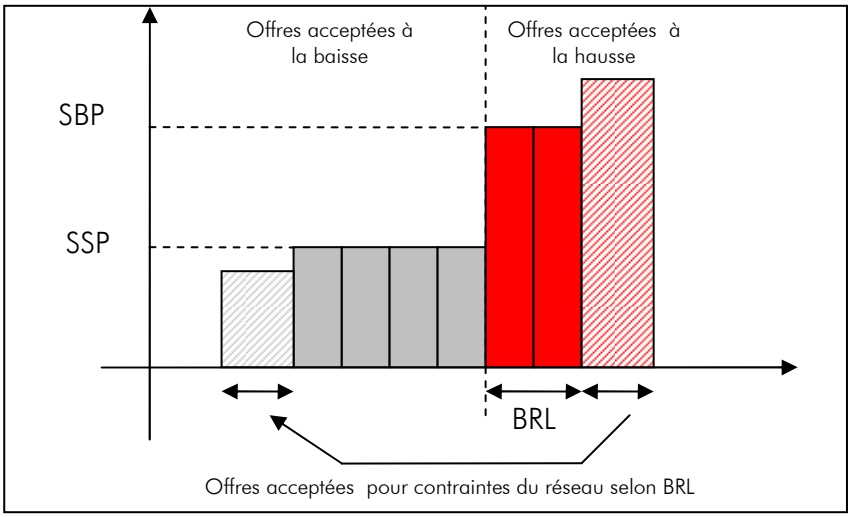

Figure 3-18 : Définition des prix des écarts avant P78

Pendant les deux premières années de fonctionnement du mécanisme d'ajustement avec cette méthode, les différences entre le SBP et le SSP ont été très grandes. Ces grandes différences venaient de la valeur très élevée donnée au BRL (180 MWh). En effet, des offres acceptées réellement pour des contraintes du réseau (par exemple des offres acceptées pour résoudre une congestion avec normalement des prix très différents) sont quand même considérées pour le calcul des prix des écarts, car la valeur haute du BRL ne laisse que très peu d'offres comme acceptées par « des contraintes du réseau ». L'effet direct a donc été des différences marquées entre le SBP et le SSP, comme on peut l'apprécier sur la figure 3-19.[180]

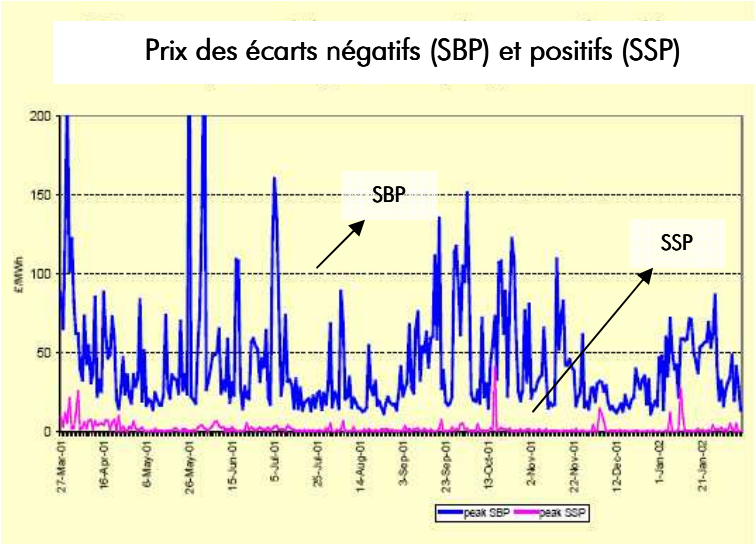

Figure 3-19 : SBP et SSP entre 2001 et 2002 (Newbery [2002])

[180] La valeur moyenne du SBP correspondait à 4 fois la valeur du prix de référence de la bourse UKPX (Henney [2002]).

Les prix des écarts, notamment le SBP, ont été réputés très volatiles et difficiles à prédire (Henney [2002], Newbery-McDaniel [2003], Roques *et al.* [2004]). Ces caractéristiques des prix des écarts créent des effets asymétriques sur les écarts négatifs, affectées par des valeurs très hautes et volatiles du SBP, et sur les écarts positifs, affectés par des valeurs faibles (mais toujours positives) et moins volatiles.

Les grandes différences des prix des écarts et leur asymétrie ont provoqué le phénomène de « sur-programmation » ou d' « *over-contracting* » (Cornwall [2002], Newbery-McDaniel [2003]). En fait, les participants cherchent à réduire au maximum la probabilité de se trouver en écart négatif et ils programment (ou opèrent des transactions) pour un volume plus grand que leur consommation. En effet, le système est la plupart du temps « long », c'est-à-dire il est en excès d'énergie et il doit accepter des offres à la baisse pour trouver l'équilibre. La figure 3-20 montre les volumes moyens mensuels du déséquilibre en Angleterre (un déséquilibre positif représente le volume d'offres à la baisse acceptées) entre juillet 2001 et septembre 2002. On constate que le système a été « long » pour tous les mois de cette période.

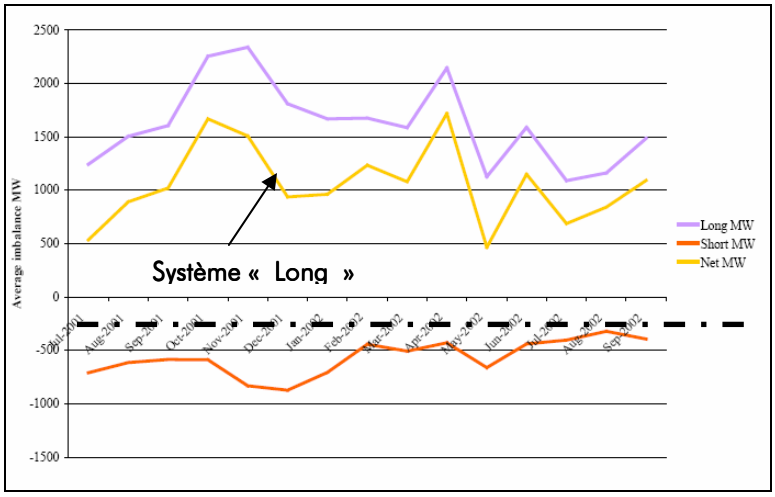

Figure 3-20 : Volume de déséquilibre sur le mécanisme d'ajustement Anglais (source ILEX [2002])

Prix des écarts en direction contraire au déséquilibre après P78

Le changement des règles « P78 » a modifié la détermination des prix des écarts dans le sens contraire au déséquilibre (« *reverse price* »). En revanche, les prix dans le sens du système (« *main prices* ») sont calculés au prix moyen pondéré des offres acceptées. Après le changement des règles, les « *reverse prices* » prennent la valeur du prix de référence (« *Market Index Price* »), calculé sur la base des transactions nouées sur la bourse d'électricité APX UK[181].

[181] Etant donné qu'APX UK n'est pas organisé comme une enchère mais comme un design du type « *continuous trading* », il n'existe pas un prix *day-ahead* bien défini. Le prix de référence est ici calculé d'après une moyenne pondérée des différentes transactions nouées sur APX UK (ELEXON [2006a]).

Le Tableau 3-5 résume la nouvelle configuration des prix des écarts.

Tableau 3-5 : Prix des écarts sur le NETA

	Système long	Système court
Ecart négatif	SBP= Prix de référence	SBP = Prix moyen pondéré (offres à la hausse)
Ecart positif	SSP = Prix moyen pondéré (offres à la baisse)	SSP= Prix de référence

Par ailleurs, un autre changement des règles a eu lieu à la mi 2002, la « *gate closure* » a été placée à une heure du temps réel (avant elle était placée à 3 heures et demi du temps réel).[182]

La figure 3-21 montre les différences entre les SBP et le SSP avant et après la modification P78. On peut remarquer qu'à partir du changement de la règle, les différences entre les prix des écarts ont diminué considérablement. Cependant, les différences entre les prix des écarts continuent à être non-négligeables (voir tableau 3-6) et peuvent atteindre des valeurs importantes (jusqu'au delà de 80 €/MWh ou ~50 £/MWh).[183]

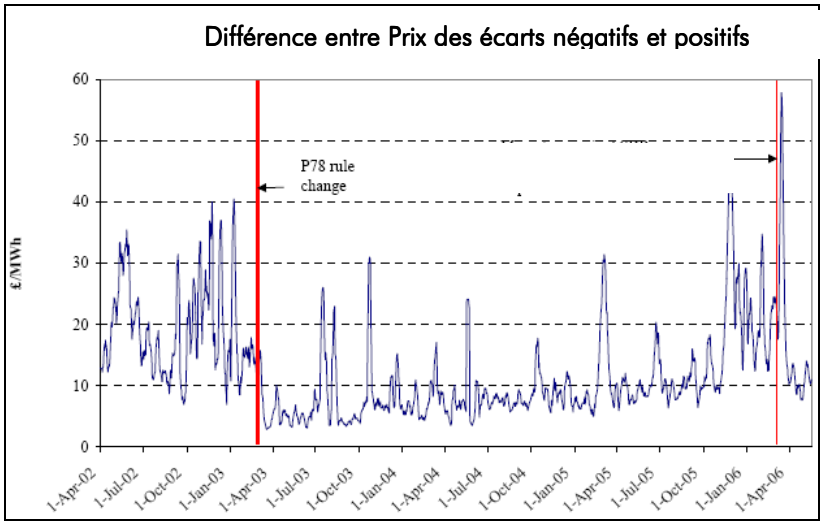

Figure 3-21 : Différence entre SBP et SSP avant et après P78 (moyenne mobile hebdomadaire) (Newbery [2006])

[182] Ce changement de règle a été largement supporté par les producteurs éoliens. Une *gate closure* plus proche au temps réel permettait une précision plus fine des modèles de prévision de la production éolienne et ainsi un positionnement contractuel des producteurs éoliens plus convenable vis-à-vis du mécanisme d'ajustement.

[183] Un nouveau changement important de règles devait être mis en œuvre fin 2006. La détermination des prix des écarts se ferait alors par le prix de l'« offre marginale » et non plus par le prix moyen. Le prix dérivé de l'offre marginale permettra de donner des signaux de rareté nécessaires dans des périodes de pointe et super-pointe. Ce changement pourrait intensifier les différences entre le SBP et le SSP (Roques *et al.* [2005], Ofgem [2006]).

Tableau 3-6 : Prix des écarts en Angleterre entre Oct.2004 et Juin 2006 (Elexon [2006b])

	Prix des Ecarts Négatifs (SBP) [£/MWh]	Prix des Ecarts Positifs (SSP) [£/MWh]	Différence SBP-SSP [£/MWh]	Différence SBP-SSP %[184]
2004 (Oct-Dec)	28,4	19,0	9,4	40%
2005	43,0	29,8	13,2	36%
2006 (Janv-Juin)	52,4	36,0	16,4	37%

Actuellement le phénomène d'*over contracting* continue sur le mécanisme d'ajustement Anglais. Cela peut être vérifié sur le rapport Elexon [2006c] où on reporte, par exemple, que plus de 84 % du temps le système a été long en août 2006. La figure 3-22 représente le déséquilibre net du système entre Juin 2005 et Mai 2006. On voit clairement que le système a été long la plupart du temps.

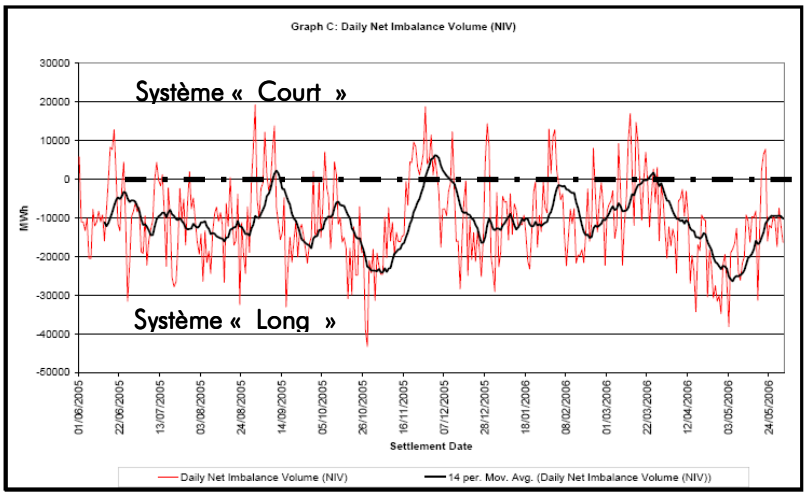

Figure 3-22 : Volume de déséquilibre sur le mécanisme d'ajustement Anglais (Elexon [2006b])

En conclusion, le design du module du temps réel en Angleterre correspond à un mécanisme d'ajustement organisé comme une enchère discriminante et un système de règlement des écarts à double prix. Les règles spécifiques qui gouvernent le calcul des prix des écarts provoquent des différences importantes entre les prix des écarts négatifs et positifs (37 % de la moyenne des prix des écarts) et pénalisent les écarts. Ces règles spécifiques affectent principalement aux prix des écarts négatifs (SBP) en impactant asymétriquement les écarts. Ces asymétries de prix des écarts incitent les participants à l'*over-contracting*, c'est-à-dire à se positionner contractuellement en excès d'énergie par rapport à la position physique.

[184] Cette colonne représente l'importance proportionnelle de la différence (absolue) des prix des écarts par rapport au niveau de prix moyen. Les valeurs de cette colonne sont calculées comme la différence des prix (colonne 3) divisée par la moyenne des prix des écarts (moyenne colonne 1 et colonne 2).

2.3.3 Le mécanisme d'ajustement en Belgique

L'équilibrage du système en Belgique est géré par ELIA (Gestionnaire de réseau Belge) par le biais d'un mécanisme d'ajustement fonctionnant sur la base d'une enchère discriminante. Les offres d'ajustement sont obligatoires pour tous les producteurs avec une puissance nominale de plus de 75 MW (CREG [2005]).

Les caractéristiques du règlement des écarts ont évolué depuis 2004. Au début, les prix des écarts étaient basés principalement sur les valeurs des prix *day ahead* des bourses d'électricité des Pays-Bas et de la France (APX et PowerNext)[185] et sur des tarifs administrés. Ces valeurs de références étaient modifiées par des coefficients de pénalité. Les prix des écarts étaient différenciés selon la direction du déséquilibre du système, selon la période de la journée et le mois. Ils étaient aussi différenciés par la taille relative de l'écart. Un seuil était défini pour chaque responsable d'équilibre.[186] Cela permettait d'utiliser des prix des écarts différents pour les écarts inférieurs et supérieurs au seuil.[187] Le tableau 3-7 donne un exemple du système des prix des écarts.

Tableau 3-7 : Système des prix des écarts en 2005[188]

	Système long	Système court
Ecart négatif	Min (1,1*Prix APX ; 75 Euros/MWh)	Min (1,75*Prix APX ; 75 Euros/MWh)
Ecart positif	Min (0,25*Prix APX ; 30 Euros/MWh)	Min (0,9*Prix APX ; 30 Euros/MWh)

La figure 3-23 montre l'évolution des prix des écarts pour les années 2004 et 2005. Les prix des écarts négatifs et positifs pour les écarts inférieurs et supérieurs au seuil sont représentés (axe des ordonnées à gauche). De plus, la différence entre les prix des écarts (négatif – positif) pour les écarts en-dessous du seuil est représentée (axe des ordonnées à droite). Nous remarquons, tout d'abord, que les prix des écarts n'ont pas eu une variation très importante au long de l'année 2004. Cela vient de ce que les prix des écarts n'étaient pas basés sur les conditions du temps réel mais sur des prix *Day-Ahead*. Nous remarquons aussi que les différences entre ces prix ont été considérables (plus de 30 €/MWh en moyenne pour les écarts inférieurs au seuil et encore plus importantes pour les écarts au dessus du seuil).

[185] Il n'y avait pas de bourse d'électricité en Belgique.
[186] En Belgique le terme « Responsable d'Accès » (*Acces Responsible Party* – ARP) est utilisé à la place de « Responsable d'équilibre ».
[187] Le seuil correspondait à 10 % du total des soutirages de chaque responsable d'équilibre.
[188] Ces prix correspondent à des écarts inférieurs au seuil, pour la période « jour » et le mois de janvier (Elia [2005]).

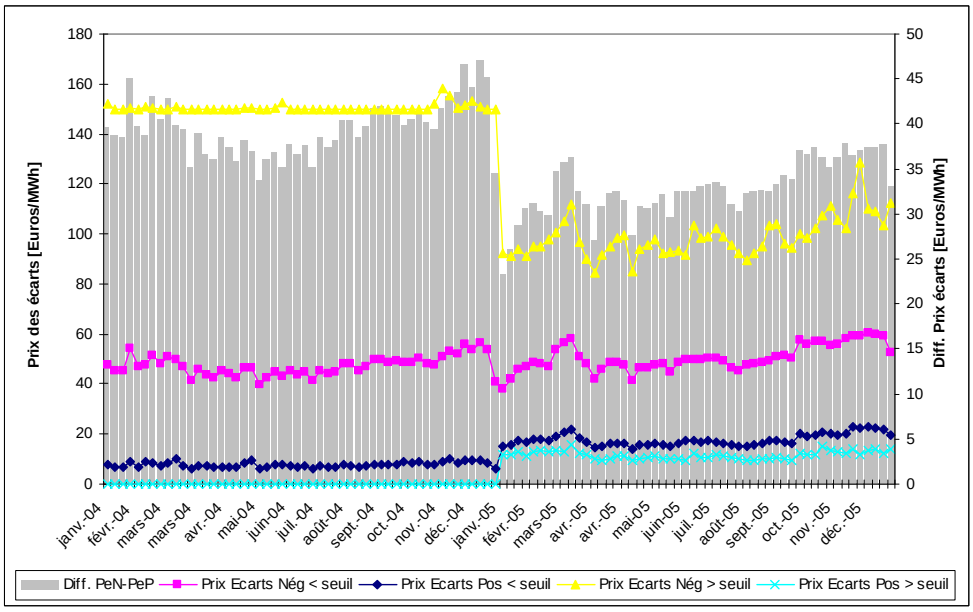

Figure 3-23 : Prix des écarts en Belgique 2004-2005 (Moy. mobile hebdomadaire - Source www.elia.be)

En 2006 de nouvelles règles pour le calcul des prix des écarts sont entrées en vigueur. Les prix des écarts sont déterminés maintenant sur la base des offres d'ajustement acceptées pour l'équilibrage. Ce nouveau système de règlement des écarts combine des prix moyens des offres, des prix de l'offre marginale, des coefficients de pénalité et des prix de référence venant de la bourse d'électricité (Belpex). Le tableau 3-8 montre un exemple de ce nouveau système de règlement.

Tableau 3-8 : Prix des écarts en Belgique à partir de 2006[189]

	Système long	Système court
Ecart négatif	1,1*Prix *day-ahead* (Belpex)	Max(1,15*PMH ;Prix offre marginale)
Ecart positif	Min(0,85*PMB ; Prix offre marginale)	0,9*Prix *day-ahead* (Belpex)

La figure 3-24 montre l'évolution des prix des écarts positifs et négatifs à partir du début de 2006. Les différences entre ces prix sont aussi représentées sur l'axe des ordonnées à droite. Nous remarquons une plus importante variation des prix des écarts venant de la prise en compte des coûts d'équilibrage en temps réel (maintenant calculés sur la base des offres acceptées). Nous remarquons aussi que les différences entre les prix restent importantes (plus de 30 €/MWh en moyenne).

[189] PMH et PMB sont les prix moyens des offres acceptées à la hausse et à la baisse respectivement. Ces prix des écarts correspondent aux écarts de janvier à septembre 2006 pour des PMB >0 et pour un volume d'offres acceptées à la baisse < 450 MW (Elia [2006]).

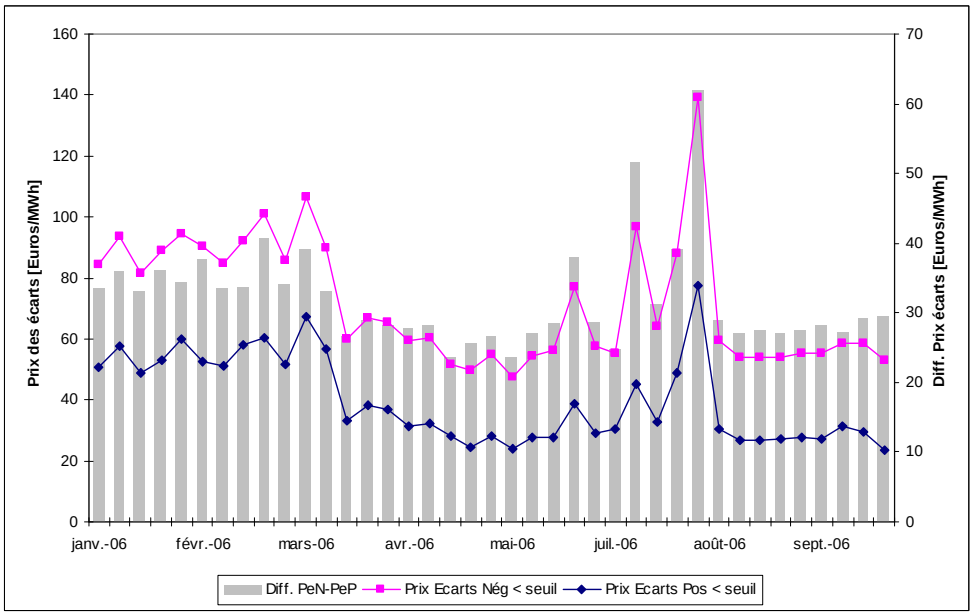

Figure 3-24 : Prix des écarts en Belgique 2006 (moyenne mobile hebdomadaire – source www.elia.be)

Le tableau 3-9 résume les valeurs moyennes et les différences des prix des écarts. Nous observons que l'importance relative de la différence entre prix a diminué considérablement après la modification des règles (malgré la faible diminution de la valeur absolue). Néanmoins, l'intensité des pénalités reste importante sur ce mécanisme comme en témoignent les différences entre les prix des écarts négatifs et positifs (57 % de la moyenne des prix des écarts).

Tableau 3-9 : Prix des écarts en Belgique 2004-2006 (source www.elia.be)

	Prix Ecarts Négatifs (PeN) [€/MWh]	Prix Ecarts Positifs (PeP) [€/MWh]	Différence PeN-PeP [€/MWh]	Différence PeN-PeP %[190]
2004	47,2	7,7	39,5	144%
2005	50,5	17,7	32,8	96%
2006 (janv-sept)	71,8	39,8	32,0	57%

En conclusion, le design du module du temps réel en Belgique correspond à un mécanisme d'ajustement sur la base d'une enchère discriminante et d'un système de règlement des écarts à double prix. Les règles pour le calcul des prix des écarts ont évolué. Cette évolution est allée d'un système de prix des écarts calculés sur la base des prix *day-ahead* et sans correspondance aux conditions du temps réel vers un système de détermination des prix des écarts basé sur les offres d'ajustement acceptées. Le prix des écarts sont calculés sur la base des prix moyens des offres acceptées ou sur l'offre marginale. Des pénalités explicites sont utilisées pour décourager les écarts et l'intensité des pénalités est importante sur ce mécanisme.

[190] Cf. note de bas de page n° 184.

2.3.4 Le mécanisme d'ajustement en France

En 2003, RTE, le Gestionnaire du réseau français, a mis en œuvre un mécanisme d'ajustement pour la gestion de l'équilibre en temps réel (CRE [2004b]). RTE compense les déséquilibres en sélectionnant des offres, après les avoir interclassées selon un critère de préséance économique (ou *merit order*). De par la loi, la totalité de la puissance non utilisée et techniquement disponible est mise à disposition de RTE par les producteurs dans leurs offres d'ajustement (CRE [2005b]). Ce mécanisme fonctionne en combinant des « guichets » autorisant la modification de programmes de production ou des positions contractuelles : c'est une « *gate closure* » par guichets.

Le système de règlement des écarts utilise comme base les prix moyens des offres acceptées pour les écarts ayant la même direction que le système, à savoir le prix moyen d'ajustement à la hausse (PMH) et le prix moyen d'ajustement à la baisse (PMB) (RTE [2006]). Ces prix moyens sont affectés par un coefficient de pénalité (1+k) pour calculer les prix des écarts. Le prix des écarts négatifs est augmenté en multipliant PMH par le coefficient de pénalité (1+k). Le prix des écarts positifs est diminué en divisant PMB par le coefficient de pénalité (1+k). Les prix des écarts dans la direction contraire à celle du système sont fixés aux prix *day ahead* de la bourse d'électricité PowerNext pour l'intervalle de livraison correspondant. Le tableau 3-10 résume le système de calcul de prix des écarts pour les deux conditions du système (long et court).

Tableau 3-10 : Prix des écarts en France

	Système long	Système court
Ecart négatif	Prix *day-ahead* (PowerNext)	PMH * (1+k)
Ecart positif	PMB/ (1+k)	Prix *day-ahead* (PowerNext)

Depuis le début du fonctionnement, les règles du mécanisme d'ajustement ont évolué principalement sur deux paramètres[191] : le nombre de guichets disponibles pour modifier les programmes, et la valeur du coefficient de pénalité. Le nombre de guichets infrajournaliers est passé de 6 à 12 entre 2003 et 2005. Le délai de neutralisation (durée entre la *gate closure* et l'intervalle de livraison) est passé en 2005 à 2 heures (CRE [2006c]). Le coefficient de pénalité (k) a été diminué 3 fois depuis 2003 (CRE [2004a], CRE [2005a], CRE [2006a], CRE [2006b]).

La figure 3-25 montre les prix des écarts négatifs et positifs entre 2004 et 2006, et les différences entre les prix des écarts négatifs et positifs (axe des ordonnées à droite). Nous remarquons que la différence des prix des écarts est toujours considérable malgré les diminutions du coefficient de pénalité (k).

[191] Les règles ont évolué aussi pour la participation d'acteurs étrangers au mécanisme. Cette modification a permis l'amélioration des conditions de concurrence grâce aux offres d'ajustement venant des pays voisins (notamment la Suisse et l'Allemagne) (CRE [2006c]).

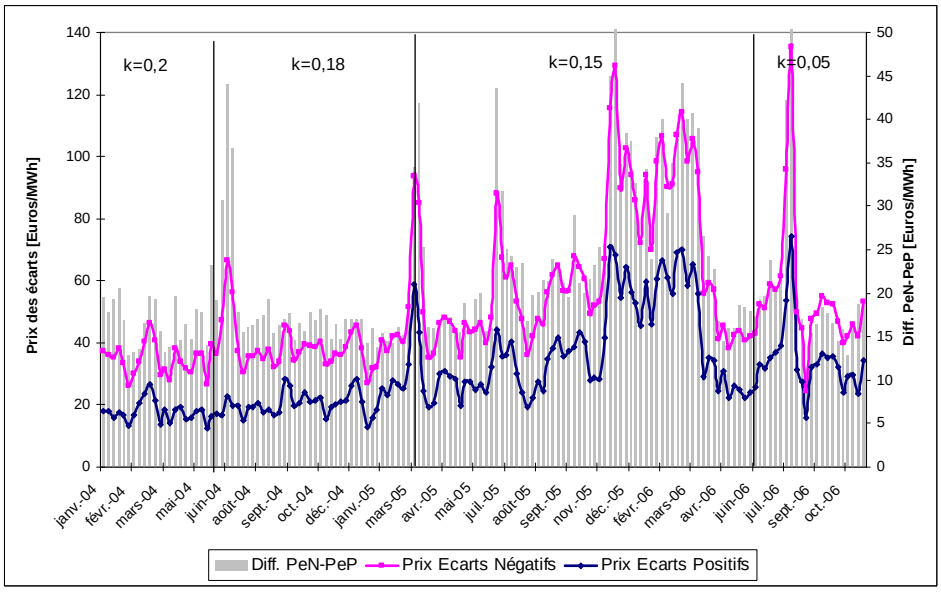

Figure 3-25 : Prix des écarts en France 2004-2006 (moy. mobile hebdomadaire – source: RTE web site)

Le tableau 3-11 présente les valeurs moyennes des prix des écarts, et les différences entre ces valeurs, pour les périodes correspondant aux différents coefficients de pénalité (k).

Tableau 3-11 : Prix des écarts en France 2004-2006 (source : www.rte-france.com)

Période	k	Prix Ecarts Négatifs (PeN) [€/MWh]	Prix Ecarts Positifs (PeN) [€/MWh]	Différence PeN-PeP [€/MWh]	Différence PeN-PeP %[192]
Janvier 2004-Juin 2004	0,20	37,2	17,9	19,3	70%
Juillet 2004-Mars 2005	0,18	40,7	23,0	17,7	55%
Avril 2005-Juin 2006	0,15	64,0	38,6	25,4	50%
Depuis Juillet 2006	0,05	55,5	34,7	20,8	46%

Malgré la diminution du coefficient de pénalité, la valeur absolue des différences n'a pas présenté de tendance claire, notamment négative. Cela est dû en partie à l'utilisation d'un coefficient multiplicatif de la pénalité. En effet, les différences entre les prix des écarts dépendent du niveau moyen de prix. Ainsi, entre 2005 et 2006, une remonté considérable des prix d'équilibrage (et des prix de l'énergie en général) s'est traduite par de fortes différences des prix des écarts. En revanche, nous trouvons une diminution continue des différences des prix des écarts en pourcentage (%) (dernière colonne) lors des modifications successives du coefficient de pénalité (figure 3-26).

[192] Cf. note de bas de page n° 184.

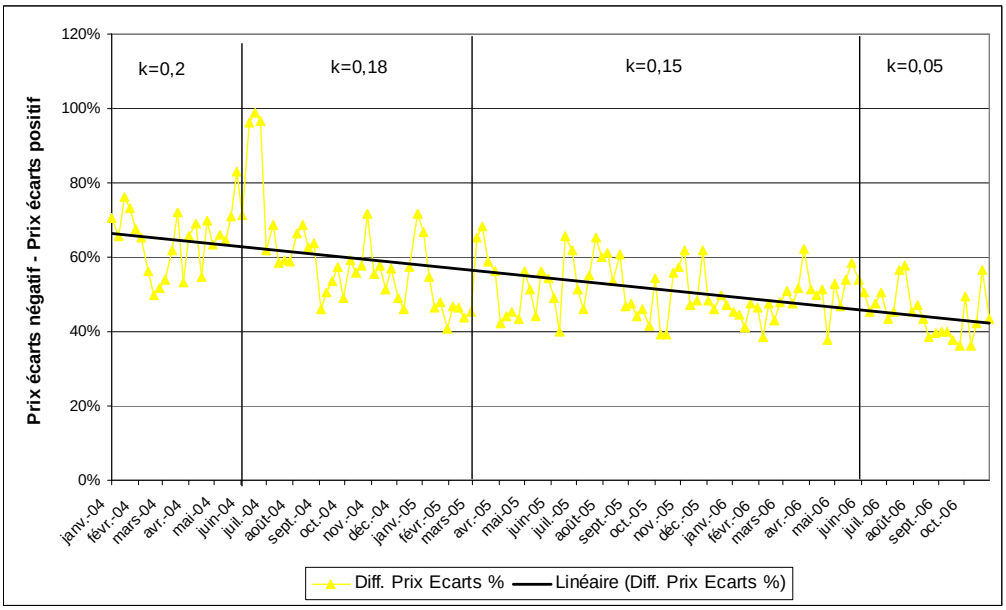

Figure 3-26 : Evolution de la différence PeN – PeP « proportionnelle » ou en pourcentage

En conclusion, le design du module du temps réel de la France correspond à notre définition de mécanisme d'ajustement. Ce mécanisme s'organise autour d'une enchère discriminante et sur la base d'un système de règlement des écarts à double prix. Les prix des écarts sont ajustés par des pénalités explicites (coefficient k). Le caractère multiplicatif du coefficient de la pénalité combiné à l'augmentation des prix de l'énergie expliquent que, malgré des successives diminutions du coefficient k, l'intensité de pénalités sur les écarts continue à être importante.

2.3.5 Le mécanisme d'ajustement aux Pays Bas

En 2001, Tennet, le Gestionnaire du réseau néerlandais, mit en œuvre un mécanisme d'ajustement pour la gestion de l'équilibre en temps réel. Avec ce mécanisme, Tennet compense les déséquilibres en sélectionnant des offres [193], après les avoir interclassées selon un critère de préséance économique (*merit order*). Les offres acceptées sont réglées au prix de l'offre marginale. Ce mécanisme fonctionne donc sous les principes d'une enchère à prix uniforme (Newbery-McDaniel [2003]). L'intervalle de livraison ou règlement est de 15 minutes et la *gate closure* est placée une heure avant chaque intervalle de livraison (« *rolling gate closure* »). Le mécanisme utilise comme base un système de règlement des écarts à double prix. Mais les règles spécifiques pour la détermination des prix des écarts ont évolué depuis 2001 et, actuellement, il a un fonctionnement très proche d'un « marché ».

[193] Ces offres sont divisées en « *regulating power* », réserves, et urgences. La totalité de la puissance non utilisée et techniquement disponible est mise à disposition de ce mécanisme dans les offres d'ajustement des producteurs (Huisman-Huurman [2004]).

Le système de règlement des écarts utilise comme base les prix des offres marginales acceptées. Ces prix des offres marginales sont ensuite affectés par un facteur de pénalité (*incentive component* IC) pour calculer les prix définitifs des écarts (le prix des écarts négatifs est augmenté en sommant « IC », et le prix des écarts positifs est diminué en soustrayant « IC »). Par simplicité, supposons, tout d'abord, que le facteur de pénalité est fixé à zéro (IC=0). Quand le système est équilibré dans une seule direction pendant un intervalle de livraison (acceptation des offres à la hausse seulement, ou acceptation des offres à la baisse seulement) les prix des écarts négatifs et positifs sont égaux au prix d'offre marginale acceptée. On constate que le fonctionnement expliqué jusqu'à ici correspond aux caractéristiques d'un « marché ». Néanmoins, si, pendant un intervalle de livraison, le système doit être équilibré dans les deux directions (« *double-sided regulation* »), le prix des écarts négatifs sera fixé au prix de l'offre marginale acceptée à la hausse, et le prix des écarts positifs sera fixé au prix de l'offre marginale acceptée à la baisse. Dans ces situations nous retrouvons les caractéristiques d'un « mécanisme » et son règlement des écarts à double prix. Bien évidement, si le facteur de pénalité (IC) est fixé différent de zéro, dans toutes les situations d'équilibrage du système, le règlement des écarts utilisera deux prix différents pour les écarts positifs et négatifs. Ce facteur de pénalité IC est recalculé toutes les semaines par Tennet en considérant le comportement des acteurs (déséquilibres individuels) et le déséquilibre total du système (Tennet [2005]). Le tableau 3-12 résume le système de calcul de prix des écarts pour les trois conditions du système (« long », « court », « double »).[194]

Tableau 3-12 : Prix des écarts au Pays Bas[195]

	Système équilibré seulement à la hausse (« long »)	Système équilibré seulement à la baisse (« court »)	Système équilibré à la hausse et à la baisse (« double »)
Ecart négatif	PMargB + IC	PMargH + IC	PMargH + IC
Ecart positif	PMargB – IC	PMargH – IC	PMargB – IC

Le design du règlement des écarts à double prix représenté dans le tableau 3-12 n'a pas changé depuis 2001 (Tennet [2001]). Cependant, deux faits ont transformé ce design pour l'approcher vers un design du type « marché ». Ces deux faits concernent d'un côté le facteur de pénalité IC et, d'un autre côté, les règles spécifiques pour définir si le système est « doublement équilibré » ou non. D'un côté, le facteur de pénalité IC, qui a atteint en début de fonctionnement en 2001 la valeur maximale de 11 €/MWh, a diminué considérablement et est actuellement égal à zéro (voir tableau 3-13). D'un autre côté, les règles spécifiques concernant la définition du système comme « doublement équilibré » (« *double-sided regulation* ») ont été modifiées en deux occasions (Octobre 2005 et Avril

[194] Notez qu'ici le système est considéré « long » (« court ») quand seulement des offres à la baisse (hausse) sont acceptées. Alors qu'en général, les autres designs considèrent le système comme « long » (« court ») quand le volume total d'offres acceptées à la baisse (hausse) est supérieur au volume total des offres acceptées à la hausse (baisse).

[195] Où PMargB est le prix de l'offre marginale appelée à la baisse, et PMargH est le prix de l'offre marginale appelée à la hausse.

2006)[196]. Ces deux modifications ont permis de diminuer la probabilité (ou la fréquence) de considérer le système comme doublement équilibré et la fréquence avec laquelle deux prix des écarts différenciés sont utilisés.

Tableau 3-13 : Facteur de Pénalité « IC »
(*Incentive Component* – source : www.tennet.nl)

Année	IC moyen [€/MWh]
2001	1,88
2002	0,46
2003	0,04
2004	0,00
2005	0,00
2006	0,00

La figure 3-27 et le tableau 3-14 montrent l'évolution des prix des écarts ainsi que la différence entre ces prix entre 2004 et 2006.

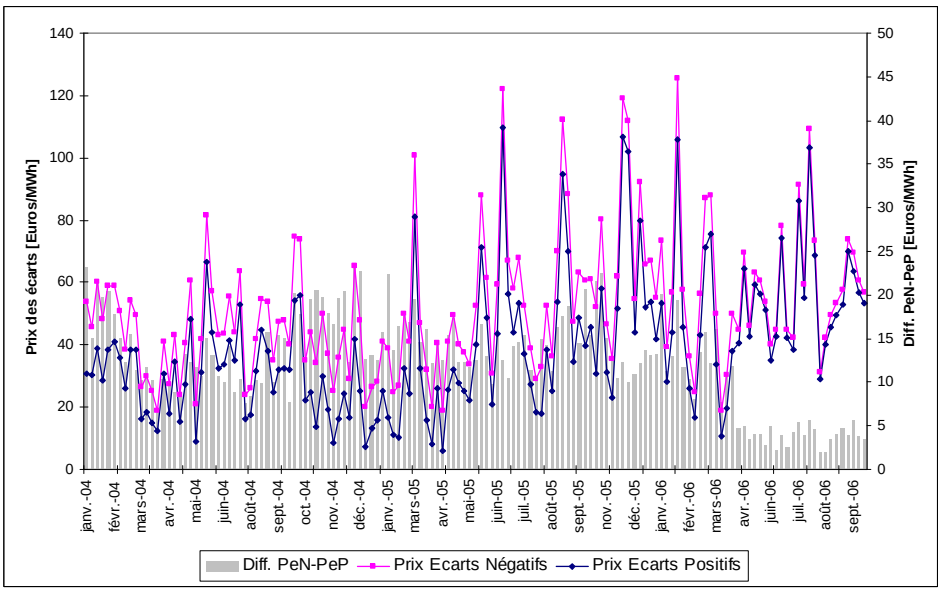

Figure 3-27 : Prix des écarts aux Pays Bas 2004-2006 (moy. mobile hebdomadaire – source www.tennet.nl)

On peut constater premièrement que les différences entre les prix des écarts négatifs et les écarts positifs ont été moins importantes que dans les autres designs étudiés précédemment (différence en pourcentage des prix des écarts de 7% pour les Pays Bas, comparée à 57% pour la Belgique et de 46% et 37% pour la France et l'Angleterre). On peut constater, deuxièmement, que la différence des prix des écarts est devenue très petite après avril 2006. Le fonctionnement actuel de ce mécanisme est donc très proche du design de type « marché ».

[196] Voir sur www.tennet.nl, « *news* » correspondant aux dates 21/10/2005 (*Change to Imbalance Price System*) et 20/04/2006 (*Change to the LFC regulation bid selection*).

Tableau 3-14 : Prix des écarts aux Pays-Bas 2004-2006 (source www.tennet.nl)

	Prix Ecarts Négatifs (PeN) [€/MWh]	Prix Ecarts Positifs (PeP) [€/MWh]	Différence PeN-PeP [€/MWh]	Différence PeN-PeP %[197]
Janv 2004 - Oct 2005	47,5	33,0	14,5	36%
Nov 2005 - Avr 2006	62,2	50,1	12,1	22%
Mai 2006 - Sept 2006	59,2	55,3	3,9	7%

On conclut que le design au Pays-Bas dont les règles de base correspondent à un mécanisme d'ajustement avec un système de règlement des écarts à double prix et la possibilité d'utilisation de pénalités, fonctionne à l'heure actuelle d'une manière très proche d'un « marché ». L'intensité des pénalités est faible. Ce module se rapproche donc d'un prix unique de l'énergie en temps réel.

Au long de cette sous-section 2.3, nous avons présenté les caractéristiques particulières de 5 designs du module du temps réel. Ces différentes variantes du design sont résumées dans le tableau 3-15 pour chacun des cas étudiés. Chacun de ces designs se rapproche plus ou moins d'un des deux types de designs définis dans la sous–section 2.2 : le « marché du temps réel » et le « mécanisme d'ajustement ».

Tableau 3-15 : Caractéristiques de design du module du temps réel

	Etape n° 1	Etape n° 2		Etape n°3					
	Equilibrage	Type d'enchère		Règlement des écarts		Prix des écarts		Pénalité Explicite	
	Merit Order	Prix uniforme	*Pay as Bid*	Prix unique	Double Prix[198]	Prix Marginal	Prix Moyen	Non	Oui
Belgique	X		X		X (57%)	X			X
France	X		X		X (46%)		X		X
Angleterre	X		X		X (37%)		X[199]	X	
Pays Bas	X	X			X (7%)	X			X
Texas	X	X		X (0%)[200]		X		X	

Indépendamment des détails spécifiques du design, nous remarquons que les variantes mises en œuvre dans les différents pays forment une gamme qui se répartit selon l'intensité de pénalités sur les écarts. Si on compare les designs étudiés par la différence entre le prix des écarts négatifs et le prix des écarts positifs, nous pouvons les placer sur un axe qui part d'un marché du temps réel où un seul prix de l'énergie est utilisé pour toutes les transactions du temps réel. Ainsi, le design du module du temps réel au Texas donne l'origine à cet axe. Ensuite, nous trouvons le design des Pays-Bas qui fonctionne à l'heure

[197] Cf. note de bas de page n° 184.
[198] Entre parenthèses la différence des prix des écarts en pourcentage entre 2004 et 2006.
[199] Prix marginal prévu pour la fin du 2006.
[200] Prix différenciés quand les coûts des réserves de remplacement sont alloués aux écarts négatifs. Nous n'avons pas trouvé d'information sur les valeurs réelles de ces coûts. Potomac [2004] rapporte que ces coûts ne sont pas importants.

actuelle d'une manière très proche à un « marché » (différences entre les prix des écarts de seulement 7%) mais qui a une conception correspondant à un mécanisme d'ajustement avec un système de règlement des écarts à double prix et la possibilité d'utilisation de pénalités. Plus loin, on trouve le mécanisme d'ajustement de l'Angleterre, qui est un exemple de mécanisme sans pénalités explicites mais avec des règles spécifiques qui créent également des prix des écarts pénalisants. Encore plus loin, le mécanisme Français basé sur un système de règlement des écarts à double prix ajustés par des pénalités explicites (coefficient k) et qui malgré les successives diminutions du coefficient de pénalité continue à avoir des différences entre les prix des écarts. Enfin, nous trouvons le design du mécanisme d'ajustement Belge où des pénalités explicites sont utilisées et où la différence entre les prix des écarts négatifs el prix des écarts positifs est proportionnellement la plus importante (57 % de la moyenne entre les différences de prix).

L'existence de cette gamme de designs montre la faisabilité de différents types d'option de design. Le fait qu'il existe plusieurs possibilités de design pour un module nous amène directement sur la question de les comparer. Dans la sous-section suivante (2.4) nous analyserons les différents aspects afin de pouvoir comparer et choisir le design.

2.4 Le choix du design : incitations et conséquences économiques

Nous avons étudié les différentes variantes de design d'un module du temps réel. Ces variantes peuvent s'approcher d'un design du type « marché » ou d'un design du type « mécanisme ». La principale différence entre ces deux types de design est la forme de règlements des transactions du temps réel et notamment l'utilisation de pénalités pour le règlement des écarts. Le choix de la forme de règlement des transactions du temps réel, et en conséquence d'une intensité des pénalités, n'est pas anodin puisqu'il fixera les incitations données aux participants du marché et influencera ainsi leur comportement en provoquant différentes conséquences économiques. Chaque type de design aura en conséquence différentes propriétés économiques.

Comme nous le verrons, ces conséquences économiques sont de différents ordre et peuvent être présentées en trois volets : l'impact du design du module du temps sur la séquence de marchés, la contribution à sécurité du réseau, la résistance au *gaming* et au pouvoir de marché. Un design optimal devrait procéder à la comparaison des différentes sur ces trois volets. Cependant, parmi les différentes conséquences possibles ce sont celles de l'impact du design du module de temps réel sur la séquence de marchés qui constituent le cœur de l'analyse économique de l'organisation des marchés comme substitut à l'entreprise intégrée. Elles seront donc, pour cette raison, placées au cœur de notre modélisation du quatrième chapitre.

2.4.1 L'impact du design du module du temps réel sur la séquence de marchés

Le module du temps réel est le dernier module de la séquence des marchés d'énergie. Ce module donne à cette séquence des marchés son seul mécanisme de formation des prix de l'énergie « en temps réel ». Cette énergie « en temps réel » étant elle-même la seule forme d'énergie physiquement échangeable entre les participants du marché, ses prix vont donc donner une base et un fondement « réels » à toute la chaîne des échanges *forward* et des marchés *forward* jusqu'à la *gate closure* (cf. section 1).

Etant donné l'impossibilité du stockage et les incertitudes présentes sur tout système électrique, les incitations données par le module du temps réel sont très volatiles car elles expriment les conditions du système en temps réel. La volatilité et l'incertitude créent des risques pour les acteurs du marché. Les acteurs prendront leurs décisions sur les marchés *forward* et leurs décisions de production en considérant ces risques selon l'information disponible au moment de prendre les décisions. En conséquence, sur un environnement où le comportement des acteurs de marché face au risque est important, les incitations données aux participants et l'information disponible sont les deux facteurs fondamentaux à considérer pour le design du module du temps réel.

