L'architecture de marchés électriques : l'indispensable marché du temps réel d'électricité

Marcelo Saguan, Yannick Perez et Jean-Michel Glachant

Édition électronique

URL : <https://journals.openedition.org/rei/4053>

DOI : 10.4000/rei.4053

ISSN : 1773-0198

Éditeur

De Boeck Supérieur

Édition imprimée

Date de publication : 15 septembre 2009

Pagination : 69-88

ISSN : 0154-3229

Référence électronique

Marcelo Saguan, Yannick Perez et Jean-Michel Glachant, « L'architecture de marchés électriques : l'indispensable marché du temps réel d'électricité », *Revue d'économie industrielle* [En ligne], 127 | 3e trimestre 2009, mis en ligne le 15 septembre 2011, consulté le 02 juin 2022. URL : <http://journals.openedition.org/rei/4053> ; DOI : <https://doi.org/10.4000/rei.4053>

L'ARCHITECTURE DE MARCHÉS ÉLECTRIQUES : L'INDISPENSABLE MARCHÉ DU TEMPS RÉEL D'ÉLECTRICITÉ

Mots-clés : Marché du temps réel, gestion des externalités, analyse modulaire.

Key words : Real-time Market, Externality Management, Modular Analysis.

I. — INTRODUCTION

L'introduction de la concurrence dans le secteur électrique, longtemps organisé autour d'un monopole intégré en production-transport allant parfois jusqu'à la distribution, fait apparaître des problèmes de coordination qu'il convient de traiter par la mise en place de dispositifs adéquats (Staropoli 2001, Glachant 2003 et Glachant & Finon 2003). Après deux décennies d'expériences, l'importance de ces dispositifs – organisationnels et institutionnels – oubliée pendant un temps, revient sur le devant de la scène après les épisodes spectaculaires qu'ont constitués les différents black-out observés en Europe et dans le monde (Hogan 2002, Newbery 2006, Sioshansi & Pfaffenberger 2006).

- (1) Supélec & ADIS-Groupe Réseaux Jean-Monnet, université de Paris-Sud 11.
- (2) Maître de conférences en Sciences économiques, ADIS-Groupe Réseaux, Jean-Monnet université de Paris-Sud 11 et Coordinator of the Loyola de Palacio Chair on EU Energy Policy at the Robert Schuman Centre for Advanced Studies, European University Institute Florence.
- (3) Professeur en Sciences économiques, ADIS-Groupe Réseaux Jean-Monnet, université de Paris-Sud 11 et Director of the Florence School of Regulation, Director of the Loyola de Palacio Chair on EU Energy Policy at the Robert Schuman Centre for Advanced Studies, European University Institute Florence.

Sans entrer dans des débats qui dépassent largement le cadre de cet article, les grands problèmes encore en suspens concernent par exemple la coordination de la croissance des réseaux face aux nouveaux enjeux du développement des énergies renouvelables, du développement des centrales au gaz et du renouvellement du nucléaire. Sur un plan social, la question du traitement de la problématique de la pauvreté énergétique reste elle aussi largement inexplorée théoriquement et empiriquement. Beaucoup de travaux restent à conduire dans cette industrie pour disposer d'un modèle concurrentiel complet de gestion de l'ensemble des problématiques électriques.

Dans cet article, nous nous focalisons sur les marchés et les mécanismes mis en œuvre pour assurer la rencontre entre la production et la consommation à court terme. Bien sûr, cette coordination doit être assurée à différents horizons temporels, du très court terme (quand la production doit équilibrer la consommation à chaque instant) au très long terme (*via* les décisions d'investissement). Il existe à l'heure actuelle des débats sur la nécessité de la mise en place des marchés ou des mécanismes spécifiques pour assurer la coordination à long terme (Green 2006, Joskow 2006). Par simplicité dans le cadre de cet article, nous postulons que les marchés et mécanismes de court terme mis en œuvre sont suffisants pour assurer la coordination de long terme. Pour une discussion approfondie des problématiques de long terme, nous renvoyons à Finon & Pignon (2008) qui en discutent plus largement.

À court terme, l'introduction de la concurrence a conduit au remplacement des anciens outils, que nous présenterons dans la section suivante, tels que l'*unit commitment* et l'*optimal dispatch* par des marchés ou des mécanismes plus ou moins décentralisés (Stoft 2002, Smeers 2004a, Smeers 2004b). Le bien « électricité » est alors défini comme un bien combinant trois composantes complémentaires : l'énergie en elle-même, le transport de cette énergie vers les consommateurs, et les réserves en cas d'aléas de consommation ou de production en temps réel (Wilson 2002). Comme le bien électrique est fortement contraint techniquement, ces trois composantes doivent être parfaitement coordonnées en temps réel sans possibilité de file d'attente. Or, étant donné le caractère dynamique de la production et de la consommation électriques, cette coordination doit être anticipée à court terme un jour avant la rencontre en temps réel.

