

Trois (des) Caractéristiques Importantes pour le Design de Marchés Electriques

Marcelo Saguan
marcelo.saguan@supelec.fr

Working paper n° 2

Version 0.0 (version incomplète et provisoire)

-Ne pas citer !

(Des fautes d'orthographe ont été ajoutées volontairement pour éviter toute citation)

1 Mars 2005

1 Introduction

Le design de marchés électriques efficaces est une discipline très complexe. Cette discipline est complexe car un design de marché est composé par plusieurs modules ou composantes, ces composantes sont superposées dans le temps et elles sont très liées entre elles.

Normalement, pour comprendre le design de marché électrique on part dans les extrêmes. Soit on essaye d'analyser tous les facteurs et phénomènes ensembles, ce qui rend l'affaire très compliquée, soit on simplifie trop les designs et à la fin tous se ressemblent et on peut penser qu'ils aboutissent vers un même résultat.

Nous avons essayé de faire un compromis et nous avons choisi 3 caractéristiques de design de marché entre J-1 et J (*day ahead* et temps réel). Nous avons déjà limité le problème à court et très court terme. A ce stade temporel, l'objectif final d'un design de marché électrique est d'atteindre l'efficacité productive (minimisation du coût total de fonctionnement) et allocative (maximisation du surplus des consommateurs et des producteurs) pour un système électrique donné (moyens de production, réseau de transport, consommateurs). Alors, le choix des caractéristiques de design a été guidé en examinant les caractéristiques « physiques » particulières des systèmes électriques : le caractère irréversible des décisions « physiques » prises par certains moyens de production dans un environnement incertain ; les incertitudes entre J-1 et J ; les différents types de réseaux et les incertitudes dans la mesure des externalités (par exemple, calcul de la capacité de transport, PTDF) et l'internalisation en J-1. Sur ces caractéristiques physiques, différents designs de marché peuvent être envisagés pour essayer d'accomplir les objectifs d'efficacité mais tous les designs ne peuvent atteindre ces objectifs similairement.

Nous proposons donc premièrement, de caractériser un design de marché (entre J-1 et J) en répondant 3 questions : 1) Le design en temps réel repose-t-il sur un mécanisme d'ajustement ou un marché du temps réel ? ; 2) L'internalisation des externalités (provoquées par les congestions) se réalise-t-elle d'une manière zonale ou nodale ? Et se réalise-t-elle en temps réel, avant le temps réel (J-1) ou sur les deux périodes ? ; 3) Le marché *forward* en J-1 est-il plus centralisé (organisés) ou plus décentralisé (non-organisés) ? Dans ces questions y sont compris les caractéristiques de design choisies et les options pour chaque caractéristique.

Une fois définies les caractéristiques de design nous analyserons sommairement les conséquences du choix de ces options. Bien sûr, les conséquences vont dépendre fortement des configurations « physiques » auxquels nous voudrions appliquer les différents designs.

Finalement nous analysons pour quatre designs de marché en fonctionnement nos trois caractéristiques et les résultats apparents obtenus pour pouvoir en sortir quelques conclusions générales.

Ce document est organisé de la manière suivante : dans la section 2 nous expliquons quelques caractéristiques « physiques » de tout système électrique, dans la section 3 nous

abordons les 3 caractéristiques de design choisies, dans la section 4 nous étudions ces trois caractéristiques pour de cas ou des marchés en fonctionnement, finalement dans la section 5 la conclusion.

Note : Nous ne considérons pas le problème du pouvoir de marché dans ce rapport.

2 Caractéristiques « physiques » d'un système électrique

Avant de rentrer dans les caractéristiques de design de marché électriques nous ferons ici un bref rappel sur les quelques caractéristiques « physiques » générales du système électrique. Avec le terme « physiques » nous voulons représenter le caractère exogène du design. Pour un système électrique particulier, ces caractéristiques physiques sont fixées et ne peuvent pas être modifiées (au moins à court terme). C'est-à-dire que l'on doit faire avec.