Le choix du design du module du temps réel modifie les incitations données aux participants ainsi que les risques présents sur ce module. L'application de pénalités sur certaines transactions modifiera les incitations données par le module du temps réel et, en conséquence, le comportement des participants sur les marchés *forward* de la séquence. Le nouveau comportement des acteurs face aux nouvelles incitations et aux risques provoquera des conséquences sur toute la séquence de marchés. Le phénomène de « sur-couverture » (ou « *over-contracting* ») (Newbery-McDaniel [2003], cf. sous section 2.1) est un bon exemple des conséquences sur la séquence de marchés d'énergie d'un système de règlement des écarts avec pénalités asymétriques, c'est-à-dire pénalisant plus les écarts négatifs que les écarts positifs. Etant donné l'asymétrie des effets provoqués, les participants achètent sur les marchés *forward* un volume supérieur à leur niveau moyen de consommation ; cela se traduit par des positions contractuelles globales présentées à la *gate closure* au dessus du niveau de consommation. En effet, les agents préfèreront être, en général, en écart positif et avoir une faible probabilité de se trouver en écart négatif pour diminuer leur exposition au risque.

L'ampleur des incitations et des risques introduits dans le module du temps réel dépend fortement de l'incertitude, c'est-à-dire du niveau d'information disponible pour les participants sur les marchés *forward*. La position temporelle de la *gate closure*, ou la durée entre la *gate closure* et le moment du temps réel, est le paramètre déterminant pour évaluer

le niveau d'information disponible pour les participants avant la *gate closure*, et donc l'ampleur des incitations et des risques (Figure 3-28).[201]

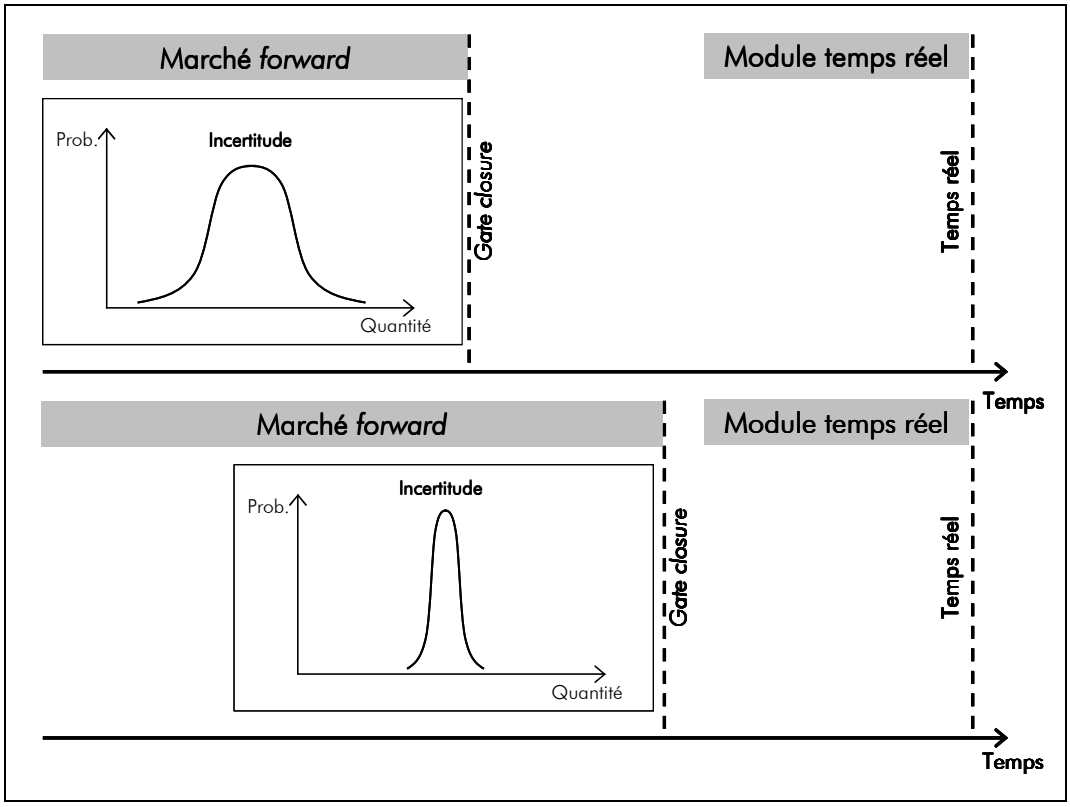

Figure 3-28 : Position temporelle de la *gate closure*

La position de la *gate closure* détermine le niveau de l'incertitude présente dans le dernier marché *forward*, celui qui est juste avant la *gate closure*. Ceci marque l'influence des incitations données par les signaux de prix du temps réel. Car les signaux de prix du temps réel affectent les transactions du temps réel ou les déviations par rapport aux transactions *forward*. Ainsi, si les incitations données pénalisent les transactions du temps réel, une *gate closure* près du temps réel permettrait aux participants de réaliser d'autres transactions sur le marché *forward* et de modifier leurs positions contractuelles et leurs programmes avec moins d'incertitude. En théorie, sans coûts de transaction et avec un marché décentralisé parfaitement liquide, si la *gate closure* était placée juste avant le temps réel, quand presque tous les aléas auraient disparu, le module du temps réel n'aurait pas de transactions à gérer et son design et ses incitations n'auraient pas d'influence sur les autres marchés. Cependant cela suppose que le module du temps réel pourrait être complètement décentralisé, ce qui n'est pas réalisable. Les caractéristiques de liquidité et d'absence de coûts de transaction sont trop exigeantes dans la réalité. Généralement les marchés *forward* infra journaliers

[201] L'information disponible pour les participants dépendra aussi d'autres caractéristiques du design du module du temps réel comme la quantité d'information fournie aux participants et de leur capacité à se servir de ces informations pour prendre des décisions. En effet, si les mécanismes des incitations sont plus complexes, ils créeront un besoin informationnel plus grand et donc plus de capacité de traitement des informations par les participants (Green-McDaniel [1999]).

(fonctionnant juste avant la *gate closure*) ne sont pas assez liquides. Les participants ne pourront pas facilement trouver des contreparties leur offrant des contrats additionnels pour modifier leurs positions contractuelles et leurs programmes journaliers. En conséquence, la position « effective » de la *gate closure* peut être, en pratique, plus éloignée du temps réel que la position de fait ou définie administrativement dans le design du module du temps réel. Par ailleurs, la proximité de la *gate closure* au temps réel est limitée par le temps minimal après la fermeture des marchés *forward* dont le Gestionnaire du Réseau de Transport a besoin pour analyser l'information qu'il y a recueillie, pour la confronter avec ses propres prévisions, avec l'état général du réseau et du système, afin d'en inférer les ressources nécessaires pour gérer le système.[202]

En conclusion, les pénalités et la position temporelle de la *gate closure* moduleront les incitations données aux participants et leur impact sur la séquence de marchés. La variation de ces paramètres de design peut provoquer aussi d'autres conséquences comme par exemple des effets asymétriques sur les différents participants et des effets sur l'efficacité productive.

Le choix des règles qui gouvernent un module du temps réel devrait être fait sur une base non-discriminatoire, et ainsi avoir un impact symétrique pour tous les types de participants (ILEX [2002]). La propriété de non-discrimination est importante pour éviter de créer de barrière à l'entrée et faciliter ainsi l'amélioration des conditions de concurrence sur le marché à long terme. Par exemple, un type de design créera des barrières à l'entrée s'il avantage les participants intégrés verticalement ou les participants de grande taille et pénalise les participants non intégrés verticalement ou de petite taille (LE [2004]). L'impact asymétrique des règles peut concerner aussi le GRT. Par exemple, les deux types de design pour le module du temps réel, marché du temps réel et mécanisme d'ajustement, n'impliquent pas le même bilan économique pour le GRT. Le design du type marché implique un bilan neutre pour les revenus du GRT. Au contraire, le design du type « mécanisme » implique généralement des bilans positifs pour les revenus du GRT. Ces aspects particuliers nous amènent à nous interroger sur la préférence du GRT entre les différents designs, alors que ce GRT est généralement un de principaux acteurs dans la définition des règles de marchés.

Enfin, les incitations données par les différents designs peuvent aussi modifier les comportements des participants face à l'équilibrage, et avoir ainsi des effets sur l'efficacité productive. L'équilibre (ou le déséquilibre) est un phénomène qui concerne tout le système électrique et pas seulement les déséquilibres individuels des parties. En effet, le GRT doit gérer le déséquilibre net (long ou court) du système. Ce déséquilibre net du système sera toujours moins important, en termes relatifs, que la somme des déséquilibres individuels

[202] Le temps minimal nécessaire pour le GRT dépendra d'abord de l'importance des « problèmes » qu'il doit gérer par le biais d'actions sur le module du temps réel (ce sont les contraintes qui ne sont pas prises en compte par les marchés *forward*, comme par exemple les congestions).

(Joskow-Schmalensee [1983]). Il existe donc des économies d'échelle pour fournir les services d'équilibrage d'une manière centralisée au niveau du système. Les incitations données par les différents designs auront un impact sur la distribution de la responsabilité d'équilibrage entre le GRT et les participants, et ceci provoquera des effets sur le coût global de l'équilibrage et donc sur l'efficacité productive.

Un design qui pénalise d'une manière importante les écarts, incite les responsables d'équilibre à l'auto-équilibrage et à réduire les possibilités d'écarts individuels.[203] Imaginons un cas extrême où l'intensité de la pénalité appliquée aux déséquilibres est très grande. Dans un tel système d'incitations, les participants préfèrent garder leurs propres moyens d'équilibrage. Car « acheter » de l'énergie en temps réel sur le mécanisme d'ajustement peut coûter très cher. Si tous les participants suivent cette stratégie d'« autogestion » de l'équilibrage, il y aura un excès de capacités disponibles dans le système par rapport au niveau nécessaire si l'équilibrage se réalisait d'une manière centralisée. La décentralisation de l'équilibrage fait disparaître les possibles économies d'alea ou d'échelle fournies par un fonctionnement centralisé. L'excès de capacités disponibles est ainsi source d'inefficacités productives. D'abord parce que garder de la capacité disponible implique de coûts supplémentaires par démarrage et fonctionnement à charge vide des moyens de production. Ensuite, deuxième source d'inefficacité, les moyens de production fonctionnant à des régimes inefficaces. Etant donné les incitations, particulièrement plus fortes sur les écarts négatifs que sur les écarts positifs, un producteur préfère programmer un plus grand nombre d'unités de production pour éviter d'être exposé à des prix très hauts des écarts négatifs. En cas de panne d'une des unités de production, le producteur peut reprogrammer rapidement les autres unités de production pour remplacer l'énergie non-produite par l'unité en panne. L'utilisation des unités de production à des régimes « non-optimaux » provoque des surcoûts de production et ainsi des inefficacités productives.[204] Une troisième source d'inefficacités viendrait de l'utilisation trop élevée par rapport à l'optimum, de moyens de productions très inflexibles (cf. Chapitre 4).

En conclusion, le choix d'un design pour le module du temps réel devrait considérer son impact sur la séquence de marchés d'énergie. La forme de règlement des écarts, notamment l'application de pénalités, définira les incitations données aux participants et influencera sur leur comportement. Premièrement, les incitations changeront le comportement des acteurs sur les marchés de la séquence. Deuxièmement certaines règles

[203] Les possibilités d'auto-équilibrage dépendent de la manière dont les écarts sont calculés. Le calcul des écarts pour chaque responsable d'équilibre peut se réaliser ou bien en considérant toutes activités confondues (production, consommation et *trading*) ou bien en considérant les activités séparément. Si le calcul des écarts se réalise au niveau du responsable d'équilibre (toutes activités confondues), les écarts venant de la production et de la consommation peuvent se compenser, ce qui permet facilement l'auto-équilibrage. Si le calcul des écarts est fait par activité, les écarts venant de la production ne peuvent s'annuler avec les écarts venant de la consommation qu'avant la *gate closure*.

[204] Henney [2002] rapporte des inefficacités de ce type en Angleterre où le design correspond à un mécanisme d'ajustement (cf. sous-section 2.3.2). En effet, une capacité importante de moyens de production au charbon peut être amenée à fonctionner en sous régime.

peuvent affecter les acteurs asymétriquement. Troisièmement, les différents designs impliquent différents bilans économiques pour le GRT (acteur principal pour la définition des règles). Quatrièmement, les choix de règles peuvent impacter l'efficacité productive. Le chapitre suivant (N°4) sera consacré à la modélisation de la séquence de marché d'énergie et à la comparaison quantitative des différents types de design. Nous étudierons cette question plus en détail en déterminant les conséquences provoquées par les deux types de design, « marché » et « mécanisme », sur la séquence des marchés d'énergie.

Nous venons de voir vu que les différents designs peuvent modifier les propriétés économiques générales de la séquence de marchés d'énergie. Cet effet est le principal effet économique à retenir dans la construction d'une séquence de marchés d'énergie comme substitut d'une séquence de contrôles d'une entreprise intégrée. Cependant, il est absolument vrai que, du point de vue de l'exploitation technique du système électrique par le gestionnaire du réseau, ces différents designs ont aussi des propriétés différentes concernant la sécurité du réseau.

2.4.2 La sécurité du réseau

Les deux variantes de design du module du temps réel peuvent être comparées du point de vue de la sécurité du réseau. Les « mécanismes d'ajustement » ont une nature de mécanisme « fort » de sécurité, auxquels on ne devrait avoir recours que pour des cas exceptionnels et avec une priorité forte à l'objectif de sécurité. En revanche, dans un design du type « marché » on peut acheter et vendre « librement » de l'énergie et l'objectif unique n'est celui de la sécurité.

Un des principaux arguments courants en faveur des mécanismes d'ajustement, et contre un « marché », est que l'utilisation d'un prix unique de marché pourrait mettre en danger la sécurité du système électrique. Les participants ne seraient pas assez incités à rester en équilibre au delà de la *gate closure*, et ainsi accroîtraient l'exposition du système électrique au risque de défaillance, en augmentant le volume d'énergie d'équilibrage traité au moment du temps réel. En pratique, le fait de « pénaliser » ou de décourager les écarts en temps réel revient aussi à transférer du GRT vers les acteurs de marché une partie de la responsabilité de l'équilibrage. Puisqu'on en attend un équilibrage plus important *ex ante* : des équilibrages supplémentaires seront mis en œuvre par les participants des marchés *forward* avant la *gate closure*, et sera constaté par le GRT après la *gate closure*.

La conception d'un cadre complet d'incitations pour le module du temps réel devrait aussi prendre en compte le facteur de la sécurité du réseau. Un cadre de cette nature consisterait à déterminer et à allouer, de manière marginale, outre les coûts d'équilibrage venant de l'acceptation d'offres d'ajustement, les coûts engendrés par une diminution ou augmentation de la sécurité du réseau en conséquence des actions individuelles des participants. Un tel système d'incitations conduirait les participants à prendre des décisions efficaces et « sûres », en considérant leurs deux principales options, celle de l'auto-

équilibrage et celle de laisser au GRT la tâche de l'équilibrage. En pratique, il existe des difficultés, théoriques de mise en œuvre, pour la détermination et l'allocation efficaces des coûts de variation de la sécurité. D'abord il est très difficile de déterminer ces coûts exactement. De plus, ces coûts peuvent être provoqués par diverses actions des participants, et pas seulement par des « écarts ». Généralement les designs ne considèrent que les écarts « contractuels », c'est-à-dire les écarts calculés comme la différence entre les positions contractuelles et les injections et soutirages en temps réel. Mais, outre les coûts d'équilibrage en énergie proprement dits et venant des écarts contractuels, il existe aussi des coûts provoqués par des informations imprécises données par les participants et utilisées par le GRT, mais non prises en compte dans les règlements du temps réel (par exemple sur les programmes de production) (ILEX [2002]). Or les informations imprécises peuvent diminuer le niveau de sécurité du réseau.[205]

Face à ces difficultés réelles pour une allocation efficace des coûts d'équilibrage, les designs des modules du temps réel (« marché » et « mécanisme ») ont choisi des conceptions différentes pour déterminer et allouer les coûts d'équilibrage.

D'un côté, les marchés du temps réel, avec leur prix unique de l'énergie en temps réel, considèrent que les coûts provoqués ou évités par un déséquilibre (négatif ou positif) correspondent à la valeur de l'offre marginale acceptée pour équilibrer le système. Les designs du type « marché » distinguent en général les informations physiques (programmes de production et de soutirage) des informations contractuelles. Cette différenciation se fait, par exemple, en permettant de réaliser des offres virtuelles (*virtual bidding*), ou en permettant de présenter des programmes déséquilibrés (au Texas : des programmes de productions qui sont différents des positions contractuelles). Grâce à cette différenciation, dans certains cas, quand le GRT détecte qu'il n'y aura pas assez de moyens d'équilibrage disponibles et donc de possibles problèmes de sécurité, il appelle certains moyens en réserve spéciale pour assurer un niveau adéquat de capacité disponible de production en temps réel. Le cas du Texas est un bon exemple pour montrer qu'un design du type « marché » ne pose pas nécessairement des problèmes de sécurité sur le réseau. Lorsque le module du temps réel au Texas a commencé à fonctionner comme un marché, le volume net de déséquilibre, celui qui est importante pour la gestion du GRT a augmenté (cf. sous-section 2.3.1). Mais il est resté dans des valeurs gérables en temps réel par le GRT (~3 % de la consommation).

D'un autre côté, les mécanismes d'ajustement considèrent généralement que tous les écarts provoquent des coûts supplémentaires de sécurité, ou que les bénéfices apportés par un écart de direction contraire au système sont moins importants que la valeur de l'énergie en temps réel. La prise en compte et l'allocation de ces coûts supplémentaires se traduisent donc par des règles spécifiques, ou par l'intégration de pénalités pour la

[205] Au Texas, la Commission de Régulation réalise des rapports annuels où les participants sont classés par niveau de précision dans l'information donnée (« *ressource plan* ») au GRT (ERCOT). Ce classement servirait comme une incitation à transmettre des informations correctes.

détermination des prix des écarts. Les règles spécifiques concernent l'allocation des coûts des capacités de production disponibles sous le contrôle du GRT (offres d'ajustement, réserves, etc.) pour faire face aux possibles déséquilibres. Or, l'introduction de pénalités, ajoutées aux coûts d'équilibrage proprement dits (offres d'ajustement acceptées), surpassent en général les coûts supplémentaires provoqués par les déséquilibres (ILEX [2002], Menanteau-Finon [2003]).

Les effets agrégés provoqués par les comportements individuels des participants, en réponse aux incitations pénalisant les écarts, correspondent, tout d'abord, à une diminution effective des déséquilibres nets du système post *gate closure*. De plus, la plupart du temps il y a alors sur le système un volume important de capacités de production disponibles (Cornwall [2002], Henney [2002]). Les deux effets, en se combinant, diminuent le niveau d'incertitude et donc les difficultés du GRT pour équilibrer le système en temps réel. La sécurité du réseau, au sens où le GRT devra gérer moins d'incertitudes pour l'équilibrage, est donc favorisée par ce type d'incitation (Roques *et al.* [2004]). Cependant, ce gain en sécurité du réseau n'est pas gratuit. Comme nous l'avons expliqué auparavant, les incitations qui pénalisent les écarts peuvent provoquer des conséquences importantes sur la séquence de marchés d'énergie. Par exemple, l'excès de moyens de capacité disponible peut entrainer des inefficacités productives ou des effets asymétriques entre les participants de différentes tailles ou envergure.

En conclusion, la sécurité du réseau est effectivement un deuxième volet qui devrait être considéré dans le choix d'un design optimal du module du temps réel. Mais ce travail est en soi un énorme chantier, celui de la valeur économique de la sécurité, bien au-delà du domaine propre de notre thèse (Joskow-Tirole [2006]). Un troisième volet des effets qui devraient être considérés est le *gaming* et le pouvoir de marché.

2.4.3 Le *gaming* et le pouvoir de marché

Les possibilités de manipulation des marchés (*gaming*) et les risques d'abus de pouvoir de marché sont deux faces d'un volet supplémentaire à prendre en compte au moment de choisir le design du module du temps réel et ses incitations.

La manipulation de marché consiste ici à réaliser certaines transactions *forward* et en temps réel, tout en présentant de fausses informations physiques au GRT, afin de maximiser le profit de l'ensemble de ces actions (Wolak [2005]). La manipulation de marché est parfois confondue avec l'arbitrage temporel. L'arbitrage temporel consiste en une prise de position contractuelle sur les marchés *forward* et sur le temps réel, mais sans falsifier des informations physiques données au GRT. Le *gaming* et les informations fausses données au GRT peuvent provoquer de problèmes de sécurité du réseau, et doivent être effectivement évités. Néanmoins les stratégies de *gaming* entre les marchés *forward* et le module du temps réel

ne sont pas facilement mises en œuvre et, en plus, elles sont rarement profitables (Sweeting [2000]).[206]

Les systèmes d'incitation basés sur le règlement des écarts à double prix des mécanismes d'ajustement sont reconnus comme « *anti-gaming* », car ils limiteraient les possibilités de manipulation de marché (Boogert-Dupont [2005])[207]. Le problème de règles anti-*gaming* à double prix et pénalités est que, pour éviter des possibilités de *gaming*, ils introduisent des distorsions à l'arbitrage temporel entre les marchés *forward* et le module du temps réel. Or, l'arbitrage temporel est bénéfique pour l'efficacité du marché (voir section 1). Dans certains cas, les rigidités imposées par le design du module du temps réel font que les participants, afin de profiter des arbitrages, sont obligés de donner des fausses informations au GRT. C'est le cas, par exemple, de l'exigence que les positions contractuelles soient couvertes exactement (équilibrées) par des programmes de production physiques. Dans un tel design, c'est tout arbitrage entre les marchés *day ahead* et le marché du temps réel qui est perçu comme du *gaming*, comme dangereux pour la sécurité du réseau car les participants doivent « falsifier » des informations physiques afin de pouvoir profiter des opportunités d'arbitrage. Il est certes important que les designs des modules du temps réel prennent en compte les différents types d'informations et de transactions entre les marchés pour éviter le *gaming*, mais en conservant les possibilités d'un arbitrage temporel qui serait bénéfique au système (Wolak [2005]).

Un autre argument contre les marchés du temps réel vient d'une structure de marché qui ne serait pas concurrentielle. Souvent, il y a peu d'acteurs qui participent activement à ce marché, et ces acteurs peuvent profiter de leur pouvoir de marché en retirant unilatéralement des capacités ou en offrant des prix très supérieurs à leurs coûts marginaux. Bien que cet argument soit valide, il n'a pas été démontré qu'un système pénalisant les transactions sur le temps réel supprime le problème du pouvoir de marché. Et, en cas d'abus de pouvoir de marché de la part de « vendeurs » actifs sur ce marché (e.g. producteurs), les « acheteurs » (e.g. *retailers*) sur le module du temps réel seraient doublement pénalisés, car ils devraient payer les prix élevés de l'énergie en temps réel plus les pénalités imposées. Une solution alternative aux problèmes de pouvoir de marché sur ce module serait de permettre la participation d'un plus grand nombre d'acteurs (en les incitant à proposer des offres) ou en autorisant les participants des zones adjacentes situées dans les

[206] Borenstein *et al.* [2004] rendent compte de la manipulation des marchés *day ahead* et du temps réel en Californie par ENRON (« *Dead-Star* » et « *Fath Boy* » gaming). Mais ils considèrent que cette manipulation a été rendue possible par un ensemble très particulier de conditions de design et de structures de marché. Ces conditions correspondent à un fort pouvoir de marché des *retailers*, combiné avec l'obligation d'achat sur le marché *day-ahead* et l'asymétrie d'information et pouvoir de marché pour les arbitrageurs (Borenstein *et al.* [2004]).

[207] Boogert-Dupont [2005] étudient les possibilités d'arbitrage entre la séquence de marchés d'énergie aux Pays-Bas (marché *day-ahead* -Bourse APX- et le mécanisme d'ajustement géré par Tennet). Ils trouvent qu'il n'est pas profitable d'appliquer des stratégies d'arbitrage pour le cas hollandais. Ces résultats montrent premièrement que toutes les possibilités d'arbitrage ont été utilisées (étant donné le design). Du fait que le design de type « mécanisme » aux Pays Bas est très proche d'un marché, ils montrent, deuxièmement, que même un design du type marché empêcherait le *gaming* sur la séquence des marchés d'énergie.

pays limitrophes par intégration de leurs modules du temps réel (cf. Chapitre 2, section 3).[208]

En conclusion, le choix d'un de design optimal pour le module du temps réel nécessiterait la prise en considération de trois volets : l'impact du design du module du temps réel sur la séquence de marchés, sur la sécurité du réseau, sur le *gaming* et le pouvoir de marché. L'impact du design du module du temps réel sur la séquence de marchés est le plus important pour la logique de notre thèse parce que le module du temps réel définit les signaux de prix au moment où la livraison du bien se réalise. Du fait des incertitudes, ces signaux transmettent des risques sur le reste de la séquence de marché. Pour cette raison, ces conséquences seront donc placées au cœur de notre modélisation du quatrième chapitre.

2.5 Conclusions de la section

Le module du temps réel est fondamental dans toute architecture du marché et il nécessite une grande d'attention pour sa conception. En effet, c'est en temps réel que les échanges du bien « électricité » se réalisent, et la coordination entre les composantes du bien électricité (énergie, transport et réserves) est déterminante pour le fonctionnement de l'industrie. Le design de ce module n'est pas simple car il s'agit d'un « pseudo marché » combinant des caractéristiques d'outil de contrôle et des caractéristiques de marché (Boucher *et al.* [2006]). Cette deuxième section a porté sur les caractéristiques de design d'un module du temps réel en se concentrant principalement sur la composante « énergie ».

Tout d'abord, nous avons découpé le module du temps réel en trois étapes caractéristiques afin de dégager les paramètres de design déterminants. La première étape est celle de la *gate closure* et de l'équilibrage du système (acceptation des offres d'ajustement) ; la deuxième étape couvre le règlement et la détermination des prix pour les transactions « actives » du temps réel ; finalement, la troisième étape correspond au règlement et à la détermination des prix pour les transactions « passives » du temps réel ou écarts. Pour chacune de ces étapes nous avons présenté les variantes possibles de design et ceci nous a permis de caractériser deux types de designs : d'une part, le « marché du temps réel » et, d'autre part, le « mécanisme d'ajustement ». La principale différence entre ces deux types de design est la méthode de règlements des transactions du temps réel (transactions actives et écarts). Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel. Le prix de l'énergie en temps réel correspond à l'offre marginale acceptée par le GRT pour équilibrer le système. Un mécanisme

[208] Cette alternative a été suivie par le Régulateur Français en autorisant la participation d'acteurs étrangers au mécanisme d'ajustement afin d'améliorer les conditions de concurrence (CRE [2006c]).

d'ajustement se reconnaît par l'usage de systèmes plus complexes pour le règlement des transactions du temps réel. Ces formes de règlement incluent généralement des prix différenciés pour les différents types de transactions et découragent (« pénalisent ») certaines transactions, notamment les écarts.

Nous avons ensuite réalisé des études de cas de 5 designs de modules du temps réel mis en œuvre : Texas, Angleterre, Belgique, France et Pays Bas. Les variantes mises en œuvre dans les différents pays forment une gamme de designs qui se répartissent entre les deux cas, « marché du temps réel » et « mécanisme d'ajustement », en fonction de l'intensité des pénalités appliquées. Pour chacun des cas, nous avons étudié dans quelle mesure les mécanismes d'ajustement s'éloignent d'un « marché du temps réel » en utilisant une mesure approximée de l'intensité des pénalités qui est la différence entre les prix des écarts négatifs et des écarts positifs. Si on compare les designs étudiés selon leur intensité des pénalités, nous distinguons, d'un côté, les cas du Texas et des Pays-Bas se rapprochant d'un marché du temps réel, sans pénalités, et, où un seul prix de l'énergie est utilisé pour toutes les transactions du temps réel. D'un autre côté, nous identifions les cas de l'Angleterre, la France et la Belgique comme des designs du type mécanisme d'ajustement dont les différences se basent sur l'intensité des pénalités.

Le choix de la méthode de règlement des transactions du temps réel, et en conséquence, celui du niveau d'intensité des pénalités, n'est pas anodin puisqu'il fixera les incitations données aux participants du marché et influencera ainsi leur comportement en générant différentes conséquences économiques. Chaque type de design aura en conséquence différentes propriétés économiques. Cela fait nous amène directement à la question de leur comparaison. C'est pourquoi, nous avons enfin analysé les possibles conséquences économiques de différents designs présentées en trois volets : l'impact du design sur la séquence des marchés, la contribution à sécurité du réseau, la résistance au *gaming* et au pouvoir de marché. Parmi ces différentes conséquences possibles, ce sont celles de l'impact du design sur la séquence de marchés qui constituent le cœur de l'analyse économique de l'organisation des marchés comme substitut à l'entreprise intégrée. En effet, les pénalités appliquées pour le règlement des écarts changeront les incitations données aux participants et influenceront leur comportement. Ces incitations données, très volatiles par nature, créent des risques pour les acteurs du marché. Les acteurs prendront leurs décisions sur les marchés *forward* et leurs décisions de production en considérant ces risques selon l'incertitude au moment de prendre les décisions. L'ampleur des incitations et des risques introduits dans le module du temps réel dépend fortement de l'incertitude sur les marchés *forward*. La position temporelle de la *gate closure* (la durée entre la *gate closure* et le moment du temps réel), et, l'intensité des pénalités seront donc les paramètres déterminants pour évaluer l'ampleur des incitations et des risques et en conséquence pour apprécier l'impact du design du module du temps réel sur la séquence de marchés.

Conclusions du chapitre

L'architecture d'un marché d'électricité est composée de plusieurs modules. Le module du temps réel est au centre de toute architecture de marché et il nécessite ainsi un traitement attentif quant à son design. Ce module doit coordonner en temps réel les trois composantes fondamentales de l'électricité : l'énergie, le transport et les réserves. Dans ce chapitre nous nous sommes concentrés sur la composante « énergie » qui est la « vraie commodité » dans l'électricité. Cette composante « énergie » est commercialisée sur une séquence de marchés.

L'objectif de ce chapitre a été de montrer quels sont les paramètres de design à prendre en compte sur le module du temps réel caractérisé de pièce fondamentale de la séquence de marchés d'énergie. Ce chapitre est organisé en deux parties. La première partie a été consacrée à l'étude de la séquence de marchés d'énergie et des liens qui existent entre ces différents marchés. La deuxième partie a porté sur les caractéristiques de design du module du temps réel qui fonctionne comme la dernière place du marché pour acheter ou vendre de l'énergie.

La séquence de marchés d'énergie est composée du module *forward* et s'achève par le module du temps réel. Le module *forward* d'énergie remplit deux rôles principaux : un rôle d'information pour faciliter la coordination et un rôle d'outil de couverture des risques du temps réel. Le module du temps réel doit assurer l'équilibrage physique en temps réel et émettre des signaux économiques cohérents avec les conditions du temps réel. A court terme, la séquence de marchés d'énergie peut être représentée par deux étapes principales, l'étape *forward* (marché *day-ahead* et marchés infrajournaliers) et l'étape du temps réel. Ces étapes sont séparées par la *gate closure*. Les transactions réalisées dans chacune de ces étapes correspondent au même bien physique, livré en temps réel, mais elles sont réglées séparément. Dans ce système de règlement, les transactions *forward* sont réglées au prix *forward* et les transactions du temps réel (déviations des injections ou des soutirages par rapport aux transactions *forward*) sont réglées au prix du temps réel.

Les marchés composant cette séquence sont liés car ils traitent le même bien livré en temps réel. Ces liens économiques sont créés par la possibilité d'un arbitrage temporel entre les étapes. Sur une séquence de marchés d'un bien stockable, la relation qui lie les différents marchés placés à différents horizons de temps est $p^F = E[p^{RT}]$, c'est-à-dire le prix *forward* est équivalent à l'espérance mathématique du prix du temps réel. Cette relation typique d'un bien stockable n'est pas applicable directement à l'électricité, mais elle est néanmoins utile pour comprendre la base des liens entre les marchés de la séquence et l'importance d'étudier les caractéristiques du design du dernier marché de la séquence. Puisque les prix et les transactions *forward* sont fondés principalement sur les prix des transactions du temps réel, les caractéristiques de designs du temps réel déterminant les prix

des différentes transactions du temps réel provoqueront des effets sur toute la séquence de marchés.

La deuxième partie du chapitre a porté donc sur les caractéristiques de design d'un module du temps réel. Afin de dégager les paramètres importants de design, nous avons, tout d'abord, présenté les trois étapes caractéristiques dans tout module du temps réel : la *gate closure* et l'équilibrage du système ; le règlement des transactions « actives » (offres d'ajustement) et, enfin, le règlement des transactions « passives » du temps réel ou écarts. Pour chacune de ces étapes nous avons présenté les variantes possibles de design. Nous avons caractérisé de lors deux types de designs : d'une part, le « marché du temps réel » et, d'autre part, le « mécanisme d'ajustement ». La principale différence entre ces deux types de design est la forme de règlement des transactions du temps réel. Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel. Un mécanisme d'ajustement utilise des formes de règlement incluant des prix différenciés pour les différents types de transactions et pénalisant certaines transactions, notamment les écarts. Afin de tester empiriquement cette typologie, nous avons ensuite réalisé des études de cas de 5 designs de modules du temps réel mis en œuvre : Texas, Angleterre, Belgique, France et Pays Bas. Les variantes de mises en œuvre dans les différents pays forment une gamme de designs qui se répartissent entre les deux cas extrêmes selon l'intensité de leurs pénalités. D'un côté, les designs du Texas et des Pays-Bas proche d'un design du type « marché » sans pénalités et d'un autre côté les designs de l'Angleterre, la France et la Belgique correspondent au design de « mécanisme d'ajustement » avec différentes intensités de pénalités.

Le choix de la forme de règlement des transactions du temps réel, et en conséquence, de l'intensité des pénalités, n'est pas anodin, il fixera les incitations données aux participants du marché et influencera leur comportement en provoquant différentes conséquences économiques. Chaque type de design aura donc des propriétés économiques différentes. Le fait qu'il existe plusieurs possibilités de design pour un module nous amène directement sur la question de leur comparaison. La confrontation des alternatives de design du module du temps réel est d'un réel intérêt empirique dans l'Union Européenne. Remarquons ici la différence marquée entre les designs d'un sous-ensemble de cas étudiés, la France, la Belgique et le Pays-Bas. Premièrement, cette comparaison est utile car actuellement, ces pays cherchent à améliorer leurs architectures de marché. Deuxièmement, ces trois pays ont intégré leurs modules *forward* et sont en train d'évaluer l'intégration de leurs modules du temps réel (cf. Chapitre 2, section 3). Or, les propositions d'intégration comptent principalement sur une mise en commun des offres d'ajustement (étapes n°1 et n°2 du design d'un module du temps réel) mais sans harmonisation sur la méthode de règlement des écarts.

Comparer les différents designs suppose la considération des conséquences économiques de différent ordre qui peuvent être séparées en trois volets : l'impact du design du module du temps réel sur la séquence des marchés, la contribution à la sécurité

du réseau et la résistance au *gaming* et au pouvoir de marché. Parmi ces différentes conséquences possibles, celles concernant l'impact du design sur la séquence de marchés constituent le cœur de l'analyse économique de l'organisation des marchés comme substitut à l'entreprise intégrée.

Dans un environnement où le comportement des acteurs de marché face au risque est important, les incitations données aux participants et l'incertitude sont les deux facteurs fondamentaux à considérer pour évaluer l'impact des différents designs du module du temps réel sur la séquence de marchés. Ces incitations changeront le comportement des acteurs sur les marchés de la séquence. L'utilisation de pénalités et la position temporelle de la *gate closure* sont les paramètres déterminants du design d'un module du temps réel pour évaluer l'ampleur des incitations et des risques. Ces différents designs peuvent affecter les acteurs de manière asymétrique ou peuvent impacter l'efficacité productive. Le chapitre suivant (N°4) sera consacré à la modélisation de la séquence de marché d'énergie et à la comparaison des différents types de design du module du temps réel. Nous étudierons cette question en déterminant les conséquences provoquées par les deux types de design, « marché » et « mécanisme » selon leur niveau de pénalité sur la séquence des marchés d'énergie.

Chapitre 4

Modélisation de la séquence des marchés d'énergie et du module du temps réel

Introduction 192

Section 1: Modélisation du module du temps réel (une zone de contrôle) 194
1.1 Définition de paramètres et variables 197
1.2 Modèle de base : présentation 198
1.3 Modèle de base : simulations numériques et discussion 214
1.4 Extension du Modèle à deux technologies de production..... 224
1.5 Conclusions de la section 233

Section 2: Modélisation de l'intégration d'architectures (deux zones de contrôle).235
2.1 Définition de paramètres et variables 237
2.2 Modélisation des architectures « séparées » et « intégrées » 238
2.3 Simulations numériques et discussion 247
2.4 Conclusions de la section 256

Conclusions du chapitre 259

Introduction

Les chapitres précédents ont souligné l'importance du module du temps réel et de son design dans une architecture de marchés. Ce module est le dernier de la séquence des marchés d'énergie et il gère la seule place où l'énergie physique est échangée entre acteurs de marché. Tous les autres marchés, fonctionnant avant le temps réel, sont des marchés *forward*, où on négocie bien des prix et des volumes, mais sans échange physique d'énergie. Pour cette raison, les caractéristiques de design module du module du temps réel auront des conséquences sur le reste de la séquence de marché.

Il existe deux variantes principales de design pour ce module qui sont : le « marché du temps réel » et le « mécanisme d'ajustement ». Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel. Le prix de l'énergie en temps réel correspond ainsi à l'offre marginale acceptée par le GRT pour équilibrer le système. Par contre, un mécanisme d'ajustement est plus complexe pour le règlement, incluant généralement des prix différenciés pour les différents types de transactions et décourageant les écarts en les « pénalisant » de manière asymétrique.

Ce chapitre propose d'un cadre d'analyse formalisé de la séquence des marchés d'énergie permettant de comparer les différentes alternatives de design pour le module du temps réel.

Deux questions primordiales guident alors les développements de ce chapitre. La première question est la comparaison d'alternatives de design pour un seul pays ou une seule zone de contrôle. Nous chercherons à évaluer les conséquences économiques des diverses alternatives de design sur une séquence de marchés d'énergie. Notre cadre d'analyse formalisé permettra, tout d'abord, de comprendre l'influence des paramètres clé de design et d'avoir une base analytique pour comparer les différents arrangements. Le choix de variantes de design provoque une série des conséquences qui devraient être prises en compte pour sélectionner le design le plus adapté pour une zone déterminée.

Cette comparaison des alternatives de design est d'un réel intérêt empirique dans l'Union européenne. Les différents designs qui sont mis en œuvre forment une gamme qui se répartit entre deux pôles extrêmes : le « marché » et le « mécanisme ». Actuellement, les pays cherchent à améliorer leurs architectures de marché et à les harmoniser (cf. Chapitre N°3). Par exemple, la Belgique et la France recourent à des mécanismes d'ajustement avec pénalités, tandis qu'une solution proche du marché de temps réel se trouve aux Pays-Bas.

Cependant, le design d'une architecture « nationale » de marché avec une seule zone de contrôle n'est qu'une partie du problème. Car ces architectures de marché « nationales » peuvent être insérées dans des marchés régionaux déjà mûrs ou en formation (Cf. Chapitre 2). La deuxième question porte donc sur l'intégration de plusieurs

architectures de marché, et particulièrement l'intégration des modules de temps réel. Ceci est aussi d'une importance élevée en Europe. Comme le marché unique européen d'électricité est toujours en construction, une évaluation des effets économiques de l'intégration d'architectures est nécessaire (Glachant-Lévêque [2005]). C'est encore plus nécessaire quand ces architectures ne présentent pas un design entièrement identique, notamment pour l'intégration des modules du temps réel. Concrètement, la France, la Belgique et les Pays-Bas ont réalisé la coordination de leurs bourses d'électricité (PXs ou marché *forward day ahead*) et leurs dispositifs d'allocation des interconnections (cf. Chapitre 2, section 3). Le fait que le fonctionnement de ces marchés *forward* est lié aux modalités de design des modules du temps réel renforce l'intérêt d'une modélisation des interactions entre modules intégrés imparfaitement harmonisés (cf. Chapitre 3, section 2).

La modélisation de la séquence des marchés d'énergie que nous proposons se base sur un modèle d'équilibre à deux étapes développé par Bessembinder-Lemmon [2000]. Les modèles d'équilibre en présence d'incertitudes se présentent comme l'alternative la plus pertinente pour les marchés d'électricité (Siddiqui [2002]). Cette modélisation prend en compte les caractéristiques intrinsèques de l'électricité comme la non-stockabilité, les incertitudes et la valorisation des risques par des participants qui opèrent sur les différents marchés de la séquence. Nous définirons les caractéristiques de design des modules de temps réel en introduisant une pénalité sur les écarts, ainsi que des positions variables de la *gate closure* pour modéliser les différentes variantes de design et leurs conséquences.

Ce chapitre sera organisé en deux sections. La première section présente la modélisation et analyse les conséquences économiques des différentes variantes de design du module du temps réel dans le cadre d'un seul marché isolé, avec une seule zone de contrôle. La deuxième section établira les conséquences économiques des différentes variantes de design entre plusieurs architectures de marché, ainsi que celles de l'intégration d'architectures entre plusieurs zones de contrôle.

Section 1: Modélisation du module du temps réel (une zone de contrôle)

Cette section a pour objet la modélisation de la séquence des marchés d'énergie avec différentes alternatives de designs pour le module du temps réel. Nous chercherons à évaluer les effets de ces diverses alternatives de design sur une séquence de marchés d'énergie pour un pays ou une zone de contrôle.

La modélisation de la séquence de marchés d'énergie avec interactions entre marchés *forward* et en temps réel dans un contexte concurrentiel et en présence d'incertitude a été étudiée par Bessembinder-Lemmon [2000] et Siddiqui [2002]. Dans les deux cas, on utilise des modèles d'équilibre avec représentation formelle des fonctions d'utilité valorisant le risque des participants.[209] Ce type de modèle présente un cadre très pertinent pour les marchés d'électricité. Car les autres modèles, de type « *cost and carry* » ne sont pas applicables pour l'électricité qui n'est pas stockable (cf. Chapitre 3, section 1).

Bessembinder-Lemmon [2000] utilisent un modèle d'équilibre à deux étapes pour étudier un marché en concurrence parfaite avec des agents ayant de l'aversion pour le risque (*risk adverse*). Ce modèle d'équilibre considère deux étapes (ou deux marchés) : une première étape de « marché *forward* » est suivie par l'étape de marché du temps réel. Chaque participant à ces marchés, acheteur (*retailer*) ou vendeur (*generator*) d'énergie, *forward* ou en temps réel, doit faire face à des incertitudes puisqu'il doit prendre des décisions sur le premier marché avant de connaître toutes les informations qui seraient pertinentes. La demande est stochastique et complètement inélastique en temps réel. Les incertitudes dans l'étape *forward* viennent principalement de ce que l'état de la demande en temps réel est inconnu. En temps réel, une fois l'état de la demande connu, un marché est utilisé pour équilibrer le système. En effet, les producteurs pouvant prendre des décisions de production en temps réel réagissent au prix du marché et équilibrent le système. Tous les acteurs de marché, présentent une aversion au risque. L'aversion au risque est formalisée par le biais d'une fonction d'utilité du type « moyenne-variance » avec un coefficient d'aversion au risque constant. Les acteurs de marché cherchent à maximiser leur utilité espérée dès la clôture du premier marché, qui fonctionne ainsi comme un marché de couverture des risques propres à la nature du second marché. Les résultats analytiques de l'équilibre développé dans ce modèle relient le prix *forward* aux caractéristiques statistiques du prix du temps réel. Ces résultats sont ensuite utilisés pour expliquer les *premium forward* ; c'est-à-dire, la différence entre le prix *forward* et l'espérance du prix du temps réel dans certains marchés (PJM et Californie).

---

[209] Green-McDaniel [1999] modélisant aussi la séquence de marchés d'énergie dans un contexte concurrentiel et sous la présence d'incertitude. Ils étudient l'interaction entre le marché *forward* et un module du temps réel dans un cadre de concurrence parfaite et des agents neutres au risque. Deux types de tarification du mécanisme d'ajustement (*pay-as-bid* et prix marginal) sont abordés.

Chez Siddiqui [2002], le modèle de Bessembinder et Lemmon est complété par l'introduction d'un marché *forward* de réserves. Sur la base d'un modèle d'équilibre à deux étapes avec des acteurs de marché ayant une aversion au risque, trois marchés sont représentés : les marchés *forward* d'énergie et de réserves, et le marché du temps réel gérant à la fois l'énergie et les réserves en temps réel. Siddiqui considère deux cas pour modéliser la demande d'énergie en temps réel : une demande stochastique et inélastique (comme chez Bessembinder et Lemmon) et une demande stochastique et élastique. Les résultats analytiques de l'équilibre développé dans ce modèle, reliant les prix *forward* de l'énergie et des réserves, et les caractéristiques statistiques du prix du temps réel sont ensuite utilisés dans le marché Californien pour expliquer le *premium forward* (différence entre le prix *forward* et l'espérance du prix du temps réel).