Pour ce faire, des marchés *forward de court terme du bien électrique* sont utilisés afin de préparer la coordination en temps réel de la fourniture électrique. Cette conception technico-économique du bien électrique conduit à la mise en œuvre d'architectures de marchés complexes, composées de quatre sous-marchés pouvant être construits de plusieurs manières différentes (Saguan 2007). Les quatre sous-marchés, ou « modules » dans la littérature sur le *market design* (4), sont alors les suivants : 1° le module *forward* d'énergie,

(4) Que nous traduirons par architecture de marché.

2° le module *forward* de transport, 3° le module *forward* des réserves et 4° le module du temps réel (5). Le fait que chaque module ait de multiples options internes d'organisation suggère une grande diversité possible d'architectures (6). Cette grande diversité révèle aussi la complexité inhérente des architectures de marchés électriques. En principe, dans un cadre simplifié (7), le choix de l'architecture devrait être moins important, car plusieurs architectures alternatives permettraient d'atteindre le « coût minimal de production ». Toutefois, les caractéristiques réelles des systèmes électriques comprennent des structures de marchés concentrées, des coûts de transactions, des externalités, des fonctions de coût non convexes, des biens publics, etc. Il convient de souligner que l'état de l'art actuel sur l'architecture de marchés électriques ne fournit pas de base théorique « robuste » pour pouvoir trancher définitivement entre les différentes solutions possibles (Wilson 2002). De plus, les différentes architectures de marché sont le résultat des choix historiques conditionnés par des situations institutionnelles variées (Glachant & Finon 2003 ; Glachant & Lévêque 2009).

À défaut de pouvoir identifier l'architecture de marché optimale, cet article fournit d'abord une taxonomie des architectures de marché utile pour comprendre les principaux choix possibles. Une fois cette taxonomie réalisée, l'objectif sera de montrer qu'il existe un noyau dur commun indispensable à tous les designs, un plus petit commun multiple. Nous démontrerons que l'« architecture minimale » est celle qui définit l'organisation des tâches et de la fourniture d'électricité en « temps réel ». Cela vient de l'importance du temps réel pour un bien comme l'électricité, bien difficilement stockable et dont les échanges physiques ne prennent effectivement place qu'en temps réel. Nous montrerons alors que les enjeux de coordination sont cruciaux à cet horizon temporel.

Cet article est organisé de la manière suivante. Après avoir synthétisé les principaux apports de l'analyse des options possibles concernant l'analyse des quatre principaux modules qui gèrent la fourniture électrique (section 2), nous vérifierons, empiriquement, l'importance du module de gestion électrique en temps réel dans trois systèmes électriques de référence (ceux de l'Angleterre (NETA), de la Pennsylvania, du Maryland et du New Jersey (PJM) et celui du Texas) (section 3). Nous démontrerons enfin l'importance du marché du temps réel comme étant l'architecture minimale devant être mise en place (section 4). La dernière section conclura (section 5).

- (5) Seuls les modules *forward* peuvent prendre la forme de purs marchés où l'offre rencontre la demande pour définir spontanément un prix.
- (6) On pourrait en principe dériver de nouvelles architectures encore inconnues à partir de la seule combinatoire des options pour chaque module.
- (7) C'est-à-dire un monde avec atomisticité d'acteurs, avec information complète, coûts de transaction nuls, des fonctions de coût convexes, une absence d'externalité, et de caractéristiques de bien public ainsi qu'une demande d'électricité inélastique.

II. — LES CHOIX D'ORGANISATION DES PRINCIPAUX MODULES DES MARCHÉS ÉLECTRIQUES

Dans l'ancien système verticalement intégré, la question de la coordination entre les unités de production et la consommation en temps réel était assurée par une séquence de contrôles-commandes, l'*unit-Commitment*, et l'*optimal Dispatch* qui assurait cette nécessaire coordination au fonctionnement du système électrique (figure 1) (Wood & Wollenberg, 1996). Cette séquence finissait avec des contrôles automatiques pour opérer les derniers réglages et ajustements.

FIGURE 1 : Schéma de la séquence des outils de contrôles-commande pour l'électricité

L'outil d'*unit commitment* était utilisé la veille du jour considéré comme l'horizon de programmation. Il servait 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 considérait 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* était utilisé de manière très rapprochée du temps réel et servait à déterminer le niveau de production définitif des unités qui sont disponibles effectivement en temps réel. Une ré-optimisation était nécessaire car les données introduites dans l'*unit commitment* n'étaient pas des valeurs déterministes. L'*optimal dispatch* fonctionnait donc en temps proche du temps réel où les prévisions sur les conditions du temps réel étaient très précises. Cet outil assurait que le système était équilibré et que les limites de capacité de transport n'étaient pas violées avec les nouvelles valeurs des injections et des soutirages.