2.1 Typologie des générateurs

Une des caractéristiques des technologies actuelles de production d'électricité et la possibilité ou non de modifier les décisions de production jusqu'au très proche du temps réel. Pour caricaturer cette caractéristique nous définirons deux types de générateurs : les générateurs « LENTS » et les générateurs « RAPIDES ». Les générateurs LENTS sont ceux qui doivent choisir leur niveau (quantité) de production une période avant de produire (par exemple en J-1). Après ce choix ils ne peuvent plus modifier ce niveau de production. En revanche les générateurs RAPIDES prennent leurs décisions de production au début de chaque période, c'est-à-dire qu'ils peuvent changer leurs décisions de production en tout moment. Bien que les technologies de production de l'électricité soient beaucoup plus variées que celles représentées ici, nous essayons de représenter de manière simple (caricaturale) les effets de ces deux types de générateurs sur les différents designs de marchés. Il faut remarquer que cette caractéristique est liée aux décisions à court terme et elle est différente à la « flexibilité horaire » de production. Par exemple certaines technologies de production nucléaire (par exemple, France) sont assez flexibles (elles pourraient suivre une courbe de charge journalière) mais il faut les programmer avec anticipation et donc les quantités décidées deviennent fixes à court terme.

Cette caractéristique de « temporalité » ou de « irréversibilité » est assez importante dans un système électrique en raison de deux phénomènes bien connus : les incertitudes (présentes sur le temps réel) et la non stockabilité de l'électricité. S'il n'y avait pas d'incertitudes, les décisions des générateurs LENTS seraient de même type que celles des générateurs RAPIDES. Si l'électricité était économiquement stockable, les problèmes de RAPIDES et LENTS ne se poseraient pas non plus.

Une autre hypothèse est que le coût marginal de production des générateurs LENTS est généralement inférieur au coût marginal de production des générateurs RAPIDES. Cela pourrait être vérifié en comparant les coûts marginaux de production des différentes technologies par rapport à la flexibilité de fonctionnement. L'exemple typique serait de comparer le coût marginal de production d'une unité type « turbine à gaz » (RAPIDE) avec une unité de production à charbon (LENT).

Il y a une exception forte à cette hypothèse. C'est la technologie hydraulique où les coûts marginaux de productions sont très bas alors que cette technologie est aussi la plus flexible. De plus, la technologie de production hydraulique (avec réservoir) permet en quelque sorte de stocker de l'énergie et de rendre l'électricité un peu plus stockable et similaire aux autres

commodités. La plus grande flexibilité et stockabilité rendent les systèmes plus hydrauliques moins sensibles aux incertitudes (à court terme) et à la contrainte production-consommation en temps réel. Cependant, ces systèmes avec une grande capacité de production hydraulique sont plutôt rares, seulement certaines régions du globe peuvent profiter de ces caractéristiques physiques privilégiées (par exemple, Scandinavie, Brésil).

2.2 Incertitudes sur la production et sur la consommation

Les incertitudes sont présentes tant du côté de la consommation que du côté de la génération. La variabilité stochastique de la demande (normalement considérée comme l'erreur de prédiction de la demande) a des effets importants sur les designs de marché. Par exemple, pour chaque acteur individuel cette erreur peut atteindre jusqu'à 15 % de sa charge totale [Newbery 2002]. D'autres incertitudes sont aussi présentes dans la génération. Deux cas typiques de ces incertitudes sont : les cas où les unités de génération ne peuvent pas atteindre leur niveau stipulé de production (par exemple, les éoliennes) ; un autre extrême est celui où l'unité n'arrive pas à démarrer comme prévu ou doit être arrêtée complètement pour des problèmes techniques.

La combinaison des incertitudes avec les caractéristiques physiques de production permettrait de créer des types d'acteurs plus réels sur un marché électrique. Cette combinaison pourrait exprimer un caractère « physique » (par exemple une technologie de production lente ou rapide avec plus ou moins d'incertitude) ou un caractère « contractuel » (par exemple des générateurs lents intégrés verticalement avec une consommation incertaine).

2.3 Types de réseau et incertitudes pour la mesure des externalités (congestions)

Les réseaux de transport d'électricité sont la base de tout marché électrique. Chaque réseau de transport a des caractéristiques propres. Ces caractéristiques vont jouer un rôle important pour le design de marché car elles définiront le degré et le type d'externalités entre les agents. Ici nous nous intéresserons à deux points : les types de réseaux et les incertitudes présentes dans la mesure des externalités.