Notre modélisation de la séquence des marchés d'énergie se base sur le modèle d'équilibre à deux étapes développé par Bessembinder-Lemmon [2000]. Notre contribution porte principalement sur la modélisation des caractéristiques du module du temps réel (Bessembinder et Lemmon ne considèrent que la variante « marché »). Pour ce faire, nous ajoutons quelques modifications. Premièrement, nous introduisons la possibilité d'une pénalité en temps réel dans le modèle de base. Nous simplifions la variété empirique des différents arrangements pour le module du temps réel en ne distinguant que deux genres de design : un « marché de temps réel », sans pénalités, et un « mécanisme d'ajustement » avec pénalités. Dans ce cadre, nous définissons ces pénalités qui transforment un « marché en temps réel » en un « mécanisme d'ajustement », comme un paramètre affectant le prix des écarts positifs et négatifs en énergie mesurés au moment du temps réel. Le GRT compare les volumes contractés sur le marché de la première étape avant la *gate closure* (J-1 ou infrajournalier) avec les mesures réelles de consommation et de production effectives à la deuxième étape. Nous définissons aussi comme un paramètre la position temporelle de la « *gate closure* » avec laquelle le GRT clôt définitivement les échanges des marchés *forward* et ouvre la période finale où l'équilibrage du temps réel s'effectue sous son autorité. La position de la *gate closure* est simulée en réglant le niveau d'incertitude à l'étape *forward*. Dans ce modèle de base, nous ne considérons que des producteurs de type « flexible ». C'est-à-dire des producteurs qui peuvent prendre leurs décisions de production jusqu'au moment du temps réel (comme pour le cas de Bessembinder et Lemmon).

Dans une variante additionnelle de modélisation (modèle « étendu »), nous envisageons des producteurs utilisant deux technologies différentes de production. D'une part, une technologie « flexible », qui sera toujours capable de réagir aux aléas révélés après la fermeture du marché *forward*. D'autre part, une technologie « inflexible », qui sera donc incapables de réagir aux aléas révélés après la fermeture du marché *forward*.

Cette section est articulée autour de 5 sous-sections. Dans la sous-section 1.1, nous définissons les variables et les paramètres. Dans la sous-section 1.2, nous présentons le modèle de base correspondant avec une seule technologie de production (producteurs flexibles). Dans la sous-section 1.3, nous réalisons des simulations numériques du modèle

de base pour étudier les conséquences économiques des différentes variantes de design du module du temps réel selon les pénalités appliquées. Dans la sous-section 1.4, nous introduisons l'extension du modèle de base qui consiste à incorporer une nouvelle technologie de production, les générateurs inflexibles. Enfin, la sous-section 1.5 présente les conclusions issues des modèles et des simulations numériques.

1.1 Définition de paramètres et variables

Paramètres

- D_{ω} la demande globale d'électricité dans l'état du monde ω,
- $D_{j,\phi}$ la demande d'électricité du revendeur j dans le niveau ϕ,
- p_j la probabilité du niveau bas de la demande ($\phi = low$) du revendeur j,
- σ_{FG} le coefficient de coût marginal des producteurs flexibles,
- N_{FG} le nombre de producteurs flexibles,
- θ_{IG} le coefficient de coût marginal des producteurs inflexibles,
- N_{IG} le nombre de producteurs inflexibles,
- k le coefficient de pénalité,
- A_{FG} le coefficient d'aversion au risque des générateurs flexibles,
- A_R le coefficient d'aversion au risque des revendeurs,
- P^C le prix d'achat des consommateurs aux revendeurs.

Variables

Les variables de quantité:

- $X^F_{FG_i}$ la quantité vendue sur le marché *forward* par le producteur flexible i,
- $X^F_{IG_l}$ la quantité vendue sur le marché *forward* par le producteur inflexible l,
- $X^F_{R_j}$ la quantité achetée par le revendeur j sur le marché *forward*,
- $X^{RT}_{R_j,\omega}$ la quantité achetée ou vendue par le revendeur j au moment du temps réel pour l'état du monde ω,
- $X^{RT}_{FG_i,\omega}$ la quantité achetée ou vendue par le producteur flexible i au moment du temps réel pour l'état du monde ω,
- $X_{FG_{i,\omega}}$ la quantité produite par le producteur flexible i pour l'état du monde ω,
- X_{IG_l} la quantité produite par le producteur inflexible l.

Les variables de prix:

- P^F le prix du marché *forward*,
- P^{RT}_{ω} le prix du temps réel pour l'état du monde ω,
- $IP^{RT}_{j,\omega}$ le prix des écarts du revendeur j pour l'état du monde ω.

1.2 Modèle de base : présentation

Nous présentons dans cette sous-section notre modèle de base pour étudier les conséquences des différents designs du module du temps réel sur la séquence des marchés d'énergie. Nous prenons comme base le modèle d'équilibre à deux étapes de Bessembinder-Lemmon [2000] et nous lui ajoutons quelques modifications.

Les deux périodes considérées par ce modèle correspondent respectivement aux étapes *forward* et du temps réel. Sur chacune de ces étapes, séparées par la *gate closure*, les acteurs peuvent réaliser des transactions (figure 4-1). Dans notre modèle de base, on peut trouver trois types d'agents économiques : les producteurs « flexibles », les revendeurs (« *retailer* »), et le GRT.

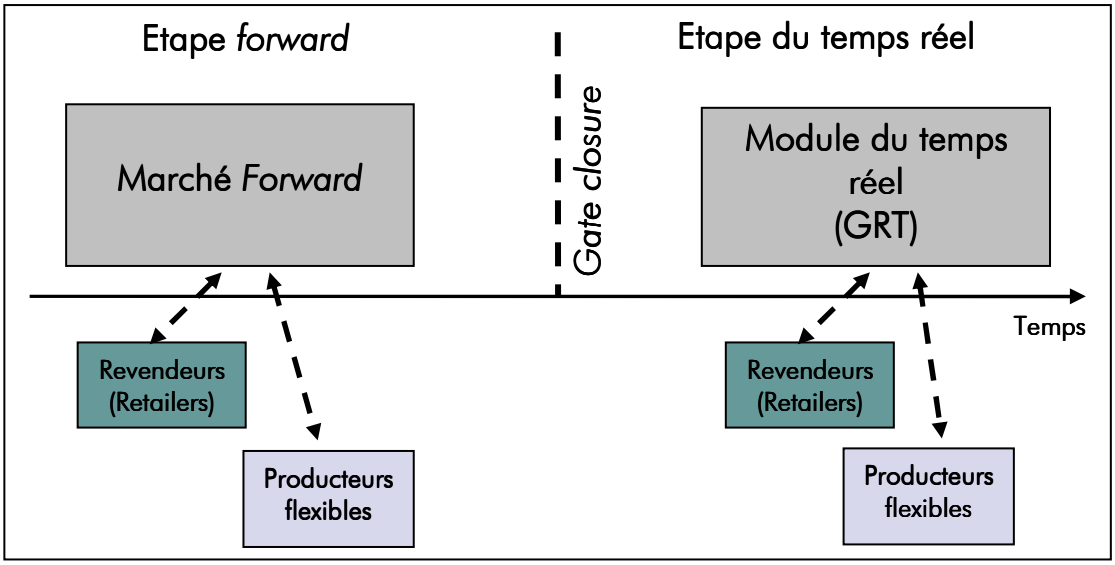

Figure 4-1 : Equilibre à deux étapes, modèle de base

Par hypothèse, nos marchés *forward* et en temps réel sont parfaitement concurrentiels. Nous analysons les décisions de production comme indépendantes dans le temps, avec deux séquences successives d'équilibre des deux marchés, en nous fondant sur l'hypothèse que l'absence de stockage de l'électricité rend les marchés indépendants les uns des autres au cours du temps (cf. Chapitre 3).[210]

La demande est stochastique et dépend de l'état du monde en temps réel décrit par la variable aléatoire ω. Cette variable est inconnue des agents quand ils prennent leurs décisions sur le marché *forward*, mais elle sera révélée au moment du temps réel. La variable aléatoire ω suit une fonction de distribution de probabilité discrète (voir plus bas). L'incertitude présente sur l'étape *forward* provient exclusivement de la nature stochastique de

[210] C'est dire que nous ne considérons pas les contraintes inter temporelles des moyens de production comme les caractéristiques particulières des fonctions de coût de production (non-convexités).

la demande.[211] Nous faisons ainsi l'hypothèse qu'il n'y a plus d'incertitude subsistante au moment du temps réel. En conséquence, les seules décisions prises sous incertitude sont celles du marché *forward*. Nous faisons aussi l'hypothèse que chaque agent connaît la fonction de distribution de la variable aléatoire ω, et que les producteurs flexibles ont une capacité de production suffisante pour faire face à toutes les demandes possibles (il n'existe pas de problèmes structurels de capacité de production, ni de problèmes de constitution de réserves).

Prendre des décisions sur le marché *forward* est donc risqué. Et les agents ayant de l'aversion pour le risque, ceci suscite une demande de contrats *forward* pour se couvrir du risque encouru jusqu'au moment du temps réel.[212] Pour modéliser ce comportement, nous faisons l'hypothèse que chaque agent maximise l'utilité espérée de sa fonction de gains, qui est de la forme :

$$E_{\omega}[U(\pi_{\omega})] \equiv E[\pi_{\omega}] - \frac{A}{2} Var[\pi_{\omega}] \ .$$ [213]

A est le coefficient d'aversion au risque et, $E[\pi_{\omega}]$ et $Var[\pi_{\omega}]$ sont respectivement la valeur espérée et la variance du profit. Cette forme de fonction est une simplification de la fonction d'utilité moyenne-variance (*mean-variance*) avec une aversion au risque constante et indépendante du profit moyen (*linear mean-variance utility function with constant absolute risk aversion*, Markowitz [1952]). Etant donné l'impossibilité de stockage de l'électricité, le profit des acteurs peut être très variable. C'est ce qui favorise l'aversion au risque des acteurs. L'aversion au risque des acteurs signifie ici que les acteurs sont prêts à avoir un profit moyen moins important à condition que ce profit soit moins variable.

Pour l'étape du temps réel, nous simplifions la variété empirique des différents arrangements pour le module du temps réel en ne distinguant que deux genres de designs : un design de « pur marché de temps réel », donc sans pénalités, et un design de « mécanisme d'ajustement » avec pénalités. Ces deux designs du module du temps réel utilisent des enchères à prix uniforme. En conséquence, le niveau de pénalités appliqué aux écarts constatés en temps réel est le seul paramètre qui transforme un marché de temps réel en mécanisme d'ajustement. La position temporelle de la *gate closure* est à son tour représentée par un paramètre qui est l'ampleur de la déviation possible de la demande finale (= l'ampleur de l'écart type de la demande) au moment du temps réel.

---

[211] Nous considérons, ici, le terme « incertitude » comme les incertitudes probabilisables donnant lieu aux problèmes de risques. Nous ne traitons pas ici les problèmes issus des incertitudes radicales (Knight [1921]).
[212] Dans un environnement sans aversion au risque et en concurrence parfaite, les agents n'auraient pas intérêt à participer sur le marché *forward* (Siddiqui [2002]).
[213] Ce type de fonction d'utilité a été étudié, par exemple, dans Rolfo [1980], Hirshleifer-Subramanyam [1993], Bessembinder-Lemmon [2000] et Siddiqui [2002].

Présentation des agents sur les marchés et du GRT

Producteurs flexibles

Il y a N_{FG} producteurs flexibles identiques. Chaque producteur flexible i peut opérer des transactions pour son électricité (énergie) sur le marché *forward* ($X^{F}_{FG_i}$) ou au moment du temps réel ($X^{RT}_{FG_i,\omega}$). Les producteurs flexibles peuvent prendre et changer leurs décisions de production jusqu'au temps réel (caractéristique de producteur « flexibles »). Leur fonction de coût de production est de type quadratique: $CT_{FG_i}(X_{FG_i}) = \sigma_{FG} X_{FG_i}{}^{2} / 2$ où σ_{FG} est le coefficient de coût marginal et X_{FG_i} est la quantité d'énergie produite en temps réel.[214] Naturellement, cette fonction de coût de production quadratique implique une fonction de coût marginal linéaire : $CMg_{FG_i}(X_{FG_i}) = \sigma_{FG}\, X_{FG_i}$.

Revendeurs

Les revendeurs achètent l'électricité sur le marché *forward* et au moment du temps réel, et la revendent à leurs clients à un prix contractuel fixe, déterminé à l'avance dans un contrat multi périodes : P^{C}.

Dans cette modélisation simple, nous considérons qu'il n'existe que deux revendeurs, A et B. Les revendeurs n'ont pas de contrôle sur le niveau réel de consommation de leurs clients dont la demande est stochastique et inélastique au moment du temps réel. Le volume exact de la demande d'électricité dont chaque revendeur j sera responsable au moment du temps réel, $D_{j,\phi(\omega)}$, est encore inconnu au moment de la prise de décision sur le marché *forward*. Chaque revendeur j sert une demande qui possède donc deux niveaux possibles ϕ : un niveau bas de demande ($D_{j,low}$) et un niveau haut de demande ($D_{j,high}$), avec les probabilités p_j et $(1-p_j)$ respectivement. Par simplicité nous supposons que la distribution de probabilité des demandes des deux revendeurs est définie par les mêmes probabilités p et $(1-p)$. C'est-à-dire que $p_j = p_A = p_B = p$. Nous faisons l'hypothèse que ces deux demandes sont indépendantes : elles ne sont pas corrélées. Les valeurs espérées des demandes des deux revendeurs sont alors : $\overline{D}_A$ et $\overline{D}_B$. La demande globale sera obtenue par combinaison de ces deux demandes individuelles des revendeurs. Elle aura donc au plus quatre états du monde possibles (ω) (figure 4-2). Sur la figure 4-3 on peut voir la fonction de distribution de la demande globale.[215]

---

[214] Remarquer que nous ne représentons pas les coûts « non-convexes » comme le coût de démarrage.

[215] Remarquer que cette fonction de distribution de probabilité est asymétrique (p>0,5). Dans nos exemples numériques, une distribution asymétrique est utilisée pour prendre en compte l'effet de la courbure ou convexité de l'offre d'électricité qui est ignorée dans notre modèle linéaire de coût marginal.

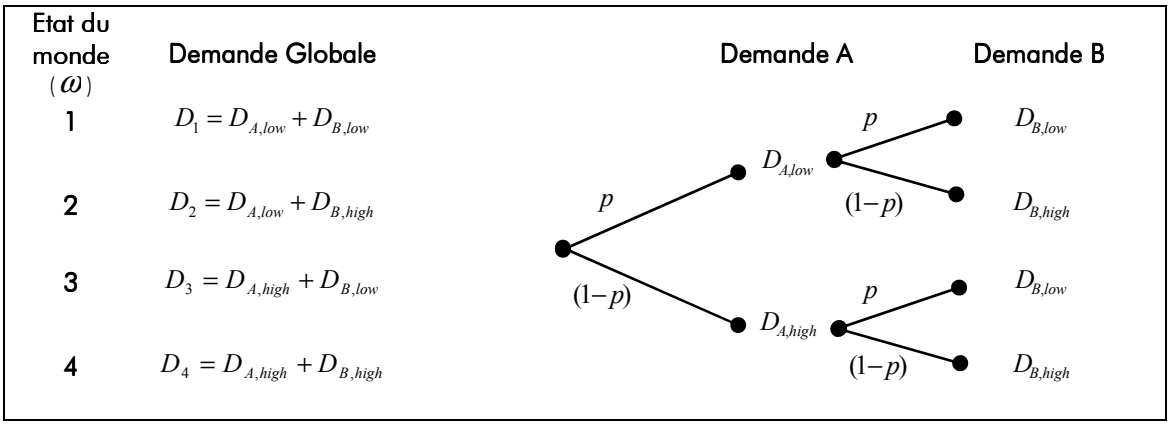

Figure 4-2 : Etats du monde de la demande globale

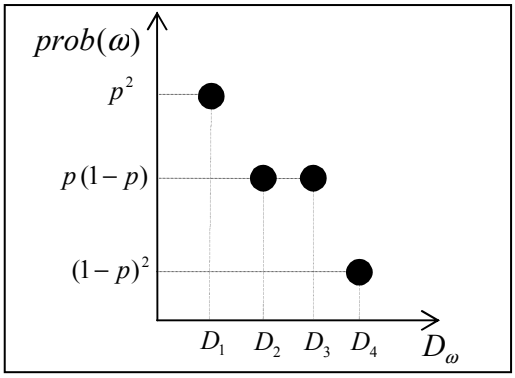

Figure 4-3 : Distribution de probabilités de la demande globale

Comme les achats *forward* des revendeurs ne correspondent jamais exactement à la consommation effective de leurs clients, les revendeurs auront des positions excédentaires ou déficitaires au moment du temps réel. Afin de respecter leurs obligations, ils achèteront ou vendront les quantités correspondantes via le module du temps réel, géré par le GRT. Pour des écarts positifs, le GRT paie les revendeurs au prix du temps réel (sur un marché du temps réel) ou à ce prix diminué par $1/k$ (sur un mécanisme d'ajustement). Pour les écarts négatifs, les revendeurs paient au GRT le prix des écarts, qui est égal au prix du temps réel (sur un marché du temps réel) ou à ce prix multiplié par k (sur un mécanisme d'ajustement). Quand $k = 1$, le prix des écarts est égal au prix de l'énergie en temps réel, et l'arrangement du module du temps réel est de type « marché du temps réel ». Quand $k > 1$, l'arrangement est de type « mécanisme d'ajustement ».

Le gestionnaire du réseau (GRT)

Le GRT est responsable de l'équilibrage du système électrique et, en conséquence, il gère l'équilibre au moment du temps réel. Pour ce faire, le GRT prend les offres proposées par les producteurs flexibles et équilibre le système au moindre coût en connaissant la demande du système. Bien que le GRT utilise les offres proposées d'une manière optimale, son bilan économique peut être positif, notamment dans le cas des designs avec pénalités. Dans ce cas, les recettes venant des offres à la baisse acceptées et pour les écarts négatifs

seront plus grandes que les dépenses provoquées par les offres à la hausse acceptées et les écarts positifs.

Cette section se poursuit en résolvant les problèmes d'optimisation des deux types de participants aux marchés (producteurs et revendeurs) et en déterminant l'équilibre des marchés. Comme nous avons deux marchés (*forward* et temps réel), il y a deux étapes au problème d'optimisation des agents. En principe, ces agents prennent d'abord position sur le marché *forward*, sur la base de leurs anticipations des conditions au moment du temps réel. Ensuite, au moment du temps réel, quand l'état de la demande est révélé, ces agents réalisent leurs transactions du temps réel en absence de toute incertitude. Dans notre démarche de modélisation, nous commençons par la prise de décision des agents au moment du temps réel, étant entendu qu'ils considèrent alors comme donnés leurs positions sur le marché *forward* et les prix de ce marché *forward* (sous-partie 1.2.1). Une fois que nous avons déterminé les positions optimales et les prix au moment du temps réel, pour chaque état du monde, nous pouvons revenir en arrière dans le temps pour déterminer les positions optimales (sous-partie 1.2.2) et les prix et quantités d'équilibre sur le marché *forward* (sous-partie 1.2.3).

1.2.1 Etape du temps réel

Au moment du temps réel, l'état du monde ω est réalisé. Il n'y a donc plus d'incertitude. De plus, dans cette étape du temps réel, les positions du marché *forward* ($X_{FG_i}^{F\ *}, X_{R_i}^{F^*}$), ont déjà été prises et le prix *forward* (P^{F^*}) est déjà déterminé.[216] Ils peuvent donc être traités comme des grandeurs fixes. En conséquence, nous pouvons calculer les positions du temps réel et le prix du temps réel, sachant que chaque agent cherche à maximiser son profit π_ω.

1.2.1.1 Producteurs flexibles

Le profit du producteur flexible i en fonction de $X_{FG_i,\omega}^{RT}$ peut être exprimé par:

$$\pi_{FG_i,\omega}(X_{FG_i,\omega}^{RT}) = P^{F^*} X_{FG_i}^{F\ *} + P_\omega^{RT} X_{FG_i,\omega}^{RT} - \frac{\sigma_{FG}}{2} X_{FG_i,\omega}^{\ 2}$$

Le premier terme représente la contribution au profit des ventes *forward*, le deuxième terme la contribution des ventes ou des achats (transactions actives : acceptation des offres à la hausse ou à la baisse) du temps réel pour l'état du monde ω, et le troisième terme représente le coût de production pour l'état du monde ω.

[216] Les variables désignées par * sont considérées fixes.

La quantité produite par le producteur flexible doit être égale aux quantités vendues sur le marché *forward* plus (moins) les quantités vendues (achetées) au moment du temps réel, $X_{FG_i,\omega} = X_{FG_i}^{F\ *} + X_{FG_i,\omega}^{RT}$.

Pour déterminer la stratégie optimale du producteur flexible, nous dérivons les conditions nécessaires du premier ordre. On cherche le maximum du profit $\pi_{FG_i,\omega}$ en fonction de la quantité vendue ou achetée en temps réel pour chaque état du monde $X_{FG_i,\omega}^{RT}$.[217] Ce maximum est atteint pour :

$$\forall \omega ; \forall i \quad \frac{\partial \pi_{FG_i,\omega}(X_{FG_i,\omega}^{RT})}{\partial X_{FG_i,\omega}^{RT}} = 0$$

Soit :[218]

$$\forall \omega ; \forall i \quad P_{\omega}^{RT} - \sigma_{FG}\,(X_{FG_i,\omega}^{RT} + X_{FG_i}^{F\ *}) = 0$$

Et finalement :

$$\forall \omega ; \forall i \quad X_{FG_i,\omega}^{RT} = \frac{P_{\omega}^{RT}}{\sigma_{FG}} - X_{FG_i}^{F\ *} \quad \text{.....} (1)$$

Notez que l'équation **(1)** peut être reformulée de la manière suivante pour représenter la courbe d'offre du producteur flexible sur le module du temps réel.[219]

$$P_{\omega}^{RT}(X_{FG_i,\omega}^{RT}) = \sigma_{FG}\,(X_{FG_i,\omega}^{RT} + X_{FG_i}^{F\ *}) = \sigma_{FG}\, X_{FG_i,\omega}$$

Logiquement, étant donné l'hypothèse de concurrence parfaite (comportement des participants comme preneurs de prix), la courbe d'offre correspond avec leurs coûts marginaux.

1.2.1.2 Revendeurs

Les revendeurs doivent acheter (ou vendre) au moment du temps réel la différence entre la demande effective réalisée $D_{j,\phi(\omega)}$ et leurs achats antérieurs sur le marché *forward* $X_{R,j}^{F}$. En conséquence, les quantités achetées (ou vendues) au moment du temps réel sont :

$$\forall \omega ; \forall j \quad X_{R_j,\omega}^{RT} = X_{R_j}^{F\ *} - D_{j,\phi(\omega)} \quad \text{.....} (2)$$

Ces ventes ou achats des revendeurs représentent des transactions « passives » (ou des « écarts ») (cf. Chapitre 3, section 2).

---

[217] Notez que la variable « prix du temps réel » est considérée par le producteur comme exogène. Etant donné l'hypothèse de concurrence parfaite, les producteurs ne peuvent pas influencer le prix par des actions individuelles.

[218] On peut vérifier que c'est un maximum car $\frac{\partial^2 \pi}{\partial X^2} < 0$.

[219] Cela suppose que les participants relisent leurs offres d'ajustement une fois qu'il y a plus d'incertitude.

Le prix des écarts dépend du signe des écarts et est défini par : [220]

$$IP_{j,\omega}^{RT} = \begin{cases} P_{\omega}^{RT}\, k & si\ X_{R_j,\omega}^{RT} \leq 0\ \ (écart\, négatif) \\ P_{\omega}^{RT}\, \dfrac{1}{k} & si\ X_{R_j,\omega}^{RT} > 0\ \ (écart\ positif) \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

où k est le coefficient de pénalité ($k \geq 1$). Remarquer que si $k = 1$ le prix des écarts est égal au prix de l'énergie. La figure 4-4 donne un exemple de calcul de prix des écarts pour le cas du système « court » (écart net négatif) et pour un coefficient de pénalité supérieur à 1. Sur cette figure, nous remarquons que le prix des écarts négatifs (IP_{neg}) est supérieur au prix de l'énergie correspondant au prix de l'offre marginale acceptée (P^{RT}) et que le prix des écarts positifs (IP_{pos}) est inférieur au prix de l'énergie.

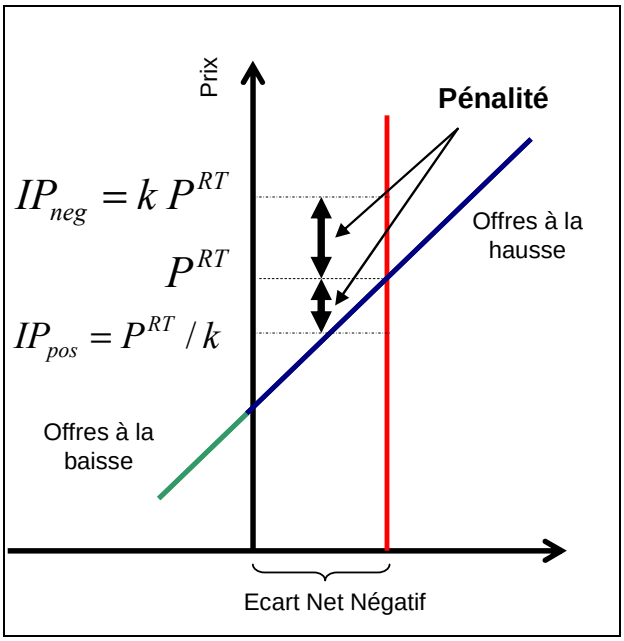

Figure 4-4 : Exemple de prix d'énergie et des écarts au moment du temps réel (cas de système « court »)

1.2.1.3 Gestionnaire du réseau (GRT)

Le GRT est responsable de la gestion de l'équilibre et fait respecter la contrainte d'équilibrage pour chaque état du monde. Pour ce faire, le GRT opère les transactions actives nécessaires (acceptation d'offres) afin de couvrir le déséquilibre net du système venant des écarts.

[220] Par simplicité, nous considérons qu'il n'y a pas de différenciation entre les écarts qui sont dans la même direction que le système et ceux qui ont une direction contraire. C'est-à-dire qu'il n'y a que deux prix d'écarts (prix des écarts positifs et prix des écarts négatifs) et ces prix ne dépend pas de la direction d'équilibrage du système (cf. Chapitre 3).

Les conditions d'équilibre du module du temps réel (que se soit un « marché » ou un « mécanisme ») sont donc définies par :[221]

$$\forall \omega \quad \sum_i X^{RT}_{FG_i,\omega} = -\sum_j X^{RT}_{R_j,\omega} \quad \text{.....} (4)$$

L'équation **(4)** représente l'équilibre des transactions du temps réel. $X^{RT}_{FG_i,\omega}$ est le volume de transactions « actives » du temps réel (offres acceptées) pour le générateur flexible i et pour l'état du monde ω et $X^{RT}_{R_j,\omega}$ et le volume de l'écart du revendeur j et pour l'état du monde ω.

Comme nous le verrons après (cf. équation **(6)**), le GRT accepte les offres proposées à la hausse ou à la baisse pour équilibrer le système à moindre coût. Or, son bilan économique peut être positif pour les designs avec pénalités. Dans ce cas, les recettes venant des offres à la baisse acceptées et celles des écarts négatifs seront supérieures aux dépenses provoquées par les offres acceptées à la hausse et les écarts positifs (Cf. Chapitre 3).

1.2.2 Etape *forward*

En revenant maintenant au moment où les positions ont été prises sur le marché *forward*, nous pouvons déterminer les conditions d'équilibre de ce marché et les quantités optimales vendues ou achetées par chaque agent participant à ce marché.

Les conditions d'équilibre de ce marché *forward* sont exprimées par l'équation suivante :

$$\sum_j X^F_{R_j} = \sum_i X^F_{FG_i} \quad \text{.....} (5)$$

L'équation **(5)** exprime que la quantité totale vendue par les producteurs sur le marché *forward* (somme des quantités vendues par chaque producteur i, $X^F_{FG_i}$) est égale à la quantité totale achetée par les revendeurs (somme des quantités achetées par chaque revendeur j, $X^F_{R_j}$).

Des équations **(1)**, **(2)**, **(4)** et **(5)** nous pouvons déduire que :

$$\forall \omega \quad P^{RT}_{\omega} = D_{\omega} \frac{\sigma_{FG}}{N_{FG}} \quad \text{.....} (6)$$

où $D_{\omega} = \sum_j D_{j,\phi(\omega)}$ est la demande globale pour l'état du monde ω.

[221] Le signe négatif du deuxième terme de cette équation vient de la concordance entre les signes des écarts et le langage utilisé sur les mécanismes d'ajustement pour définir les écarts positifs et les écarts négatifs du point de vue du GRT.

L'équation (6) représente le prix de l'énergie en temps réel P_{ω}^{RT} pour chaque état du monde ω. Ce prix de l'énergie du temps réel correspondrait à l'offre marginale acceptée par le GRT pour équilibrer le système à moindre coût (voir figure 3). P_{ω}^{RT} dépend logiquement de la demande globale pour l'état du monde ω et de la courbe d'offre agrégée sur le module du temps réel.

Nous pouvons développer maintenant les décisions optimales des participants sur le marché *forward*. Pour cela, nous utiliserons les fonctions de profit des participants afin de construire les fonctions d'utilité espérée et de déterminer ainsi les couvertures optimales de chaque participant.

1.2.2.1 Producteurs flexibles

Sur le marché *forward*, nous pouvons exprimer le profit des générateurs flexibles comme suit:

$$\pi_{FG_i,\omega}(X_{FG_i}^F) = P^F X_{FG_i}^F + P_{\omega}^{RT} X_{FG_i,\omega}^{RT} - \frac{\sigma_{FG}}{2}\left(X_{FG_i}^F + X_{FG_i,\omega}^{RT}\right)^2$$

Le premier terme représente la contribution au profit des ventes *forward*, le deuxième terme la contribution des ventes ou des achats (acceptation des offres à la hausse ou à la baisse) du temps réel pour l'état du monde ω, et le troisième terme représente le coût de production pour l'état du monde ω.

L'objectif d'un producteur flexible i est maintenant de choisir $X_{FG_i}^F$, la quantité qu'il est prêt à vendre sur le marché *forward* pour un prix *forward* donné P^F, de manière à maximiser son utilité espérée :

$$E_{\omega}[U(\pi_{FG_i,\omega})] \equiv E[\pi_{FG_i,\omega}] - \frac{A_{FG}}{2} Var[\pi_{FG_i,\omega}] \ .$$

Où $E[\pi_{FG,\omega}]$ et $Var[\pi_{FG_i,\omega}(X_{FG_i}^F)]$ sont respectivement la valeur espérée et la variance du profit du producteur flexible :

$$E[\pi_{FG,\omega}] = \sum_{\omega} prob(\omega)\ \pi_{FG_i,\omega} \text{ et } Var[\pi_{FG_i,\omega}] = \sum_{\omega} prob(\omega)\ \left(\pi_{FG_i,\omega} - E[\pi_{FG_i,\omega}]\right)^2 .$$

A_{FG} est le coefficient d'aversion au risque du producteur flexible.

Il faut remarquer que le producteur flexible connaît à ce moment ses décisions optimales sur le module du temps réel pour chaque état du monde et qu'il les intégrera au moment de faire son choix sur le marché *forward* (cf. sous-section 1.2.1).

Alors les conditions nécessaires de premier ordre pour assurer la maximisation de l'espérance de la fonction d'utilité sont[222] :

$$\forall i \quad \frac{\partial E\left[U(\pi_{FG_i,\omega}(X^F_{FG_i}))\right]}{\partial X^F_{FG_i}} = \frac{\partial\left(E[\pi_{FG_i,\omega}(X^F_{FG_i})] - \frac{A_{FG}}{2} Var[\pi_{FG_i,\omega}(X^F_{FG_i})]\right)}{\partial X^F_{FG_i}} = 0$$

De cette équation et de **(1)** et **(6)** nous pouvons déduire :

$$X^F_{FG_i} = \frac{P^F - E[P^{RT}_\omega]}{A_{FG}\ Var[P^{RT}_\omega]} + \frac{Cov[\rho_{FG_i,\omega}, P^{RT}_\omega]}{Var[P^{RT}_\omega]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Où,

$E[P^{RT}_\omega]$ et $Var[P^{RT}_\omega]$ sont respectivement la valeur espérée et la variance du prix du temps réel,

$\rho_{FG_i,\omega}$ est le profit « sans couverture » (*unhedged*) du producteur i :

c.à.d. avec $X^F_{FG_i} = 0$, on a $\rho_{FG_i,\omega} \equiv \pi_{FG_i,\omega}(X^F_{FG_i} = 0) = \frac{1}{2}\frac{\sigma_{FG}}{N_{FG}{}^2} D_\omega{}^2$

et $Cov[\rho_{FG_i,\omega}, P^{RT}_\omega]$ est la covariance entre le profit « sans couverture » et le prix du temps réel.

L'équation **(7)** représente la quantité optimale que le producteur flexible i est prêt à vendre sur le marché *forward* pour un prix *forward* P^F donné. Dit autrement, cette équation représente l'offre du producteur flexible i sur le marché *forward*. En effet, en reformulant l'équation **(7)** nous obtenons :

$$P^F(X^F_{FG_i}) = \underbrace{E[P^{RT}_\omega] - A_{FG}\ Cov[\rho_{FG_i,\omega}, P^{RT}_\omega]}_{cf.\ figure\ 4-5} + A_{FG}\ Var[P^{RT}_\omega] \quad X^F_{FG_i}$$

Nous pouvons interpréter cette équation de la manière suivante (voir figure 4-5). Pour un prix *forward* donné, un producteur serait prêt à vendre une quantité plus grande sur le marché *forward* (déplacement à droite de son offre) pour deux motifs : le premier à cause d'une action d'arbitrage. Par exemple, si la valeur espérée du prix du temps réel diminue, le producteur préférerait vendre plus sur le marché *forward* et moins en temps réel pour arbitrer entre les deux marchés. Le deuxième motif représente une action de « couverture ». Par exemple, un producteur serait prêt à vendre une quantité plus grande sur le marché *forward* si son profit « sans couverture » est plus corrélé positivement avec le prix du temps réel. La covariance mesure statistiquement comment le profit « sans couverture » du producteur augmente (ou diminue) quand le prix du temps réel augmente (ou diminue), et elle donne donc une indication de l'intérêt de vendre plus ou moins sur le marché *forward* par rapport aux risques encourus sur le marché du temps réel. En effet, plus cette covariance

[222] Hirshleifer-Subramayam [1993] montrent que les fonctions d'utilité linéaires sur le profit espéré et sa variance sont concaves selon la quantité *forward* et trouvent donc son maximum pour les conditions de premier ordre.

est grande plus le producteur voudra vendre de l'énergie *forward* et éviter de prendre des risques en temps réel.

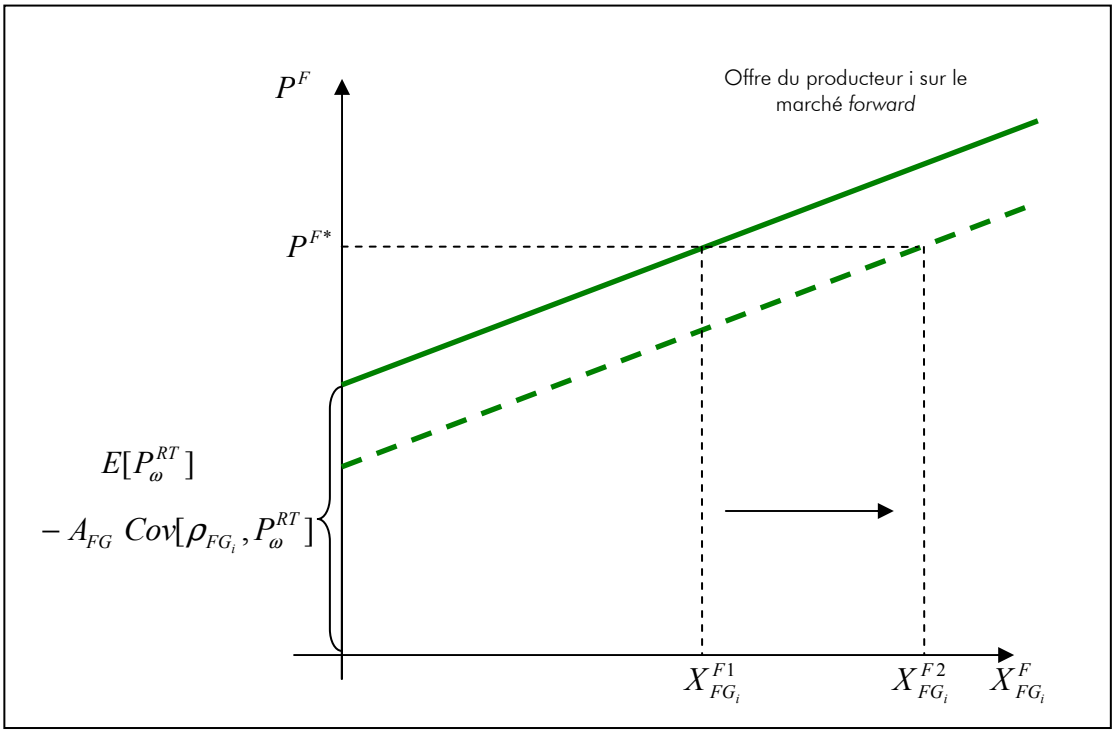

Figure 4-5 : Offre du producteur flexible i sur le marché *forward*

1.2.2.2 Revendeurs

Nous pouvons de manière similaire exprimer le profit d'un revendeur j :

$$\pi_{R_j,\omega}(X_{R_j}^F) = P^C D_{j,\phi(\omega)} - P^F X_{R_j}^F + IP_{j,\omega}^{RT} X_{R_j,\omega}^{RT}$$

Le premier terme de cette expression représente la contribution au profit des « reventes » à ses clients. Les clients consomment une quantité $D_{j,\phi(\omega)}$ pour chaque état du monde et paient au revendeur un prix fixe P^C. Le deuxième terme représente le « coût » du revendeur pour des achats ($X_{R_j}^F$) sur le marché *forward*. Le troisième terme représente le coût (écart négatif) ou le gain (écart positif) des transactions « passives » du temps réel du revendeur pour l'état du monde ω.

L'objectif d'un revendeur j est de choisir $X_{R_j}^F$, la quantité qu'il est prêt à acheter sur le marché *forward* pour un prix *forward* donné P^F, de manière à maximiser l'espérance de son utilité : $E_\omega[U(\pi_{R_j,\omega})] \equiv E[\pi_{R_j,\omega}] - \frac{A_R}{2} Var[\pi_{R_j,\omega}]$. $E[\pi_{R_j,\omega}]$ et $Var[\pi_{R_j,\omega}]$ sont respectivement la valeur espérée et la variance du profit pour le revendeur j. Et A_R est le coefficient d'aversion au risque des revendeurs.

Les conditions nécessaires de premier ordre pour assurer la maximisation de l'espérance de la fonction d'utilité sont [223]:

$$\forall j \quad \frac{\partial E[U(\pi_{R_j,\omega}(X^F_{R_j}))]}{\partial X^F_{R_j}} = 0$$

En utilisant ces équations et **(2)** nous pouvons déduire que:

$$\forall j \quad X^F_{R_j} = \frac{E\left[IP^{RT}_{j,\omega}\right] - P^F}{A_R\ Var[IP^{RT}_{j,\omega}]} - \frac{Cov[\rho_{R_j,\omega}, IP^{RT}_{j,\omega}]}{Var[IP^{RT}_{j,\omega}]} \quad \text{.....(8)}$$

Où,

$E\left[IP^{RT}_{j,\omega}\right]$ et $Var[IP^{RT}_{j,\omega}]$ sont respectivement le prix anticipé des écarts et la variance de ce prix des écarts,

$\rho_{R_j,\omega}$ est le profit « sans couverture » (*unhedged*) du revendeur j :

c.à.d. $\rho_{R_j,\omega} \equiv \pi_{R_j,\omega}(X^F_{R_j} = 0) = P^C\ D_{j,\omega} - IP^{RT}_{j,\omega}\ D_{j,\omega}$

et $Cov[\rho_{R_j,\omega}, IP_{j,\omega}]$ est la covariance entre le profit « sans couverture » du revendeur j et le prix des écarts.

Pour le cas particulier sans pénalités ($k = 1$) nous avons:

$$\forall j \quad X^F_{R_j} = \frac{E\left[P^{RT}_{\omega}\right] - P^F}{A_R\ Var[P^{RT}_{\omega}]} - \frac{Cov[\rho_{R_j,\omega}, P^{RT}_{\omega}]}{Var[P^{RT}_{\omega}]} \quad \text{.....(9)}$$

L'équation **(9)** représente la quantité optimale que le revendeur j est prêt à acheter sur le marché *forward* pour un prix *forward* P^F donné.[224] Dit autrement, cette équation représente la demande du *revendeur* j sur le marché *forward*. En effet, en reformulant l'équation **(9)** nous obtenons :

$$P^F(X^F_{R_{ji}}) = \underbrace{E[P^{RT}_{\omega}] - A_R\ Cov[\rho_{R_j,\omega}, P^{RT}_{\omega}]}_{cf.\ figure\ 4-6} - A_R\ Var[P^{RT}_{\omega}]\ X^F_{R_{ji}}$$

Nous pouvons interpréter cette équation de la manière suivante (figure 4-6). Pour un prix *forward* donné, un revendeur serait prêt à acheter une quantité plus grande sur le marché *forward* (déplacement à droite de sa demande) pour deux motifs : le premier à cause d'une action d'arbitrage. Par exemple, si la valeur espérée du prix du temps réel augmente, le revendeur préférerait acheter plus sur le marché *forward* et moins en temps réel pour arbitrer entre les deux marchés. Le deuxième motif représente une action de « couverture ». Un revendeur serait prêt à acheter une quantité d'autant plus grande sur le marché *forward* si son profit « sans couverture » est d'autant plus corrélé négativement

[223] Cf. note de bas de page n° 222.
[224] La même analyse peut être faite pour l'équation (8).

avec le prix du temps réel. La covariance, en quelque sorte, mesure statistiquement comment le profit « sans couverture » du revendeur diminue (ou augmente) quand le prix du temps réel augmente (ou diminue), et elle donne donc une indication d'acheter plus ou moins sur le marché *forward* par rapport aux risques encourus sur le temps réel. En effet, plus la covariance sera grande en valeur absolue (covariance négative), le revendeur voudra acheter plus sur le marché *forward* et éviter de prendre des risques en temps réel.

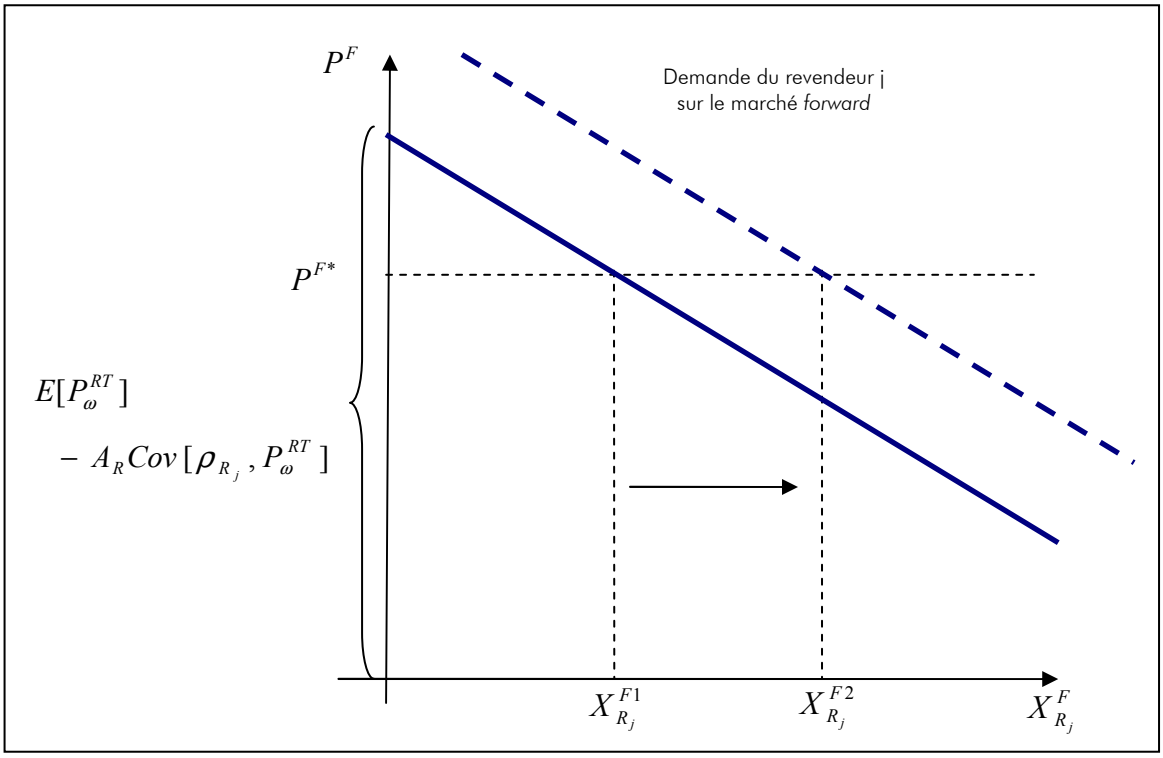

Figure 4-6 : Demande du revendeur j sur le marché *forward*

1.2.3 Equilibre des marchés

Nous pouvons maintenant utiliser les équations (3), (5) et (6) avec les positions optimales des participants sur le marché *forward* (équations (7) et (8)) pour déterminer le prix d'équilibre du marché *forward* P^{F*}. Autrement dit, nous nous servons du système d'équations (3)-(8) pour isoler le prix *forward* d'équilibre P^{F*}.