Les contrôles automatiques se mettaient en route après réception des ordres envoyés par l'outil d'*optimal dispatch*. Ces contrôles ne réalisaient pas d'optimisation au sens économique ; ils se fondaient principalement sur des caractéristiques techniques. Les deux contrôles automatiques les plus importants étaient les réglages primaire et secondaire, dont l'objectif est d'assurer l'équi-

FIGURE 2 : Schéma du remplacement de contrôles par des marchés

libre entre production et consommation sans intervention volontaire du coordinateur (8).

L'introduction de la concurrence par la suppression complète des outils de contrôles-commandes, sans construction préalable d'une architecture de marché, se heurte aux propriétés économiques qui dérivent des caractéristiques spécifiques de l'électricité (9) et leurs effets économiques impliquent une coordination essentiellement en temps réel. Le besoin de coordination en temps réel sur le réseau électrique limite le remplacement complet des outils de contrôles-commandes par des marchés électriques *forward*. Ce besoin de coordination exige la construction d'une architecture de marché capable d'en assurer la coordination en temps réel. L'art de la conception d'une architecture de marché électrique consiste alors à choisir quels outils de contrôles-commandes peuvent être remplacés par quelle forme de mécanisme de marché (figure 2) (10). En pratique, le résultat final de l'introduction de la concurren-

- (8) 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 qui a dérégulé la fréquence sur le réseau. 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'elle 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.
- (9) 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.
- (10) Par simplicité, nous présentons ici seulement deux étapes dans la séquence de marchés/contrôle commande. Des contrôles-commandes étaient utilisés entre les périodes de J-1 et le temps réel et ces outils ont été remplacés par des marchés appelés *intra-journaliers*, car ils commercialisent des produits électricité livrables le même jour, quelques heures après la réalisation de la transaction.

ce est donc toujours une combinaison entre contrôles-commandes et marchés (Boucher & Smeers 2002). Il est donc nécessaire de souligner ici l'importance du concept de séquence de marchés à court terme et la nécessité d'identifier trois composantes nécessaires et complémentaires du bien électricité, à savoir l'énergie, le transport et les réserves.

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 ces deux composantes permet de considérer les externalités et les caractéristiques d'infrastructure des réseaux électriques. 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 en soutire à 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, confrontée aux caractéristiques spécifiques de l'électricité, impose un traitement particulier pour établir logiquement la limite de remplacement entre anciens contrôles et nouveaux 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. Pour lui, de purs marchés, où l'offre et la demande se rencontreraient et s'équilibreraient spontanément par le biais d'un prix, ne suffiraient pas pour combiner et coordonner instantanément (en moins d'une seconde) les trois composantes et donc éviter l'effondrement. La « réalité » physique restreint donc le remplacement des contrôles par des marchés. Comme corollaire, des purs marchés pour les trois composantes ne sont réalisables qu'en anticipation du temps réel, donc « à terme », comme le permettent les marchés *forward* (Wilson, 2002, Boucher *et al.*, 2006). C'est pourquoi on parlera de marché « physique ». Pour Wilson (2002), l'architecture doit prévoir trois 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 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. Ce pseudo-marché physique du temps réel, doit alors être géré par le gestionnaire du réseau, qui joue donc un rôle prépondérant dans l'architecture de marché.

On comprend dès lors qu'une architecture de marché sera composée de quatre 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 (11). Les fonctions essentielles assurées par ces marchés peuvent être réalisées par plusieurs mécanismes.

D'abord, les options de design pour le *module d'énergie forward* se différencient par leur degré de centralisation. Elles se répartissent alors 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* (Stoft, 2002).

Les congestions – à savoir les externalités négatives supportées par les infrastructures de transport – sont gérées par le module de transport. Les options de design du *module de transport forward* se placent sur deux axes: i) la séparation entre le module de transport et de l'énergie et ii) le degré d'internalisation des externalités anticipé *forward*. 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 second axe considère le degré d'internalisation des externalités du transport sur l'étape *forward*, qui dépend de la différenciation spatiale du bien électrique. Le degré d'internalisation des externalités sera défini par la forme à laquelle on intègre, dans les modèles économiques, les contraintes réelles de transport. Ainsi on peut intégrer toutes les contraintes de transport et définir des prix différents pour chaque nœud du réseau (12). Il existe aussi des solutions simplifiées consistant à utiliser des modèles économiques de réseau simplifiés permettant, en théorie, de choisir le degré d'internalisation d'externalités. Saguan (2007) distingue ainsi trois représentations géographiques du réseau: le modèle nodal, le modèle zonal et la zone unique. Plus le modèle économique du réseau se rapproche de la réalité physique du réseau, plus les externalités seront internalisées efficacement dans les décisions des consom-