Types de réseau

D'abord un réseau peut se différencier par sa « capacité » de transport relative à l'utilisation (intensité des congestions). Les réseaux sans congestions sont des réseaux qui n'ont (presque) pas de contraintes techniques. La dimension des infrastructures de transport est suffisante pour accepter toutes les variantes (au moins toutes les plus probables) des programmes de production et consommation. Si le réseau possède une large capacité de transport (relativement à l'utilisation) elle présentera peu de problèmes de congestion (et d'externalités). La topologie (ou la forme) de ce type de réseau n'a pas une très grande importance. C'est-à-dire que, du point de vue du design, le fait qu'un réseau soit très ou peu maillé n'est pas fondamental quand l'intensité de congestion est faible. Mais si la capacité de transport relative est faible, le « type » de réseau est important pour le choix du design. Ici,

nous introduisons encore deux types de réseaux caricaturaux : les réseaux « zonables » et les réseaux « non-zonables ». Deux caractéristiques principales rendent un réseau « zonable » ou « non-zonable » : 1) la topologie (arborescente/radiale ou maillée) et 2) la distribution (localisation) de la production et la consommation.

Les réseaux dits « zonables » sont des réseaux hypothétiques où l'on peut déterminer un nombre limité de zones¹ stables (disons inférieur à 10) qui possèdent deux caractéristiques particulières : il n'y a pas de congestions à l'intérieur de la zone ; la distribution de la production et de la consommation à l'intérieur de la zone n'a pas d'influence sur les interconnexions avec les autres zones². Les réseaux dits « radiaux » ou arborescents (couloir de lignes reliant zones de production et zones de consommation) peuvent rentrer dans cette classification. Dans cette catégorie, l'idée de « stabilité » des congestions est aussi ajoutée. Ce point correspond à la localisation et à la variabilité de la production et de la consommation, c'est-à-dire que les congestions ont toujours le même comportement (cas typique nord-sud). De même, ce type de réseau hypothétique permet au gestionnaire du réseau de prédire assez précisément les capacités de transport.

Les réseaux « non-zonables » sont ceux qui ne correspondent pas à la définition ci-dessus. Pour pouvoir accomplir les deux conditions mentionnées ci-dessus, le nombre de zones nécessaires est très grand ou directement afin de pouvoir les vérifier chaque zone devrait être composée par un seul nœud. Généralement les réseaux très maillés avec fortes congestions sont « non-zonables ». De plus, l'idée de changement des endroits congestionnés dans le temps est ajoutée dans ce type de réseau. Dès lors, on peut supposer que les capacités de transport sont difficilement prédictibles ex-ante.

Dans la version finale des exemples de types de réseaux seront ajoutés.

Incertitudes dans la mesure des externalités

Comme nous le verrons par la suite, un point clé du design de marché est l'internalisation des externalités. Cela suppose que l'on peut les mesurer avec une précision suffisante. Mais une telle précision de mesure est seulement possible en temps réel. Les mesures réalisées avant le temps réel ont un certain degré d'incertitude et cette incertitude est plus importante pour les réseaux définis comme « non-zonables ».

Pour mesurer les externalités il est nécessaire de connaître les injections et soutirages définitifs (finaux) pour tous les nœuds ainsi que les limites de capacité de transports des différents éléments du réseau, les coefficients d'utilisations des éléments congestionnés (par exemple, PTFD)³ et finalement la valorisation de cette utilisation pour chaque agent. Tout cela peut changer entre J-1 et J. Un changement sur la consommation dans un endroit peut changer la distribution des flux sur tout le réseau, modifier les lignes qui sont congestionnées et la capacité de transport des lignes utilisées. Les limites de capacité des différents éléments peuvent changer à cause de plusieurs facteurs. Pour les limites physiques le facteur le plus

¹ Une zone est un ensemble de nœuds liés par des lignes de transport.

² Techniquement cela reviendrait à dire que tous les PTFDs des nœuds à l'intérieur de la zone par rapport à toutes les interconnexions inter-zones sont égaux.