1.2.3.1 Equilibre des marchés sans pénalités (k=1)

En résolvant le cas sans pénalités ($k = 1$) nous obtenons:

$$P^F = E[P_\omega^{RT}] + A\left(\frac{N_{FG}}{\sigma_{FG}}\left(E[P_\omega^{RT}] - P^C\right)Var[P_\omega^{RT}] + \frac{N_{FG}}{2\sigma_{FG}}Skew[P_\omega^{RT}]\,Var[P_\omega^{RT}]^{\frac{3}{2}}\right) \quad .....(10)$$

Où $E[P_\omega^{RT}]$, $Var[P_\omega^{RT}]$ et $Skew[P_\omega^{RT}]$ sont respectivement la valeur espérée, la

variance et le degré d'asymétrie (*skewness*) [225] du prix du temps réel. $A = \left(N_{FG} / A_{FG} + N_R / A_R \right)^{-1}$ représente le coefficient d'aversion au risque global et dépend du nombre de participants et de leurs coefficients d'aversion au risque. N_{FG} et N_R sont respectivement le nombre de producteurs flexibles et de revendeurs (dans notre cas $N_R = 2$).

Ce résultat est équivalent au résultat de Bessembinder-Lemmon [2000]. Sur un « marché du temps réel » où il n'y a pas de pénalités au moment du temps réel, le prix *forward* de l'électricité dépend des anticipations du prix du temps réel, des conditions statistiques de la demande globale, et des paramètres des coûts de production (variance et asymétrie -*skewness*- des prix du temps réel). Nous remarquons ici que la relation entre le prix *forward* et le prix du temps réel pour l'électricité est basée en partie sur l'espérance mathématique du prix du temps réel (cf. Chapitre 3, section 1). Cependant, elle dépend aussi d'autres facteurs, notamment ceux issus de l'aversion au risque. Si le coefficient d'aversion était zéro, le prix *forward* équivaudrait à l'espérance mathématique du prix du temps réel.

Bessembinder-Lemmon [2000] tirent principalement des conclusions de l'équation (10) en la reformulant pour mesurer le « *premium* » (la différence entre le prix *forward* et l'espérance mathématique du prix du temps réel).

Premium

$$P^F - E[P_\omega^{RT}] = A \left(\underbrace{\frac{N_{FG}}{\sigma_{FG}} \left(E[P_\omega^{RT}] - P^C \right) Var[P_\omega^{RT}]}_{terme1} + \underbrace{\frac{N_{FG}}{2\sigma_{FG}} Skew[P_\omega^{RT}]\, Var[P_\omega^{RT}]^{\frac{3}{2}}}_{terme2} \right)$$

La « taille » du *premium* dépend logiquement du coefficient d'aversion au risque global (A). Plus il y a de l'aversion au risque, plus le *premium* est grand. Ce *premium* peut être positif (le prix *forward* est supérieur à l'espérance du prix du temps réel) ou négatif (le prix *forward* est inférieur à l'espérance du prix du temps réel). Dans cette équation, on note que le signe du *premium* dépend principalement de deux facteurs : la variance et le coefficient d'asymétrie (ou *skewness*) des prix du temps réel. Le premier terme dépend exclusivement de la variance du prix du temps réel et son signe est donné par $E[P_\omega^{RT}] - P^C$ qui est généralement négatif.[226] Le deuxième terme dépend de la variance du prix du temps

[225] L'asymétrie (ou *skewness*) du prix du temps réel dépend de deux facteurs : la courbure ou convexité de la fonction d'offre en temps réel (dans notre cas la fonction de coût marginal de l'industrie), et la symétrie de la fonction de distribution de la demande (Bessembinder-Lemmon [2000]). La fonction de coût marginal agrégée dans l'électricité est réputée être convexe, à cause de la rapide augmentation du coût marginal venant de la variété de technologies des moyens de production connectés sur le réseau (voir chapitre 1).

[226] Etant donné que généralement pour inciter à un revendeur averse au risque à rentrer dans l'industrie, le prix fixe payé par ses clients doit être plus grand que la valeur espérée du prix du temps réel.

réel, mais aussi du coefficient d'asymétrie *skewness* du prix du temps réel (généralement positif dans le cas de l'électricité). Ainsi ces deux termes sont de signes opposés (terme 1 négatif, et terme 2 positif). Alors, pour des valeurs de variance de P_{ω}^{RT} faibles, le terme négatif (terme 1) sera plus important que le terme positif (terme 2) et le premium sera négatif. Pour des valeurs de la variance du P_{ω}^{RT} grandes, le terme positif (terme 2) sera plus important que le terme négatif (étant donné la différence d'exposants) et le premium sera positif (voir figure 4-7).

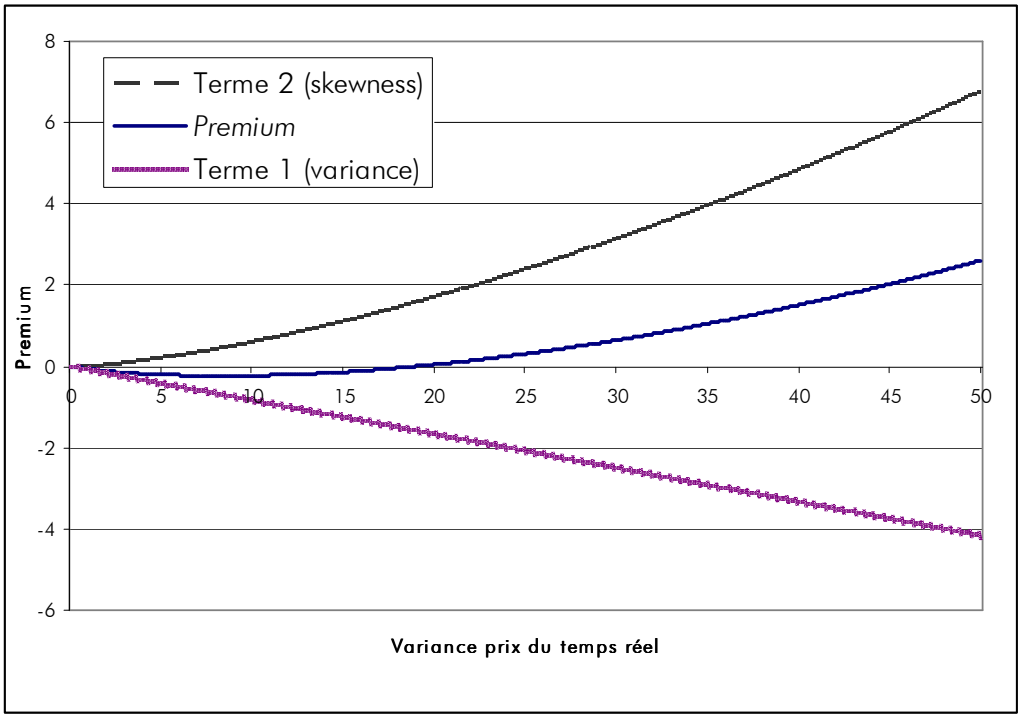

Figure 4-7 : *Premium forward* en fonction de la variance du prix du temps réel (*Skewness* >1)

Conceptuellement, le fait que le *premium* soit plus grand et de signe positif quand le *skewness* du prix du temps réel est positif peut être expliqué en regardant les équations de profit « sans couverture » pour les deux acteurs de ce marché. Une fonction de distribution de prix asymétrique (positivement) correspond à de faibles probabilités pour les valeurs hautes des prix. La réalisation possible de ces cas extrêmes provoque sur les acteurs des effets complètement différents. Pour les producteurs, la réalisation d'un de ces cas extrêmes signifie un profit très grand. Les événements extrêmes engendrent pour lui des coûts d'opportunité très grands s'il vend sur le marché *forward*. Il aura donc une préférence pour vendre relativement moins (ou la même quantité à un prix plus important) sur le marché *forward*. En revanche, la situation du revendeur est complètement différente. La réalisation d'un de ces cas extrêmes provoque pour lui des pertes (ou diminution du profit) très grandes. Il voudra donc acheter plus sur le marché *forward* pour se couvrir de ces possibles cas extrêmes. La tension provoquée sur le marché *forward* modifiera le prix *forward* jusqu'à que les quantités achetées et vendues soient équilibrées (figure 4-8).

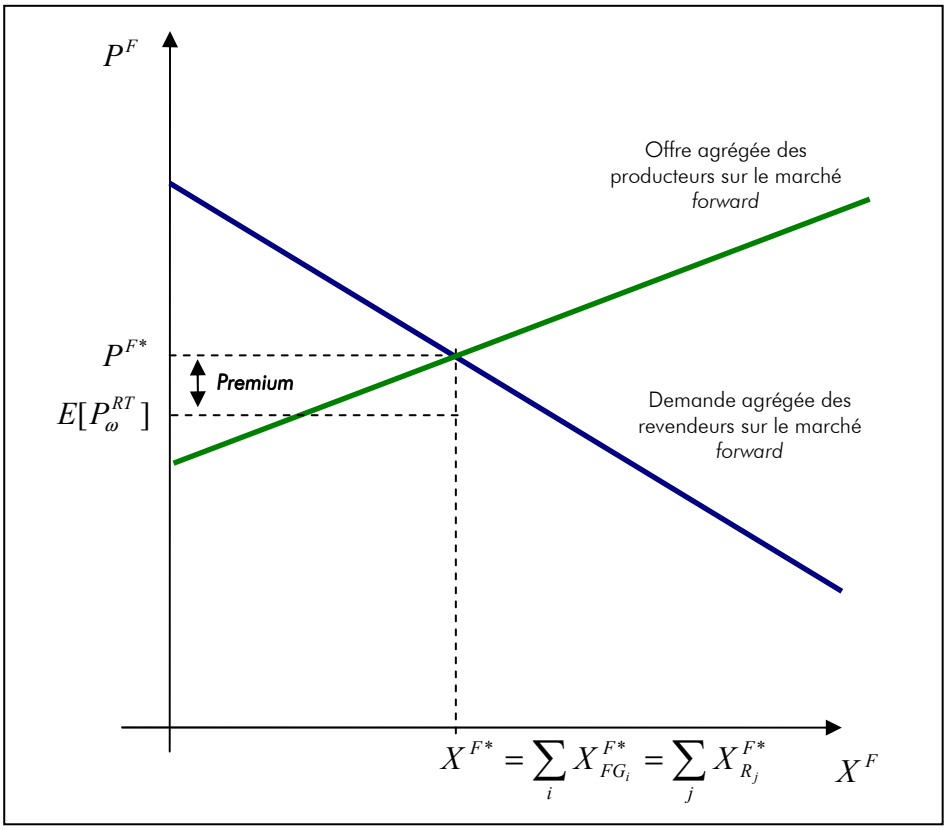

Figure 4-8 : Equilibre du marché *forward* (cas d'un *premium* positif)

1.2.3.2 Equilibre des marchés avec pénalités (k>1)

Pour résoudre les cas où il existe une pénalité ($k > 1$), la détermination du prix *forward* est plus complexe et il est difficile de dégager une équation analytique simple du type de **(10)**. Etant donné la complexité de l'équation, et son intraitabilité de manière analytique (le modèle a beaucoup de paramètres), nous étudierons les conditions d'équilibre par le biais de simulations numériques. Les simulations numériques se présentent comme une alternative aux études analytiques quand l'impossibilité d'obtention, ou l'intraitabilité, des résultats analytiques est considéré (Walliser [2005]). Les simulations numériques peuvent servir de guide pour une démonstration ultérieure en permettant d'en explorer les caractéristiques fondamentales et qualitatives des résultats. Notre objectif est réaliser une analyse comparative des effets des différents types de design, et les simulations numériques nous permettront de comprendre et d'individualiser les effets du design des autres effets qui dépendent des paramètres spécifiques du modèle.

De plus, pour résoudre les cas où il existe une pénalité ($k > 1$), on doit faire une hypothèse concernant le signe des écarts. En effet, pour calculer P^{F^*} et les quantités d'équilibre ($X_{R_j}^{F^*}$ et $X_{FG_i}^{F^*}$), il faut déterminer pour chaque état du monde si un revendeur a un écart négatif (sa demande est supérieure à la quantité achetée sur le marché *forward* $D_j > X_{R_j}^F$) ou positif (sa demande est inférieure à la quantité achetée sur le marché *forward*

$D_j < X_{R_j}^F$). Ne connaissant pas $X_{R_j}^{F\,*}$ au moment de déterminer les signes des écarts, nous supposons d'abord que $D_{j,low} \leq X_{R_j}^{F\,*} \leq D_{j,high}$. Ainsi les prix des écarts peuvent être définis pour chaque état du monde. Une fois le calcul réalisé dans les simulations numériques, cette hypothèse sera vérifiée.

Dans la prochaine sous-section, nous en présentons des simulations numériques réalisées sous *Mathematica®*.

1.3 Modèle de base : simulations numériques et discussion

Nous recourrons ici à des simulations numériques pour étudier les conséquences économiques de l'utilisation de pénalités au moment du temps réel.

Nous étudions trois cas différents :

- ***Benchmark*** : marché de temps réel (pas de pénalité ; $k = 1$),
- **Mécanisme N°1** : mécanisme d'ajustement avec pénalité faible ($k = 1{,}2$),
- **Mécanisme N°2** : mécanisme d'ajustement avec pénalité forte ($k = 1{,}4$).

Chacun de ces trois cas est étudié pour deux types de *gate closure* : éloignée du moment du temps réel, ou bien proche du temps réel. Nous représentons ces différentes positions temporelles par leurs conséquences sur l'ampleur de l'incertitude affectant la prise de décision et nous les exprimons en modifiant l'ampleur de la déviation standard de la demande ($Std_{Sys1} = 10$) ou ($Std_{Sys2} = 20$).

Cette sous-section se poursuit en décrivant les paramètres des simulations (1.3.1), en donnant les premiers résultats des simulations numériques (1.3.2), et enfin, en analysant ces résultats (1.3.3).

1.3.1 Les paramètres des simulations

Les paramètres des simulations doivent d'abord être déterminés. Pour ce faire nous les avons repris des simulations réalisées par Bessembinder-Lemmon [2000]. Etant donné que nous ne réaliserons qu'une analyse comparative des effets des différents types de design (cas n°1 et n°2 comparés au benchmark), les paramètres du modèle joueront un rôle secondaire, car ils resteront constants pour les différents cas de comparaison. Cependant, une étude de sensibilité de ces paramètres est présentée en annexe. Tous les paramètres sont regroupés sur le Tableau 4-1.

Tableau 4-1 : Paramètres des simulations numériques

Description	Symbole	Valeur
Nombre de producteurs flexibles	N_{FG}	10
Probabilité de l'état bas de la demande	p	0,8
Coefficient d'aversion au risque des producteurs flexibles	A_{FG}	0,1
Coefficient d'aversion au risque des revendeurs	A_R	0,02
Prix fixe payé par les consommateurs	P^C	35
Coefficient de coût des producteurs flexibles	σ_{FG}	3
Esperance de la demande pour le revendeur du type A (Moins exposé)	$\overline{D}_A$	200/3
Esperance de la demande pour le revendeur du type A (Plus exposé)	$\overline{D}_B$	100/3
Déviation standard de la demande du revendeur du type A	Std_A	$Std_{Sys}/\sqrt{2}$
Déviation standard de la demande du revendeur du type B	Std_B	$Std_{Sys}/\sqrt{2}$

Pour nos simulations, nous considérons qu'il y a 10 producteurs flexibles identiques (N_{FG}), chacun avec un coefficient de coût σ_{FG} égal à 3. Il y a 2 revendeurs (revendeur « type A » et revendeur « type B »). Les coefficients d'aversion au risque des participants, A_{FG} et A_R, sont choisis de manière à avoir un coefficient d'aversion au risque global (A) de 0,005 (Bessembinder-Lemmon [2000]).[227]

La valeur espérée de la demande globale du système est la somme des valeurs espérées des demandes individuelles des revendeurs, $\overline{D}_{Sys} = \overline{D}_A + \overline{D}_B$ et la déviation standard de la demande globale est $Std_{Sys} = \sqrt{Std_A + Std_B}$. Une valeur de $\overline{D}_{Sys} = 100$ sera utilisée et la valeur espérée du prix du temps réel sera donc égale à 30.[228] Le prix fixe payé par le consommateur P^C est fixé à 35.

Les caractéristiques des paramètres de demande ont été choisies pour représenter différents types d'agents participants aux marchés ($\overline{D}_A > \overline{D}_B$ et $Std_A = Std_B = Std_{Sys}/\sqrt{2}$) Notamment, des revendeurs de type « A » pourraient représenter des acheteurs nets de grande taille (bénéficiant ainsi d'effets de foisonnement pour leurs demandes) ou des revendeurs verticalement intégrés avec des producteurs. Les autres revendeurs, de type « B », représentent des acheteurs nets de petite taille et non intégrés verticalement. C'est pourquoi le rapport entre la déviation standard et la valeur espérée de la demande est plus

---

227 $A = \left(N_{FG}/A_{FG} + N_R/A_R \right)^{-1}$

228 De l'équation **(6)** nous pouvons déduire que $E[P_\omega^{RT}] = \frac{\sigma_{FG}}{N_{FG}} E[D_\omega] = \frac{\sigma_{FG}}{N_{FG}} \overline{D}_{Sys}$

grande pour les revendeurs de type B que pour les revendeurs de type « A ». Ainsi nous avons : $Std_B / \overline{D}_B > Std_A / \overline{D}_A$.

Les producteurs représentent des gros vendeurs nets. En effet, ils peuvent, dans notre modèle, facilement éviter les inconvénients d'une panne dans une seule de leurs centrales, et ils ne font face à aucun risque ou incertitude interne à leur entreprise sur leurs propres décisions de production.

La distribution de la demande a été choisie asymétrique (coefficient de *skewness* positif : p>0,5) pour prendre en compte la courbure (ou convexité) importante de la courbe d'offre des producteurs qui n'est pas représentée dans nos modèles (coût marginal linéaire résultat des fonctions de coût quadratiques).

Pour rendre compte des différentes positions possibles de la *gate closure*, les états possibles des demandes individuelles ont été exprimés comme des fonctions de la valeur espérée et d'une déviation standard. Ainsi pour le revendeur j , la valeur espérée et la déviation standard des demande seront $\overline{D}_j = E[D_{j,\phi(\omega)}]$ et $Std_j = \sqrt{Var[D_{j,\phi(\omega)}]}$. Les différents états possibles de ces demandes seront alors exprimés par :

-Demande basse : $D_{j,low} = \overline{D}_j - \sqrt{(1-p)p}\, Std_j / p$

-Demande haute : $D_{j,high} = \overline{D}_j + \sqrt{(1-p)p}\, Std_j / (1-p)$.[229]

1.3.2 Premiers résultats

Les premiers résultats correspondant à notre modèle de base sont présentés dans les tableaux 4-2 et 4-3. Le tableau 4-2 montre les résultats pour le cas de *gate closure* « proche » du temps réel, et le tableau 4-3 pour le cas de *gate closure* « loin » du temps réel.

Dans chaque tableau, nous présentons les résultats pour le *Benchmark* (k=1) puis pour les cas de pénalité faible et forte (k=1,2 et k=1,4). Le prix *forward* d'équilibre et la valeur espérée du prix du temps réel y sont représentés, ainsi que les quantités *forward* d'équilibre, vendues et achetées par les différents participants. Les profits et utilités espérées sont aussi présentés. Enfin, nous indiquons l'utilité totale (somme des utilités de tous les participants) et le coût total espéré. L'efficacité proprement dite et l'efficacité productive sont calculées sur la base de l'utilité (ou le coût total) du *benchmark*.[230]

---

[229] Si la distribution de probabilité de la demande était symétrique (p=0,5), ces deux états serait représentés simplement par $D_{j,low} = \overline{D}_j - Std_j$ et $D_{j,high} = \overline{D}_j + Std_j$.

[230] Efficacité = 100 (Ut. tot. Cas x) / (Ut. Tot. Benchmark) ; Efficacité productive = 100 (1 – Coût total Cas x / Coût total benchmark). L'utilité totale inclue le profit du GRT après ajustement par l'aversion au risque (A_grt=0,01).

Tableau 4-2 : Résultats du modèle de base, pour une *gate closure* proche du temps réel

Gate closure proche du temps réel (Std_{sys}=10)		Benchmark	Cas 1	Cas 2
		Sans pénalité	Pénalité faible	Pénalité forte
Prix	Prix *forward*	29,5	29,6	30,1
	E[Prix temps réel]	30,0	30,0	30,0
Quantité *Forward*	Revendeur type A	66,5	67,3	69,9
	Revendeur type B	33,1	34,0	36,6
	Producteurs flexibles	99,6	101,3	106,5
Profit espéré	Revendeur type A	352,3	307,9	236,3
	Revendeur type B	168,6	129,3	73,1
	Producteurs flexibles	1464,1	1478,5	1526,2
Utilité espérée	Revendeur type A	348,5	303,0	230,2
	Revendeur type B	164,8	124,3	67,1
	Producteurs flexibles	1461,5	1476,7	1525,0
Profit espéré GRT		0,0	69,3	149,3
Coût et efficacité	Utilité espérée totale	1974,8	1963,5	1953,6
	Efficacité	100,00%	99,43%	98,93%
	Coût de production	1515,0	1515,0	1515,0
	Efficacité productive	100,00%	100,00%	100,00%

Tableau 4-3 : Résultats du modèle de base, pour une *gate closure* loin du temps réel

Gate closure loin du temps réel (Std_{sys}=20)		Benchmark	Cas 1	Cas 2
		Sans pénalité	Pénalité faible	Pénalité forte
Prix	Prix *forward*	28,9	31,6	33,8
	E[Prix temps réel]	30,0	30,0	30,0
Quantité *Forward*	Revendeur type A	70,5	74,2	77,3
	Revendeur type B	37,1	40,8	44,0
	Producteurs flexibles	107,6	115,0	121,2
Profit espéré	Revendeur type A	350,2	70,8	-189,8
	Revendeur type B	147,2	-43,6	-228,8
	Producteurs flexibles	1442,7	1740,5	2024,3
Utilité espérée	Revendeur type A	311,5	21,1	-241,9
	Revendeur type B	108,5	-93,2	-280,9
	Producteurs flexibles	1423,0	1719,1	1985,9
Profit espéré GRT		0,0	172,2	334,2
Coût et efficacité	Utilité espérée totale	1843,0	1795,4	1740,2
	Efficacité	100,00%	97,42%	94,42%
	Coût de production	1560,0	1560,0	1560,0
	Efficacité productive	100,00%	100,00%	100,00%

En comparant les quantités d'équilibre achetées par les revendeurs sur le marché *forward*, nous vérifions que l'hypothèse prise pour la détermination des prix des écarts est valide ($D_{j,low} \leq X_{R_j}^{F\,*} \leq D_{j,high}$, cf. sous-section 1.2.3.2). Voir tableaux 4-4 et 4-5. Par

ailleurs, nous avons vérifié que les autres cas ($X_{R_j}^{F\,*} < D_{j,low} < D_{j,high}$, $D_{j,low} < D_{j,high} < X_{R_j}^{F\,*}$) n'étaient pas possibles.

Tableau 4-4 : Demandes individuelles et quantités achetées pour chaque revendeur

$Std_{sys} = 10$	Revendeur type A	Revendeur type B
$D_{j,low}$	63,1	29,8
$D_{j,high}$	80,8	47,5
$X_{R_j}^{F\,*}$ (pénalité faible)	67,3	34,0
$X_{R_j}^{F\,*}$ (pénalité forte)	69,9	36,6

Tableau 4-5 : Demandes individuelles et quantités achetées pour chaque revendeur

$Std_{sys} = 20$	Revendeur type A	Revendeur type B
$D_{j,low}$	59,6	26,3
$D_{j,high}$	95,0	61,6
$X_{R_j}^{F\,*}$ (pénalité faible)	74,2	40,8
$X_{R_j}^{F\,*}$ (pénalité forte)	77,3	44,0

1.3.3 Discussion des résultats

Quatre conséquences économiques apparaissent dans ces simulations numériques : 1° une distorsion des prix et des quantités du marché *forward* ; 2° des modifications asymétriques des utilités des participants aux marchés ; 3° une hausse des revenus du GRT ; et 4° des inefficacités.

1.3.3.1 La distorsion des prix du marché *forward* et l'*over-contracting*

L'utilisation de pénalités en temps réel modifie les coûts d'opportunité des participants sur le marché *forward*. Ce qui entraîne des distorsions du prix du marché *forward*. La Figure 4-9 présente les modifications des prix du marché *forward* pour les différents cas que nous étudions.

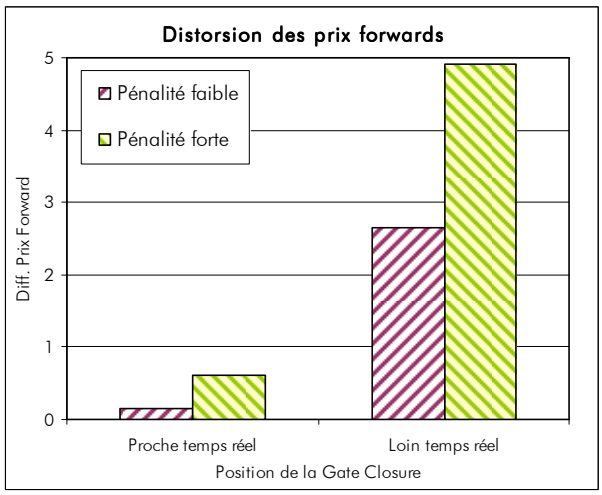

Figure 4-9 : Influence d'une pénalité sur le prix *forward*, modèle de base

Un autre effet des pénalités est le phénomène d'*over-contracting*. Les Figures 4-10 et 4-11 l'illustrent par le ratio entre les achats *forward* des revendeurs et la demande individuelle anticipée. Quand une pénalité est appliquée on peut voir que les revendeurs cherchent toujours à acheter plus que la demande anticipée. Bien sûr, l'*over-contracting* est plus important pour un revendeur plus exposé (pour un revendeur de type B).

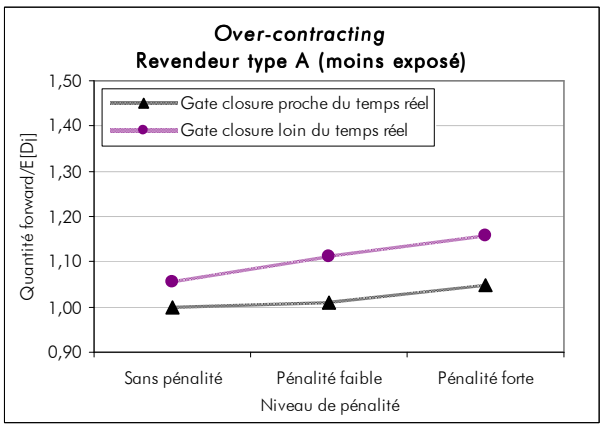

Figure 4-10 : Influence d'une pénalité sur la quantité achetée par le revendeur type A (moins exposé)

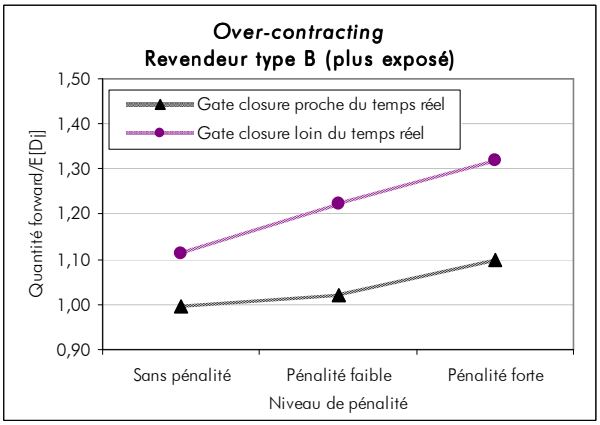

Figure 4-11 : Influence d'une pénalité sur la quantité achetée par le revendeur type A (moins exposé)

Remarquons ici que des preuves empiriques de l'effet d'*over-contracting* ont été retrouvées dans des mécanismes d'ajustement avec de fortes pénalités. C'est notamment le cas de l'Angleterre où l'*over-contracting* a été détecté (Cornwall [2002], Newbery-McDaniel [2003], cf. Chapitre 3, sous-section 2.3).

La distorsion du prix *forward* et l'effet d'*over-contracting* peuvent être expliqués en regardant les équations qui représentent les courbes d'offre et de demande des différents participants sur le marché *forward* (cf. sous-section 1.2.2). Les pénalités accroissent la covariance entre le profit « sans couverture » des revendeurs et le prix des écarts ainsi que la volatilité des prix des écarts. Cela produit une augmentation de la demande des revendeurs sur le marché *forward*. En effet, les revendeurs préfèrent acheter plus sur le marché *forward* pour couvrir leurs profits et pour avoir une exposition plus faible aux écarts négatifs. Ceci crée une tension sur le marché *forward* qui aboutit à une distorsion du prix et de la quantité sur ce marché (figure 4-12).

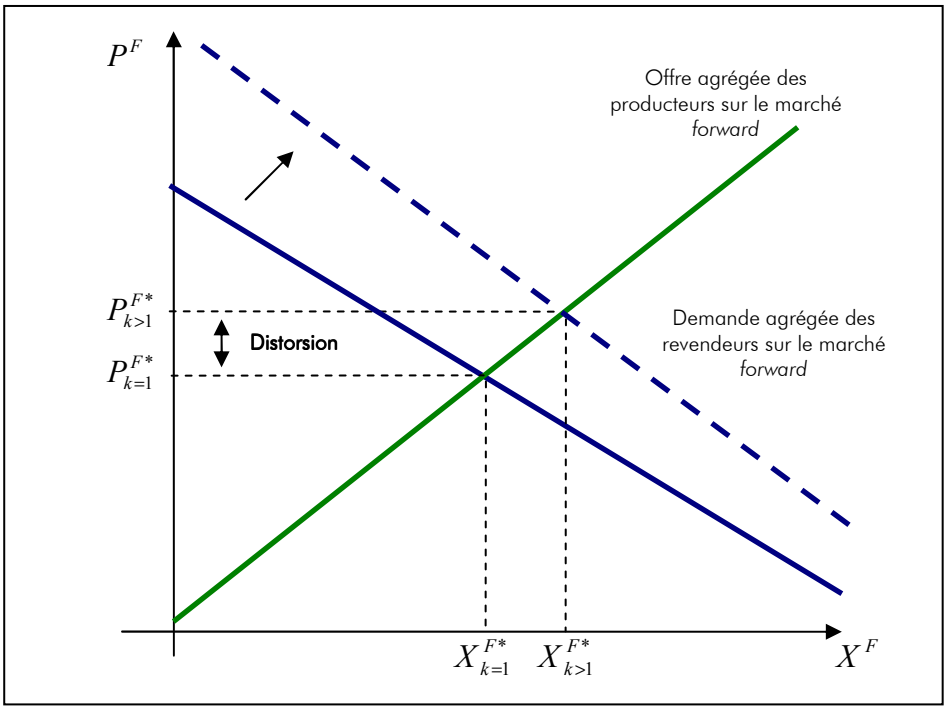

Figure 4-12 : Influence d'une pénalité en temps réel sur l'équilibre du marché *forward*

La distorsion des prix sur le marché *forward* peut modifier la distribution d'utilité entre les participants au marché et créer une barrière à l'entrée pour certains agents. De plus, ces distorsions peuvent créer une apparence d'exercice de pouvoir de marché par réapparition d'un mark up « prix / coûts » même quand le marché est concurrentiel (Smeers [2005b]).

1.3.3.2 Changements asymétriques des utilités des participants aux marchés

L'introduction de pénalités n'affecte pas de manière identique tous les participants aux marchés. On peut le constater en regardant comment leur utilité espérée change avec l'introduction de pénalités. Les Figures 4-13 et 4-14 représentent ces changements par rapport au *benchmark* (qui est le cas sans pénalités).

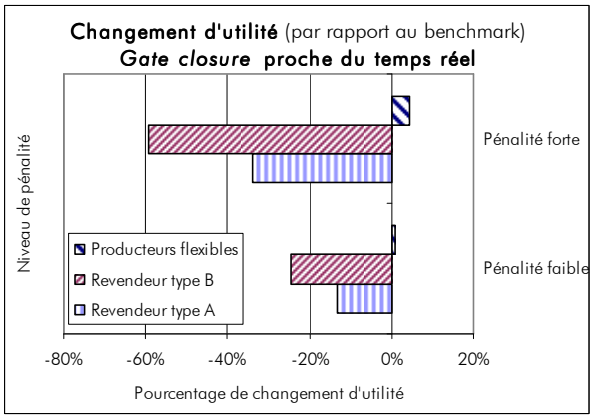

Figure 4-13 : Influence d'une pénalité sur le changement d'utilité des participants, *gate closure* proche du temps réel

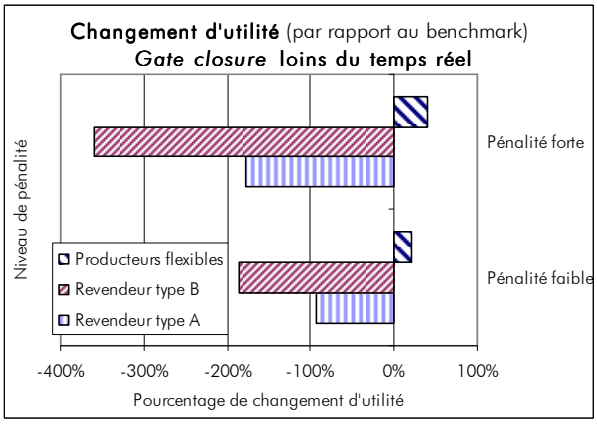

Figure 4-14 : Influence d'une pénalité sur le changement d'utilité des participants, *gate closure* loin du temps réel

Deux conséquences principales sont constatées. La première conséquence est une redistribution d'utilité entre les revendeurs et les producteurs. Les acheteurs nets sur le marché *forward* sont des revendeurs, et leur utilité diminue. Tandis que les vendeurs nets sont des producteurs, et leur utilité augmente. Ce transfert d'utilité pourrait être justifié en le considérant comme une forme de rémunération de la « flexibilité » des producteurs car, finalement, ce sont eux qui permettent l'équilibre du système en temps réel. Mais cette argumentation pose deux questions. La première est de déterminer le niveau « optimal » de pénalité, afin de rémunérer correctement les technologies flexibles. La deuxième vient de ce que ce transfert d'utilité pourrait aller aussi à des producteurs qui ne présentent pas cette flexibilité. Nous reviendrons sur cette possibilité dans la sous section suivante (1.4) qui considère deux types de technologies de production.

La seconde conséquence est un impact asymétrique entre les revendeurs de type A et B. Les pénalités affectent plus les petits agents non intégrés verticalement (les revendeurs de type B) que les agents de grande taille ou intégrés. Les revendeurs de type B (à la fois petits et non intégrés) voient leur utilité décroître deux fois plus, proportionnellement, que les revendeurs de type A (qui sont de grande taille ou intégrés verticalement en production). L'utilisation de pénalités crée ainsi un type de barrière à l'entrée pour des agents de petite taille ou non intégrés en production. Plus généralement, nous pourrions penser que le « mécanisme d'ajustement » est défavorable pour les agents devant faire face à plus d'incertitude (revendeurs ou agrégateurs avec de petites bases de clientèle, petits producteurs, producteurs éoliens, etc.). Cette barrière peut inciter de tels agents à ne pas entrer dans l'industrie et réduire ainsi la dynamique concurrentielle (Newbery [2005a], LE [2004]).

1.3.3.3 Accroissement des revenus du GRT

L'introduction d'une pénalité crée un flux de revenus pour le GRT (figure 4-15). Ce revenu du GRT (familièrement appelé « *Beer fund* ») s'accroît avec le niveau des pénalités et quand la position de la *gate closure* s'éloigne du temps réel.

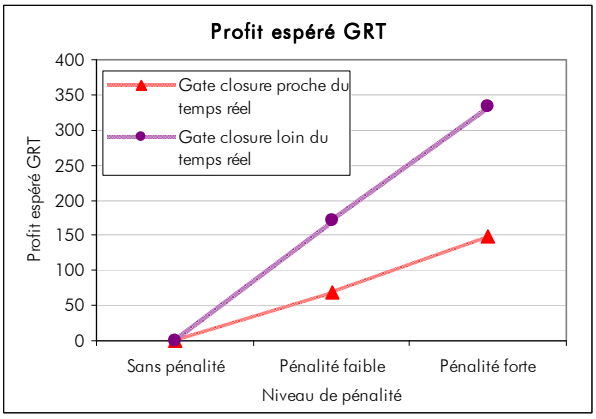

Figure 4-15 : Influence d'une pénalité sur le profit espéré du GRT

Un tel mécanisme de revenu ne donne pas de bonnes incitations au GRT pour choisir le meilleur *design* de module du temps réel. Le fait que le revenu du GRT s'accroisse automatiquement quand le niveau de pénalité augmente, et quand la *gate closure* s'éloigne, ne donne aucune indication utile sur l'amélioration de la sécurité du système que le GRT pourrait obtenir en pénalisant d'avantage les participants aux marchés.

Il est important de remarquer que l'utilité des producteurs s'accroît lui aussi avec le niveau des pénalités. Comme, dans certains pays, les GRT, les gros producteurs et les producteurs verticalement intégrés peuvent être assez proches les uns des autres et être tous très influents sur les choix de règles du *Market Design*, on peut craindre que de mauvais choix initiaux d'un design du module du temps réel ne soient suivis par un verrouillage prolongé sur ces mauvaises bases. Et qu'il soit ensuite difficile d'améliorer les dispositifs mis

en place. En conséquence, une attention particulière doit être portée sur la régulation du revenu du GRT provenant des pénalités. Cette régulation peut être orientée pour l'allocation de ce revenu pour payer certains coûts spécifiques du GRT (réserves, coûts venant des non-convexités, etc.). Une autre possibilité de régulation serait la redistribution de ce revenu entre les participants du marché, en proportion de la quantité d'énergie injectée ou soutirée. Cependant cette dernière possibilité pourrait augmenter l'asymétrie de l'impact des pénalités sur l'utilité trouvée précédemment (cf. 1.3.3.2, Cornwall [2002]).

1.3.3.4 Inefficacités

Dans la figure 4-16, nous représentons l'efficacité qui est calculée en comparant l'utilité totale (somme des utilités des tous les participants) à l'utilité du cas de référence. Nous y trouvons des inefficacités venant de la diminution de l'utilité totale. Ces inefficacités qui apparaissent proviennent du fait que les pénalités accroissent la volatilité des profits. Les participants aux marchés, ayant de l'aversion au risque, voient donc leur utilité espérée décroître.

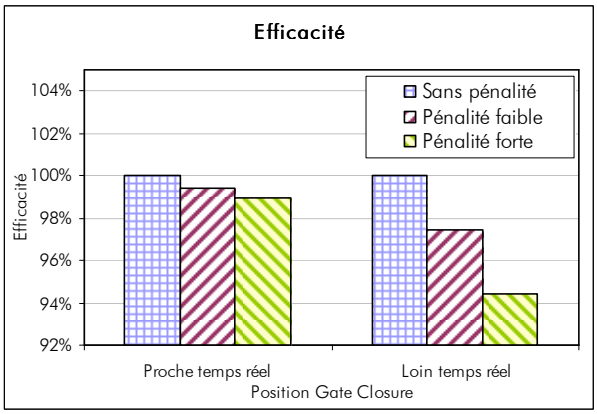

Figure 4-16 : Influence d'une pénalité sur l'efficacité

En revanche, dans notre modèle de base, où tous les producteurs sont flexibles, l'efficacité productive n'est pas endommagée par l'introduction de pénalités. En effet, puisque les producteurs flexibles déterminent les quantités de production en se basant exclusivement sur le prix du temps réel, les pénalités sur les écarts n'impactent pas leurs décisions de production. Notez que la prise en compte des moyens de production plus complexes, qui prennent leurs décisions de production en présence d'incertitudes ou en se basant sur des prix *forward,* pourraient provoquer des inefficacités productives (cf. sous-section 1.4). De la même manière, la prise en compte des coûts de démarrage (et des coûts à charge vide) des moyens de production, non représentés dans notre modèle, pourrait modifier les résultats d'inefficacité productive venant d'un excès de capacité de production disponible. Finalement, l'inclusion d'acteurs du marché plus complexes, comme par exemple des revendeurs possédant des moyens de production propres et de la capacité à l'auto-équilibrage, pourrait aussi déboucher sur des inefficacités productives lors de l'application d'une pénalité.

En conclusion, notre modèle de base nous a permis de montrer que les différents designs pour un module du temps réel ont différentes propriétés économiques si on considère leur impact sur la séquence de marchés de l'énergie. Selon notre modèle de base l'utilisation de pénalités, c.à.d. le design du mécanisme d'ajustement, change les incitations données en temps réel et provoque ainsi quatre conséquences sur la séquence de marchés : une distorsion sur le prix et sur les quantités *forward* d'équilibre (*over-contracting*) ; un impact asymétrique défavorisant les participants aux marchés de petits taille ou non-intégrés verticalement ; un impact sur le revenu du GRT et enfin, une diminution de l'utilité totale ou inefficacité. Toutes ces conséquences augmentent quand la *gate closure* s'éloigne du temps réel, puisque les incertitudes sur le marché *forward* augmentent et amplifient ainsi les incitations données par le mécanisme d'ajustement. Nous avons réalisé une analyse de sensibilité des résultats (cf. annexe 2). De cette analyse nous remarquons que, dans toutes les simulations réalisées avec les nouveaux paramètres, le « signe » ou la direction des différents effets d'une pénalité n'ont pas été modifié. En revanche, le poids ou la valeur absolue des résultats dépendent des différents paramètres.

Le modèle de base n'a considéré que le cas des producteurs flexibles. Ces producteurs sont libres de prendre leurs décisions de production jusqu'au temps réel. C'est-à-dire qu'ils sont complètement flexibles. Cependant, en réalité, les moyens de productions réels ne possèdent pas tous cette flexibilité.

1.4 Extension du Modèle à deux technologies de production

Dans la présente sous-section nous présentons une extension du modèle où nous introduisons une nouvelle technologie de production, inflexible, à côté de celle flexible de notre modèle de base. Les producteurs inflexibles doivent choisir le volume de leur production dans le cadre du marché *forward* parce que leurs unités de production ne sont plus flexibles au-delà de cette échelle de temps. Ces producteurs prennent donc leurs décisions de production en observant le prix sur le marché *forward* et dans le contexte d'incertitude de ce marché.

Dans notre extension du modèle de base, il y a N_{IG} producteurs inflexibles identiques. En raison de la nature de leur technologie de production, ils doivent prendre leurs décisions de production avant la fermeture du marché *forward* (*gate closure*). Et ils ne pourront plus modifier ces décisions. En conséquence, ils vendent leur électricité uniquement sur le marché *forward*. Leur fonction de coût de production est de type quadratique: $CT_{IG_l}(X_{IG_l}) = \theta_{IG} X_{IG_l}{}^2 / 2$ où θ_{IG} est le coefficient de coût marginal et X_{IG_l} est la quantité d'énergie produite par le producteur inflexible l.

Dans cette sous-section, nous présenterons d'abord la modélisation de l'extension du modèle de base (1.4.1). Ensuite, nous réaliserons les simulations numériques (1.4.2). Enfin, nous discuterons les résultats et conclurons (1.4.3).

1.4.1 Modélisation

Nous gardons les hypothèses du modèle de base. Nous ajoutons une hypothèse concernant le comportement des producteurs inflexibles. Nous supposons que les producteurs inflexibles ne prennent pas volontairement de positions déséquilibrées (en écart) à l'horizon du temps réel. En conséquence, ils produisent exactement la quantité qu'ils ont vendue sur le marché *forward* ($X_{IG_l} = X^F_{IG_l}$).

Pour faciliter la compréhension, nous reprenons les équations développées précédemment qui seront donc encore valables dans ce modèle étendu à deux technologies de production.

1.4.1.1 Etape du temps réel

Les équations représentant le comportement des producteurs flexibles et des revendeurs ne changement pas en temps réel.

Producteurs flexibles

$$\forall \omega ; \forall i \quad X^{RT}_{FG_i,\omega} = \frac{P^{RT}_{\omega}}{\sigma_{FG}} - X^F_{FG_i} \quad \text{.....(11)}$$

Revendeurs

$$\forall \omega ; \forall j \qquad X^{RT}_{R_j,\omega} = X^F_{R_j} - D_{j,\phi(\omega)} \quad \text{.....(12)}$$

$$IP^{RT}_{j,\omega} = \begin{cases} P^{RT}_{\omega}\, k & si\ X^{RT}_{R_j,\omega} \le 0\ (écart\, négatif) \\ P^{RT}_{\omega}\, \dfrac{1}{k} & si\ X^{RT}_{R_j,\omega} > 0\ (écart\ positif) \end{cases} \quad \text{.....(13)}$$

GRT

Etant donné que les producteurs inflexibles ne participent pas sur les transactions du temps réel, l'équation représentant l'équilibre est toujours :

$$\forall \omega \quad \sum_i X^{RT}_{FG_i,\omega} = -\sum_j X^{RT}_{R_j,\omega} \quad \text{.....(14)}$$

1.4.1.2 Etape *forward*

En revanche, sur l'étape *forward,* il faut changer les équations représentant la condition d'équilibre du marché *forward* et le comportement des participants.

Pour la condition d'équilibre du marché *forward* l'équation **(5)** devient :

$$\sum_i X^F_{FG_i} + \sum_l X^F_{IG_l} = \sum_j X^F_{R_j} \quad \text{.....(15)}$$

Cette nouvelle équation **(15)** exprime que la quantité totale vendue par les producteurs (somme des quantités vendues par chaque producteur flexible i, $X^F_{FG_i}$ et par

chaque producteur inflexible l, $X_{IG_l}^F$) est égale à la quantité totale achetée par les revendeurs (somme des quantités achetées par chaque revendeur j, $X_{R_j}^F$).

Des équations (11), (12), (14) et (15) on peut déduire que :

$$\forall \omega \quad P_\omega^{RT} = \left(D_\omega - \sum_l X_{IG_l}^F \right) \frac{\sigma_{FG}}{N_{FG}} \quad \text{.....(16)}$$

où $D_\omega = \sum_j D_{j,\phi(\omega)}$ est la demande globale pour l'état du monde ω.