- (11) Nous utiliserons ci-après le terme de marché « *forward* de court terme » pour distinguer la période allant de 24 heures avant la consommation jusqu'à quelques instants avant le temps réel, en opposition au marché du temps réel où le bien est effectivement disponible.
- (12) Notons que théoriquement, ceci peut être réalisé de deux manières selon le choix de l'option d'architecture concernant la séparation des modules transport et énergie. Si le module de transport est intégré au module énergie, des prix nodaux seront calculés et internaliseront une grande partie des externalités. Si les modules de transport et énergie sont séparés, une définition très précise de droits d'utilisation des capacités de transport (théoriquement des droits pour utiliser chaque ligne) permettrait une différenciation géographique des prix de l'énergie pouvant internaliser une grande partie des externalités.

mateurs et des producteurs. Si le modèle économique fait disparaître trop d'informations, les congestions seront gérées par des mécanismes qui ne prennent pas en compte la configuration réelle du réseau et ainsi introduisent des inefficacités d'usage et de localisation des consommations et des productions.

Dans la figure 3, nous représentons schématiquement les options de design qui concernent le degré d'internalisation des congestions et des externalités de réseau en fonction de la prise en compte des dimensions géographiques. Ainsi la réalité physique du réseau peut être capturée par trois représentations économiques de cette réalité : la plus simple consiste à faire totalement disparaître les caractéristiques physiques du réseau en créant une zone unique de contrôle des congestions. Dans ce modèle très simple, les externalités ne sont pas internalisées par les acteurs et les décisions économiques se font sans que leur impact sur le réseau soit connu. Dans la représentation économique zonale, la même réalité physique est découpée en deux zones reliées par une ligne qui supporte « principalement » la congestion (13). Cette représentation intermédiaire permet alors de signaler aux acteurs qu'il existe une congestion principale et cela conduit les acteurs à prendre en compte cette contrainte dans leurs décisions de production et de consommation. Enfin, dans la représentation nodale, le réseau physique est parfaitement représenté économiquement. Dans ce modèle, toutes les décisions économiques intègrent les externalités de réseau efficacement.

FIGURE 3 : Schéma du degré d'internalisation des externalités

(13) Ici la question de qui définit les zones est essentielle à l'efficacité de ce mode de représentation économique du réseau physique.

Troisièmement, le *module des réserves forward* présente différentes options de design pour la forme d'organisation de l'achat et de la production des réserves (14). La production de réserves peut être organisée par enchères (par exemple journalières) ou contractuellement entre le Gestionnaire de Réseaux de Transport (GRT) et les fournisseurs de réserves.

Enfin, les designs du *module du temps réel*, qui intègrent les transactions du temps réel (aussi appelées « écarts »), c'est-à-dire les différences entre les productions/consommations en temps réel et les transactions *forward*, se différencient selon deux axes principaux : i) la définition des signaux de prix du temps réel pour les différentes transactions du temps réel et ii) le degré d'internalisation des externalités de transport en temps réel. En premier lieu, pour la définition des signaux de prix du temps réel dans les transactions du temps réel, on distingue deux options de design : un design du type « marché du temps réel » et un design du type « mécanisme d'ajustement » (Saguan, 2007).

Dans un *marché du temps réel*, un prix unique de l'énergie en temps réel est utilisé pour toutes les transactions du temps réel : un achat d'énergie en temps réel ou « écart négatif » (une consommation supérieure à l'achat d'énergie réalisé sur le marché *forward* ou une production inférieure à la vente d'énergie sur le marché *forward*) et une vente d'énergie en temps réel ou « écart positif » (une consommation inférieure à l'achat d'énergie réalisé sur le marché *forward* ou une production supérieure à la vente d'énergie sur le marché *forward*) seront valorisés à un même prix qui devrait correspondre au coût marginal de l'équilibrage du système.

Dans un *mécanisme d'ajustement*, plusieurs prix sont définis pour les transactions du temps réel. *A minima*, il existe un prix différent pour les écarts négatifs et pour les écarts positifs. Aucun de ces prix ne correspond au coût marginal d'équilibrage du système qui serait obtenu par l'usage d'un marché du temps réel. Ces prix sont calculés sur la base du coût d'équilibrage par le gestionnaire du réseau et sont normalement ajustés avec des pénalités. L'objectif des pénalités est de limiter au maximum les transactions du temps réel (le prix des écarts négatifs est normalement supérieur au coût marginal d'équilibrage et le prix des écarts positifs est inférieur au coût marginal d'équilibrage).