L'équation (16) exprime que le prix de l'énergie en temps réel P_ω^{RT} pour chaque état du monde ω. Ce prix de l'énergie du temps réel correspondrait à l'offre marginale acceptée par le GRT pour équilibrer le système à moindre coût. P_ω^{RT} dépend logiquement de la demande globale pour l'état du monde ω, de la quantité totale produite par les producteurs inflexibles, et de la courbe d'offre agrégée des producteurs flexibles.

Pour le comportement des participants, nous avons à caractériser les producteurs flexibles, les revendeurs et les producteurs inflexibles :

Producteurs flexibles

L'équation (7) devient:

$$\forall i \quad X_{FG_i}^F = \frac{P^F - E[P_\omega^{RT}]}{A_{FG}\ Var[P_\omega^{RT}]} + \frac{Cov[\rho'_{FG_i,\omega}, P_\omega^{RT}]}{Var[P_\omega^{RT}]} \quad \text{.....(17)}$$

où $\rho'_{FG_i,\omega}$ est le profit « sans couverture » des producteurs flexibles, c.à.d. avec $X_{FG_i}^F = 0$ d'où $\rho'_{FG_i,\omega} = \pi_{FG_i,\omega}(X_{FG_i}^F = 0) = \frac{1}{2}\sigma_{FG}(D_\omega - \sum_l X_{IG_l}^F)^2$.

Revendeurs

L'équation (8) ne change pas :

$$\forall j \quad X_{R_j}^F = \frac{E\left[IP_{j,\omega}^{RT}\right] - P^F}{A_R\ Var[IP_{j,\omega}^{RT}]} - \frac{Cov[\rho_{R_j,\omega}, IP_{j,\omega}^{RT}]}{Var[IP_{j,\omega}^{RT}]} \quad \text{.....(18)}$$

Producteurs inflexibles

L'objectif du producteur inflexible l est donc de sélectionner $X_{IG_l}^F$ (ou X_{IG_l}) pour maximiser son profit. Notez que le producteur inflexible ne recourt pas à une fonction d'utilité avec aversion au risque pour maximiser ses profits car ses décisions ne sont pas « risquées ». Par hypothèse, ce producteur réalise toutes ses transactions sur le marché *forward*.

Sa fonction de profit est donnée par : $\pi_{IG_l}(X_{IG_l}^F) = P^F X_{IG_l}^F - \frac{\theta_{IG}}{2} X_{IG_l}^{F\ 2}$.

La condition nécessaire de premier ordre pour maximiser son profit est donc :

$$\frac{\partial \pi_{IG_l}(X^F_{IG_l})}{\partial X^F_{IG_l}} = 0 = P^F - \theta_{IG} X^F_{IG_l}$$

donc $X^F_{IG_l} = \dfrac{P^F}{\theta_{IG}}$ .....(19)

1.4.1.3 Equilibre de marchés

On peut maintenant utiliser les équations (13), (15) et (16) avec les positions optimales sur le marché *forward* (équations (17), (18) et (19)) pour déterminer les équilibres de marchés. (P^F , P^{RT}_{ω}). Autrement dit, nous nous servons du système d'équations (13)-(19) pour isoler le prix *forward* d'équilibre P^{F*} et les prix du temps réel pour chaque état du monde P^{RT}_{ω} .

Comme dans le modèle de base, pour résoudre le cas où il existe une pénalité ($k > 1$), on doit faire une hypothèse concernant le signe des écarts. Nous supposons d'abord que $D_{j,low} \leq X^{F\,*}_{R_j} \leq D_{j,high}$. Ainsi les prix des écarts peuvent être définis pour chaque état du monde.

Pour trouver l'équilibre, c'est-à-dire, pour déterminer le prix *forward*, les prix du temps réel et les quantités *forward* d'équilibre, nous réaliserons des simulations numériques sous *Mathematica®*.

1.4.2 Simulations numériques

Nous recourrons à des simulations numériques pour étudier les conséquences économiques de l'utilisation de pénalités au moment du temps réel.

Nous étudions trois cas différents (comme dans le modèle de base) :

- ***Benchmark*** : marché de temps réel (*pas de pénalité ;* k=1),
- **Mécanisme N°1** : mécanisme d'ajustement avec pénalité faible (k=1,2),
- **Mécanisme N°2** : mécanisme d'ajustement avec pénalité forte (k=1,4).

Dans notre modèle étendu où, à côté des producteurs flexibles, interviennent des producteurs inflexibles qui doivent prendre toutes leurs décisions de production avant la fermeture du marché *forward*, nous ne calculons les résultats que pour une position intermédiaire de la *gate closure* (Std_{Sys1} =15). Par hypothèse les producteurs inflexibles possèdent une technologie de production déterminée pour laquelle ils doivent prendre leurs décisions de production juste avant la *gate closure*. En conséquence, la position de la *gate closure* doit correspondre à la technologie des producteurs inflexibles, et la comparaison

entre différentes positions de la *gate closure* ne serait donc pas pertinente. Dans notre modèle étendu aux technologies inflexibles de production, nous n'étudions donc qu'une seule position de la *gate closure*.

Les paramètres sont les mêmes que pour le modèle de base sauf quelques modifications réalisées afin de garder un certain degré de comparabilité. Dans notre modèle étendu (avec producteurs flexibles et producteurs inflexibles), nous avons fixé le nombre de producteurs inflexibles à $N_{IG}=10$. Le coefficient de coût des producteurs flexibles est changé de $\sigma_{FG}=3$ à $\sigma_{FG}=6$, et le coefficient de coût pour les producteurs inflexibles est fixé à $\sigma_{IG}=6$. De cette manière, la courbe d'offre globale « hypothétique » (coût marginal agrégé de production), correspond toujours à la courbe d'offre utilisée pour le modèle de base. Les premiers résultats du modèle étendu sont présentés dans le Tableau 4-6.

Tableau 4-6 : Résultats du modèle étendu à deux technologies de production

Gate closure intermédiaire ($Std_{sys}=15$)		Benchmark	Cas 1	Cas 2
		Sans pénalité	Pénalité faible	Pénalité forte
Prix	Prix *forward*	29,9	31,5	32,8
	E[Prix temps réel]	30,1	28,5	27,2
Quantité forward	Revendeur type A	70,6	72,5	74,1
	Revendeur type B	37,2	39,2	40,7
	Producteurs flexibles	57,9	59,2	60,2
	Producteurs inflexibles	49,9	52,5	54,6
Profit espéré	Revendeur type A	270,1	85,2	-74,9
	Revendeur type B	101,5	-31,4	-148,8
	Producteurs flexibles	813,7	922,0	1019,8
	Producteurs inflexibles	747,1	827,0	895,6
Utilité espérée	Revendeur type A	212,6	22,4	-138,0
	Revendeur type B	44,0	-94,2	-211,9
	Producteurs flexibles	790,9	893,6	977,9
	Producteurs inflexibles	747,1	827,0	895,6
Profit espéré GRT		0,0	125,9	227,9
Coût et efficacité	Utilité espérée totale	1794,7	1762,1	1712,8
	Efficacité	100,00%	98,18%	95,43%
	Coût de production	1567,5	1571,3	1580,4
	Efficacité productive	100,00%	99,76%	99,18%

1.4.3 Discussion des résultats

L'utilisation de pénalité provoque une augmentation des prix *forward*. Sur figure 4-17 nous représentons les différences entre les prix *forward* pour les cas avec et sans pénalité.

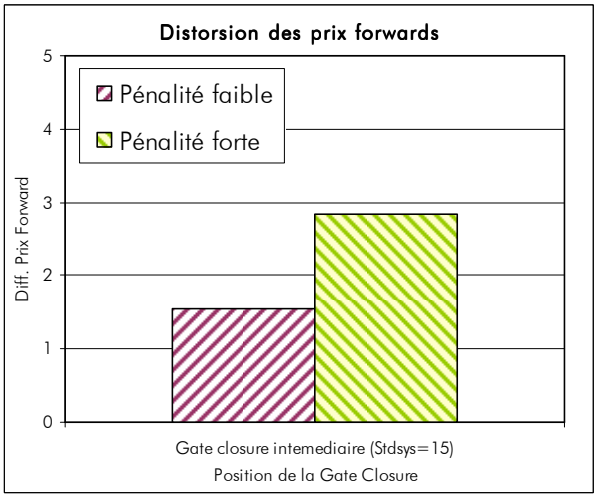

Figure 4-17 : Influence d'une pénalité sur le prix *forward* dans le modèle étendu

Le phénomène d'*over-contracting* se répète aussi pour le modèle avec deux technologies de production. Sur la figure 4-18, nous représentons les rapports entre les quantités achetées par les revendeurs et la valeur espérée de leur demande. Ce rapport augmente quand la pénalité est plus grande. C'est-à-dire que les revendeurs achètent d'avantage sur le marché *forward* afin de se couvrir des coûts élevés d'équilibrage sur le mécanisme d'ajustement.

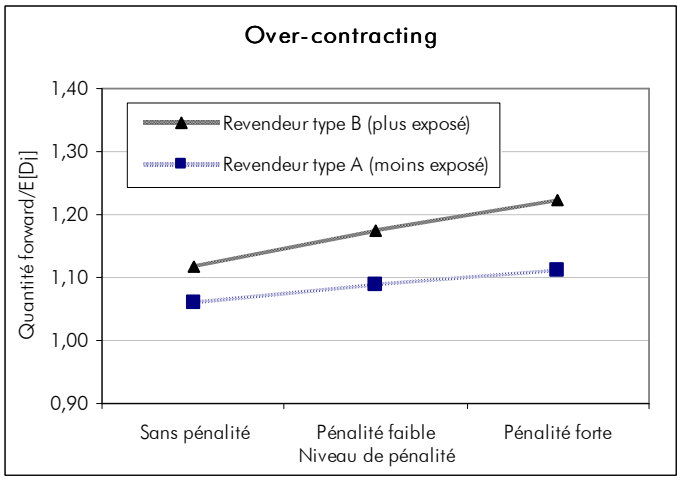

Figure 4-18 : Influence d'une pénalité sur les quantités achetées par les revendeurs dans le modèle étendu

Logiquement, dans notre extension du modèle, le profit du GRT augmente avec l'utilisation des pénalités (figure 4-19).

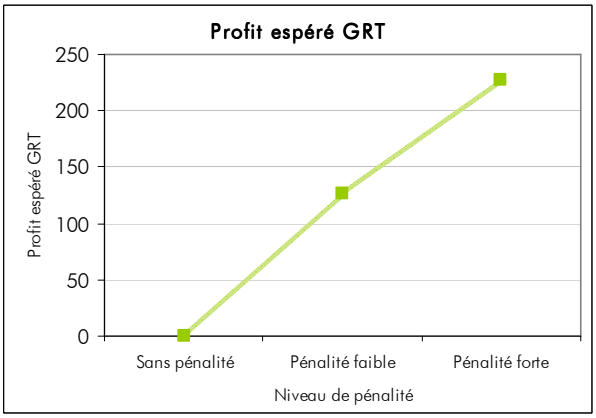

Figure 4-19 : Influence d'une pénalité sur le profit espéré du GRT dans le modèle étendu

L'introduction de pénalités produit des effets asymétriques sur les participants aux marchés. On peut le constater en regardant comment leur utilité espérée change avec l'introduction de pénalités. La Figures 4-20 représente ces changements par rapport au *benchmark* qui est le cas sans pénalités. On constate que des transferts d'utilité se produisent entre les revendeurs et les producteurs. Le revendeur le plus exposé souffre d'une diminution d'utilité deux fois plus grande que le revendeur moins exposé. Ces résultats sont similaires à ceux du modèle de base.

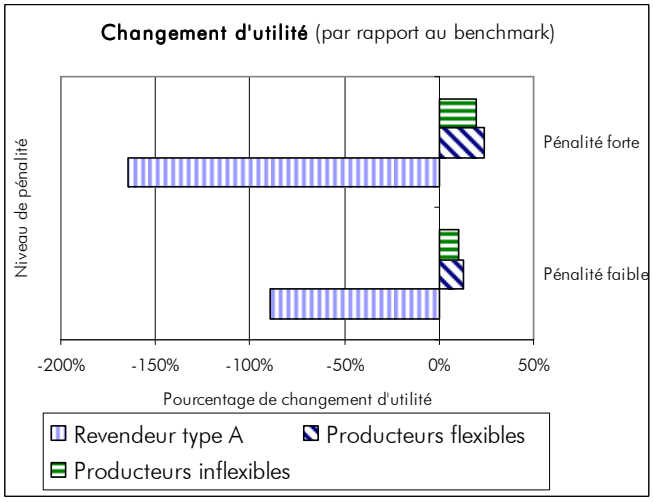

Figure 4-20 : Influence d'une pénalité sur les changements d'utilité dans le modèle étendu[231]

Cependant, dans notre modèle étendu, l'utilité transférée est partagée entre les deux types de producteurs, flexibles et inflexibles. Il est important de remarquer que l'utilité des producteurs inflexibles augmente après l'introduction d'une pénalité. Cette augmentation est presque de même proportion que pour les producteurs flexibles. On constate ainsi que l'argument d'utiliser une pénalité pour « payer » la flexibilité des technologies n'est plus valide.

[231] Pour favoriser la clarté de cette figure, nous n'y représentons pas le changement d'utilité du revendeur du type B.

Une autre conséquence de l'introduction de pénalités dans le modèle à deux technologies est que les prix du temps réel et les prix des écarts sont affectés par l'excès de production des producteurs inflexibles. La figure 4-21 représente les prix *forward* d'équilibre et la moyenne du prix du temps réel pour les cas sans pénalité et avec pénalité. Contrairement aux résultats obtenus pour le modèle de base, lorsqu'une pénalité est introduite dans le modèle à deux technologies, la moyenne du prix du temps réel diminue. Ceci est dû à l'augmentation de la quantité produite par les producteurs inflexibles. Puisque les producteurs inflexibles prennent leurs décisions de production sur la base du prix *forward*, la distorsion provoquée par la pénalité pousse les producteurs inflexibles à produire d'avantage (figure 4-22)[232].

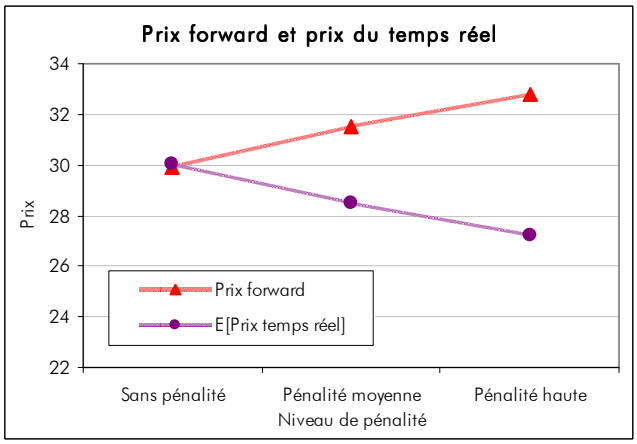

Figure 4-21 : Influence d'une pénalité sur les prix *forward* et du temps réel

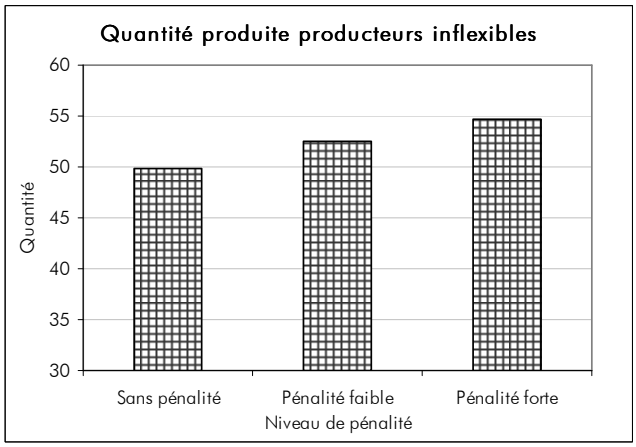

Figure 4-22 : Influence d'une pénalité sur les quantités produites par les producteurs inflexibles

La Figure 4-23 montre l'influence des pénalités sur l'efficacité dans le modèle à deux technologies de production (flexible et inflexible). Des inefficacités productives, des augmentations du coût de production, surgissent ici parce que les producteurs inflexibles prennent de mauvaises décisions de production en raison des distorsions de prix sur le

[232] Notez que, par définition, les producteurs inflexibles produisent en temps réel exactement la quantité qu'ils ont vendue sur le marché *forward*. Ils ne se mettent pas volontairement en écart et ne participent pas en temps réel avec la gestion de l'équilibre.

marché *forward*. Ces inefficacités productives proviennent par l'augmentation du coût total espéré de production (cf. figure 4-24). Il est important de remarquer que les producteurs inflexibles suivent principalement les prix du marché *forward* pour prendre leurs décisions de production.

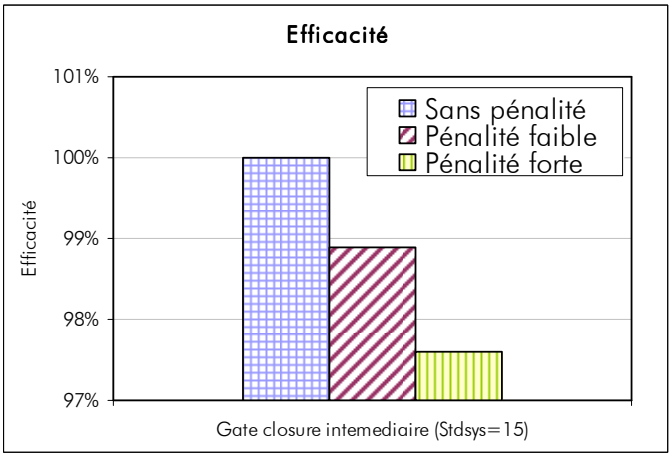

Figure 4-23 : Influence d'une pénalité sur l'efficacité dans le modèle étendu.

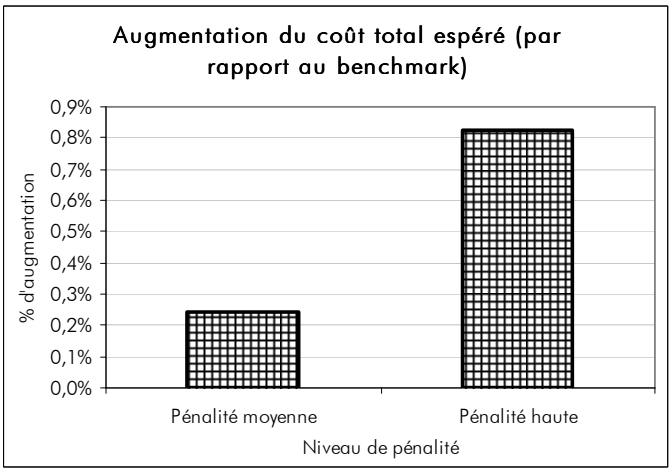

Figure 4-24 : Influence d'une pénalité sur le coût total espéré dans le modèle étendu.

En conclusion, notre extension du modèle considérant deux technologies de productions (flexible et inflexible) nous a permis de montrer que les différents designs pour un module du temps réel ont des différentes propriétés économiques concernant leur impact sur une séquence de marchés d'énergie. De la même manière que lors des simulations réalisées avec le modèle de base, l'utilisation de pénalités dans ce cas-ci modifie les incitations données en temps réel et provoque ainsi quatre conséquences sur la séquence de marchés :

- 1° une distorsion des prix du marché *forward* et *over-contracting* ;
- 2° des modifications asymétriques du bien-être des participants aux marchés ;
- 3° une hausse des revenus du GRT ;
- et 4° des inefficacités (réduction de l'utilité totale).

De plus, nous avons trouvé de nouvelles conséquences à l'utilisation des pénalités en présence de deux technologies de production :

- 5° une distorsion sur les prix du temps réel et sur les quantités produites en temps réel ;
- 6° une augmentation du bien-être des producteurs inflexibles ;
- 7° des inefficacités productives (augmentation du coût total espéré de production).

Toutes ces conséquences devraient augmenter pour une *gate closure* plus éloignée du temps réel.

1.5 Conclusions de la section

Dans cette section nous avons utilisé un modèle d'équilibre à deux étapes, *forward* et temps réel, basé sur le travail de Bessembinder et Lemmon pour étudier les conséquences économiques des deux variantes de design du temps réel : le marché du temps réel et le mécanisme d'ajustement.

Les principales hypothèses reprises à Bessembinder et Lemmon sont : la concurrence parfaite sur chacune des deux étapes (les participants se comportent comme des preneurs de prix) ; l'aversion au risque des participants ; l'incertitude présente seulement à l'étape *forward* et provenant de la nature stochastique de la demande, avec des capacités de production jamais défaillantes en temps réel.

Notre modélisation y a apporté les caractéristiques propres au design du temps réel. Nous avons simplifié la variété empirique des différents designs pour le module du temps réel existants en ne distinguant que deux genres : un design de « pur marché de temps réel », donc sans pénalités, et un design de « mécanisme d'ajustement » avec pénalités. Nous avons aussi défini comme paramètre de design la position dans le temps de la « *gate closure* », et nous l'avons représentée par le niveau d'incertitude présente à l'étape *forward*. Enfin, nous avons introduit deux types de technologies de production (flexible et inflexible).

Nous avons d'abord considéré un modèle de base où trois types d'agents économiques interagissent sur les marchés *forward* et du temps réel : les producteurs « flexibles » pouvant prendre leur décision de production en temps réel, les revendeurs, et le GRT. Nous avons alors étudié les conséquences économiques de l'utilisation de pénalités dans le module du temps réel (c.a.d. le choix entre un marché et un mécanisme). En réalisant quelques simulations numériques nous avons trouvé quatre conséquences économiques principales : 1°) une distorsion des prix et des quantités du marché *forward* ; 2°) des modifications asymétriques de l'utilité des participants aux marchés, affectant plus les agents de petite taille ou déintégrés ; 3°) un accroissement des revenus du GRT ; et 4°) des inefficacités. L'importance de ces conséquences grandit quand la position temporelle de la *gate closure* est plus éloignée du temps réel ce qui augmente le niveau d'incertitude.

Puis, nous avons présenté une extension du modèle de base avec deux technologies de production, flexible et inflexible. Les producteurs inflexibles doivent prendre leurs décisions de production avant la *gate closure* et ne peuvent plus la modifier. Avec des simulations numériques nous avons retrouvé les quatre conséquences économiques apparues dans le modèle de base. Et nous avons aussi trouvé de nouvelles conséquences : 5° une distorsion sur les prix et les quantités du temps réel ; 6° une augmentation du bien-être des producteurs inflexibles ; 7° des inefficacités productives (augmentation du coût total espéré de production).

Compte tenu de ces premiers résultats, il est bien clair maintenant que les designs des modules du temps réel ne peuvent plus être conçus exclusivement comme ceux de mécanismes techniques de sécurité et sans prendre en compte l'impact de ces règles sur la séquence de marchés d'énergie. Les designs du temps réel sont bien des arrangements institutionnels où les règles du jeu sont déterminées pour tous les agents, avec plusieurs conséquences effectives à l'horizon du temps réel, mais aussi sur les marchés d'énergie *forward* (*Day Ahead* ou infrajournaliers). En choisissant la position temporelle de la *gate closure*, on définit le niveau d'incertitude sur l'étape *forward*, c'est-à-dire la structure d'information accessible aux agents pour leurs décisions sur les marchés d'énergie *forward*. En combinant les positions de la *gate closure* et les niveaux de pénalités, on définit le système d'incitations qui s'applique aux décisions en incertitude des autres agents ayant de l'aversion pour le risque. De surcroît, ces règles du jeu affectent de manière asymétrique les revendeurs et les producteurs, les revendeurs de petite taille et ceux de grande taille, les producteurs flexibles et les producteurs inflexibles. L'utilisation des pénalités provoque un transfert d'utilité des revendeurs vers les producteurs (flexibles et inflexibles). Les pénalités peuvent constituer des barrières à l'entrée pour une catégorie entière d'acteurs : tous ceux qui sont déintégrés et de petite taille. D'autre part, l'utilisation de pénalité accroît directement le revenu du GRT ce qui peut engendrer de mauvaises incitations puisqu'il est un des principaux acteurs dans la définition du *Market Design* du temps réel. Enfin, l'utilisation de pénalités provoque deux formes d'inefficacités : une diminution d'utilité totale mais aussi une augmentation du coût de production en présence de la technologie inflexible. Ces premiers résultats montrent que les designs qui s'écartent du celui du « marché du temps réel » devraient être évités ou combinés à des positions temporelles de la *gate closure* proches du temps réel.

Dans cette section nous nous sommes concentrés sur la comparaison d'alternatives pour le module du temps réel pour une seule zone de contrôle et sans considérer l'interaction avec les zones de contrôle adjacentes. Or, les zones de contrôle sont rarement isolées et il est important de considérer l'intégration des architectures de marché, notamment lorsque l'harmonisation entre ces architectures n'est pas parfaite (cf. Chapitre 2, section 3). C'est pourquoi dans la section suivante, nous analyserons les conséquences économiques de l'intégration des architectures de marché entre zones de contrôle adjacentes.

Section 2: Modélisation de l'intégration d'architectures (deux zones de contrôle)

Cette deuxième section a pour objet d'étudier les effets économiques de l'intégration d'architectures de marché entre zones adjacentes, en partant de ce que notre réflexion a apporté à l'analyse des modules du temps réel. Cette question est d'importance en Europe. Puisqu'un marché européen unique d'électricité est toujours en construction, divers projets d'intégration entre zones sont en marche et les états cherchent à améliorer leurs *market designs* nationaux et à les harmoniser. Dans cette perspective, un cadre d'évaluation des effets économiques de l'intégration d'architectures est particulièrement nécessaire.

L'intégration d'architectures en Europe s'est centrée sur l'intégration des marchés *forward* d'énergie dans les bourses d'électricité (cf. chapitre 2, section 3). Récemment l'intégration de modules du temps réel a commencé à être considérée plus sérieusement. Différentes formes d'organisation de l'intégration sont utilisées, ou envisagées comme applicables pour l'intégration des modules du temps réel (ETSO [2005], ETSO [2006a], ETSO [2006c], Frontier-Consentec [2005]). Indépendamment des différences de mise en œuvre, ces formes d'organisation ont un objectif commun et cherchent à établir une coordination diminuant au maximum les « effets de frontières » entre les GRT (cf. Chapitre 2, section 3). Les GRT ainsi coordonnés se comportent globalement comme s'ils étaient un GRT unique, équilibrant le système global de la façon la plus économique. En d'autres termes, après la mise en commun des offres proposées dans les différentes zones, l'acceptation d'offres d'ajustement se fait de manière coordonnée, idéalement en suivant un « *dispatch* » optimal. Bien qu'un certain degré d'harmonisation entre les designs soit désirable pour l'application de ces formes de coordination entre architectures, une harmonisation complète de tous les modules n'est pas une condition indispensable à un fonctionnement acceptable pour un premier pas vers l'intégration.

Concrètement, la France, la Belgique et les Pays-Bas ont coordonné leurs trois bourses d'électricité (leurs PXs : marchés *forward day ahead*) en y intégrant un dispositif commun d'allocation des interconnections (cf. *Market Coupling*, Chapitre 2, section 3). Cependant, les trois modules du temps réel y sont restés organisés sur une base nationale ; bien que, dans chacun de ces pays, le fonctionnement du marché *forward* de court terme reste directement lié au fonctionnement du module du temps réel. Il en résulte une configuration d'intégration très spéciale : une intégration effective sur le premier horizon de temps, celui du J-1, sans intégration complète sur l'autre horizon –le temps réel-, dans un cadre de marché où toutes les transactions en énergie restent conçues sur l'ensemble des deux étapes. Cette configuration spéciale renforce l'intérêt d'une évaluation des conséquences économiques de l'intégration. Cette évaluation est particulièrement nécessaire parce que les designs utilisés pour les modules de temps réel n'ont pas été harmonisés. Les Pays-Bas utilisent un design du temps réel semblable à celui du type « marché », tandis que

la France et la Belgique recourent à un design du type « mécanisme » appliquant de fortes pénalités (cf. Chapitre 3).

Pour évaluer les conséquences économiques de ces procédés d'intégration des marchés d'énergie, nous reprendrons le cas de base des modèles d'équilibre à deux étapes développés dans la section précédente. Nous analyserons l'intégration dans une configuration simple, avec deux zones de contrôle interconnectées par une seule ligne. Deux types de participants sont présents dans chacune des deux zones : des producteurs et des revendeurs. Dans le cadre du modèle de base, une seule technologie de production est considérée : celle des producteurs « flexibles ». Notre but, dans cette ultime section, étant de comparer des cas typiques d'intégration des modules d'énergie, nous ne chercherons pas à reproduire toutes les caractéristiques empiriques du cas franco-belgo-néerlandais. Pour bien nous centrer sur les caractéristiques des marchés d'énergie à intégrer, nous ne traitons pas le module de transport et nous supposons donc qu'il n'existe pas de congestion significative à l'interconnexion. Le design de marché électrique proposé pour chaque zone est composé de deux modules d'énergie. Un module de marché *forward* (PX) et un autre du temps réel. Nous comparerons alors deux genres de designs. Le premier est un design sans intégration aucune, où les deux modules d'énergie (*forward* et temps réel) restent tous deux séparés du même module dans l'autre zone. Le second design est celui d'une intégration complète où les modules *forward* et du temps réel des deux zones sont fusionnés. Dans ce cadre de modélisation, nous étudierons par simulations numériques deux profils différents d'intégration. Dans un premier profil d'intégration, les caractéristiques des marchés à intégrer sont identiques (fonction de coût, incertitude, niveau d'aversion au risque, design/pénalité, etc.). Alors que, dans un autre profil d'intégration, les caractéristiques des marchés à intégrer diffèrent avant l'intégration. Ces différences viennent notamment d'une harmonisation incomplète entre leurs designs, associant une zone avec pénalité (mécanisme) avec une zone sans pénalité (marché). Notre but sera alors de comparer les conséquences économiques d'une intégration avec et sans harmonisation parfaite des designs des modules du temps réel.

Dans cette section nous aborderons tout d'abord, la définition des paramètres et variables (sous-section 2.1). Ensuite, nous présenterons la modélisation des architectures « séparées » et celle des architectures « intégrées » (sous-section 2.2). Puis, par simulations numériques nous évaluerons les conséquences économiques des différents profils d'intégration des marchés (sous-section 2.3). Enfin nous conclurons (sous-section 2.4).

2.1 Définition de paramètres et variables

Paramètres :

- z la zone,
- $D_{z,\phi}$ la demande d'électricité du revendeur de la zone z pour le niveau ϕ,
- p_z la probabilité du niveau bas de la demande ($\phi = low$) du revendeur de la zone z,
- $\sigma_{FG,z}$ le coefficient de coût marginal du producteur flexible dans la zone z,
- k_z le coefficient de pénalité de la zone z,
- A le coefficient d'aversion au risque pour les producteurs flexibles et pour les revendeurs,
- P_z^C Prix d'achat des consommateurs au revendeur de la zone z.

Variables de quantité :

- $X_{FG,z}^F$ quantité vendue sur le marché *forward* par le producteur flexible dans la zone z,
- $X_{R,z}^F$ quantité achetée sur le marché *forward* par le revendeur dans la zone z,
- $X_{R,z,\omega}^{RT}$ quantité achetée ou vendue en temps réel par le revendeur dans la zone z pour l'état du monde ω,
- $X_{FG,z,\omega}^{RT}$ quantité achetée ou vendue en temps réel par le producteur flexible dans la zone z pour l'état du monde ω,
- $X_{FG,z,\omega}$ quantité produite par le producteur flexible dans la zone z pour l'état du monde ω.

Variables de prix :

- P_z^F prix du marché *forward* dans la zone z,
- $P_{z,\omega}^{RT}$ prix du temps réel dans la zone z pour l'état du monde ω,
- $IP_{z,\omega}^{RT}$ prix des écarts pour le revendeur de la zone z pour l'état du monde ω.

2.2 Modélisation des architectures « séparées » et « intégrées »

Afin d'étudier les effets de l'intégration de deux architectures de marché entre deux zones, nous utilisons le cadre développé dans la section précédente et qui est fondé sur le modèle de base des équilibres à deux étapes. L'architecture de marché proposée pour chaque zone est composée de deux modules d'énergie : un module de marché *forward* (PX) et un module du temps réel.[233]

La seule technologie de production qui sera considérée est celle des producteurs « flexibles » (Modèle de base, cf. section 1). Etant donné qu'il existe ici une seule technologie de production, nous utilisons par la suite le terme « producteur » à la place de celui de « producteur flexible ».

Nous considérerons un système simple représenté par la fig. 4-25. Ce système est composé de deux zones de contrôle z (zone 1 et zone 2). Dans chaque zone il y a un producteur et un revendeur (dans la zone 1 : producteur 1 et revendeur 1 ; dans la zone 2 : producteur 2 et revendeur 2). Nous faisons l'hypothèse d'une concurrence parfaite (les acteurs ont un comportement de preneur de prix) sur les deux modules de marché. Ces deux zones de contrôle sont reliées par une seule ligne d'interconnexion. Nous faisons l'hypothèse que la capacité de transport de cette ligne n'est jamais saturée. C'est-à-dire que nous étudions des cas sans congestion. Ceci correspond à deux zones interconnectées par une ligne ayant une limite de capacité suffisamment grande pour être toujours supérieure aux flux transitant (quel que soit l'état du monde).

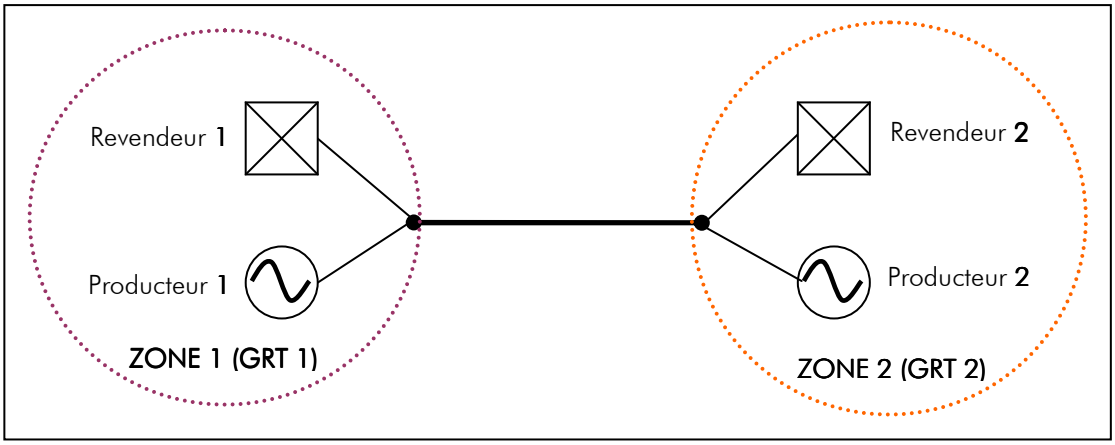

Figure 4-25 : Schéma du système de deux zones de contrôle

Dans cette configuration du système, nous considérons deux cas d'intégration d'architectures de marché pour les modules d'énergie (*forward* et temps réel) de chaque zone. Nous aboutissons donc à deux modèles de référence (voir figure 4-26). Un premier modèle de référence sans intégration aucune, où les deux modules de marchés d'énergie (*forward* et temps réel) restent tous deux séparés du même module dans l'autre zone. Et un

[233] Afin de simplifier, nous considérons que la *gate closure* est placée au même moment dans les deux marchés.

second modèle de référence d'intégration où, cette fois, les deux modules de marchés d'énergie (*forward* et temps réel) ont été tous les deux intégrés avec le même module dans l'autre zone. Dans ce cadre simple de modélisation, le fait que les modules des deux architectures soient séparés signifie qu'il n'est pas possible de réaliser des transactions de zone à zone. En revanche, quand les modules des deux architectures sont intégrés, des transactions de zone à zone sont possibles.

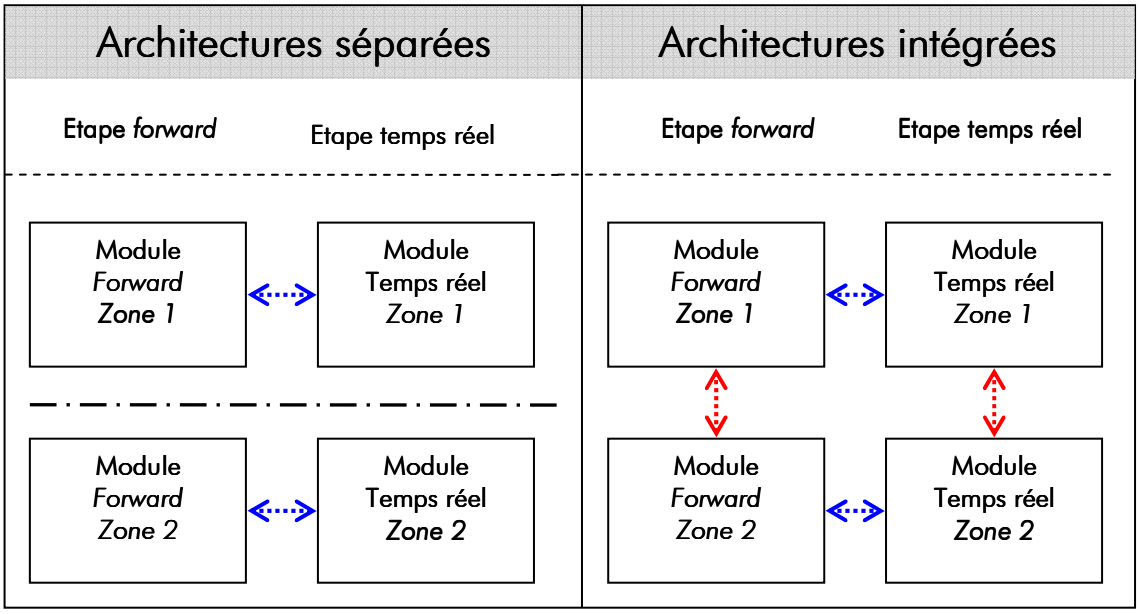

Figure 4-26 : Schéma des différents degrés d'intégration des architectures.

Nous avons donc le cadre suivant :

- **Modèle d'architectures « séparées »**
  - Modules d'énergie *forward* séparés
  - Modules du temps réel séparés
- **Modèle d'architectures « intégrées »**
  - Modules d'énergie *forward* intégrés
  - Modules du temps réel intégrés

Les acteurs économiques (producteurs et revendeurs) peuvent vendre ou acheter de l'énergie sur les modules (marché *forward* ou temps réel) de leurs zones correspondantes. En effet, nous supposons que les acteurs de chaque zone ne réalisent des transactions que dans leur zone. C'est l'architecture du design qui se charge de réaliser l'intégration entre les zones, quand l'intégration existe. Pour le cas d'architectures intégrées, nous supposons qu'il existe une coordination entre les différents marchés afin de simuler un fonctionnement coordonné (par exemple, comme NordPool ou comme le *Market Coupling* entre les Pays Bas, la Belgique et la France, cf. Chapitre 2, section 3). C'est-à-dire que le design de l'intégration des architectures réalise tous les transactions d'arbitrage possibles entre les zones. De lors, étant donné l'absence de congestions, l'intégration des architectures de marché suppose l'égalité de prix entre les zones (pour le prix *forward* et pour les prix du temps réel dans tous les états du monde).

Comme dans notre modèle de base de la section précédente, les acteurs subissent un certain degré d'incertitude au moment de prendre leurs décisions sur le marché d'énergie *forward*. Par aversion au risque, les acteurs maximisent une fonction d'utilité composée d'un terme correspondant au profit moyen et d'un terme qui est négativement proportionnel à la variance du profit. Dans cette section, nous utilisons un même coefficient d'aversion au risque A pour tous les acteurs du marché (producteurs et revendeurs).

Les producteurs de chaque zone z ont une fonction du coût total quadratique représentée par : $CT_{FG,z}(X_{FG,z,\omega}) = \sigma_{FG,z}/2\, X_{FG,z,\omega}{}^2$ où $\sigma_{FG,z}$ est le coefficient de coût marginal du producteur dans la zone z. Alors la fonction du coût marginal du producteur dans la zone est z : $CMg_{FG,z}(X_{FG,z,\omega}) = \sigma_{FG,z}\, X_{FG,z,\omega}$.

Les revendeurs ont la responsabilité d'acheter l'énergie nécessaire pour fournir la demande de leurs consommateurs. Chaque revendeur fait face à une demande stochastique et inélastique. Cette demande a deux états de réalisation possibles : l'état low où le niveau de la demande est bas ($D_{z,low}$) avec une probabilité de p_z, et l'état $high$ où le niveau de la demande est haut ($D_{z,high}$) avec une probabilité de $1-p_z$ (figure 4-27). Cette énergie est revendue aux consommateurs à un prix fixe P_z^C. Ces états de la demande par zone sont reliés à la variable aléatoire ω.

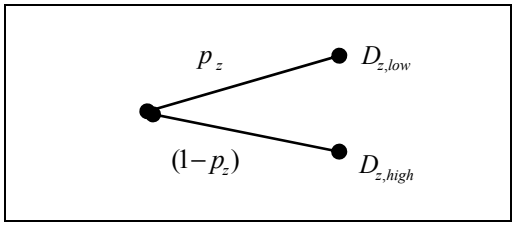

Figure 4-27 : Etats de la demande du revendeur de la zone z.

La tableau 4-7 montre les différents états des demandes par zone par rapport à la variable ω. Les différents états la demande « agrégée » des deux zones y sont aussi représentés. [234]

Tableau 4-7 : Etats du monde de la demande par zone et « agrégée »

Etat du monde (ω)	Demande zone 1	Demande zone 2	Demande « agrégée »
1	$D_{1,low}$	$D_{2,low}$	$D_1 = D_{1,low} + D_{2,low}$
2	$D_{1,low}$	$D_{2,high}$	$D_2 = D_{1,low} + D_{2,high}$
3	$D_{1,high}$	$D_{2,low}$	$D_3 = D_{1,high} + D_{2,low}$
4	$D_{1,high}$	$D_{2,high}$	$D_4 = D_{1,high} + D_{2,high}$

[234] La demande « agrégée » renvoie à la somme des demandes par zone pour chaque état du monde, c'est-à-dire, la demande qu'il y aurait si les zones sont intégrées en temps réel.

Cette section se poursuit en résolvant les problèmes d'optimisation des deux types de participants au marché (producteurs et revendeurs) en déterminant l'équilibre des marchés. Dans notre démarche de modélisation nous commençons par la prise de décision des agents au moment du temps réel, étant entendu qu'ils considèrent alors comme donnés leurs positions sur le marché d'énergie *forward* et les prix de ce marché *forward* (sous-partie 2.2.1). Une fois que nous avons déterminé les positions optimales et les prix au moment du temps réel pour chaque état du monde, nous pouvons revenir en arrière dans le temps pour déterminer les positions optimales (sous-partie 2.2.2) et les prix et quantités d'équilibre sur le marché *forward* (sous-partie 2.2.3). Quand cela est nécessaire, nous présenterons les équations correspondant aux deux cas différents d'intégration (architectures « séparées » ou « intégrés ») en l'indiquant explicitement.

2.2.1 Etape du temps réel

Au moment du temps réel, l'état du monde ω est réalisé. Il n'y a donc plus d'incertitude. De plus, dans cette étape du temps réel, les positions du marché *forward* ($X_{FG,z}^{F\ *}$, $X_{R,z}^{F\ *}$) ont déjà été prises et le prix *forward* ($P_z^{F^*}$) est déjà déterminé.[235] Ces grandeurs peuvent donc être traitées comme des grandeurs fixes. En conséquence, nous pouvons calculer les positions du temps réel et le prix du temps réel sachant que chaque agent cherche à maximiser son profit π_ω.

2.2.1.1 Producteurs flexibles

Le profit du producteur flexible de la zone z en fonction de $X_{FG,z,\omega}^{RT}$ peut être exprimé par :

$$\pi_{FG,z,\omega}(X_{FG,z,\omega}^{RT}) = P_z^{F^*} X_{FG,z}^{F\ *} + P_{z,\omega}^{RT} X_{FG,z,\omega}^{RT} - \frac{\sigma_{FG,z}}{2} X_{FG,z,\omega}^{\ 2}$$

Le premier terme représente la contribution au profit des ventes d'énergie *forward*, le deuxième terme la contribution des ventes ou des achats (acceptation des offres à la hausse ou à la baisse) du temps réel pour l'état du monde ω et le troisième terme représente le coût de production pour l'état du monde ω.

La quantité d'énergie produite par le producteur flexible doit être égale aux quantités vendues sur le marché *forward* plus (moins) les quantités vendues (achetées) au moment du temps réel, $X_{FG,z,\omega} = X_{FG,z}^{F} + X_{FG,z,\omega}^{RT}$.

[235] Les variables désignées par * sont considérées fixes.

Pour déterminer la stratégie optimale du producteur flexible, nous dérivons les conditions nécessaires du premier ordre : [236]

$$\forall \omega ; \forall z \quad \frac{\partial \pi_{FG,z,\omega}(X^{RT}_{FG,z,\omega})}{\partial X^{RT}_{FG,z,\omega}} = 0 = P^{RT}_{z,\omega} - \sigma_{FG,z}\,(X^{RT}_{FG,z,\omega} + X^{F}_{FG,z})$$

donc :

$$\forall \omega ; \forall z \quad X^{RT}_{FG,z,\omega} = \frac{P^{RT}_{z,\omega}}{\sigma_{FG,z}} - X^{F}_{FG,z} \quad \text{.....} (20)$$

Notez que l'équation (20) peut être reformulée de la manière suivante pour représenter la courbe d'offre du producteur flexible sur le module du temps réel :

$$P^{RT}_{z,\omega}(X^{RT}_{FG,z,\omega}) = \sigma_{FG,z}\,(X^{RT}_{FG,z,\omega} + X^{F}_{FG,z}) = \sigma_{FG,z}\, X_{FG,z,\omega}$$

Logiquement, étant donné l'hypothèse de concurrence parfaite (les participants se comportent comme des preneurs de prix) la courbe d'offre correspond au coût marginal.

2.2.1.2 Revendeurs

Le revendeur n'a pas de contrôle sur les variations de volume de la consommation d'énergie. En temps réel, l'état de la demande est défini. La demande est inélastique au prix ; le revendeur de la zone z vend l'électricité aux consommateurs à un prix fixe P^{C}_{z}. Les revendeurs achètent (ou vendent) au moment du temps réel la différence entre la demande effective réalisée $D_{z,\phi(\omega)}$ et leurs achats antérieurs sur le marché d'énergie *forward* $X^{F}_{R,z}$. En conséquence, les quantités achetées (ou vendues) au moment du temps réel sont :

$$\forall \omega ; \forall z \quad X^{RT}_{R,z,\omega} = X^{F}_{R,z} - D_{z,\phi(\omega)} \quad \text{.....} (21)$$

Ces ventes ou achats des revendeurs représentent des transactions « passives » ou écarts. Le prix des écarts dépend du signe des écarts et il est défini par :

$$IP^{RT}_{z,\omega} = \begin{cases} P^{RT}_{z,\omega}\, k_z & si\ X^{RT}_{R,z,\omega} \le 0\ \ (écart\,négatif) \\ P^{RT}_{z,\omega}\, \dfrac{1}{k_z} & si\ X^{RT}_{R,z,\omega} > 0\ \ (écart\ positif) \end{cases} \quad \text{.....} (22)$$

où k_z est le coefficient de pénalité de la zone z $(k_z \ge 1)$.

2.2.1.3 Gestionnaires du réseau (GRT)

La manière dont les GRTs réalisent l'équilibrage dépend de la forme d'organisation et du degré d'intégration des architectures de marché.

---

[236] Notez que le variable « prix du temps réel » est considérée par le producteur comme exogène. Etant donné l'hypothèse de concurrence parfaite, les producteurs ne peuvent pas influencer le prix par des actions individuelles.

Architectures séparées

Si les architectures de marché ne sont pas intégrées, chaque GRT est responsable de la gestion de l'équilibre pour sa zone de contrôle et pour chaque état du monde. Pour ce faire, ce GRT opère les transactions actives nécessaires (acceptation d'offres de producteur de sa zone) afin de couvrir le déséquilibre net de sa zone de contrôle. Il n'y a pas d'échanges entre les zones. Ainsi, pour ce cas des modules de marchés séparés les conditions d'équilibre du module du temps réel (marché ou mécanisme) sont donc définies par : [237]

$$\forall \omega; \forall z \quad X_{FG,z,\omega}^{RT} = -X_{R,z,\omega}^{RT} \quad \text{.....} (23)$$

L'équation **(23)** représente l'équilibre des transactions du temps réel pour chacune des deux zones et pour chaque état du monde. $X_{FG,z,\omega}^{RT}$ est le volume de transactions « actives » du temps réel (offres acceptées) pour le producteur flexible de la zone z et pour l'état du monde ω et $X_{R,z,\omega}^{RT}$ et le volume de l'écart du revendeur de la zone z et pour l'état du monde ω.

Architectures intégrées

Si les architectures de marché sont intégrées, les deux GRT se coordonnent pour équilibrer le système composé par les deux zones de contrôle. Les offres de deux zones sont mises en commun et les GRTs choisissent donc les meilleures offres des producteurs placés dans les deux zones. Il se peut donc que des flux d'énergie aient lieu entre les zones en temps réel. Ainsi, pour ce cas des modules de marchés intégrés les conditions d'équilibre du module du temps réel (marché ou mécanisme) sont donc définies par :

$$\forall \omega \quad \sum_z X_{FG,z,\omega}^{RT} = \sum_z -X_{R,z,\omega}^{RT} \quad \text{.....} (24)$$

L'équation **(24)** représente l'équilibre des transactions du temps réel pour l'ensemble des deux zones et pour chaque état du monde. La somme du volume des offres acceptées pour équilibrer le système correspond à la somme des volumes des écarts des deux zones.

2.2.2 Etape *forward*

En revenant maintenant au moment où les positions sont prises sur le marché d'énergie *forward*, nous pouvons déterminer les conditions d'équilibre des marchés et les quantités optimales vendues ou achetées par chaque agent participant à ces marchés.

[237] Le signe négatif du deuxième terme de cette équation vient de la concordance entre les signes des écarts et le langage utilisé sur les mécanismes d'ajustement pour définir les écarts positifs et les écarts négatifs du point de vue du GRT.

Architectures séparées

Pour le cas de marchés d'énergie séparés, les conditions d'équilibre des marchés *forward* sont exprimées par l'équation suivante :

$$\forall z \quad X_{FG,z}^{F} = X_{R,z}^{F} \quad \text{.....} (25)$$

L'équation **(25)** représente que la quantité d'énergie vendue par le producteur de la zone z sur le marché *forward*, $X_{FG,z}^{F}$ est égale à la quantité achetée par le revendeur de la zone z sur le marché *forward*, $X_{R,z}^{F}$. Il y donc pas d'échange entre les marchés *forward* de chaque zone.

Des équations **(20)**, **(21)**, **(23)** et **(25)** nous pouvons déduire que :

$$\forall \omega \forall z \quad P_{z,\omega}^{RT} = \sigma_{FG,z} \, D_{z,\phi(\omega)} \quad \text{.....} (26)$$

où $D_{z,\phi(\omega)}$ est la demande dans la zone z pour l'état du monde ω.

L'équation **(26)** représente le prix de l'énergie en temps réel $P_{z,\omega}^{RT}$ pour chaque état du monde ω. Notez que ce prix correspond à l'offre marginale acceptée par chaque GRT pour équilibrer sa zone. Il n'y a pas d'échange d'énergie en temps réel entre les zones et les prix du temps réel de chaque zone peuvent donc être différents.

Architectures intégrées

Pour le cas de marchés d'énergie intégrés, les conditions d'équilibre des marchés *forward* sont exprimées par l'équation suivante :

$$\sum_{z} X_{R,z}^{F} = \sum_{z} X_{FG,z}^{F} \quad \text{.....} (27)$$

L'équation **(27)** exprime que la quantité vendue par les producteurs des deux zones sur le marché *forward*, $\sum_{z} X_{FG,z}^{F}$, est égale à la quantité achetée par les revendeurs des deux zones sur le marché *forward*, $\sum_{z} X_{R,z}^{F}$. Il peut y avoir donc des échanges entre les marchés *forward* des différentes zones.

Des équations **(20)**, **(21)**, **(24)** et **(27)** (et en supposant que les prix sont égaux dans les deux marchés) nous pouvons déduire que :

$$\forall \omega \quad P_{1,\omega}^{RT} = P_{2,\omega}^{RT} = \frac{\sigma_{FG,1} \, \sigma_{FG,2}}{\sigma_{FG,1} + \sigma_{FG,2}} (D_{1,\phi(\omega)} + D_{2,\phi(\omega)}) \quad \text{.....} (28)$$

où $D_{z,\phi(\omega)}$ est la demande d'énergie dans la zone z pour l'état du monde ω.

L'équation **(28)** représente le prix de l'énergie en temps réel $P_{z,\omega}^{RT}$ pour les deux zones et pour chaque état du monde ω. Notez bien que les prix sont égaux, qu'il peut y avoir des échanges d'énergie entre les zones pour les différents états du monde, et que tout

ceci correspond à un fonctionnement coordonné optimal réalisé en temps réel par les GRT. Ce prix correspond donc à l'offre marginale acceptée pour équilibrer le système composé des deux zones.

Nous pouvons développer maintenant les stratégies optimales des participants sur le marché d'énergie *forward*. Pour cela, nous utiliserons les fonctions de profit des participants afin de construire les fonctions d'utilité espérée et de déterminer ainsi les stratégies de couverture optimales de chaque participant.

2.2.2.1 Producteurs flexibles

Sur le marché *forward*, nous pouvons exprimer le profit des producteurs comme suit:

$$\pi_{FG,z,\omega}(X^F_{FG,z}) = P^F_z X^F_{FG,z} + P^{RT}_{z,\omega} X^{RT}_{FG,z,\omega} - \frac{\sigma_{FG,z}}{2}\left(X^F_{FG,z} + X^{RT}_{FG,z,\omega}\right)^2$$

L'objectif d'un producteur est maintenant de choisir $X^F_{FG,z}$, la quantité qu'il est prêt à vendre sur le marché *forward* pour un prix *forward* donné P^F_z, de manière à maximiser son utilité espérée :

$$E_\omega[U(\pi_{FG,z,\omega})] \equiv E[\pi_{FG,z,\omega}] - \frac{A}{2} Var[\pi_{FG,z,\omega}] \ .$$

Où, $E[\pi_{FG,z,\omega}]$ et $Var[\pi_{FG,z,\omega}(X^F_{FG,z})]$ sont respectivement la valeur espérée et la variance du profit du producteur. A est le coefficient d'aversion au risque.

Alors les conditions nécessaires de premier ordre sont :

$$\forall z \quad \frac{\partial E\left[U(\pi_{FG,z,\omega}(X^F_{FG,z})\right]}{\partial X^F_{FG,z}} = \frac{\partial\left(E[\pi_{FG,z,\omega}(X^F_{FG,z})] - \frac{A}{2} Var[\pi_{FG,z,\omega}(X^F_{FG,z})]\right)}{\partial X^F_{FG,z}} = 0$$

De cette équation et de **(20)** et **(26)** ou **(28)** nous pouvons déduire :

$$\forall z \quad X^F_{FG,z} = \frac{P^F_z - E[P^{RT}_{z,\omega}]}{A\, Var[P^{RT}_{z,\omega}]} + \frac{Cov[\rho_{FG,z,\omega}, P^{RT}_{z,\omega}]}{Var[P^{RT}_{z,\omega}]} \quad \text{.....(29)}$$

où $E[P^{RT}_{z,\omega}]$ et $Var[P^{RT}_{z,\omega}]$ sont respectivement la valeur espérée et la variance du prix du temps réel. Notez que cette équation est valable pour les deux cas, séparé et intégré. Néanmoins, les valeurs des termes de cette équation dépendront du cas précis considéré.

2.2.2.2 Revendeurs

Nous pouvons, de manière similaire, exprimer le profit du revendeur de la zone z :

$$\pi_{R,z,\omega}(X^F_{R,z}) = P^C_z D_{z,\omega} - P^F_z X^F_{R,z} + IP^{RT}_{z,\omega} X^{RT}_{R,z,\omega}$$

Le premier terme de cette expression représente la contribution à son profit apportée par les « reventes » à ses clients. Les clients consomment une quantité $D_{z,\omega}$ pour chaque état du monde et paient au revendeur un prix fixe P_z^C. Le deuxième terme représente le « coût » pour le revendeur de ses achats ($X_{R,z}^F$) sur le marché *forward*. Le troisième terme représente le coût (écart négatif) ou le gain (écart positif) des transactions « passives » du temps réel du revendeur pour l'état du monde ω.

L'objectif d'un revendeur de la zone z est de choisir $X_{R,z}^F$, la quantité d'énergie qu'il est prêt à acheter sur le marché *forward* pour un prix *forward* donné P_z^F, de manière à maximiser l'espérance de son utilité : $E_\omega[U(\pi_{R,z,\omega})] \equiv E[\pi_{R,z,\omega}] - \frac{A}{2} Var[\pi_{R,z,\omega}]$. $E[\pi_{R,z,\omega}]$ et $Var[\pi_{R,z,\omega}]$ sont respectivement la valeur espérée et la variance du profit pour le revendeur de la zone z. Et A est le coefficient d'aversion au risque des revendeurs.

Les conditions nécessaires de premier ordre sont:

$$\forall z \quad \frac{\partial E[U(\pi_{R,z,\omega}(X_{R,z}^F))]}{\partial X_{R,z}^F} = 0$$

En utilisant cette équation et **(21)** nous pouvons déduire que:

$$\forall z \qquad X_{R,z}^F = \frac{E[IP_{z,\omega}^{RT}] - P_z^F}{A\, Var[IP_{z,\omega}^{RT}]} - \frac{Cov[\rho_{R,z,\omega}, IP_{z,\omega}^{RT}]}{Var[IP_{z,\omega}^{RT}]} \quad \text{.....} (30)$$

où $E[IP_{z,\omega}^{RT}]$ et $Var[IP_{z,\omega}^{RT}]$ sont respectivement le prix anticipé des écarts et la variance de ce prix des écarts. Notez que cette équation est valable pour les deux cas, séparé et intégré. Néanmoins, les valeurs des termes de cette équation dépendront du cas précis considéré.

2.2.3 Equilibre des marchés

Un équilibre des marchés existe si l'on peut trouver des valeurs de P_z^F, $X_{FG,z}^F$, $X_{R,z}^F$ qui vérifient certains sous-ensembles des équations développées précédemment.

Les équations **(29)** et **(30)** représentent les fonctions d'offre et de demande de participants sur le marché *forward* (cf. section 1). L'offre et la demande et les possibles arbitrages entre les marchés d'énergie des différentes zones dépendent du degré d'intégration des marchés. A l'intersection entre l'offre *forward* et la demande *forward* on pourra déterminer le prix et les quantités d'équilibre. Si les marchés d'énergie ne sont pas intégrés, l'équilibre de l'offre et de la demande sur le marché *forward* se réalisera par zone. Si les marchés d'énergie sont intégrés, le prix d'équilibre sera dérivé des fonctions d'offres et de demande des deux zones mises en commun.

Architectures séparées

Pour le cas des marchés séparés, nous pouvons utiliser les équations **(22)**, **(25)** et **(26)** avec les positions optimales des participants sur le marché d'énergie *forward* (équations **(29)** et **(30)**) pour déterminer les prix et les quantités d'équilibre du marché *forward* (P_z^F, $X_{FG,z}^F$ et $X_{R,z}^F$). Par exemple, pour le cas d'architectures séparées sans pénalités (avec $k_z = 1$) on a :

$$P_z^{F,sép} = E[P_{\omega,z}^{RT,sép}] + \frac{A}{2}\left(\frac{1}{\sigma_{FG,z}}\left(E[P_{\omega,z}^{RT,sép}] - P_z^C\right) Var[P_{\omega,z}^{RT,sép}] + \frac{1}{2\sigma_{FG,z}} Skew[P_{\omega,z}^{RT,sép}]\, Var[P_{\omega,z}^{RT,sép}]^{\frac{3}{2}}\right)$$

Où $E[P_{\omega,z}^{RT,sép}]$, $Var[P_{\omega,z}^{RT,sép}]$ et $Skew[P_{\omega,z}^{RT,sép}]$ sont respectivement la valeur espérée, la variance et le degré d'asymétrie (*skewness*) du prix du temps réel résultant des **marchés du temps réel séparés**.

Architectures intégrées

Pour le cas de marchés intégrés, nous pouvons utiliser les équations **(22)**, **(27)** et **(28)** avec les positions optimales des participants sur le marché *forward* (équations **(29)** et **(30)**) pour déterminer les prix et les quantités d'équilibre du marché *forward* (P_z^F, $X_{FG,z}^F$ et $X_{R,z}^F$). Par exemple, pour le cas d'architectures intégrées sans pénalités (avec $k_z = 1$) et en considérant deux marchés « identiques » on a :

$$P_z^{F,\text{int}} = E[P_{\omega,z}^{RT,\text{int}}] + \frac{A}{2}\left(\frac{1}{\sigma_{FG,z}}\left(E[P_{\omega,z}^{RT,\text{int}}] - P_z^C\right) Var[P_{\omega,z}^{RT,\text{int}}] + \frac{1}{2\sigma_{FG,z}} Skew[P_{\omega,z}^{RT,\text{int}}]\, Var[P_{\omega,z}^{RT,\text{int}}]^{\frac{3}{2}}\right)$$

Où $E[P_{\omega}^{RT,\text{int}}]$, $Var[P_{\omega}^{RT,\text{int}}]$ et $Skew[P_{\omega}^{RT,\text{int}}]$ sont respectivement la valeur espérée, la variance et le degré d'asymétrie (*skewness*) du prix du temps réel résultant des **marchés du temps réel intégrés**.

Afin d'étudier les conséquences économiques de l'intégration des marchés nous réalisons par la suite des simulations numériques développées avec l'outil *Mathematica*®.

2.3 Simulations numériques et discussion

L'objectif de cette sous-section est d'étudier, par le biais des simulations numériques, les conséquences économiques de l'intégration de deux architectures de marché. Nous comparerons les résultats de l'équilibre des marchés dans le cas de deux architectures totalement séparées et pour deux architectures complètement intégrées (c'est-à-dire, les marchés *forward* et en temps réel sont tous deux intégrés). Nous considérons ici des cas où les caractéristiques « structurelles » des zones sont identiques. Les paramètres qui définissent les participants pour chaque zone sont de même valeur (coût marginal, incertitude, aversion au risque, etc.).

Dans cette configuration, nous étudions les conséquences économiques de l'intégration de deux architectures de marché selon le degré d'harmonisation des designs du module du temps réel. Bien qu'une partie des modules du temps réel soit intégrée (partie correspondante à l'acceptation des offres), une autre partie, qui est celle du règlement des écarts, reste indépendante dans chacune des zones et peut alors être réalisée différemment (cf. Chapitre 3).

Les deux cas que nous proposons d'étudier sont donc les suivants :

- **Cas « harmonisé »** : cas de l'intégration de deux zones avec des designs identiques pour le module du temps réel (marché du temps réel – marché du temps réel, aucune des deux zones n'applique donc de pénalité $k_1 = k_2 = 1$) ;
- **Cas « non-harmonisé »** : cas de l'intégration de deux zones avec des designs différents pour le temps réel, l'un avec « mécanisme » et l'autre avec « marché ». Dans une zone (zone 1) une pénalité est utilisée pour le règlement des écarts, tandis que dans l'autre zone (zone 2), le règlement est effectué sans aucune pénalité ($k_1 > 1$ et $k_2 = 1$).

La pénalité sur le module du temps réel sera donc le paramètre de différenciation entre les deux zones. Les simulations sont alors réalisées pour deux niveaux de pénalité sur la zone 1 : Sans pénalité ($k_1 = 1$) et avec pénalité ($k_1 = 1{,}3$).[238]

2.3.1 Paramètres

Tout d'abord, de la même manière que dans la section 1, les états possibles des demandes individuelles par zone ont été exprimés comme des fonctions de la valeur espérée et d'une déviation standard. Ainsi pour le revendeur de la zone z la valeur espérée et la déviation standard de sa demande seront $\overline{D}_z = E[D_{z,\phi(\omega)}]$ et $Std_z = \sqrt{Var[D_{z,\phi(\omega)}]}$. Les différents états possibles de ces demandes seront alors exprimés par :

-Demande basse (*low*) : $D_{z,low} = \overline{D}_z - \sqrt{(1-p_z)p_z}\, Std_z / p_z$

-Demande haute (*high*) : $D_{z,high} = \overline{D}_z + \sqrt{(1-p_z)p_z}\, Std_z / (1-p_z)$.

Comme nous l'avons anticipé, nous ne considérons ici que des cas où les caractéristiques « structurelles » des zones sont identiques. Nous entendons par là que les paramètres qui définissent les participants pour chaque zone sont donc de même valeur. Les producteurs des deux zones (producteurs 1 et 2) ont la même fonction de coût marginal ($\sigma_{FG,1} = \sigma_{FG,2}$) ; les revendeurs correspondants à chaque zone (revendeurs 1 et 2) connaissent le même niveau d'incertitude ($Std_1 = Std_2$), la même distribution de la demande ($p_1 = p_2$) et le même prix fixe payé par les consommateurs ($P_1^C = P_2^C$). Le coefficient

[238] Ce niveau de pénalité corresponde à un niveau intermédiaire entre la pénalité « faible » (k=1,2) et la pénalité forte (k=1,4) utilisés dans la section 1.

d'aversion au risque (A) des acteurs est le même pour tous les acteurs dans les deux zones. Ce choix nous permet de nous concentrer sur les effets de l'intégration venant de l'équilibrage et des différences de design.

Nous reprenons ici les paramètres utilisés dans la section précédente adaptées pour cette nouvelle configuration par zone (1 producteur et 1 revendeur). Le coefficient de coût marginal, σ_{FG} , est fixé égal à 0,3 pour les deux zones.[239] La valeur espérée de la demande dans chaque zone est de $\overline{D}_1 = \overline{D}_2$ =100 et la déviation standard est $Std_1 = Std_2 = 15$.[240] En conséquence, la valeur espérée du prix du temps réel sera égale à 30 pour les deux zones.[241] Le prix fixe payé par le consommateur P_z^C est calé à 35. Le coefficient d'aversion au risque des participants, A, est fixé à 0,01.[242] La distribution de la demande a été choisie asymétrique (coefficient de *skewness* positif : p=0,8) (cf. section 1.3.1). L'ensemble des paramètres sont représentés dans le tableau 4-8.

Tableau 4-8 : paramètres des simulations numériques

Description	Symbole	Valeur
Probabilité de l'état bas de la demande	p	0,8
Coefficient d'aversion au risque des producteurs flexibles et revendeurs	A	0,01
Prix fixe payé par les consommateurs	P_z^C	35
Coefficient du coût des producteurs flexibles	$\sigma_{FG,z}$	0,3
Espérance de la demande pour les revendeurs	$\overline{D}_z$	100
Déviation standard de la demande	Std_z	15

2.3.2 Premiers résultats et discussion

Les résultats correspondants à l'intégration de deux modules de type « marché », où les designs des modules de temps réel sont donc similaires, sont présentés dans le tableau 4-9. Les résultats d'une intégration quand les designs des modules de temps réel ne sont pas complètement harmonisés (l'un est de type « mécanisme » et l'autre de type « marché ») sont présentés dans le tableau 4-10.[243]

[239] Notez qu'ici nous avons un seul producteur par zone. La courbe d'offre agrégée est donc équivalente à celle des simulations de la section 1 avec 10 producteurs et un coefficient de coût marginal fixé égal à 3.

[240] Ceci correspond à une valeur intermédiaire entre la *gate closure* proche du temps réel et celle éloignée du temps réel dans la section 1.

[241] De l'équation **(26)** nous pouvons déduire que $E[P_{z,\omega}^{RT}] = \sigma_{FG,z}\, E[D_{z,\omega}] = \sigma_{FG,z}\, \overline{D}_z$

[242] Notez que le coefficient d'aversion au risque « global » par zone, $\left(N_{FG}/A + N_R/A \right)^{-1}$,est égal à celui utilisé dans la section 1, c'est-à-dire 0,005 (cf. section 1.3.1).

[243] L'utilité totale inclue le profit du GRT après ajustement par l'aversion au risque (A_grt=0,01).

Tableau 4-9 : Intégration de deux modules de type « marché du temps réel » (sans pénalité)

zone 1 : Sans pénalité zone 2 : Sans pénalité		Architectures Séparées	Architectures Intégrées
Prix *forward*	Prix *forward* zone 1	29,5	29,4
	Prix *forward* zone 2	29,5	29,4
Quantité *Forward*	Producteur zone 1	108,5	100,1
	Revendeur zone 1	108,5	100,1
	Producteur zone 2	108,5	100,1
	Revendeur zone 2	108,5	100,1
Utilité espérée	Producteur zone 1	1473,5	1458,0
	Revendeur zone 1	491,3	513,5
	Producteur zone 2	1473,5	1458,0
	Revendeur zone 2	491,3	513,5
Profit espéré GRT zone 1		0,0	0,0
Profit espéré GRT zone 2		0,0	0,0
Coût et efficacité	Utilité espérée totale	3929,5	3942,9
	Coût de production	3067,5	3033,8

Tableau 4-10 : Intégration d'un « mécanisme » (pénalité en zone 1) avec un « marché » (zone 2)[244]

zone 1 : Pénalité (k1=1,3) zone 2 : Sans pénalité		Architectures Séparées	Architectures Intégrées
Prix *forward*	Prix *forward* zone 1	29,9	29,6
	Prix *forward* zone 2	29,5	29,6
Quantité *Forward*	Producteur zone 1	110,6	101,9
	Revendeur zone 1	110,6	105,4
	Producteur zone 2	108,5	101,9
	Revendeur zone 2	108,5	98,3
Utilité espérée	Producteur zone 1	1519,8	1476,1
	Revendeur zone 1	305,6	374,6
	Producteur zone 2	1473,5	1476,1
	Revendeur zone 2	491,3	495,6
Profit espéré GRT zone 1		138,2	119,8
Profit espéré GRT zone 2		0,0	0,0
Coût et efficacité	Utilité espérée totale	3918,6	3919,2
	Coût de production	3067,5	3033,8

Nous présentons et discutons trois types de résultats : les prix et les quantités *forward* d'équilibre, le coût total de production et l'utilité totale et, enfin, les changements individuels d'utilité des participants (producteurs et revendeurs des deux zones et GRTs). Nous comparerons ces résultats pour chacun des deux cas étudiés : i) pour l'intégration entre deux modules de temps réel « harmonisés » (« marché » avec « marché ») ; et, ii) pour

[244] Nous supposons qu'après l'intégration tout le profit venant de l'utilisation des pénalités est pour GRT de la zone 1. Une définition supplémentaire des règles serait nécessaire pour assigner une distribution du profit différent entre les différents GRTs.

l'intégration entre deux modules de temps réel « non-harmonisés » (« mécanisme » avec « marché »).

Prix et quantités *forward*

Dans le cas de l'intégration de deux designs du type marché, nous pouvons remarquer une légère diminution des prix *forward* après l'intégration (figure 4-28). D'autre part, nous remarquons une réduction plus importante dans les quantités sur les marchés *forward* (figure 4-29). Ces deux effets combinés sont dus aux modifications des courbes d'offres et de demande des participants sur le marché *forward*. En fait, la modification du comportement des participants sur le marché *forward* provient directement de l'intégration des marchés du temps réel entre les deux zones. En effet, l'intégration des marchés du temps réel permet un échange d'énergie entre les zones (une fois l'état du monde connu) ; cela modifie les prix du temps réel et les quantités vendues par les producteurs en temps réel. Ces modifications pour les différents états du monde changent les stratégies optimales des participants face au risque (courbes d'offre et de demande sur le marché *forward*) et impactent finalement les prix et les quantités *forward* d'équilibre.

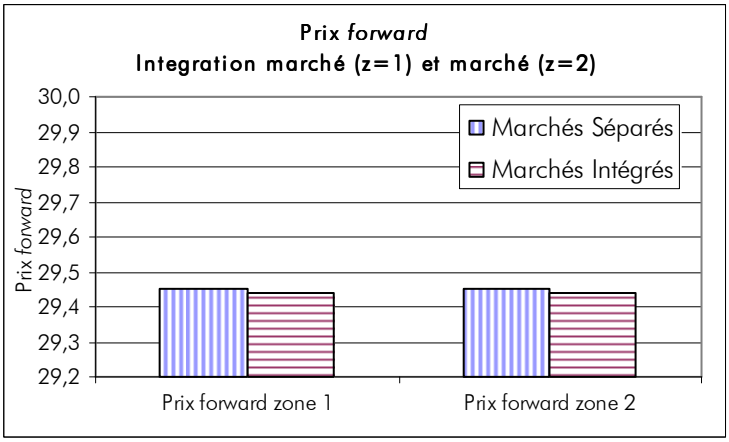

Figure 4-28 : Prix *forward* avant et après intégration de designs harmonisés

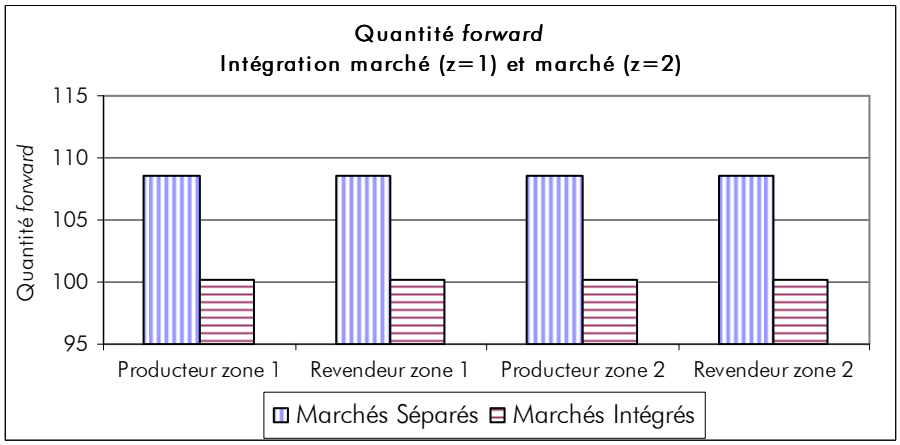

Figure 4-29 : Quantités *forward* avant et après intégration de designs harmonisés

Il est important de remarquer ici que les effets sur le marché *forward* ne viennent que de l'intégration des marchés du temps réel. Ces effets ne se produiraient pas si on n'intégrait que leurs marchés *forward*. Les deux zones ont des caractéristiques structurelles identiques (coût marginal, demande, aversion au risque, etc.) et, sans intégration, les prix *forward* d'équilibre des deux zones coïncident. Dès lors, aucun effet sur les marchés *forward* ne se produirait après une intégration partielle (c'est-à-dire, la seule intégration des marchés *forward*).

La figure 4-29 permet aussi de voir qu'après l'intégration complète d'architectures (marchés *forward* et du temps réel) entre les zones il n'y a pas d'échange entre zones sur les marchés *forward* : la quantité vendue sur le marché *forward* sur une zone est équivalente à la quantité achetée sur la zone. Cela vient aussi du fait que les deux zones présentent les mêmes caractéristiques structurelles. Même si les conditions du temps réel changent après l'intégration (prix et quantités du temps réel différents) et produisent des effets sur les marchés *forward*, ces effets impactent de manière similaire les participants des deux zones sur l'étape *forward*. En conséquence les effets d'intégration sur les prix et quantités *forward* sont symétriques sur les deux zones.

En revanche, pour le cas de l'intégration entre deux designs non-harmonisés, les effets sur les prix et les quantités *forward* ne sont pas symétriques pour les deux zones. Sur la figure 4-30, nous pouvons remarquer une diminution du prix *forward* de la zone 1 (zone du mécanisme) puis une augmentation du prix *forward* dans la zone 2 (zone du marché). En effet, avant l'intégration, le prix *forward* de la zone 1 présentait une distorsion importante provenant de l'utilisation de pénalités sur le module du temps réel (cf. section 1). Après l'intégration, la distorsion sur le prix *forward* de la zone 1 diminue. En revanche, la distorsion se transmet, en partie, sur la zone 2 où les pénalités ne sont pas pourtant appliquées.

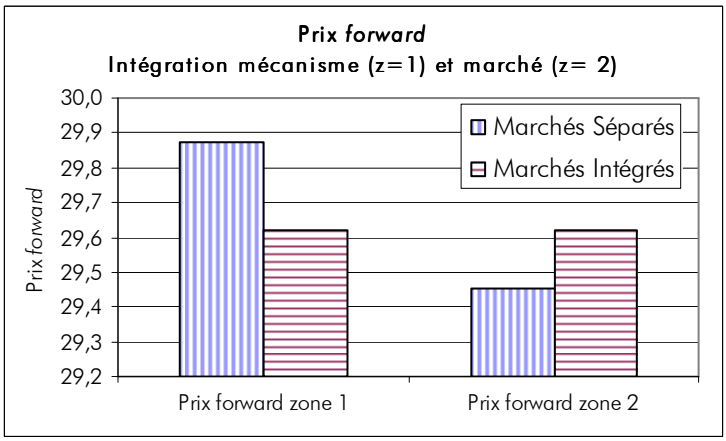

Figure 4-30 : Prix *forward* avant et après intégration de designs non-harmonisés

Sur la figure 4-31 nous représentons les quantités *forward* d'équilibre, avant et après l'intégration, pour le cas de designs non-harmonisés. Nous constatons des modifications asymétriques des quantités d'équilibre sur le marché *forward*. Par ailleurs, il y a des

échanges entre les marchés d'énergie *forward* des deux zones. En effet, le revendeur de la zone 1 achète une quantité plus grande sur le marché *forward* que celle vendue par le producteur de la zone 1. Parallèlement, dans la zone 2, le producteur vend une quantité plus grande sur le marché *forward* que celle achetée par le revendeur de cette zone. Bien que les deux zones présentent les mêmes caractéristiques structurelles (coût marginaux, demande, aversion au risque, etc.) ces échanges sur le marchés *forward* sont la conséquence des différences de design. En fait, le revendeur de la zone 1 prend en compte dans sa stratégie optimale qu'il devra payer une pénalité et il voudra donc acheter plus sur le marché d'énergie *forward* pour diminuer son exposition aux écarts négatifs.

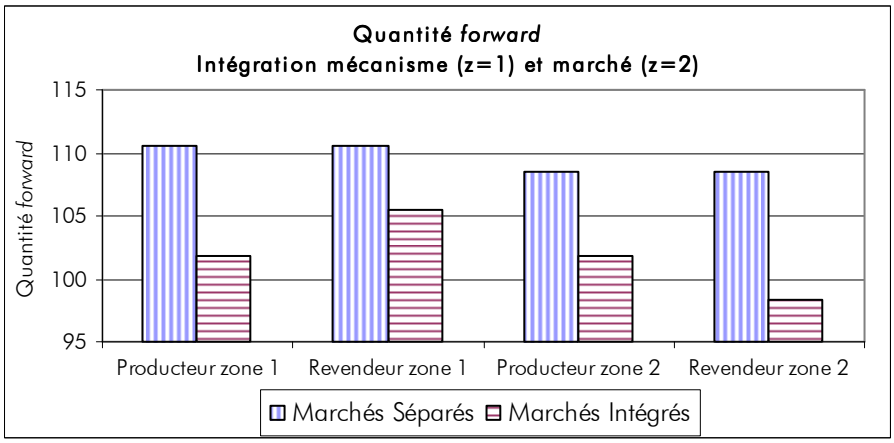

Figure 4-31 : Quantités *forward* avant et après intégration de designs non-harmonisés

Les échanges « *forward* » entre les zones, provoqués exclusivement par le manque d'harmonisation des designs du temps réel, ne produisent pas des effets directs dans notre cadre d'analyse simplifié. Cependant, ils peuvent provoquer des effets importants dans un cadre plus général. Par exemple, si on prenait en compte que la capacité de transport entre les zones est limitée ; ou s'il s'agissait d'une intégration partielle où seul les modules *forward* seraient intégrés. Dans ce cadre plus général les échanges entre marchés *forward* pourraient provoquer des distorsions sur les mécanismes d'allocation des capacités de transport. Ces échanges provoqués par des différences de design pourraient, dans un cas extrême, provoquer des congestions. De plus, si les architectures ne sont intégrées que d'une manière partielle (intégration des modules *forward* sans intégration en temps réel), ces échanges *forward* pourraient provoquer des inefficacités productives en temps réel. Ce type d'intégration partielle implique en effet le respect des échanges *forward* pour tous les états du monde en temps réel.

Les effets produits sur les prix et les quantités d'équilibre (*forward* et du temps réel) lors de l'intégration d'architectures provoqueront aussi des conséquences sur le coût total espéré de production et l'utilité totale.

Coût total espéré de production et utilité totale

Nous évaluons les gains d'efficacité découlant de l'intégration en comparant l'utilité totale ainsi que les coûts espérés de production pour les deux situations étudiées (situation de marchés séparés et situation de marchés intégrés). Sur la figure 4-32, nous représentons les gains d'utilité et la réduction des coûts pour le cas de l'intégration des designs harmonisés. Nous remarquons que l'intégration provoque une augmentation de l'utilité totale et une diminution du coût total espéré. Comme nous l'avons déjà mentionné, l'intégration de deux marchés subissant des aléas distincts (sur deux zones électriques distinctes) provoque une diminution du coût total espéré car, pour certains états de réalisation de la demande, ces aléas peuvent s'annuler mutuellement. De plus l'utilisation des moyens de production les moins coûteux est possible une fois que l'état de la demande a été défini dans les deux zones et un échange d'énergie entre les zones est possible. Par exemple, dans le cas d'un état du monde particulier où la demande d'une zone est haute et celle de l'autre basse, l'échange d'énergie entre ces zones permet i) une plus grande production à moindre coût dans la zone de demande basse et, ii) une production moins élevée dans la zone de demande haute. L'augmentation de l'utilité est ici moins importante que la diminution du coût, car l'intégration des marchés du temps réel augmente aussi la variabilité des profits des participants. En conséquence, il existe aussi des effets de diminution de l'utilité des participants, car la variabilité du profit est valorisée négativement.

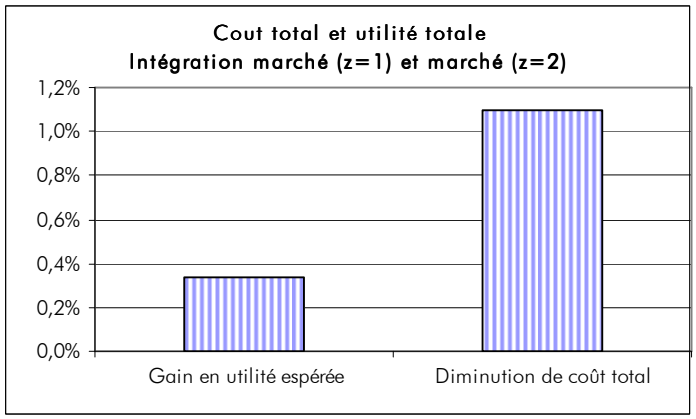

Figure 4-32 : Gain en utilité totale et diminution du coût total après l'intégration de designs harmonisés.

Sur la figure 4-33, nous représentons les gains d'utilité et la réduction des coûts pour le cas de l'intégration des designs non-harmonisés. Nous remarquons ici que l'intégration provoque une diminution du coût total espéré égale à celle du cas d'intégration de designs harmonisés. En effet, les échanges d'énergie qui se produisent en temps réel pour ce cas « non-harmonisé » sont les mêmes que ceux du cas « harmonisé ». Cependant, bien que la réduction de coûts de production soit équivalente au cas précédent, les gains d'utilité sont très réduits.

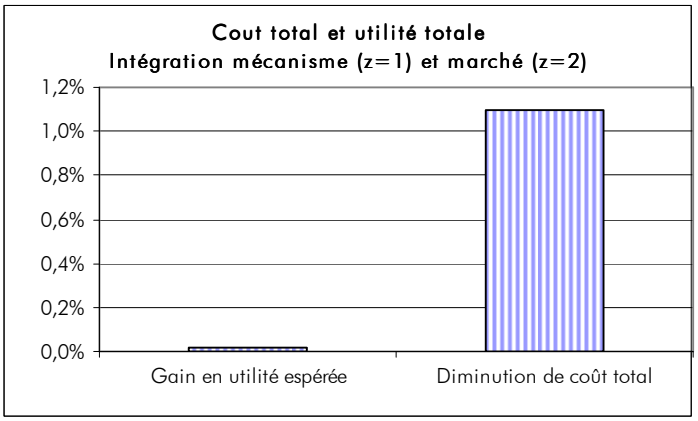

Figure 4-33 : Gain en utilité totale et diminution du coût après l'intégration de designs non-harmonisés

En conclusion, l'intégration entre zones de contrôle adjacentes est toujours bénéfique du point de vue de l'intérêt collectif. En effet, l'intégration amène toujours une réduction du coût espéré de la production et une hausse de l'utilité totale. Cependant cette hausse de l'utilité ne se répartit pas de manière uniforme entre les participants.

Changement des utilités des participants

Sur la figure 4-34 nous représentons le changement d'utilité de chaque participant après l'intégration entre deux designs harmonisés. Les transferts d'utilité entre les participants dépendent des résultats d'équilibre du marché d'énergie *forward* (prix et quantité, cf. figures 4-28 et 4-29). Dans notre simulation, la diminution du prix *forward* et des quantités *forward* d'équilibre après l'intégration des marchés provoque un transfert d'utilité des producteurs vers les revendeurs.

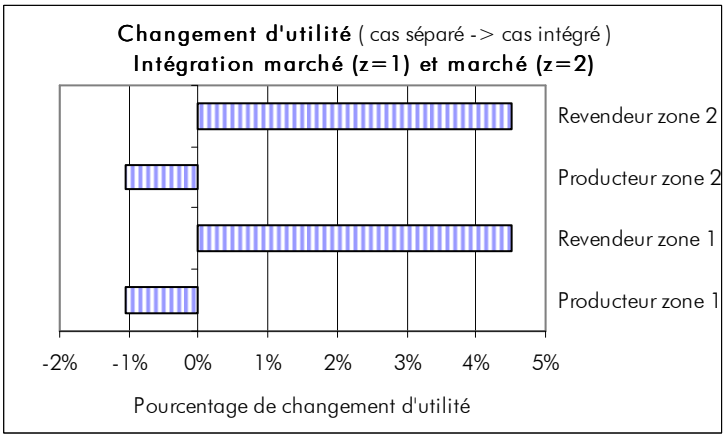

Figure 4-34 : Changement d'utilité des participants après intégration de designs harmonisés.

Tous les participants ne sont pas gagnants après l'intégration des marchés de temps réel. Mais le choix des gagnants et des perdants est strictement déterminé par les caractéristiques structurelles des zones. C'est-à-dire par la nature de leurs caractéristiques

économiques respectives. Etant donné que dans nos simulations ces caractéristiques sont identiques les effets sont symétriques pour les deux zones.

En revanche, pour le cas d'intégration de designs non-harmonisés, les effets sur les participants ne sont pas symétriques entre les zones. Sur la figure 4-35, nous pouvons observer le changement d'utilité de chaque participant après l'intégration des zones pour un module de type mécanisme (zone 1 avec pénalité) et un module de type marché (zone 2 sans pénalité). Les effets de l'intégration varient pour chaque participant. Le principal gagnant est le revendeur de la zone 1 et le principal perdant est le producteur de la zone 1 (là où les pénalités sont appliquées). En effet, étant donné la réduction de la distorsion du prix *forward* produite sur la zone 1, l'utilité du producteur de cette zone diminue. En revanche, le revendeur de cette même zone gagne en utilité après l'intégration. D'une part, la distorsion du prix *forward* de sa zone diminue et, d'autre part, il achète une quantité plus grande sur le marché *forward* diminuant ainsi le volume de ses achats « pénalisés » en temps réel.

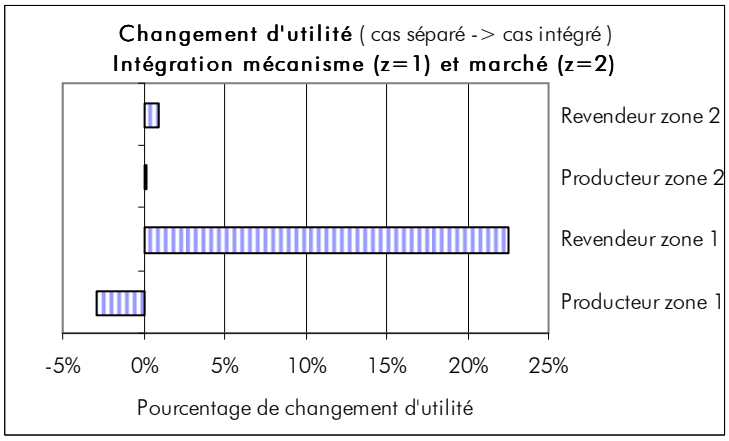

Figure 4-35 : Changement d'utilité après intégration de designs non harmonisés.

Finalement, pour le cas de l'intégration de designs non-harmonisés, nous remarquons que le profit du GRT de la zone 1 diminue après l'intégration (tableau 4-10). Cet effet pourrait donner de mauvaises incitations au GRT de cette zone et, combiné avec la perte d'utilité du producteur de sa zone, pourrait bloquer l'intégration des architectures de deux zones (cf. section 1).

2.4 Conclusions de la section

Dans cette section nous avons utilisé un modèle d'équilibre à deux étapes, *forward* et temps réel, basé sur le travail de Bessembinder et Lemmon pour étudier les conséquences économiques de l'intégration d'architectures de marché entre deux zones de contrôle différentes.

Nous avons développé la modélisation pour étudier l'intégration d'architectures de marché, notamment pour les modules énergie *forward* et du temps réel. A cette fin, nous

avons utilisé une configuration simple, avec deux zones de contrôle interconnectées et sans restriction de capacité de transport. Dans chaque zone de contrôle nous avons considéré un producteur « flexible » et un revendeur. L'équilibre de chaque zone est géré par un GRT différent. Dans cette configuration, nous avons adapté le modèle de base de notre section 1 afin d'introduire les possibilités d'intégration entre zones. Nous avons ainsi considéré deux cas extrêmes d'intégration d'architectures de marché pour les modules d'énergie (*forward* et temps réel) de chaque zone. Le premier modèle est celui d'architectures « séparées » (sans intégration aucune entre zones). Le deuxième modèle est celui d'architectures « intégrées » (modules *forward* et du temps réel intégrés).

Avec des simulations numériques nous avons étudié les conséquences de l'intégration entre deux zones sur les différentes conditions d'harmonisation. Ainsi, nous avons considérés deux profils différents des designs de modules du temps réel des architectures à intégrer. Un premier profil d'intégration où les architectures sont parfaitement harmonisées (marché-marché) et un deuxième profil d'intégration où les architectures ne sont pas parfaitement harmonisées (mécanisme et marché).

Quatre conséquences principales s'imposent au terme de nos modélisations et simulations numériques de l'intégration d'architectures de marché. Premièrement, l'intégration entre zones de contrôle adjacentes est toujours bénéfique du point de vue de l'intérêt collectif. En effet, l'intégration amène toujours une réduction du coût espéré de la production et une hausse de l'utilité totale. Cependant la hausse de l'utilité peut être moins importante que la réduction du coût espéré car l'intégration modifie l'utilité de ces agents averses au risque en augmentant la variabilité de leurs revenus.