En second lieu, le degré d'internalisation des externalités du transport en temps réel peut avoir les mêmes options de design que pour l'étape *forward* : une différenciation des prix à chaque nœud du réseau (nodal), en quelques zones (zonal) ou une seule zone (timbre-poste). Notons qu'à ce stade, les

(14) La demande de réserves est toujours administrée et gérée par le gestionnaire du réseau. Le gestionnaire du réseau est donc toujours l'acheteur des réserves et les seules options de design portent sur comment ces achats sont organisés et comment les réserves sont produites.

congestions doivent être absolument résolues pour conserver l'intégrité du système. Pour cela, les congestions provoquées par les agents économiques qui n'ont pas intégré, dans leurs décisions, les contraintes de transport (congestions résiduelles) doivent être gérées par le gestionnaire du réseau. Celui-ci utilise donc le module du temps réel pour réaliser des modifications de programmes de production (appelé *redispatching*) afin de contrôler les congestions résiduelles. Le coût de *redispatching* qui est payé au producteur afin qu'il modifie son programme de production est socialisé entre les différents utilisateurs du réseau.

Le tableau 1 récapitule les différentes options internes d'organisation pour chaque module.

TABLEAU 1 : Synthèse des options de design des modules

Module <i>forward</i> d'énergie	Degré de centralisation	<ul style="list-style-type: none"> • Marché centralisé • Marché décentralisé
Module <i>forward</i> de transport	Intégration	<ul style="list-style-type: none"> • Intégré au module énergie <i>forward</i> • Séparé du module énergie <i>forward</i>
	Degré d'internalisation d'externalités	<ul style="list-style-type: none"> • Une seule zone • Quelques zones (zonal) • Chacun des nœuds (nodal)
Module <i>forward</i> des réserves	Organisation de la production	<ul style="list-style-type: none"> • Marchés organisés • Contrats bilatéraux producteurs/GRT
Module du temps réel	Détermination des signaux de prix	<ul style="list-style-type: none"> • Marché du temps réel • Mécanisme d'ajustement
	Degré d'internalisation d'externalités	<ul style="list-style-type: none"> • Une seule zone (timbre-poste) • Quelques zones (zonal) • Chacun des nœuds (nodal)

Parmi les marchés électriques de référence, au niveau mondial, l'Angleterre, la PJM et le Texas sont particulièrement importants pour comprendre l'importance du marché du temps réel dans la coordination des activités électriques.

III. — ANALYSE EMPIRIQUE DES MODÈLES DE RÉFÉRENCE AU NIVEAU INTERNATIONAL

Pourquoi choisir l'Angleterre, la Pennsylvanie, le Maryland et le New Jersey (PJM) et le Texas comme cas d'étude ? Avant tout, il convient de noter que les trois expériences sont emblématiques des tentatives conduites à travers le monde et, qu'à bien des égards, elles ont valeur d'exemple. Ainsi, PJM est devenue, après la crise californienne de 2000-2001, le modèle de référence le plus largement cité comme « modèle de design centralisé robuste », et la référence empirique du *Standard Market Design* proposé par le régulateur fédéral

américain (15). À l'inverse, le Texas 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 le dernier cas est celui de l'Angleterre. Modèle de référence par son côté précurseur (1989) et par son processus très innovant de « réforme de la réforme » conduit en 2001, elle nous offre un éclairage intéressant car soulignant que l'on peut réussir à modifier une architecture de marchés très décentralisée, fondée sur des marchés bilatéraux et un module du temps réel, à partir d'une ancienne architecture de marché aussi centralisée que l'*Electricity Pool* de 1990 à 2001.

Nous pensons donc que ces trois cas sont révélateurs de l'importance du marché du temps réel. Nous partirons donc du cas anglais, qui est probablement le plus familier au lecteur, puis nous traiterons le cas texan, et nous finirons par le modèle proposé par PJM au reste du monde.

3.1. Le cas de l'Angleterre (NETA/BETTA)

Entre 1990 et 2001, malgré l'influence grandissante du régulateur et son interventionnisme régulier, les imperfections des règles de marché initiales sont apparues irrémédiables et ont conduit à un changement radical et complet du design du marché électrique. Sur le banc des accusés, les abus de pouvoir de marché, de nature verticale, horizontale ou locale ont été dénoncés (Newbery, 2000). Il convient ici de souligner que la structure réglementaire encadrait les transactions sur le *Pool* anglais permettait l'exercice d'importants pouvoirs de marché horizontaux et locaux. Les deux producteurs dominants pouvaient alors influencer les prix sur le marché (16). Face à ce problème, un changement radical de l'architecture de marché a été longuement discuté et introduit en 2001 (Staropoli 2001, Newbery 2002, Perez 2002). Cette nouvelle architecture est connue sous le nom de NETA (*New Electricity Trading Arrangement*) et a remplacé l'ancienne architecture, très centralisée, connue sous le nom d'« UK Pool » (Sweeting 2000). À l'inverse, le NETA a une architecture de marché très décentralisée où les participants réalisent la plupart de leurs échanges de façon bilatérale. En 2005, la nouvelle architecture de marché a été étendue à l'Écosse, donnant lieu à l'architecture BETTA (*British Electricity Trading and Transmission Arrangement*).

Nous voulons ici attirer l'attention sur le fait que le design du NETA/BETTA est centré sur le module du temps réel. Ce module fonctionne comme un

(15) Federal Energy Regulatory Commission (FERC).

(16) Trois comportements stratégiques qui utilisent les quantités offertes d'électricité comme une variable d'action ont été identifiés sur le *Pool* anglais: (i) le retrait de capacité « simple » destiné à dégager une rente de rareté, (ii) le retrait de capacité destiné à manipuler le mécanisme d'incitation, (iii) l'arbitrage entre le marché spot et le marché d'ajustement.

« mécanisme d'ajustement » (*Balancing Mechanism*) et est géré par le gestionnaire du réseau NGC (17). C'est un « mécanisme » dans la mesure où ce design s'écarte logiquement de la formation d'un prix de marché unique pour tous les participants. Au moins deux prix en temps réel coexistent pour traiter les « écarts » (18). 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* (19), et d'un prix pour vendre de l'énergie en temps réel (*SSP-System Selling Price*) (20) quand la production réelle est supérieure à la quantité vendue sur les marchés *forward* (21). La définition des prix des écarts s'établit sur la base des règles spécifiques qui prennent en compte des coûts additionnels et des pénalités. Les prix des écarts diffèrent du prix unique du marché et, de fait, pénalisent plus particulièrement les écarts et surtout les écarts négatifs.

Les congestions (22) sont aussi gérées par le biais du mécanisme d'ajustement (*balancing mechanism*). Mais les prix du temps réel ne prennent 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. NGC doit donc acheter et vendre de l'énergie à des ressources localisées, et utilise pour cela le *balancing mechanism*. 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. L'essentiel des transactions est réalisé bilatéralement. Il existe une bourse organisée (APX Power UK (23)) pour les contrats *forward* de court terme, mais le volume de transactions sur cette bourse est faible (24).

- (17) *National Grid Company* (NGC) est le gestionnaire du réseau pour la zone de contrôle correspondant à l'Angleterre.
- (18) C'est-à-dire ceux qui consomment ou produisent une quantité différente à leur position contractuelle *forward* transmise au gestionnaire du réseau (NGC).
- (19) Ou consommation réelle > achats sur les marchés *forward*.
- (20) 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).
- (21) Ou la consommation réelle < achats sur les marchés *forward*.
- (22) À savoir les externalités négatives supportées par les infrastructures de transport.
- (23) Cf. (www.apxgroup.com).
- (24) 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.

Sur le NETA/BETTA il n'y a pas non plus 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, soit dans les prix des écarts en temps réel.

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

3.2. Le cas du Texas

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

L'architecture du marché texan est 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 2003) (25). Ce modèle économique de réseau simplifié permet de définir des coûts de congestion pour les éléments les plus congestionnés du réseau et de faire payer aux utilisateurs la différence des prix entre les zones. Étant 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 (26). Le coût de gestion des congestions locales est alors socialisé entre tous les participants de la zone (Potomac 2006).

(25) Depuis le début de cette architecture, le nombre de zones a varié entre quatre et cinq.

(26) Ces deux mécanismes, le marché du temps réel et le mécanisme pour régler les congestions locales, sont matérialisés par des programmes d'optimisation fonctionnant comme des enchères. Ils se réalisent séquentiellement (pour gérer les congestions locales il faut connaître le résultat du marché du temps réel) mais presque simultanément (le marché du temps réel fonctionne à 15 minutes avant le véritable temps réel et le mécanisme de congestions locales quelques minutes après).

Sur le marché texan, il n'y a pas de marché *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 complètement centralisé (Baldick & Niu 2005). Les consommateurs ont une obligation de réserves qui est définie selon les volumes de réserves totales nécessaires et calculée par ERCOT au prorata des consommations. Ils peuvent choisir différentes formes de fourniture comme programmer des réserves propres, opter pour des contrats bilatéraux, ou participer au marché organisé.

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

3.3. Le cas de PJM

L'architecture de PJM, qui gère le système électrique de plusieurs États du nord-est des États-Unis, est plus centralisée que celles 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, FERC 2002) (27).

L'architecture du modèle « temps réel » en PJM est du type « marché » : il n'existe qu'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 (modèle nodal) et représentant le réseau physique afin d'internaliser, avec précision, les externalités de congestion liées au transport.

Les modules *forward* d'énergie et de transport sont intégrés dans le marché *forward* du jour pour le lendemain (*day-ahead market*). Ce marché n'est pas obligatoire car 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, présente des programmes de production dans le marché *day-ahead* organisé pour se couvrir des coûts de transport du temps réel. Ici, le degré d'internalisation

(27) Il est important de remarquer que ce type d'architecture a été déjà mis en place dans plusieurs régions des États-Unis depuis quelques années (par exemple *NE Pool*, *NY Pool* et *MidWest Pool*) (Sioshansi & Pfaffenberger, 2006).

d’externalités est maximum, étant donné qu’il y a une différenciation 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 au marché *day-ahead* paieront la différence de prix nodaux *day-ahead* si, en temps réel, ils respectent exactement leurs engagements pris sur 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 toujours, 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’opportunité de ne pas fournir l’énergie mise en réserve (PJM 2006) (28).