Deuxièmement, en isolant le cas particulier de l'intégration « harmonisée » entre deux zones au *design* de type « marché du temps réel », il est remarquable que ce soit l'intégration en temps réel qui apporte les gains économiques. Car ce ne serait pas le cas si on n'intégrait que leurs marchés *forward*, puisque nous modélisons deux zones avec des caractéristiques structurelles identiques. Dans notre simulation, l'intégration provoque un transfert d'utilité des producteurs (perdants) vers les revendeurs (gagnants). Le choix des gagnants et des perdants est strictement déterminé par les caractéristiques structurelles des zones, c'est-à-dire par la nature de leurs caractéristiques économiques respectives. Cela correspond exactement à l'effet économique recherché par l'intégration des zones.

Troisièmement, dans l'autre cas, celui d'une intégration « non harmonisée » entre deux zones, l'une au *design* de type « marché du temps réel » et l'autre au design de type « mécanisme », on doit remarquer que l'intégration apporte également des gains économiques et qu'ils sont aussi répartis inégalement entre participants. Cependant ce ne sont plus, cette fois, les caractéristiques économiques structurelles qui président au choix des gagnants et des perdants, mais la nature des règles définissant le mécanisme du temps réel.

Quatrièmement, il existe donc une asymétrie des déterminants du choix des gagnants et des perdants à l'intégration, entre le cas avec harmonisation et le cas sans harmonisation des modules du temps réel. Il faudrait donc veiller à ce à ce que l'intégration d'architectures se produise dans des conditions d'harmonisation adéquates et à ce que la conception des règles du temps réel ne soit pas déléguée à des participants qui sont des perdants potentiels dans certaines variantes de design et d'harmonisation.

Conclusions du chapitre

Dans ce chapitre nous avons mis en œuvre un cadre d'analyse formalisé de la séquence des marchés d'énergie, couvrant les différentes alternatives de design pour le module du temps réel : le « marché du temps réel » et le « mécanisme d'ajustement ». Nous nous sommes concentrés sur deux questions. La première question liée à la comparaison des alternatives de design pour un pays ou une zone de contrôle. En effet, nous cherchions à évaluer les effets de ces diverses alternatives de design sur une séquence de marchés d'énergie. La deuxième question a porté sur l'intégration d'architectures de marché, et plus particulièrement l'intégration des modules de temps réel selon ses différentes alternatives de design et son influence sur la séquence de marché.

Dans une première section, nous avons d'abord modélisé la séquence des marchés d'énergie sur la base du modèle d'équilibre à deux étapes développé par Bessembinder-Lemmon [2000]. Nous avons repris leurs hypothèses principales : la concurrence parfaite sur chacune des deux étapes (les participants se comportent comme des preneurs de prix) ; l'aversion au risque des participants ; l'incertitude présente seulement à l'étape *forward* et venant de la demande, avec des capacités de production jamais défaillantes en temps réel. Notre modélisation y a apporté les caractéristiques propres au design du temps réel. Nous avons simplifié la diversité empirique des différents arrangements pour le module du temps réel existants en isolant deux designs : un « marché de temps réel », sans pénalités, et un « mécanisme d'ajustement » avec pénalités. Nous avons aussi fixé la position dans le temps de la « *gate closure* » comme paramètre de design, et l'avons représentée par le niveau d'incertitude présent à l'étape *forward*. Enfin, nous avons introduit deux types de technologies de production (flexible et inflexible).

Nous avons d'abord considéré un modèle de base où trois types d'agents économiques interagissent sur les marchés *forward* et du temps réel : les producteurs « flexibles » pouvant prendre leur décision de production en temps réel, les revendeurs, et le GRT. Puis, nous avons présenté une extension du modèle de base avec deux technologies de production, à savoir la production flexible et la production inflexible. Les producteurs inflexibles doivent prendre leurs décisions de production avant la *gate closure* et ne peuvent plus la modifier. En réalisant quelques simulations numériques nous avons mis en évidence les conséquences économiques principales de l'utilisation de pénalités en temps réel. D'abord, l'utilisation de pénalités provoque une distorsion des prix et des quantités sur le marché *forward*. Cette distorsion peut s'étendre aux prix et aux quantités d'équilibre en temps réel en présence de moyens de production inflexibles. D'autre part, ces distorsions ne touchent pas tous les participants de la même manière. Lors de l'utilisation d'une pénalité, un transfert d'utilité se produit toujours des revendeurs vers les producteurs. Les producteurs, flexibles et inflexibles, représentent toujours les gagnants de l'utilisation de pénalités ce qui affaiblit l'argument d'utilisation d'une pénalité comme moyen de rémunérer la « flexibilité ». En revanche, les revendeurs, faisant face aux incertitudes, sont toujours les perdants. De

surcroît, les revendeurs plus affectés par les pénalités sont ceux de petite taille ou déintégrés (revendeurs les plus exposés). Cela constitue ainsi une barrière à l'entrée qui entrave directement la dynamique concurrentielle. De plus, l'utilisation de pénalités accroît directement les revenus du GRT. Cet effet peut fournir au GRT des mauvaises incitations car il est l'acteur central du design de règles, et le responsable de la sécurité du réseau, souvent invoquée pour justifier le recours à des pénalités. Enfin, l'application de pénalités provoque une diminution de l'utilité totale. Cette forme d'inefficacité est amplifiée par des inefficacités productives (augmentation du coût de production) due à la technologie de production inflexible. L'importance de toutes ces conséquences économiques de l'utilisation de pénalités augmente quand la position temporelle de la *gate closure* s'éloigne du temps réel. D'après ces résultats, il est bien clair maintenant que les designs des modules du temps réel ne sont pas neutres vis-à-vis des résultats des marchés de gros d'énergie et de la dynamique concurrentielle sur ces marchés. En conséquence, les designs appliquant de pénalités devraient être évités dans la mesure du possible, ou bien combinés avec de *gate closure* proche du temps réel afin de minimiser les effets négatifs des pénalités.

Dans une deuxième section nous avons étudié la question de l'intégration d'architectures de marché, et plus particulièrement l'intégration des modules de temps réel selon ses différentes alternatives de design et son influence sur la séquence de marché. Nous avons donc adapté le modèle de base développé dans notre section 1 pour étudier les conséquences économiques de l'intégration d'architectures de marché entre deux zones de contrôle différentes. A cette fin, nous avons utilisé une configuration simple, avec deux zones de contrôle interconnectées et sans restriction de capacité de transport. Dans chaque zone de contrôle nous avons considéré un producteur « flexible » (unique technologie de production considérée) et un revendeur. Dans cette configuration, nous avons considéré deux cas extrêmes d'intégration d'architectures de marché pour les modules d'énergie (*forward* et temps réel) de chaque zone : un modèle d'architectures « séparées » (sans intégration aucune entre zones) et un modèle d'architectures « intégrées » (modules *forward* et du temps réel intégrés).

Grâce à des simulations numériques nous avons mis en évidence les conséquences de l'intégration entre deux zones sur les différentes conditions d'harmonisation. Ainsi, deux profils d'harmonisation de design des modules du temps réel ont été considérés : un premier profil d'intégration où les architectures sont parfaitement harmonisées (marché-marché) et un deuxième profil d'intégration où les architectures ne sont pas parfaitement harmonisées (mécanisme et marché).

Quatre conséquences principales émergent de ces simulations. Premièrement, l'intégration entre zones de contrôle adjacentes est toujours bénéfique pour l'intérêt collectif. Elle conduit toujours à une réduction du coût espéré de la production et à une hausse de l'utilité totale. Les gains économiques viennent directement de l'intégration des modules du temps réel. Cependant, dans un environnement où les agents sont averses au risque, l'augmentation de l'utilité peut être moins importante que la réduction du coût espéré

puisque l'intégration modifie la variabilité de leurs revenus. Deuxièmement, pour le cas particulier de l'intégration « harmonisée » entre deux zones au *design* de type « marché du temps réel », le choix des gagnants et des perdants est strictement déterminé par les caractéristiques structurelles des zones, ce qui correspond exactement à l'effet économique recherché par l'intégration des zones. Troisièmement, dans l'autre cas, celui d'une intégration « non harmonisée » entre deux zones, l'une au *design* de type « marché du temps réel » et l'autre au design de type « mécanisme », ce ne sont plus les caractéristiques économiques structurelles qui président le choix des gagnants et des perdants, mais bien la nature des règles du temps réel. Quatrièmement, il existe donc une asymétrie des déterminants du choix des gagnants et des perdants à l'intégration, entre les cas avec ou sans harmonisation des modules du temps réel. Il faudrait donc veiller d'une part, à ce que l'intégration d'architectures se produise dans des conditions d'harmonisation adéquates, et d'autre part, à ce que la conception du design du temps réel ne soit pas déléguée à des participants qui soit, peuvent tirer profit personnellement de certaines variantes de design et d'harmonisation, soit peuvent bloquer l'intégration s'ils anticipent d'être perdants.

Alors que les pays de l'Union européenne cherchent actuellement à améliorer leurs *market designs* et à les harmoniser, nous voulons souligner que les conséquences économiques que nous avons détectées et analysées ne devraient pas rester ignorées des décideurs, que ce soit les GRTs ou les Régulateurs.

Conclusion Générale

Notre recherche s'inscrit dans l'analyse économique des architectures de marché électrique. La construction d'une architecture de marché est une condition nécessaire pour la création d'un marché de gros d'électricité. Le manque de maîtrise rationnelle dans le choix des options optimales de design pousse ainsi à une recherche plus approfondie sur les différents designs, prenant en compte les spécificités du bien électricité.

Cette thèse s'appuie sur un cadre d'analyse modulaire permettant de séparer les problèmes du *market design* en autant de modules distincts. L'analyse modulaire a montré le rôle clé du module du temps réel qui constitue le noyau central de toute architecture de marché électrique et où tous les échanges physiques entre acteurs de marché se réalisent. Dès lors, notre thèse contribue à l'analyse économique du *market design* du temps réel.

Nous avons développé un cadre formalisé basé sur des modèles économiques d'équilibre à deux étapes appréhendant les caractéristiques séquentielles des marchés électriques. Ce cadre formalisé a permis d'analyser les deux grandes variantes de design appliquées au module du temps réel dans une zone de contrôle : le « marché » et le « mécanisme ». Grâce aux simulations numériques, nous avons montré que le design du module du temps réel n'est pas neutre vis-à-vis de la séquence des marchés d'énergie et de la dynamique concurrentielle. Les designs qui s'écartent du type « marché » et utilisent des systèmes de prix pénalisant les transactions du temps réel devraient être évités, dans la mesure du possible, car ils provoquent des distorsions, des inefficacités et peuvent créer des barrières à l'entrée. Parallèlement, la fermeture des marchés *forward*, déterminée par la position temporelle de la *gate closure*, devrait être le plus proche possible du temps réel afin de diminuer l'ampleur de ces distorsions. Ce cadre formalisé a permis ensuite d'évaluer les conséquences de l'intégration de modules du temps réel entre deux zones de contrôle. Les résultats des simulations numériques ont montré que l'intégration des modules du temps réel est fondamentale pour la création de marchés régionaux. De surcroît, l'intégration de ces modules doit être suivie d'une harmonisation adéquate des designs, afin que les effets de l'intégration proviennent des caractéristiques économiques des systèmes électriques intégrés et non des règles du design.

Au moment où certains pays de l'Union européenne cherchent à améliorer ou à harmoniser leurs *market designs*, des conséquences économiques de cette nature ne devraient pas rester ignorées des autorités responsables, Gestionnaires du Réseau de Transport (GRT) ou Régulateurs.

La **première partie** de cette thèse à été consacrée au développement d'un cadre d'analyse modulaire afin de pouvoir identifier et caractériser les différentes composantes d'une architecture de marché électrique. Le développement d'un cadre d'analyse modulaire demande aussi d'identifier la nature des mécanismes de coordination nécessaires à l'industrie électrique.

Dans le **premier chapitre** de cette thèse nous avons étudié les caractéristiques spécifiques de la production, du transport et de la consommation dans un système électrique. Nous avons alors montré la nature des mécanismes de coordination nécessaires pour faire face à ces caractéristiques spécifiques, notamment l'environnement incertain et le stockage difficile du bien électricité. Pour exploiter un système électrique de manière optimale à court terme, une entreprise intégrée verticalement doit utiliser une séquence de contrôles-commandes commençant en J-1 et finissant en temps réel (l'*unit commitment*, l'*optimal dispatch*, les contrôles automatiques en temps réel). Si l'entreprise intégrée fonctionne sur un réseau de transport interconnecté dont le contrôle est partagé par d'autres entreprises, la mise en place de mécanismes de coordination supplémentaires entre les séquences de contrôle est alors nécessaire pour une exploitation optimale de l'ensemble du système.

Dans le **deuxième chapitre**, nous avons analysé la création d'un marché de gros comme le remplacement de la séquence de contrôles-commandes par une séquence de marchés. Cette analyse, en isolant les éléments essentiels d'une architecture de marché électrique, a permis de définir les composantes du bien électricité et d'identifier les modules (ou sous-marchés potentiels) constituant le cœur d'une architecture de marché. Comme le bien « électricité » est composé de trois éléments - l'énergie, le transport et les réserves - une architecture de marché est elle-même composée de quatre modules : 1° le module énergie *forward*, 2° le module de transport *forward*, 3° le module des réserves *forward* plus 4° le module du temps réel. Nous avons alors développé un cadre d'analyse modulaire pour une zone de contrôle isolée en étudiant les principales options de design pour les quatre modules. Une architecture de marché pour une zone de contrôle se construit en choisissant une option pour chacun de ces modules et pour les articulations entre ces modules. Puis, dans un cadre plus général avec des zones de contrôle multiples, nous avons utilisé l'approche modulaire pour en déduire les formes d'intégration des architectures de marché. La construction d'un marché électrique régional consiste donc en une coordination des différents modules des architectures de marché adjacentes. Cette intégration peut ne pas être complète. Elle peut être mise en œuvre module par module, et peut aussi se réaliser sans une harmonisation parfaite entre les designs de différentes zones de contrôle (à l'exemple du « *Market Coupling* » de la France, de la Belgique et des Pays-Bas).

Grâce au cadre d'analyse modulaire développé, nous avons identifié un noyau indispensable, commun à toutes les architectures possibles, et qui constitue l'architecture « minimale » des réformes électriques. La **deuxième partie** de cette thèse a été consacrée à l'étude de ce noyau indispensable qui est le module du temps réel.

Le **troisième chapitre** a analysé la séquence de marchés en se centrant sur la composante « énergie ». La séquence des marchés d'énergie commence par les marchés *forward* et s'achève par le module du temps réel, gérant lui-même la seule place où l'énergie électrique, un bien non-stockable, est physiquement échangée entre les acteurs de marché. Les signaux de prix envoyés par ce module sont donc déterminants car ils influencent les comportements des agents économiques sur toute la séquence des marchés. Dès lors, nous avons caractérisé le design du module du temps réel en distinguant deux types : d'une part, le « marché du temps réel » et, d'autre part, le « mécanisme d'ajustement ». Un marché du temps réel utilise un prix unique de l'énergie pour toutes les transactions du temps réel correspondant à la valeur marginale d'équilibrage du système. Par contre, un mécanisme d'ajustement recourt à des systèmes de prix plus complexes, incluant généralement des prix différenciés pour les différents types de transactions et décourageant les écarts des opérateurs du marché en les « pénalisant ». Nous avons confronté ces deux designs du module du temps réel à des cas pratiques, et nous avons distingué les designs existants selon l'intensité de pénalités appliquées. En Europe, la Belgique et la France recourent à des mécanismes d'ajustement avec pénalités, tandis qu'une solution proche d'un marché de temps réel peut être trouvée aux Pays-Bas. Cette asymétrie de designs entre trois pays intégrant leurs marchés de gros a montré l'intérêt des problèmes d'harmonisation de designs lors de l'intégration d'architectures.

Enfin, dans le **dernier chapitre**, nous avons construit un cadre formalisé d'étude des différents designs du module du temps réel afin d'évaluer leur impact sur la séquence des marchés d'énergie. Nous sommes partis d'un modèle d'équilibre à deux étapes, développé initialement par Bessembinder-Lemmon [2000], et dont les principales hypothèses sont : la concurrence parfaite sur chacune des deux étapes ; l'aversion au risque des participants ; l'incertitude présente seulement à l'étape *forward* et provenant de la nature stochastique de la demande, avec des capacités de production jamais défaillantes en temps réel. Notre thèse a apporté ici une contribution nouvelle à l'analyse en modélisant les caractéristiques pertinentes du module du temps réel. Premièrement, nous avons introduit la notion de pénalité en temps réel et simplifié la variété empirique des différents arrangements comme une paire de designs : un « marché de temps réel », sans pénalités, et un « mécanisme d'ajustement » avec pénalités. Deuxièmement nous avons défini comme un paramètre la position temporelle de la « *gate closure* », avec laquelle le GRT clôt définitivement les échanges des marchés *forward*. Ce qui détermine le niveau d'incertitude pour les décisions propres à l'étape *forward*. Dans cette configuration de base, nous avons considéré quatre types d'agents économiques interagissant sur les marchés *forward* et du temps réel : des producteurs « flexibles » pouvant prendre leurs décisions de production en temps réel, des producteurs inflexibles prenant leurs décisions de production sur l'étape *forward*, des revendeurs, et le Gestionnaire du réseau qui gère le module du temps réel.

Grâce à ce cadre d'analyse formalisé, nous avons pu de répondre à deux questions importantes pour la construction d'architectures de marché.

Premièrement, ***quelle est la meilleure alternative de design pour le module du temps réel dans une seule zone de contrôle*** ?

D'après nos résultats, les designs qui s'écartent du type « marché » et utilisent de pénalités devraient être évités, ou au moins combinés avec une *gate closure* proche du temps réel. Nos simulations numériques ont montré les conséquences économiques de l'utilisation d'une pénalité en temps réel pour une seule zone de contrôle. Elle provoque une distorsion des prix et des quantités d'équilibre sur le marché *forward*. Cette distorsion peut s'étendre aux prix et aux quantités d'équilibre en temps réel en présence de moyens de production inflexibles. D'autre part, les distorsions créées par l'utilisation de pénalités ne touchent pas tous les acteurs de la même manière. L'utilisation des pénalités provoque un transfert d'utilité des revendeurs vers les producteurs quelle que soit leur technologie, « flexible » ou « inflexible ». Ce transfert affaiblit l'argument d'utilisation d'une pénalité comme moyen de rémunérer la « flexibilité ». L'utilisation de pénalités produit aussi des modifications asymétriques de l'utilité des participants, affectant surtout les agents de petite taille ou déintégrés. Cela constitue ainsi une barrière à l'entrée qui entrave directement la dynamique concurrentielle. De plus, l'utilisation de pénalités provoque une diminution de l'utilité totale. Cette forme d'inefficacité est amplifiée par des inefficacités productives (augmentation du coût de production) en présence d'une technologie de production inflexible. Enfin, l'utilisation de pénalités accroit directement les revenus du GRT, acteur central du design de règles, et responsable de la sécurité du réseau qui est souvent invoquée pour recourir à des pénalités. Ceci souligne l'importance d'un cadrage par le Régulateur dans la définition des règles. D'autant plus que l'importance de toutes ces conséquences économiques de l'utilisation de pénalités augmente quand la position temporelle de la *gate closure* s'éloigne du temps réel.

Deuxièmement, ***quels sont les effets de l'intégration des modules du temps réel entre deux zones de contrôle, notamment quand ces modules sont imparfaitement harmonisés*** ?

Nous avons étudié l'intégration des modules du temps réel dans une configuration simple, avec deux zones de contrôle interconnectées et sans restriction de capacité de transport. Les deux zones possèdent des caractéristiques « structurelles » équivalentes mais peuvent harmoniser ou non leurs designs. Nous comparons donc deux profils différents d'intégration : un premier profil parfaitement harmonisé (marché du temps réel et marché du temps réel) ; et un deuxième profil imparfaitement harmonisé (mécanisme d'ajustement et marché du temps réel). Les résultats des simulations ont montré, en premier lieu, que l'intégration de modules du temps réel est toujours bénéfique du point de vue de l'intérêt collectif. Car elle réduit le coût total espéré de la production et augmente de l'utilité totale. Cependant, dans un environnement où les agents ont de l'aversion au risque, les modifications de la variabilité des revenus en temps réel, que provoque l'intégration, peuvent réduire le gain d'utilité qui aurait été obtenu sans aversion au risque. En deuxième lieu, les résultats ont montré que, lors de l'intégration, les règles de design du module du temps réel dans une zone peuvent provoquer des effets et des distorsions sur les marchés de l'autre zone. En conséquence, l'harmonisation du design joue donc un rôle clé dans la détermination des gagnants et des perdants de l'intégration des marchés. Contrairement à

l'effet économique recherché par l'intégration des zones, où le choix des gagnants et des perdants serait strictement déterminé par les caractéristiques structurelles des zones, le manque d'harmonisation du module du temps réel peut apporter ses propres effets non-désirés. Dans le cas d'intégration entre une zone au *design* de type « marché du temps réel » et une autre au design de type « mécanisme », ce ne sont plus les caractéristiques économiques structurelles qui président au choix des gagnants et des perdants mais bien la nature des règles définissant le module du temps réel. Il serait donc préférable de soigner l'harmonisation des règles lors de l'intégration et de veiller à ce que la conception du design du temps réel dans une zone n'interfère pas dans le processus d'intégration. En effet, puisque le choix de gagnants et perdants de l'intégration dépend des règles choisies, ces dernières ne devraient pas être déléguées aux agents susceptibles de tirer profit ou de bloquer l'intégration.

Les résultats de notre recherche formalisée pour étudier le *market design* du « temps réel » conduisent à plusieurs **nouvelles pistes**.

Le premier axe de recherche complémentaire serait de lever certaines hypothèses de la modélisation pour pouvoir traiter des cas plus généraux et plus réalistes. L'hypothèse de la concurrence parfaite pourrait être levée en introduisant du pouvoir de marché, par exemple une concurrence à la Cournot sur l'étape du temps réel suivant les travaux de Yao *et al.* [2004]. D'autre part, l'hypothèse d'absence de contraintes de capacité de production, qui se rattache directement à la sécurité du réseau, pourrait aussi être levée. Ceci impliquerait, d'un côté, de considérer la valeur économique de la sécurité du réseau (Joskow-Tirole [2006]) et, d'un autre côté, de prendre en compte la fourniture de réserves (Siddiqui [2002]). Une fois ces deux hypothèses levées, une comparaison complète des différentes alternatives de design du module du temps réel serait possible. Un design optimal devrait en effet prendre en compte tous les effets économiques et techniques de la sécurité du réseau et du pouvoir de marché, ou tous au moins les plus importants, que ce soit pour une seule zone de contrôle ou pour plusieurs zones de contrôle.

Une seconde recherche complémentaire consisterait à étendre la modélisation des designs du temps réel en prenant en compte d'autres caractéristiques des designs existants. Dans notre modélisation, nous avons simplifié la variété empirique des différents arrangements du temps réel en ne distinguant que deux genres de design, « marché » et « mécanisme », différenciés par le paramètre de pénalité. Cette modélisation pourrait être approfondie en étudiant différents types d'enchères (enchère à prix uniforme ou enchère discriminante) ou différentes méthodes de détermination des prix des écarts (prix marginal, prix moyen, prix *forward*, etc.) (Green-McDaniel [1999]). Cet axe de recherche pourrait s'étendre d'un seul intervalle de livraison à plusieurs intervalles de livraison. Ceci permettrait de mieux appréhender les caractéristiques spécifiques de l'électricité, notamment les liens intertemporels dans la production d'énergie pour différents intervalles de livraison (coût de démarrage, contraintes de *ramping*, etc.).

Une troisième recherche introduirait dans la modélisation des agents de marché plus complexes. Ceci permettrait de construire de « maquettes » avec des configurations différentes qui chercheraient à s'approcher de situations réelles de marchés électriques particuliers. Ce serait, par exemple, l'introduction d'acteurs intégrés combinant des moyens de production de différentes technologies, ou bien différents groupes de consommateurs. Ou, encore, l'étude de l'intégration entre zones quand les technologies majoritaires de production sont de nature différente (flexibles et inflexibles). De même, on pourrait prendre en compte des producteurs faisant face à des incertitudes en production. Un cas extrême de ce type de producteur serait l'éolien. Cette technologie a un coût marginal de production à peu près nul, mais elle est aussi la source de fortes incertitudes pour le réseau, et peut provoquer des effets économiques collatéraux lors de l'intégration de différentes zones.

En fin, une quatrième et dernière recherche développerait des solutions analytiques du modèle pour les confronter à des données réelles. Notre méthodologie actuelle des simulations numériques et les pistes développées dans cette thèse pourraient alors être utilisées comme une base pour des développements analytiques généralisables. Ces développements permettraient de tester économétriquement des donnés réelles de différents marchés (Bessembinder-Lemmon [2000], Siddiqui [2002]).

Annexe

Annexe N°1 : Abréviations

€ / US$ / £	Euro / Dollar / Livre
APX	*Amsterdam Power Exchange* (Bourse d'électricité J-1 aux Pays-Bas, cf. www.apx.nl)
APX UK	*Amsterdam Power Exchange UK* (Bourse d'électricité J-1en Angleterre, cf. www.apxgroup.com)
BELPEX	Bourse d'électricité J-1 en Belgique (cf. www.belpex.be)
BRL	*Balancing Reserve Level* (Mécanisme d'ajustement Anglais)
CRE	Commission de Régulation de l'Energie (France, cf. www.cre.fr)
CREG	Commission de Régulation de l'Electricité et le Gaz (Belgique, cf. www.creg.be)
Elexon	Entreprise chargée du règlement financier du mécanisme d'ajustement Anglais (*Balancing Mechanism*, cf. www.elexon.com)
ELIA	Gestionnaire du réseau de transport en Belgique (cf. www.elia.be)
ELSPOT	Bourse d'électricité J-1 à NORDPOOL (cf. www.nordpool.com).
ERCOT	Gestionnaire du réseau de transport correspondant à l'état du Texas aux Etats-Unis (cf. www.ercot.com).
GRT	Gestionnaire du réseau de transport
IC	*Incentive Component* (Mécanisme d'ajustement aux Pays-Bas)
ISO	*Independent System Operator*
MidWest Pool	Architecture de marché d'électricité correspondant à plusieurs états du centre-ouest des Etats-Unis (Wisconsin, Minnesota, Missouri, Michigan, etc. - cf. www.midwestiso.org)
MWh	Mégawatt heure
NE Pool	Architecture de marché d'électricité correspondant à l'état de *New England* au nord-est des Etats-Unis (cf. www.iso-ne.com)
NETA	Architecture de marché d'électricité correspondant à l'Angleterre après 2001 (*New electricity trading arrangements*)
NGC	*National Grid Company* (Gestionnaire du réseau de transport en Angleterre, cf. www.ngc.com)
NORDPOOL	Architecture de marché d'électricité correspondant aux pays scandinaves (Norvège, Danemark, Finlande, Suède, cf. www.nordpool.com)
NY Pool	Architecture de marché d'électricité correspondant à l'état de *New York* au nord-est des Etats-Unis (cf. www.nyiso.com)
OFGEM	*The Office of Electricity and Gas Markets* (Régulateur Anglais, cf. www.ofgem.gov.uk)
OMEL	Architecture de marché d'électricité correspondant à l'Espagne (cf. www.omel.es)
PeN	Prix des écarts négatifs
PeP	Prix des écarts positifs
PJM	Architecture de marché d'électricité correspondant aux états de *Pennsylvania*, *New Jersey*, *Maryland*, etc. au nord-est des Etats-Unis (cf. www.pjm.com)
PMargB	Prix marginal des offres acceptées à la baisse
PMargH	Prix marginal des offres acceptées à la hausse

PMB	Prix moyen des offres acceptées à la baisse
PMH	Prix moyen des offres acceptées à la hausse
PowerNext	Bourse d'électricité J-1 en France (cf. www.powernext.fr)
PX	*Power Exchange* (Bourse d'électricité)
RTE	Gestionnaire du réseau de transport en France (www.rte-france.net)
RTO	*Regional Transmission Organization*
SBP	*System Buying Price* (Prix des écarts négatifs, Mécanisme d'ajustement Anglais)
SSP	*System Selling Price* (Prix des écarts positifs, Mécanisme d'ajustement Anglais)
Tennet	Gestionnaire du réseau de transport au Pays-Bas (www.tennet.nl)

Annexe N°2 : Analyse de sensibilité : modèle de base

Nous étudions ici la sensibilité des résultats aux changements des paramètres. Dans ce but, nous prenons le cas de base avec des pénalités fortes (k=1,4) et une *gate closure* loin du temps réel.

Nous réalisons des simulations en faisant varier les paramètres d'aversion au risque (A_R et A_{FG}), le prix fixe payé par les consommateurs (P^C), et enfin l'asymétrie de la distribution de probabilité de la demande (p) et le coefficient de cout marginal (σ_{FG}) Ces paramètres sont modifiés de +10% et -10%. Nous comparons alors les nouveaux résultats avec les résultats du cas de base. [245]

Nous remarquons que, dans toutes les simulations réalisées avec les nouveaux paramètres, le « signe » ou la direction des différents effets d'une pénalité n'ont pas été modifié. En revanche, le poids ou la valeur absolue des résultats dépendent des paramètres. Le tableau A-1 résume les résultats de l'étude de sensibilité. Chaque ligne de ce tableau correspond aux résultats développés précédemment, à savoir : la distorsion du prix *forward* (différence entre le prix *forward* avec pénalité et sans pénalité) ; la diminution de l'utilité des revendeurs (différence entre l'utilité avec et sans pénalité) ; l'augmentation de l'utilité des producteurs (différence entre l'utilité avec et sans pénalité), l'augmentation du profit du GRT (différence du profit avec et sans pénalité) et enfin la diminution de l'utilité totale (différence entre l'utilité totale avec et sans pénalité) qui est une forme d'inefficacité. Les différentes colonnes du tableau représentent le changement des résultats par des flèches.

Tableau A-1 : Sensibilité des résultats aux changements des paramètres.[246]

	A		P^C		p		σ_{FG}	
	(-10%)	(+10%)	(-10%)	(+10%)	(-10%)	(+10%)	(-10%)	(+10%)
Distorsion Prix *Forward*	↓	↑	↓	↑	↓	↑	↓	↑
Dim. Ut. Revendeurs	↓	↑	↑	↓	↓	↑	↓	↑
Aug. Ut. Producteurs	↓	↑	↑	↓	↓	↑	↓	↑
Aug. Profit GRT	↓	↑	↑	↓	↓	↑	↓	↑
Inefficacités	↓	↑	↑	↓	↑	↑	↑	↓

Nous remarquons plusieurs résultats généraux. Les effets d'une pénalité augmentent quand l'aversion au risque des participants, A , est plus grande et ils diminuent quand l'aversion est plus petite. L'asymétrie de la fonction de distribution de la demande (qui est

[245] Les tableaux des résultats pour chacun de cas sont mis à la fin de l'annexe.

[246] A coefficient d'aversion au risque ; P^C prix fixe des consommateurs ; p asymetrie distribution de probabilité de la demande ; σ_{FG} coefficient du coût marginal de production.

aussi une mesure de la courbure ou convexité de la fonction d'offre en temps réel) affecte de la même manière les effets d'une pénalité. Plus la distribution est asymétrique, plus les effets d'une pénalité sont importants. Le paramètre du prix fixe payé par les consommateurs (P^C) a une influence sur les effets d'une pénalité. Plus P^C est grand est plus la distorsion du prix *forward* sera plus grande. En revanche le transfert d'utilité entre les producteurs et le revendeur sera proportionnellement moins important.

Tableaux des résultats de l'analyse de sensibilité

Tableau A-2 : Résultats cas de base

Cas de base		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	28,9	33,8
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	70,5	77,3
	Revendeur type B	37,1	44,0
	Producteurs flexibles	107,6	121,2
Profit espéré	Revendeur type A	350,2	-189,8
	Revendeur type B	147,2	-228,8
	Producteurs flexibles	1442,7	2024,3
Utilité espérée	Revendeur type A	311,5	-241,9
	Revendeur type B	108,5	-280,9
	Producteurs flexibles	1423,0	1985,9
Profit espéré GRT		0,0	334,2
Coût et efficacité	Utilité espérée totale	1843,0	1740,2
	Efficacité	100,00%	94,42%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-3 : sensibilité au coefficient d'aversion au risque (-10%)

A -10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	29,0	33,4
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	70,5	77,2
	Revendeur type B	37,1	43,9
	Producteurs flexibles	107,6	121,1
Profit espéré	Revendeur type A	342,5	-155,8
	Revendeur type B	143,1	-209,2
	Producteurs flexibles	1454,4	1971,2
Utilité espérée	Revendeur type A	307,7	-203,1
	Revendeur type B	108,3	-256,6
	Producteurs flexibles	1436,7	1937,2
Profit espéré GRT		0,0	333,7
Coût et efficacité	Utilité espérée totale	1852,7	1752,9
	Efficacité	100,00%	94,61%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-4 : sensibilité au coefficient d'aversion au risque (+10%)

A +10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	28,8	34,3
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	70,5	77,4
	Revendeur type B	37,1	44,0
	Producteurs flexibles	107,6	121,4
Profit espéré	Revendeur type A	357,9	-223,7
	Revendeur type B	151,2	-248,3
	Producteurs flexibles	1430,9	2077,4
Utilité espérée	Revendeur type A	315,3	-280,6
	Revendeur type B	108,7	-305,1
	Producteurs flexibles	1409,3	2034,7
Profit espéré GRT		0,0	334,7
Coût et efficacité	Utilité espérée totale	1833,3	1727,2
	Efficacité	100,00%	94,21%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-5 : sensibilité au prix fixe payé par le consommateur (PC-10%)

PC -10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	31,0	35,9
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	73,4	80,2
	Revendeur type B	40,0	46,8
	Producteurs flexibles	113,4	127,0
Profit espéré	Revendeur type A	-34,0	-608,6
	Revendeur type B	-50,4	-462,3
	Producteurs flexibles	1674,5	2307,8
Utilité espérée	Revendeur type A	-94,4	-669,0
	Revendeur type B	-110,8	-522,7
	Producteurs flexibles	1655,0	2241,7
Profit espéré GRT		0,0	353,1
Coût et efficacité	Utilité espérée totale	1449,9	1377,5
	Efficacité	100,00%	95,01%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-6 : sensibilité au prix fixe payé par le consommateur (PC+10%)

PC +10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	26,8	31,8
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	67,5	74,4
	Revendeur type B	34,2	41,1
	Producteurs flexibles	101,7	115,5
Profit espéré	Revendeur type A	722,2	217,3
	Revendeur type B	332,5	-7,1
	Producteurs flexibles	1235,4	1764,4
Utilité espérée	Revendeur type A	674,5	168,9
	Revendeur type B	284,8	-55,5
	Producteurs flexibles	1203,2	1742,0
Profit espéré GRT		0,0	315,4
Coût et efficacité	Utilité espérée totale	2162,6	2068,9
	Efficacité	100,00%	95,67%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-7 : sensibilité à l'asymétrie de la distribution de demande (p-10%)

p -10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	28,2	32,4
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	67,7	73,4
	Revendeur type B	34,4	40,1
	Producteurs flexibles	102,1	113,5
Profit espéré	Revendeur type A	392,0	-54,9
	Revendeur type B	166,9	-142,2
	Producteurs flexibles	1381,1	1830,4
Utilité espérée	Revendeur type A	353,1	-96,9
	Revendeur type B	128,1	-184,2
	Producteurs flexibles	1358,8	1804,6
Profit espéré GRT		0,0	306,6
Coût et efficacité	Utilité espérée totale	1840,0	1551,7
	Efficacité	100,00%	84,33%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-8 : sensibilité à l'asymétrie de la distribution de demande (p+10%)

p +10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	30,0	36,3
	E[Prix temps réel]	30,0	30,0
Quantité *Forward*	Revendeur type A	74,9	83,7
	Revendeur type B	41,6	50,3
	Producteurs flexibles	116,5	134,0
Profit espéré	Revendeur type A	275,1	-452,9
	Revendeur type B	107,6	-410,4
	Producteurs flexibles	1557,3	2400,9
Utilité espérée	Revendeur type A	214,8	-541,0
	Revendeur type B	47,4	-498,5
	Producteurs flexibles	1539,3	2328,2
Profit espéré GRT		0,0	402,4
Coût et efficacité	Utilité espérée totale	1801,5	1619,8
	Efficacité	100,00%	89,91%
	Coût de production	1560,0	1560,0
	Efficacité productive	100,00%	100,00%

Tableau A-9 : sensibilité au coefficient de coût marginal (Sigma-10%)

Sigma -10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	24,2	28,2
	E[Prix temps réel]	27,0	27,0
Quantité *Forward*	Revendeur type A	67,2	74,0
	Revendeur type B	33,9	40,7
	Producteurs flexibles	101,1	114,7
Profit espéré	Revendeur type A	665,8	249,6
	Revendeur type B	306,6	23,0
	Producteurs flexibles	1123,6	1541,9
Utilité espérée	Revendeur type A	624,8	210,1
	Revendeur type B	265,7	-16,5
	Producteurs flexibles	1095,8	1524,9
Profit espéré GRT		0,0	281,5
Coût et efficacité	Utilité espérée totale	1986,4	1911,7
	Efficacité	100,00%	96,24%
	Coût de production	1404,0	1404,0
	Efficacité productive	100,00%	100,00%

Tableau A-10 : sensibilité au coefficient de coût marginal (Sigma+10%)

Sigma +10%		Benchmark	Cas 2
		Sans pénalité	Pénalité forte
Prix	Prix *forward*	34,0	40,0
	E[Prix temps réel]	33,0	33,0
Quantité *Forward*	Revendeur type A	73,1	80,0
	Revendeur type B	39,8	46,6
	Producteurs flexibles	112,9	126,6
Profit espéré	Revendeur type A	-5,1	-682,3
	Revendeur type B	-38,7	-517,1
	Producteurs flexibles	1827,8	2596,4
Utilité espérée	Revendeur type A	-74,2	-753,5
	Revendeur type B	-107,8	-588,3
	Producteurs flexibles	1804,9	2519,1
Profit espéré GRT		0,0	387,0
Coût et efficacité	Utilité espérée totale	1622,8	1531,2
	Efficacité	100,00%	94,35%
	Coût de production	1716,0	1716,0
	Efficacité productive	100,00%	100,00%

Bibliographie

Abou Chacra [2005]. Abou Chacra F., « Valorisation et optimisation du stockage d'énergie dans un réseau d'énergie électrique », Thèse de Doctorat en Génie Electrique, Université Paris XI, Orsay, 2005.

Ajodhia-Hakvoort [2005]. Ajodhia V., Hakvoort R., "Economic Regulation of quality in electricity distribution networks", *Utilities Policy*, 2005, Vol. 13, Issue 3, pp. 211-221.

Aliprantis *et al.* [1990]. Aliprantis C. D., Brown D.J. et Burkinshaw O., Existence and Optimality of Competitive Equilibria, Heidelberg and New York: Springer and Verlag, 1990.

Allaz-Vila [1993]. Allaz, B. et Vila J-L, 1993, "Cournot competition, forward markets and efficiency", *Journal of Economic Theory*, 59, 1993.

Arrow-Debreu [1954]. Arrow K.J., Debreu G., "Existence of an equilibrium for a competitive economy", *Econometrica*, 1954.

Averch-Jonhnson [1962]. Averch H., Jonhnson L., "Behavior of the firm under regulatory constraint ", *American Economic Review*, Vol. 52, n°5, pp. 1052-1069.

Bailey-Eaton [2001]. Bailey, M. and C. Eaton, "Moving Toward Seamless Energy Markets: Evidence from the Northeast." *The Electricity Journal*, v. 14, #6, Julliet 2001.

Baldick [2003a]. Baldick R., "Shift factors in ERCOT congestion pricing Ross Baldick", Working paper, University of Texas (Austin), disponible sur http://www.ece.utexas.edu/~baldick/papers/shiftfactors.pdf , Mars 2003.

Baldick [2003b]. Baldick R., "Variation of Distribution Factors with Loading," *IEEE Transactions on Power Systems*, 18(4):1316–1323, November 2003.

Baldick *et al.* [2005]. Baldick R., Helman U., Hobbs B., O'Neill R., "Design of Efficient Generetion Markets", Proceedings of the IEEE, Vol. 93, N° 11, Novembre 2005;

Baldick-Niu [2005]. Baldick R., Niu H., "Lessons Learned: The Texas Experience," dans Electricity Deregulation: Where to from here? J. Griffin and S. Puller (Eds), The University of Chicago Press, 2005.

Baldwin-Clark [2000]. Baldwin C. et Clark K. Design Rules, Volume 1, The Power of Modularity, MIT Press, Cambridge MA, 2000.

Barquin *et al.* [2004]. Barquin, J., Boots, M.G., Ehrenmann, A., Hobbs, B.F., Neuhoff, K. and Rijkers, F.A.M., "Network-constrained models of liberalized electricity markets: the devil is in the details", CMI EP 32 Working Paper, disponible sur www.econ.cam.ac.uk/electricity/, 2004.

Bastard [1998]. Bastard P., « Fonctionnement et réglage des systèmes de transport et de distribution d'énergie électrique » – Cours Supélec, 1998;

Belmans [2004]. Belmans R., « La liberalización del mercado de la electricidad a examen », *Energia*, n° 178, May-June, 2004; pp. 51-55.

Benintendi-Boccard [2003]. Benintendi D., Boccard N., "An Efficient Mechanism for Cross-Border Relief", Présenté à la Conférence IDEI, Toulouse, 2003.

Benitez-Crampes [2003]. Benitez D., Crampes C., "L'allocation des capacités de transport pour l'électricité", Rapport réalisé pour la Commission de Régulation de l'Energie, Gremaq-CNRS, juillet 2003.

Bergman [2005]. L. Bergman, "Why has the Nordic electricity market worked so well?", Rapport Elforsk, disponible sur www.elforsk-marketdesign.net , juin 2005.

Bessembinder-Lemmon [2000]. Bessembinder, H. and Lemmon, M., "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets", *Journal of Finance*, Vol. 57, pp. 1347-1382, 2002.

Bjørndal-Jörnsten [2001]. Bjørndal M. et Jörnsten K., "Zonal pricing in a deregulated electricity market", *The Energy Journal*, 22(1), 51-73, 2001;

Bjørndal-Jörnsten [2004a]. Bjørndal M. et Jörnsten K., "Allocation of Resources in the Presence of Indivisibilities: Scarf's Problem Revisited." Discussion Paper, Department of Finance and Management Science, Norwegian School of Economics and Business Administration (NHH), 2004.

Bjørndal-Jörnsten [2004b]. Bjørndal M. et Jörnsten K., "Equilibrium Prices Supported by Dual Price Functions in Markets with Non-Convexities" Discussion Paper, Department of Finance and Management Science, Norwegian School of Economics and Business Administration (NHH), 2004.

Boisseleau [2004]. Boisseleau F., "The role of power exchanges for the creation of a single European electricity market: market design and market regulation", Thèse de doctorat, Université Paris IX et Université Delft, Janvier 2004.

Bolle [1992]. Bolle F., " Supply Function Equilibria and the Danger of Tacit Collusion : The Case of Spot Markets for Electricity ", *Energy Economics*, Vol.15, pp. 94-102, 1992.

Boogert-Dupont [2005]. Boogert A., Dupont D., "On the Effectiveness of the Anti-Gaming Policy Between the Day-Ahead and Real-Time Electricity Markets in the Netherlands", *Forthcoming in Energy Economics*, Mars 2005.

Borenstein *et al.* [1999]. Borenstein S., Bushnell J., Wolak, "Diagnosing Market Power in California's Deregulated Wholesale Electricity Market", Working Paper CPC99-007, F. Institute of Business and Economic Research Competition Policy Center (University of California, Berkeley), 1999.

Borenstein *et al.* [2000]. Borenstein S., Bushnell J., Stoft S., "The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry", *The RAND Journal of Economics*, Vol. 31, No. 2, pp. 294-325, 2000.

Borenstein *et al.* [2004]. Borenstein S., Bushnell J., Knittel Ch., Wolfram C., "Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets", CSEM WP 138, Decembre 2004.

Boucher *et al.* [2006]. Boucher J., Daxhelet O., Smeers Y., "An alternative benchmark of the Internal Electricity Market", Journées de la section électrotechnique du club EEA, Supélec, Mars 2006.

Boucher-Smeers [1999]. J. Boucher and Y. Smeers, "Alternative Models of Restructured Electricity Systems, Part 1 No Market Power", CORE working paper, 1999.

Boucher-Smeers [2002]. Boucher, J. and Smeers, Y., "Towards a common European Electricity Market – Path in the right direction…still far from an effective design". *Journal of Network Industries* 3(4), 375-424, 2002.

Boucher-Smeers [2003], Boucher J. & Smeers Y., "The European Regulation on Cross Border Trade : can one do without a stardard market design ?", Working Paper, Université Catholique de Louvain, Mars 2003.

Cadwalader *et al.* [1998]. Cadwalader M., Harvey S., Pope S., Hogan W. "Market Coordination of Transmission Loading Relief Across Multiple Regions", Working Paper disponible sur http://ksghome.harvard.edu/~whogan/extr1298.pdf, 1998.

Cadwalader *et al.* [1999]. Cadwalader M., Harvey S., Pope S., Hogan W., « Coordinating congestion relief across multiple regions », Working Paper disponible sur http://ksghome.harvard.edu/~whogan/isoc1099r.pdf, 1999.

Cameron-Cramton [1999]. Cameron L., Cramton P., "The Role of the ISO in U.S. Electricity Markets: A Review of Restructuring in California and PJM", *Electricity Journal*, 71-81, Avril 1999.

Cardell *et al.* [1997]. Cardell J., Hitt C., et Hogan W.W., "Market Power and Strategic Interaction in Electricity Networks," *Resource and Energy Economics*, 19(1-2), 1997, 109-137.

CE [1996]. Commission Européenne (CE), "Directive 96/92 concernant les règles communes pour le marché intérieur de l'électricité", 1996.