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 également fondé sur un modèle nodal et par son module de réserves *forward* fonctionnant sur la base des marchés journaliers organisés ou enchères.

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

TABLEAU 2 : Comparaison des architectures de marché

Architecture	Module temps réel (énergie, transport et réserves)	Module énergie <i>forward</i>	Module transport <i>forward</i>	Module réserves <i>forward</i>
Angleterre/ NETA/BETTA	Mécanisme/Zone unique	Décentralisé	–	Bilatéral entre GRT et producteurs
Texas	Marché/Zonal	Décentralisé	–	Marché organisé/ bilatéral
PJM	Marché/Nodal	Centralisé	Intégré/Nodal	Marché organisé/ bilatéral

(28) L’intégration (partielle) du marché d’énergie au marché de réserves a deux avantages principaux. Premièrement, ceci facilite la participation de producteurs dans les enchères de réserves ; en effet les producteurs sont sûrs qu’ils ne fourniront des réserves (et pas d’énergie) que quand cela sera économiquement rentable pour eux. Deuxièmement, ceci permet de gagner en efficacité en utilisant les moyens les moins chers pour fournir les réserves et l’énergie.

L'analyse empirique des architectures de marché de référence a confirmé et décrit leur grande diversité. 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 (29). Cette situation a deux conséquences logiques : la première porte sur l'importance d'une recherche plus approfondie sur l'identification de l'ensemble des différents designs possibles, objectif qui dépasse largement le cadre de cet article ; la seconde nous conduit à chercher à identifier le noyau commun à tous les designs, nécessaire pour remplir des fonctions « électriques » minimum indispensables.

Nous l'appellerons l'« architecture minimale », et nous verrons dans la section suivante pourquoi il s'agit du module du temps réel.

IV. — ARCHITECTURE MINIMALE : LE MODULE DU TEMPS RÉEL

Notre architecture minimale est bien entendu définie par rapport à l'objectif de cet article qui se focalise sur le fonctionnement du système électrique à court terme. Toutefois, il est intéressant de noter que la tendance actuelle en Europe est de négliger l'importance du marché en « temps réel » et la plupart des pays recourent, à cette échelle de temps, à des « mécanismes d'ajustement » qui distordent le prix de l'énergie du « temps réel » en imposant des pénalités aux agents économiques qui se trouvent en écart avec leurs prévisions (Newbery, 2006). 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. Si l'on cherche à définir 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 4).

FIGURE 4 : Concept d'architecture minimale

(29) Ce qui rend très complexe la détermination d'une architecture optimale est la grande variété des caractéristiques physiques des systèmes électriques.

Reprenons d'abord le premier axe. Le degré de centralisation combine deux notions : l'organisation d'une part, et la centralisation proprement dite d'autre part. L'organisation d'une architecture consiste à définir de façon explicite des règles pour encadrer les transactions dans chaque module (30). La centralisation correspond ici au fait que toutes les transactions sont formalisées et standardisées. 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 concernant l'échange. Ce sont des marchés de gré à gré. Logiquement, 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. Cependant pour certains modules, l'organisation et la centralisation sont obligatoires comme c'est le cas du module du temps réel.

Le second axe, le degré d'internalisation des externalités de transport, 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 précisément, plus le bien électricité est différencié spatialement. On trouve donc sur cet axe la finesse avec laquelle les externalités sont internalisées, allant de la définition nodale des prix, à la définition zonale et enfin à l'absence totale de différenciation spatiale. 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.

Cette architecture minimale ne demande en effet 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 *de facto* peu les acteurs d'un point de vue « physique ». Par exemple, le module *forward* d'énergie 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 établir par les autorités concernent la communication au GRT des échanges d'énergie futurs prévus entre les participants.

De même, le module *forward* de transport 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.

Enfin le module réserves *forward* peut être assuré par le GRT qui s'occupe alors de signer des contrats avec des fournisseurs de réserves.

(30) L'exemple typique de la définition de règles pour les transactions est l'utilisation d'enchères pour conduire les échanges.

A contrario, 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. 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 pris en considération 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 *ex ante*. De ce fait, le module du temps réel est la base minimale de toute architecture. Ce résultat découle 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.

V. — CONCLUSIONS

Dans cet article, nous avons développé un cadre d'analyse des architectures de marché électrique à court terme. Ainsi, 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.

Ainsi, nous avons pu vérifier la 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. Cette diversité est riche d'enseignements théorique et empirique et des travaux ultérieurs pourraient chercher à en faire un panorama plus complet.

Enfin, 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. 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. En effet, si l'on analyse les architectures de marché du point de vue de l'effort des autorités publiques pour les construire ou les modifier, notre approche permet de justifier la focalisation des efforts sur le module « du temps réel », qui devrait être considéré comme un axe prioritaire de la politique de création du marché unique de l'électricité en Europe.

RÉFÉRENCES

- BALDICK R. (2003), « Shift factors in ERCOT congestion pricing », *Working paper*, University of Texas (Austin), disponible sur (<http://www.ece.utexas.edu/~baldick/papers/shiftfactors.pdf>), mars 2003.
- BALDICK R., NIU H. (2005), « Lessons Learned: The Texas Experience », dans *Electricity Deregulation: Where to from here ?*, J. Griffin and S. Puller (Eds), The University of Chicago Press, 2005.
- BOUCHER J., SMEERS Y. (2002), « 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.
- Federal Energy Regulatory Commission (FERC) (2002), « Notice of proposed rulemaking on open transmission service and standardized market design ». Docket n° RM01-12-000.
- BOUCHER J., DAXHELET O., SMEERS Y. (2006), « An alternative benchmark of the Internal Electricity Market », Journées de la section électrotechnique du Club, EEA, Supélec, mars.
- FINON D., PIGNON V. (2008), « Electricity and long-term capacity adequacy: The quest for regulatory mechanism compatible with electricity market », *Utilities Policy* XX (2008) 1-16.
- GLACHANT J.-M., FINON D. (Eds.) (2003), *Competition In European Electricity Markets: A Cross-Country Comparison*, Edward Elgar.
- GLACHANT J.-M. (2003), « Quatre designs de réforme électrique: Grande-Bretagne, Californie, Scandinavie et Allemagne », *Economies et Sociétés*, EN n° 2-3, 2003, pp. 231-255.
- GLACHANT J.-M., LÉVÈQUE F. (2009), *Electricity Reform in Europe: Towards a Single Energy Market*, Edward Elgar.
- GREEN R. (2006), « Investment and generation capacity », dans F. Lévêque, éd. 2006, chap. 2, pp. 21-53
- HOGAN W.-W. (2002), « Electricity Market Restructuring: Reforms of Reforms ». *Journal of Regulatory Economics*, vol. 21, n° 1, pp. 103-132.
- JOSKOW P. (2006), « Market electricity markets and investment in new generating capacity ». *Working Paper*, CEEPR-MIT 06-009, April. Published in Helm D. dir. (2007), *The New Energy Paradigm*, Oxford University Press, 2007
- NEWBERY, D. (2000), « Privatization, Restructuring, and Regulation of Network Utilities », Massachusetts Institute of Technology Press. Potomac Economics, Ltd, « 2005 State of the Market Report for the Ercot Wholesale Electricity Markets », disponible sur www.ercot.com, Juillet 2006.
- NEWBERY, D. (2002) *Issues and options for restructuring electricity supply industries*, DAE Working Paper WP 0210, University of Cambridge.
- NEWBERY D. (2006), « Refining Market Design », CWPE 0615 and EPRG 0515, *Working paper*, février 2006.
- PEREZ Y. (2002), *L'économie néo-institutionnelle des réformes électriques européennes*, thèse de l'Université Paris 1 Panthéon-Sorbonne.
- PJM Interconnection (2006). « Ancillary Service, PJM 101, the basics », Training material, disponible sur (www.pjm.com), 2006.
- SAGUAN M. (2007), *L'analyse économique des architectures de marché électrique. Application au market design du temps réel*. Thèse de doctorat en Sciences économiques. Université Paris Sud XI.
- SIOSHANSI F.-P., PFAFFENBERGER W. (2006), *Electricity Markets around the World*, Elsevier 2006.
- SMEERS Y. (2004a), « L'électricité: marchés organisés et marchés de gré à gré », présentation orale – L'ouverture des marchés européens de l'électricité – IDEP (Institut d'Économie Publique) Marseille, janvier 2004.
- SMEERS Y. (2004b), « L'électricité: marchés organisés et marchés de gré à gré », *Revue d'économie publique*, 14-2004/1, décembre 2004.

- STAROPOLI C. (2001), *Organisation et efficacité des marchés de gros d'électricité, une analyse économique des marchés anglo-gallois et nordique*. Thèse université de Paris 1 Panthéon Sorbonne.
- STOFT S. (2002), *Power System Economics, Designing Market for Electricity*. New York, Wiley IEEE.
- SWEETING A. (2000), « The Wholesale Market for Electricity in England and Wales: Recent Developments and Future Reforms », *Working paper MIT*, septembre 2000.
- WILSON R. (2002), « Architecture of the power markets », *Econometrica*, 70(4), 1299-1344.
- YU J., TENG Sh., MICKEY J. (2005), « Evolution of ERCOT Market », conference EPRI, 2005.
- WOOD A.-J., WOLLENBERG B. (1996), *Power generation operation and control* – Second Edition – John Wiley & Sons, Inc.