CE [2003]. Commission Européenne (CE), "Directive 2003/54/EC of the European Parliament and of the Council concerning common rules for the internal market in electricity and repealing Directive 96/92/EC.", 2003.

CE [2004]. Commission Européenne (CE), "Third benchmarking report on the implementation of the internal electricity and gas market", Rapport CE, Bruxelles, 2004.

CE [2005]. Commission Européenne (CE), "Report on progress in creating the internal gas and electricity market", Rapport CE, Novembre 2005.

Chance [2002]. Chance D., "Risk Neutral Pricing in Discrete Time", Teaching note 96-02, Louisiana State University, Novembre 2002.

Chandley-Hogan [2006]. Chandley J., Hogan W., "A Path to Preventing Undue Discrimination and Preference in Transmission Services", Working Paper disponible sur http://ksghome.harvard.edu/~whogan , Août 2006.

Chao *et al.* [2000]. Chao H., Peck S., Oren S., Wilson R., "Flow-based transmission rights and congestion management". *The Electricity Journal*, octobre, pp.38-58.

Chao *et al.* [2005]. Chao H., Oren, S.S., Papalexopoulos, A., Sobajic, D.J. et Wilson, R., "Interface between engineering and market operations in restructured electricity systems", *Proceedings of the IEEE*, Volume: 93, Issue: 11, 1984-1997, Nov. 2005.

Chao-Peck [1996]. Chao H., Peck S., "A Market Mechanism for Electric Power Transmission". *Journal of Regulatory Economics*, n°10, pp.25-59.

Chao-Wilson [1999]. Chao H., Wilson R., "Multi-Dimensional Procurement Auction for Power Reserves: Incentive-Compatibility Evaluation and Settlement Rules", Presented at the UCEI Power Conference, Mars 1999.

Coase [1937]. Coase R. H., "The Nature of the Firm", *Economica*, 1937.

Coase [1988]. Coase R. H., - The Firm, the Market and the Law -, Chicago, IL: Univ. Chicago Press, 1988, pp.1-31.

Contreras *et al.* [2001]. Contreras J. Candiles O. De la Fuente J.I. Gomez T., "Auction Design in Day-ahead Electricity Markets", *IEEE Transactions on power Systems*, Vol. 16 N°1, février 2001.

Cornwall [2002]. Cornwall N., "The Need for Balance in NETA imbalance Pricing", *Power UK 96*, Février 2002.

Costello [2001]. Costello K., "Interregional Coordination versus RTO Mergers : A cost-benefit perspective", *The Electricity Journal*, Mars 2001.

Counsell-Evans [2003]. Counsell K., Evans L., "Day-Ahead Electricity Markets: Is There a Place for a Day-Ahead Market in the NZEM?", Working Paper, NZ Institute for the Study of Competition and Regulation, 2003.

Cramton *et al.* [2005]. Cramton P., Shoham Y., et Steinberg R. (eds.), Combinatorial Auctions, MIT Press, 2005.

Cramton-Stoft [2006a]. Cramton P., Stoft S., "The Convergence of Market Designs for Adequate Generating Capacity", Working Paper WP-2006-007, University of Maryland, Avril 2006.

Cramton-Stoft [2006b]. Cramton P., Stoft S., "Uniform-Price Auctions in Electricity Markets", forthcoming The Electricity Journal, 2006.

CRE [2004a]. Commission de Régulation de l'Energie (CRE), France, « Délibération relative à l'approbation des règles du mécanisme d'ajustement », disponible sur www.cre.fr, Juin 2004.

CRE [2004b]. Commission de Régulation de l'Energie (CRE), France, « Rapport d'activité 2003/2004 », disponible sur www.cre.fr, Juin 2004.

CRE [2005a]. Commission de Régulation de l'Energie (CRE), France, « Délibération du 17 mars 2005 relative à l'approbation des règles du mécanisme d'ajustement », disponible sur www.cre.fr, Mars 2005.

CRE [2005b]. Commission de Régulation de l'Energie (CRE), France, « Rapport d'activité 2004/2005 », disponible sur www.cre.fr, Juin 2005.

CRE [2006a]. Commission de Régulation de l'Energie (CRE), France, « Communication de la Commission de régulation de l'énergie du mars 2006 relative à l'amélioration du mécanisme d'ajustement », disponible sur www.cre.fr, Mars 2006.

CRE [2006b]. Commission de Régulation de l'Energie (CRE), France, « Délibération du 22 juin 2006 relative à l'approbation des règles relatives à la programmation, au mécanisme d'ajustement, au recouvrement des charges d'ajustement », disponible sur www.cre.fr, Juin 2006.

CRE [2006c]. Commission de Régulation de l'Energie (CRE), France, « Rapport d'activité 2005/2006 », disponible sur www.cre.fr, Juin 2006.

CRE-AEEG [2004] Commission de Régulation de l'Energie (CRE) France et Autorità per l'energia elettrica e il gas (AEEG), « Report on the Events of September, 28th, 2003, culminating in the separation of the Italian power system from the other UCTE networks", rapport collectif publié sur les sites de l'AEEG et de la CRE (www.cre.fr) Avril 2004.

CREG [2005]. Commission de Régulation de l'Electricité et du Gaz (CREG) Belgique "Rapport Annuel 2004 à la Commission Européenne", Rapport, Juillet 2005.

Curien [2000]. Curien N., Economie des réseaux. Paris : La Découverte - Repères 293, 2000.

Demsetz [1968]. Demsetz H., "Why Regulate Utilities", *Journal of Law and Economics*, vol. 11, pp 50-65.

ECE [2006]. Department of Electrical and Computer Engineering (ECE), "Modelling of Cost Rate Curves", Cours "Economic systems for electric power planning", Iowa State University, disponible sur www.ee.iastate.edu/~jdm/ee458/CostCurves.doc, 2006.

EEI [2001]. Edison Electric Institute (EEI), "Power Market Auction Design: Rules and Lessons In Market-based Control for the New Electricity Industry", Août 2001.

Ehrenmann-Smeers [2005]. Ehrenmann A., Smeers Y., "Inefficiencies in European Congestion Management Proposals", *Utilities policy*, 13(2), 135-152, 2005.

Ehrenmann-Smeers [2006]. Ehrenmann A., Smeers Y., "Critical Assessment of Congestion Management Solutions: Efficiency and Market Power", présenté au workshop "Institutional setting for the IEM: The Role and Status of Power Exchanges", Florence School of Regulation, Janvier 2006.

Elexon [2006a]. Elexon, "Market Index Definition Statement for Market Index Data Provider(s)", disponible sur www.elexon.com, Avril 2006.

Elexon [2006b]. Elexon, "Review of credit assessment price (CAP)", Rapport Elexon disponible sur www.elexon.com, Juillet 2006.

Elexon [2006c]. Elexon, "Trading Operation Report", Rapport Elexon, disponible sur www.elexon.com, Septembre 2006.

ELIA [2005]. ELIA, Gestionnaire du réseau Belge, "Tarif 2005 pour compensation des déséquilibres quat-horaires", disponible sur www.elia.be, 2005.

ELIA [2006]. ELIA, Gestionnaire du réseau Belge, "Tarifs de déséquilibre", disponible sur www.elia.be, 2006.

Elmaghraby *et al.* [2004]. Elmaghraby W. J., O'Neill R., Rothkopf M. et Stewart W., "Pricing and Efficiency in Lumpy Energy Markets," *The Electricity Journal*, vol. 17, no. 5, June 2004, pp. 54-64.

ERGEG [2006]. European Regulators Group for Electricity and Gas (ERGEG), "Guidelines of Good Practice for Electricity Balancing Markets Integration", disponible sur www.ergeg.org, juin 2006.

ETSO [2001]. European Transmission System Operators (ETSO), "Co-ordinated auctioning : a market-based method for transmission capacity allocation in meshed networks", disponible sur www.etso-net.org, Avril 2001.

ETSO [2003]. European Transmission System Operators (ETSO), "Current State of Balance Management in Europe", disponible sur www.etso-net.org, 2003.

ETSO [2004a]. European Transmission System Operators (ETSO), "Cross-border electricity exchanges on meshed AC power systems", www.etso-net.org, Avril 2004.

ETSO [2004b]. European Transmission System Operators (ETSO), "An Overview of Current Cross-border Congestion Management Methods in Europe", www.etso-net.org, septembre 2004.

ETSO [2005]. European Transmission System Operators (ETSO), "Current State of Trading Tertiary Reserves Across Borders in Europe", disponible sur www.etso-net.org, Novembre 2005.

ETSO [2006a]. European Transmission System Operators (ETSO), "Key Issues in Facilitating Cross-Border Trading of Tertiary Reserves and Energy Balancing", disponible sur www.etso-net.org, Mai 2006.

ETSO [2006b]. European Transmission System Operators (ETSO), "Current State of Balance Management in South East Europe", disponible sur www.etso-net.org, Juin 2006.

ETSO [2006c]. European Transmission System Operators (ETSO), "ETSO Comments on ERGEG Guidelines of Good Practice for Electricity Balancing Market Integration", disponible sur www.etso-net.org, Juillet 2006.

ETSO [2006d]. European Transmission System Operators (ETSO), "An Overview of Current Cross-border Congestion Management Methods in Europe", disponible sur www.etso-net.org, Mai 2006.

ETSO [2007]. European Transmission System Operators (ETSO), "Balance Management Harmonisation and Integration 4th Report", disponible sur www.etso-net.org, Janvier 2007.

ETSO-EuroPEX [2004]. European Transmission System Operators (ETSO) et Association of European Power Exchanges (Euro-PEX), "Flow-based Market Coupling : A joint ETSO-EuroPEX proposal for cross-border congestion management and integration of electricity maketsmarkets in Europe", Interim Report, disponible sur www.etso-net.org, Septembre 2004.

Euro-PEX [2003]. Association of European Power Exchanges (Euro-PEX), "Using implicit auctions to manage cross-border congestion: decentralised market coupling", disponible sur www.europex.org, Juillet 2003.

FERC [1999]. Federal Energy Regulatory Commission (FERC), "Regional Transmission Organizations". ORDER n°.2000, Final Rule, Docket No.RM99-2-000.

FERC [2002]. Federal Energy Regulatory Commission (FERC), "Notice of proposed rulemaking on open transmission service and standardized market design". Docket No.RM01-12-000.

Finon [2001]. Finon D., « L'intégration des marchés électriques européens : de la juxtaposition de marchés nationaux à l'établissement d'un marche régional », *Economies et Sociétés*, Série Economie de l'Energie, n. 8, janv.-fév. 2001, p. 55-87.

Frontier-Consentec [2005]. Frontier Economics, Consentec, "Benefits and practical steps towards the integration of intraday electricity markets and balancing mechanisms", Rapport préparé pour la Commission Européenne, disponible sur http://europa.eu.int/comm/energy/electricity/publications/index_en.htm , Décembre 2005.

Glachant [2002]. Glachant, J.-M., "Why Regulate Deregulated Network Industries ?", *Journal of Network Industries* 3: 297-311, 2002.

Glachant [2007]. Glachant, J-M., "Brown Out or Black Out in Europe in November 2006: to entertain ... with EU Colours Out?", Presentation, Conference "The Economics of Energy Markets", Toulouse, France, Janvier 2007.

Glachant *et al.* [2005]. Glachant, J.-M., Perez Y., Pignon V. et Saguan M. « Un marché européen de l'électricité ou des marchés dans l'Europe ? Regards croisés : Economistes et ingénieur », Droit de la régulation et l'Economie, Annales de l'UMR André Tunc - Université de Paris I, 2005.

Glachant-Finon [2003]. Glachant, J.M. et Finon D. (eds.), Competition in European Electricity Markets: A Cross-Country Comparison, Edward Elgar, 2003.

Glachant-Lévêque [2005]. Glachant, J.M. and Lévêque, F., "Electricity Single Market in the European Union: What to do next?" Discussion paper of the European Union research project SESSA, (http://www.sessa.eu.com/ documents/bruxellesp/SESSA_report_wp1.pdf). Working Paper CEEPR 2005-15 at MIT (web.mit.edu/ceepr), Sept 2005.

Glachant-Pignon [2005]. Glachant J.-M., Pignon V., "Nordic Congestion's Arrangement as a Model for Europe? Physical constraints and Economic Incentives", *Utilities Policy*, 2005.

Glachant-Saguan [2006]. Glachant J.-M., Saguan M., « Le commerce européen de l'électricité dans le respect des contraintes du système électrique : pour un dialogue économistes – ingénieurs », Revue E tijdschrift, Société Royale Belge des Electriciens – n° 3, 2006.

Gonen [1986]. Gonen T., Electric Power Distribution System Engineering, McGraw-Hill.

Green [1994]. Green R., "Britain's Unregulated Electricity Pool", dans *From Regulation to Competition : New Frontiers in Electricity Markets*, Einhorn, M.A. Ed, chap.4, pp. 73-97, Klewer Academic Publishers, Boston.

Green [2006]. Green R., "Investment and generation capacity", dans F. Lévêque, *Competitive Electricity Market and Sustainability*, Edward Elgar, 2006.

Green *et al.* [2005]. Green R., Lorenzoni A., Perez Y. et Pollitt M., "Benchmarking Electricity in the EU", ADIS WORKING PAPER n°2005-01.

Green-McDaniel [1999]. Green, R. et McDaniel, T.,"Modelling Reta: A model of forward trading and the balancing mechanism", Working paper, 1999.

Green-Newbery [1992]. Green, R. et Newbery, D., " Competition in the British Electricity Spot Market ", *Journal of Political Economy*, Vol. 100, N°5, pp. 929-953, 1992.

Hadsell-Shawky [2006]. Hadsell L., Shawky H., "Electricity Price Volatility and the Marginal Cost of Congestion: An Empirical Study of Peak Hours on the NYISO Market, 2001-2004", *The Energy Journal*, Vol 27 N°2, 2006.

Hartshorn-Harvey [2001]. Hartshorn A., Harvey S. "Assessing the short-run benefits from a combined northeast market", LECG, LLC, Octobre, 2001.

Harvey [2003]. Harvey S., « The Virtues of Virtual RTOs », présentation, http://www.ksg.harvard.edu/hepg/RTO_ISO_Market_Reports.htm, septembre 2003.

Harvey-Hogan [2000a]. Harvey S., Hogan W., "Nodal and Zonal congestion management and the exercise of market power", http://ksghome.harvard.edu/~.whogan.cbg.Ksg/zonal_jan10.pdf, 2000.

Harvey-Hogan [2000b]. Harvey S., Hogan W., "Nodal and Zonal congestion management and the exercise of market power: Further comments", http://ksghome.harvard.edu/~.whogan.cbg.Ksg/zonal_Feb11.pdf, 2000.

Harvey-Pope [2001]. Harvey S., Pope S., "Feasibility Study for a Combined Day-Ahead Market in the Northeast", http://www.lecg.com, 2001.

Hennebel [2007]. Hennebel M., « Valorisation des services système sur un réseau de transport d'électricité en environnement concurrentiel », Thèse de doctorat en Génie Electrique, Université de Paris XI, Orsay, 2007.

Henney [2002]. Henney A, « An Independent Review of NETA », EEE Ltd, London, 2002.

Hiroux-Maupas [2006]. Hiroux C., Maupas F., "The management of wind power output under a feed-in tariff model: the role of TSO", 29th IAEE International Conference, Postdam / Allemagne, Juin 2006.

Hirschhausen *et al.* [2007]. Hirschhausen Ch., Weigt H., Zachmann G., "Price Formation and Market Power in Germany's Wholesale Electricity Markets", Working Paper WP-GE-15b, Dresden University of Technologie, Janvier 2007.

Hirschhausen-Zachmann [2005]. Hirschhausen Ch., Zachmann G., "Ensuring EU Enlargement to New Member States", Papier présenté à la conference "Implementing the Internal Market of Electricity: Proposals and Time-Tables", Bruxelles, Septembre 2005.

Hirshleifer-Subramanyam [1993]. Hirshleifer, D., A. Subramayam, "Futures versus Share Contracting as Means of Diversigying Output Risk", *The Economic Journal*, 103, 620-638.

Hirst [2001]. Hirst E.,"Real-Time Balancing Operations and Markets: Key to Competitive Wholesale Electricity Markets," Edison Electric Institute, Washington, DC, and Project for Sustainable FERC Energy Policy, Alexandria, VA, April 2001.

Hirst-Kirby [1998]. Hirst E. et Kirby B., "Technical and Market Issues for Operating Reserves", Consulting in Electric-Industry Restructuring Oak Ridge, Tennessee November 1998.

Hirst-Kirby [2003]. Hirst E. et Kirby B., "Allocating Costs of Ancillary Services: Contingency Reserves and Regulation", OAK RIDGE NATIONAL LABORATORY, ORNL/TM-2003/152, http://www.ornl.gov/sci/btc/apps/Restructuring/tm2003_152.pdf, juin 2003.

Hisman-Huurman [2004]. Huisman R., Huurman Ch., "Being in Balance : Economic Efficiency in the Dutch Power Market", Erasmus Research Institute of Management (ERIM) Rapport disponible sur www.erim.eur.nl, Juillet 2004.

Hobbs *et al.* [2001]. Hobbs, B., Rothkopf M., O'Neill R., Chao H., (eds.). The Next Generation of Unit Commitment Models. . Kluwer International Series in Operations Research & Management, 2001.

Hobbs-Rijkers [2004]. Hobbs, B., Rijkers F.A.M., "Strategic Generation With Conjectured Transmission Price Responses in a Mixed Transmission Pricing System – Part I: Formulation", *IEEE Transactions on Power Systems*, Vol. 19, No 2, Mai 2004.

Hogan [1992]. Hogan W.W., "Contract Networks for Electric Power Transmission". *Journal of Regulatory Economics*, n°4, pp.211-242.

Hogan [1998]. Hogan W., « Nodes and zones in electricity markets : seeking simplied congestion pricing », dans *Designing Competitive electricity markets*, Chao H. et Huntington H. (Eds.), Kluwer Academic Publisher, pp. 33-62, 1998.

Hogan [2000]. Hogan, W.W., "Flowgate Rights and Wrongs". Working Paper, Harvard Electricity Policy Group, 2000.

Hogan [2001]. Hogan, W.W., "Interregional coordination of electricity markets", presentation, Technical Conference on Interregional Coordination – FERC, juin 2001.

Hogan [2002a]. Hogan W. W., "Financial Transmission Rights Formulations", mimeo, Center for Business and Government, JFK School of Government, Harvard University, disponible sur http://www.whogan.com, 2002.

Hogan [2002b]. Hogan W.W., "Electricity Market Restructuring: Reforms of Reforms". *Journal of Regulatory Economics*, vol.21, n°1, pp.103-132, 2002.

Hogan [2006a]. Hogan W.W., "Reliability and scarcity pricing: operating reserves demand curves", Working Paper, Harvard Electricity Policy Group, Mars 2006.

Hogan [2006b]. Hogan W.W., "Revenue Sufficiency Guarantees and Cost Allocation", Mai 2006.

Hogan-Ring [2003]. Hogan W., Ring B., "On Minimum-Uplift Pricing for Electricity Markets." Working Paper, John F. Kennedy School of Government, Harvard University, 2003.

Hunt [2002]. Hunt S., Making Competition work in Electricity, John Wiley & Sons ed., Inc., New York, 2002.

ILEX [2002]. ILEX Energy Consulting, "Cost-reflectivity of imbalance charges", note disponible sur http://www.ilexenergy.com, 2002.

Joskow [2006a]. Joskow P. L., "Markets for Power in the United States: An Interim Assessment", *The Energy Journal* 27(1), 1-36, Janvier 2006.

Joskow [2006b]. Joskow P. L., "Competitive Electricity Markets and Investment in New Generating Capacity", WP-2006-009, Mai 2006.

Joskow [2006c]. Joskow P. L., "Designing Wholesale Electricity Markets", Conférence Market Power : What market design ?, CGEMP Université Paris Dauphine, disponible sur www.dauphine.fr/cgemp/, Mars 2006.

Joskow [2006d]. Joskow P. L., "Patterns of Transmission Investment (and related issues)", dans F. Lévêque, *Competitive Electricity Market and Sustainability*, Edward Elgar, 2006.

Joskow-Schmalensee [1983]. Joskow P. L., Schmalensee R., Markets for Power - An Analysis of Electrical Utility Deregulation, The MIT Press Ed, 1983.

Joskow-Tirole [2006]. Joskow and Tirole J., "Reliability and Competitive Electricity Markets (REVISED)", June 2006, forthcoming Rand Journal of Economics.

Kamat-Oren [2004]. Kamat, R. et Oren S., "Two Settlements Systems for Electricity Markets under Network Uncertainty and Market Power", *Journal of Regulatory Economics*, 25:1 5-37, 2004.

Karakatsani-Bunn [2005]. Karakatsani N., Bunn D., "Diurnal Reversal of Electricity Forward Premia", London Business School, Février 2005.

Kiesling-Giberson [2004]. Kiesling L., Giberson M., "Electric Network Reliability as a Public Good", Paper submitted to CMU conference: Electricity Transmission in Deregulated Markets, novembre 2004.

Kirby-Milligan [2005]. Milligan M., Kirby B. "Estimating the Ramping Capability of a Control Area or Balancing Authority and Implications for Moderate or High Wind Penetration", National Renewable Energy Laboratory, Mai 2005.

Klobasta [2006]. Klobasta M., Ragwitz M., Obersteiner C., "Strategies for an Efficient Integration of Wind Power Considering Demand Response", 29th IAEE International Conference, Potsdam – Allemagne, Juin 2006.

Knight [1921]. Knight F., Risk, uncertainty and profit, Houghton Mifflin Company published, 1921.

Krapels-Flemming [2005]. Krapels E.; Flemming P. – "Impacts of the PJM RTO Expansion"- A REPORT PREPARED FOR PJM BY ENERGY SECURITY ANALYSIS, INC., disponible sur www.pjm.com; 2005.

Kristiansen [2004]. Kristiansen T., "Markets for Financial Transmission Rights", Working Paper, Norwegian University of Science and Technology, Department of Electrical Power Engineering, 2004.

Laughlin [2006]. Laughlin K., "Regional Transmission Organizations", Présentation Harvard Electricity Policy Group, Juin 2006.

LE [2004]. London Economics (LE) "Structure and Functioning of the Electricity Market in Belgium in a European Perspective", Rapport, Octobre 2004.

Leuthold *et al.* [2005]. Leuthold F., Rumiantseva I., Weigt H., Jeske T., Hirschhausen Ch., "Nodal pricing in the German Electricity Sector – A Welfare Economics Analysis, with Particular Reference to Implementing Offshore Wind Capacities", Working Paper WP-GE-08a, Dresden University of Technologie, septembre 2005.

Lévêque [2004]. Lévêque F., Economie de la Réglementation, Editions La Découverte, Paris, 2004.

Longstaff-Wang [2004]. Longstaff F., Wang A., "Electricity Forward Prices: A High-Frequancy Empirical Analysis", *Journal of Finance,* Vol. 59, 1877-1900, 2004.

Markowitz [1952]. Markowitz, H., "Portfolio Selection", *Journal of Finance*, Vol. VII, No 1, Mars,1952.

Mas-Collel *et al.* [1995]. Mas-Colell A., Green J., Whinston M., Microeconomic Theory, 1995.

Ma-Sun [2005]. Ma X., Sun D., "Key Elements of a Successful Market Design", *Proceedings of 2005 IEEE/PES*, T.D. Conference, China, 2005.

Meeus [2006]. Meeus L., "Power exchange auction trading platform design", Thèse de doctorat, KATHOLIEKE UNIVERSITEIT LEUVEN, Juillet 2006.

Meeus *et al.* [2005]. Meeus L., Purchala K., Belmans R., « Development of the Internal Electricity Market in Europe", *The Electricity Journal*, Vol. 18, Issue 6, Juillet 2005.

Meeus *et al.* [2006]. Meeus L., Belmans R., Glachant J-M, « Regional electricity market integration France-Belgium-Netherlands », *Revue E tijdschrift*, Société Royale Belge des Electriciens – n° 3, 2006.

Menager [2002]. (eds) Bastard P., Fargue D., Laurier P., Mathieu B., Nicolas M., Roos P. Electricité : Voyage au cœur du système, Eyrolles 2002.

Menanteau-Finon [2003]. Menanteau R., Finon D., « L'intégration de la production intermittente dans les marches électriques libéralisés: des surcoûts techniques aux pénalités économiques imposées par les règles de fonctionnement », Cahier de recherche LEPII, Série EPE, n°32, Mars 2003.

MISO-PJM [2004]. Midwest ISO & PJM "Market to Market : Interregional Coordination Process", disponible sur www.miso-pjm.com, 2004.

Moreno [2005]. Moreno H., « Valorisation des services système dans un réseau de transport d'électricité », Thèse de doctorat en Génie Electrique, Université de Paris XI, Orsay, 2005.

MPSC *et al.* [2001]. Maryland Public Service Commission (MPSC), The District of Columbia Public Service Commission, et The Virginia State Corporation Commission, "Request For Rehearing in the Matter of PJM Interconnection, LLC", http://www.psc.state.md.us/psc/electric/rehearing.htm#5, 2001.

NEISO-NYISO [2003]. ISO New England, New York ISO, « Virtual Regional Dispatch : Concept, evaluation and proposal », Joint Working Paper, disponible sur www.ksg.harvard.edu/hepg/, 2003.

Neuhoff [2003]. Neuhoff, K., "Coupling Transmission and Energy Markets Reduces Market Power", 26th IAEE INTERNATIONAL CONFERENCE, PRAGUE, 2003.

Newbery [2002] Newbery D., "England's experience with NETA", présentation dans International Experience with Energy Liberalization: Lessons for Europe conference, Oviedo, Espagne, http://www.econ.cam.ac.uk/electricity, Julliet 2002.

Newbery [2005a]. Newbery D., "Refining Market Design", Discussion paper of the European Union research project SESSA, disponible sur www.sessa.eu.com, Septembre 2005.

Newbery [2005b]. Newbery D., "European Electricity Liberalisation", *The Energy Journal*, Special Issue 2005.

Newbery [2006]. Newbery D., "The Design of Wholesale Electricity Markets: Best Practice", MIT/Cambridge/Iberdrola Electricity Conference, Madrid, http://www.electricitypolicy.org.uk, Juin 2006.

Newbery-McDaniel [2003]. Newbery D., McDaniel T., "Auctions and trading in energy markets : an economic analysis", http://www.electricitypolicy.org.uk, 2003.

NORDEL [2003]. NORDEL, "Nordel Model for Balance Pricing and Settlement", disponible sur www.nordel.com, Mai 2003.

O'Neill *et al.* [2005]. O'Neill R., Sotkiewicz P., Hobbs B. F., Rothkopf M. H., Stewart W. R., "Efficient Market-Clearing Prices in Markets with Nonconvexities", *European journal of operational research*, vol. 164, no1, pp. 269-285, 2005.

Ofgem [2006]. The Office of Gas and Electricity Markets (Ofgem), "BSC modification proposal P194 'Revised Derivation of the Main Energy Imbalance Price' – Impact Assessment", Rapport, Janvier 2006.

Oren [1998]. Oren S., "Authority and Responsability of the ISO : Objectives, Options and tradeoffs", dans *Designing Competitive electricity markets*, Chao H. et Huntington H. (Eds.), Kluwer Academic Publisher, pp. 79-96, 1998.

Oren [2001a]. Oren S., "Design of Ancillary Service Markets", Hawaii International Conference On System Sciences, http://certs.lbl.gov/pdf/hawaii-01.pdf, janvier 2001.

Oren [2001b]. Oren S., "Efficient Intrazonal Transmission Pricing", Rapport, disponible sur www.puc.state.tx.us/electric/projects/23220/, 2001.

Oren [2001c]. Oren S., "Report to the public utility commission of Texas on the Ercot protocols", Rapport, www.puc.state.tx.us/electric/projects/23220/, février 2001.

Oren [2003a]. Oren S., "Ensuring Generation Adequacy in Competitive Electricity Markets", University of California Energy Institute, Working Paper EPE-007, Juin 2003.

Oren [2003b]. Oren S., "Market Design and Gaming in Competitive Electricity Systems", Présentation 23rd Arne Ryde Symposium "The Nordic Electricity Market", Octobre 2003.

Ott [2000]. Ott A., "Can Flowgates Really Work? An Analysis of Transmission the PJM Market from April 1, 1998 - April 30, 2000", PJM, http://www.ksg.harvard.edu/hepg/flowgate/flowgate.pdf, Septembre 2000.

Ott [2002]. Ott A., "Evaluation of the Increase in the Economic Efficiency of the Overall PJM Unit Commitment and Economic Dispatch Resulting from the Integration of Allegheny Power into PJM Energy Market", PJM, Décembre 2002.

Ott [2003]. Ott A., "PJM-MISO: Achieving a Virtual RTO Through a Joint Operating Agreement", présentation http://www.ksg.harvard.edu/hepg, septembre 2003.

Ott [2006]. Ott A., "The US Experience with LMP: PJM and Mid-West ISO Market Coordination", présentation http://www.ksg.harvard.edu/hepg, mai 2006.

Perez [2002]. Perez Y. « L'économie néo-institutionnelle des réformes électriques européennes », Thèse de doctorat en Sciences Economiques, Université Paris 1 Panthéon-Sorbonne, 2002.

Phulpin *et al.* [2005]. Phulpin Y., Hennebel M., Plumel S., « La traçabilité de l'électricité : une méthode équitable pour l'allocation des coûts de transmission », Electrotechnique du Futur EF'2005, Grenoble (France), Septembre 2005.

Pignon [2003]. Pignon V., « L'Harmonisation des Méthodes de Tarification du Transport d'Electricité, une Analyse Economique du Marché Unique Européen », Thèse de Doctorat en Sciences Economiques, Université Paris 1, Panthéon Sorbonne, 2003.

PJM [2004]. PJM, "MIP Based Unit Commitment", PJM MIC Meeting, disponible sur www.pjm.com/committees/mic/downloads/20040331-mip-based-unit-commitment.pdf, 2004.

PJM [2006]. PJM Interconnection, "Ancillary Service, PJM 101, the basics", Training material, disponible sur www.pjm.com, 2006.

PJM-MISO [2006]. The Joint Board for the PJM-MISO Region, "Report on Security Constrained Economic Dispatch", summit à la FERC, Docket No. AD05-13-000, Mai 2006.

Potomac [2004]. Potomac Economics, Ltd, "2003 State of the Market Report for the Ercot Wholesale Electricity Markets", disponible sur www.ercot.com, Juliet 2004.

Potomac [2006]. Potomac Economics, Ltd, "2005 State of the Market Report for the Ercot Wholesale Electricity Markets", disponible sur www.ercot.com, Juliet 2006.

Quiquempoix *et al.* [2004]. Quiquempoix O., Fliscounakis S., Bougarde E., « Prévision des pertes électriques sur le réseau THT et HT français », CIGRE, Avril 2004.

Raineri-Rudnick [1996]. Raineri R. et Rudnick H., « Analisis de normativas de calidad de servicio para empresas distribuidoras », Working Paper, Universidad Católica de Chile, Chile, 1996.

Rassenti-Smith [1998]. Rassenti S., Smith V. "Deregulating Electric Power : Market Design Issues and Experiments " dans *Designing Competitive Electricity Markets*, (Eds) Chao H., Huntington, H., pp.105-121, 1998.

Rolfo [1980]. Rolfo, J. « Optimal Hedging under Price and Quantity Uncertainty : The Case of a Cocoa Producer », *Journal of Political Economy*, 88, 100-116.

Roques *et al.* [2004]. Roques F., Newbery D., Nuttall W., « Generation Adequacy and Investment Incentives : from Pool to NETA ", Working Paper, http://www.electricitypolicy.org.uk, 2004.

Roques *et al.* [2005]. Roques F., Newbery D., Nuttall W., « Investment Incentives and Electricity Market Design : the British Experience", *Review of Network Economics*, Vol.4 Issue 2 – Juin 2005.

RTE [2003]. Gestionnaire du Réseau de Transport d'Electricité, France (RTE), « Mémento de la sûreté du système électrique » - http://www.rte-france.com, 2003.

RTE [2005]. Gestionnaire du Réseau de Transport d'Electricité, France (RTE), « Schéma de développement du réseau public de transport d'électricité », http://www.industrie.gouv.fr/energie/electric/rpte/sommaire.htm, 2005.

RTE [2006]. Gestionnaire du Réseau de Transport d'Electricité, France (RTE), « Règles relatives à la Programmation, au Mécanisme d'Ajustement et au dispositif de Responsable d'Equilibre », disponible sur www.rte-france.com, Juillet 2006.

Ruff [2001]. Ruff L., "Flowgates, Contingency-Constrained Dispatch, and Transmission Rights", *The Electricity Journal*, janvier – février, pp.34-55, 2001.

Salanié [2000]. Salanié B., Microeconomics of Market Failures, The MIT Press, 2000.

Saravia [2003]. Saravia C., « Speculative Trading and Market Performance : The effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market », CSEMWP 121, 2003.

Saussier [1997]. Saussier S., « Choix contractuels et Coûts de transaction : une analyse économique des contrats d'approvisionnement en charbon d'EDF », Thèse de doctorat en Sciences Economiques, Université Paris 1, Panthéon Sorbonne, 1997.

Schweppe *et al.* [1988]. Schweppe F.C, Caramanis M.C., Tabors R.D., Bohn R.E., Spot Pricing of Electricity, Kluwer Academic Publishers, 1988.

Shahidehpour *et al.* [2002]. Shahidehpour M., Yamin H., Li Z., Market Operation in Electric Power System, IEEE – Wiley Inter Science, 2002.

Shanker R. [2001]. Shanker R., "Is a Single Reasonable NE Market Feasible ? An Intial Analysis of Technology and Scalability ", Northeast RTO Mediation http://www.ksg.harvard.edu/hepg/RTO_ISO_Market_Reports.htm, Août 2001.

Siddiqui [2002]. Siddiqui A. S. "Equilibrium Analysis of Forward Markets for Electricity and Reserves." PhD dissertation, University of California, Berkeley, 2002.

Sioshansi [2006]. Sioshansi (eds.), Electricity Market Reform: An International Perspective, IAEE publication 2006.

Smeers [2001]. Smeers Y., "Market Incompleteness in Regional Transmission Markets", Working Paper, Université Catholique de Louvain, Janvier 2001.

Smeers [2003]. Smeers Y. "International Congestion Management", Contribution to the session on "International Congestion Management", CMI Electricity Project Transmission Workshop, Cambridge, 18–19, July 2003.

Smeers [2004a]. Smeers Y., « L'électricité: marchés organises et marchés de gré à gré », Présentation orale - L'ouverture des marches européens de l'électricité – IDEP (Institut d'Économie Publique) Marseille, janvier 2004.

Smeers [2004b]. Smeers Y., « L'électricité: marchés organises et marchés de gré à gré », *Revue d'économie publique* 14-2004/1, décembre 2004.

Smeers [2005a]. Smeers Y., "System Performance (Reliability, Congestion Costs, Production Costs", présentation en " Edf/EPRI Workshop on Electricity Market Performance", Paris, février 2005.

Smeers [2005b]. Smeers Y., "How well can one measure market power in restructured electricity systems?" Discussion paper of the European Union research project SESSA, Mai 2005, (http://www.sessa.eu.com/ documents/wp/D14.1_Smeers.pdf).

Staropoli [2001]. Staropoli C., « Organisation et efficacité des marchés de gros d'électricité – Une analyse économique des marchés anglo-gallois et nordique », Thèse de Doctorat en Sciences Economiques, Université de Paris 1 Panthéon – Sorbonne, 2001.

Stavros [2000]. Stavros R. "Transmission 2000: Can ISOs Iron Out the Seams?", *Public Utilities Fortnightly*, Mai 2001.

Stoft [1998]. Stoft S., "Gaming Intra-zonal Congestion in California", www.stoft.com, 1998.

Stoft [2002]. Stoft, S. Power System Economics. IEEE Press, Wiley-Interscience.

Sun [2005]. Sun J., "U.S. Financial Transmission Rights : Theory and Practice", WP #05008, Working Papers Series, Department of Economics, IOWA State University, Mars 2005.

Sweeting [2000]. Sweeting A., "The Wholesale Market for Electricity in England and Wales: Recent Developments and Future Reforms", Working paper M.I.T, septembre 2000.

Tennet [2001]. Tennet (Gestionnaire de réseau Néerlandais), "The Imbalance Pricing System as at 1 january 2001", Février 2001.

Tennet [2005]. Tennet (Gestionnaire de réseau Néerlandais), "The Imbalance Pricing System as at 01-01-2001, revised per 26-10-2005", disponible sur www.tennet.nl, Août 2005.

Tirole [1991]. Tirole J., The Theory of Industrial Organization, The M.I.T. Press.

Transelec [2005]. Transelec, "Memoria Annual 2005", disponible sur www.transelec.cl, 2005.

Tseng *et al.* [1999]. Tseng C., Oren S., Cheng C., Li C., Svoboda A. et Johnson R., "A Transmission-Constrained Unit Commitment Method in Power System Scheduling," *Decision Support Systems*, vol. 24, no. 3,4, pp 297-310, Janvier 1999.

UCTE [2007] Union for the Co-ordination of Transmission of Electricity (UCTE), "Final Report, System Disturbance on 4 November 2006", Rapport disponible sur www.ucte.org, Janvier 2007.

Varian [1995]. Varian H.R., Analyse Microéconomique. Traduction française De Boeck Université, 1995.

Vehvilainen [2005]. Vehvilainen L., "Applying Mathematical Finance Tools to the competitive Nordic Electricity Market", Helsinki University of Technology, Institute of Mathematics, Research Reports A475, 2005.

Von der Fehr *et al.* [2005]. Von der Fehr, N-H M., E.S. Amundsen and L. Bergman (2005), "The Nordic Market: Signs of Stress?", *The Energy Journal*, Special Issue on European Energy Liberalisation, 2005.

Von der Fehr-Harbord [1993]. Von der Fehr, N-H M., Harbord D., "Spot market competition in the UK electricity industry," *Economic Journal* 103, 531-546, 1993.

Walliser [2005]. Walliser B., "Les fonctions des modèles économiques", Working paper ENPC-CERAS, EHESS, 2005.

Williamson [1985]. Williamson O., The Economic Institutions of Capitalism, New York, The Free Press, 1985.

Wilson [1998a]. Wilson R., « Pricing Issues », dans *Designing Competitive electricity markets*, Chao H. et Huntington H. (Eds.), Kluwer Academic Publisher, pp. 97-103, 1998.

Wilson [1998b]. Wilson R., « Design Principles », dans *Designing Competitive electricity markets*, Chao H. et Huntington H. (Eds.), Kluwer Academic Publisher, pp. 159-184, 1998.

Wilson [2002]. Wilson R., "Architecture of electric power markets". *Econometrica*, 70(4), 1299-1340.

Wolak [2005]. Wolak F., "Arbitrage, Risk Management, and Market Manipulation: What Do Energy Traders Do and When it is Illegal?", Présentation, Stanford, 2005.

Wood-Wollenberg [1996]. Wood, A. J., Wollenberg B. - Power generation operation and control - Second Edition - John Wiley & Sons, Inc.

Yao *et al.* [2004]. Yao, J., S. S. Oren et I. Adler., "Computing Two-Settlement Cournot Equilibria in Electricity Markets," *Proceeding of the 37th Hawaii International Conference on Systems Sciences (HICSS 37)*. Big Island, HI, 2004.

Yu *et al.* [2005]. Yu J., Teng Sh., Mickey J., « Evolution of ERCOT Market", conference EPRI, 2005.

Zachmann [2005]. Zachmann G., "Convergence of Electricity Wholesale Prices in Europe?", working paper DIW, septembre 2005.

Table des matières

Introduction Générale 1

Chapitre 1
Système électrique et exploitation optimale par une entreprise intégrée verticalement 7
Introduction 8
Section 1: Description des éléments d'un système électrique 9
1.1 Consommation 10
1.1.1 Fluctuations de la consommation 10
1.1.2 Difficultés de prévision de la consommation 12
1.1.3 Elasticité-prix de la demande 14
1.2 Moyens de production 15
1.2.1 Types de centrales électriques 16
1.2.2 Caractéristiques des centrales électriques 17
1.3 Réseau de transport 21
1.3.1 Equilibre du système électrique 24
1.3.2 Les équations du réseau, les lois de Kirchhoff et les PTDFs 25
1.3.3 Limites de capacité de transport d'un réseau électrique 27
1.4 Conclusions de la section 29
Section 2: Exploitation optimale par une entreprise intégrée 31
2.1 Réseau de transport et zone de contrôle 32
2.1.1 Étendue du réseau de transport 33
2.1.2 Étendue de la zone de contrôle 34
2.2 La séquence de contrôles-commandes (une zone de contrôle) 34
2.2.1 *Unit commitment* (UC) 37
2.2.2 *Optimal dispatch* (OD) 40
2.2.3 Contrôles automatiques supplémentaires (en temps réel) 41
2.3 Conclusions de la section 43
Section 3: Coordination d'entreprises intégrées sur un même réseau 45
3.1 La coordination parfaite et bénéfices associés 45
3.1.1 Coordination parfaite par fusion des séquences de contrôle 45
3.1.2 Bénéfices de la coordination 46
3.2 Coordination de séquences de contrôles-commandes 47
3.3 Conclusions de la section 52
Conclusions du chapitre 53

Chapitre 2
Architectures de marché électrique 55
Introduction 56
Section 1: Contrôles-commandes *vs.* Marchés 60
1.1 Contrôles-commandes *vs.* marchés pour un bien simple 60
1.2 Contrôles-commandes *vs.* marchés pour le bien électricité 63
1.2.1 Les caractéristiques spécifiques de l'électricité et leurs conséquences 64
1.2.2 Construction d'une architecture de marché électrique 68
1.3 Conclusions de la section 70
Section 2: Architectures de marché sur une seule zone de contrôle 72
2.1 Le module d'énergie *forward* 74
2.2 Le module de transport *forward* 78
2.3 Le module des réserves *forward* 84
2.4 Le module du temps réel 86

2.5 Articulation entre modules et arbitrage 90
2.6 Analyse empirique d'architectures de marché 93
2.6.1 Le cas de l'Angleterre (Neta) 93
2.6.2 Le cas du Texas 95
2.6.3 Le cas de PJM 96
2.7 Architecture minimale : le module du temps réel 98
2.8 Conclusions de la section 100
Section 3: Architectures de marché sur plusieurs zones de contrôle 102
3.1 L'intégration d'architectures de marché de plusieurs zones de contrôle 103
3.1.1 L'intégration parfaite d'architectures 104
3.1.2 Mise en œuvre d'intégration d'architectures 107
3.2 L'intégration d'architectures par coordination des modules 110
3.2.1 Coordination des modules *forward* 111
3.2.2 Coordination des modules du temps réel 116
3.3 Conclusions de la section 118
Conclusions du chapitre 121

Chapitre 3
Séquence des marchés d'énergie et design du module du temps réel 124
Introduction 125
Section 1: Séquence des marchés d'énergie à court terme 127
1.1 Organisation de la séquence des marchés d'énergie 128
1.1.1 Marchés *forward* 129
1.1.2 Marché du temps réel 134
1.2 Liens entre marchés et arbitrage temporel 136
1.2.1 L'arbitrage entre marchés 137
1.2.2 L'arbitrage temporel en présence d'incertitudes 138
1.2.3 L'arbitrage temporel sur la séquence des marchés d'énergie 140
1.3 Conclusions de la section 142
Section 2: Design du module du temps réel 143
2.1 Etapes du module du temps réel 144
2.1.1 *Gate closure* et équilibrage du système 144
2.1.2 Détermination des prix et règlement des transactions « actives » 148
2.1.3 Détermination des prix et règlement des transactions « passives » 150
2.2 Types de design du module du temps réel 155
2.2.1 Marché du temps réel 155
2.2.2 Mécanisme d'ajustement 157
2.3 Etudes de cas 159
2.3.1 Le marché du temps réel au Texas 160
2.3.2 Le mécanisme d'ajustement en Angleterre 162
2.3.3 Le mécanisme d'ajustement en Belgique 168
2.3.4 Le mécanisme d'ajustement en France 171
2.3.5 Le mécanisme d'ajustement aux Pays-Bas 173
2.4 Le choix du design : incitations et conséquences économiques 177
2.4.1 L'impact du design du module du temps réel sur la séquence 178
2.4.2 La sécurité du réseau 182
2.4.3 Le *gaming* et le pouvoir de marché 184
2.5 Conclusions de la section 186
Conclusions du chapitre 188

Chapitre 4
Modélisation de la séquence des marchés d'énergie et du module du temps réel..... 191
Introduction 192
Section 1: Modélisation du module du temps réel (une zone de contrôle) 194
1.1 Définition de paramètres et variables 197
1.2 Modèle de base : présentation 198
1.2.1 Etape du temps réel..... 202
1.2.2 Etape *forward* 205
1.2.3 Equilibre des marchés..... 210
1.3 Modèle de base : simulations numériques et discussion 214
1.3.1 Les paramètres des simulations..... 214
1.3.2 Premiers résultats 216
1.3.3 Discussion des résultats..... 218
1.4 Extension du Modèle à deux technologies de production..... 224
1.4.1 Modélisation 225
1.4.2 Simulations numériques 227
1.4.3 Discussion des résultats..... 228
1.5 Conclusions de la section 233
Section 2: Modélisation de l'intégration d'architectures (deux zones de contrôle)..... 235
2.1 Définition de paramètres et variables 237
2.2 Modélisation des architectures « séparées » et « intégrées »..... 238
2.2.1 Etape du temps réel..... 241
2.2.2 Etape *forward* 243
2.2.3 Equilibre des marchés..... 246
2.3 Simulations numériques et discussion 247
2.3.1 Paramètres..... 248
2.3.2 Premiers résultats et discussion 249
2.4 Conclusions de la section 256
Conclusions du chapitre 259

Conclusion Générale 262

Annexe 268
Annexe N°1 : Abréviations 268
Annexe N°2 : Analyse de sensibilité : modèle de base 270

Bibliographie 276

Table des matières..... 293