³ Tous ces éléments sont disponibles dans le logiciel de calcul des prix nodaux.

connu est la variation entre la température estimée en J-1 et la température réelle en J. Mais principalement, les règles de sécurité utilisées par les gestionnaires du réseau (par exemple, règle N-1⁴) créent des liens entre la capacité de transport et les injections et soutirages dans les nœuds, alors ces limites varient quand les injections et les soutirages varient. Finalement les coefficients pour mesurer l'utilisation des éléments congestionnés (par exemple, PTDF, mesure simple de Qui utilise Quoi) peuvent varier après un changement de la topologie du réseau (soit par commande du dispatcheur soit suite à un problème dans un élément du réseau) [Hogan 2000], [Boucher-Smeers 2002], [ETSO 2004a]⁵.

⁴ Les règles de sécurité du réseau consistent à garder le système dans une plage de fonctionnement telle que le système puisse rester dans les limites physiques raisonnables après la réalisation d'un événement grave (contingence, par exemple, perte d'une ligne et/ou d'un générateur), compris dans un ensemble d'événements graves déterminés préalablement (liste des contingences).

⁵ Cet effet est largement amplifié quand on veut mesurer et internaliser les externalités (capacité d'interconnexion) entre zones d'un seul réseau maillé reliées par plusieurs lignes de transport. Par exemple, le calcul de la capacité entre deux zones se réalise normalement en utilisant un cas de base du réseau (qui représente les injections et soutirages espérés pour le lendemain) [ETSO 2004a]. Ce cas de base est modifié afin de représenter différents niveaux d'échange d'énergie entre les zones. La capacité maximale de transport est définie comme celle correspondant au maximum d'échange entre les zones respectant les limites de sécurité. Le calcul consiste à évaluer que, pour chacune des contingences préétablies, les conditions du système restent dans les limites permises (pré et post contingence). La variation stochastique de la capacité de transport entre deux zones à laquelle nous faisons référence correspond à la variation stochastique qui conduit à la situation réelle (connue en temps réel) par rapport au cas de base (utilisé en J-1). En effet, avec la même liste de contingences, deux scénarios différents donnent deux capacités de transport différentes. Ce constat nous montre que le calcul de la capacité de transport entre deux zones devient plus certain lorsqu'on s'approche du temps réel. Pour internaliser les externalités précisément et au maximum possible, on doit donc s'approcher le plus possible du temps réel.

3 Caractéristiques des designs de marchés

Trois caractéristiques de design découlent des caractéristiques « physiques » propres aux systèmes électriques (voir section précédente). Ce sont le design du marché du temps réel, le design de l'internalisation d'externalités et le design du marché *forward*. Bien que nous présentions les 3 caractéristiques de design d'une manière séparée, elles sont très liées entre elles. Pour chacune des caractéristiques nous présenterons des options ou choix simplifiés. Ces options normalement sont composées d'un compromis entre *marchés* et *contrôles* ou *commandes* (gérés par le gestionnaire du réseau). Les designs avec une forte dose de *marché* font confiance aux incitations données par la concurrence entre les parties (décentralisées) pour arriver à un objectif (efficacité). En revanche, les designs avec une forte dose en *contrôle* reposent sur le « design » des incitations données au gestionnaire du réseau.

3.1 Design du marché du temps réel

Importance du marché du temps réel

Etant donné les incertitudes présentes sur le système électrique, la non-stockabilité et de l'importance de l'équilibre production-consommation en temps réel, le marché du temps réel est fondamental dans un design de marché électrique. Le marché du temps réel est le seul marché « physique » car les producteurs et les consommateurs seront mesurés pour ce qu'ils produisent et ce qu'ils consomment. C'est en temps réel que l'on devrait connaître le vrai prix (coût marginal) de l'énergie. Finalement, le marché du temps réel est la base de tous les marchés *forward* qui le précèdent [Hirst 2001]. Les marchés *forward* ne peuvent pas remplacer complètement le *dispatch* en temps réel. La demande et l'offre d'électricité peuvent changer continuellement et de manière stochastique. Les déviations entre J-1 et J sont inévitables et nécessitent d'être gérées de manière appropriée.

Le design de marché du temps réel est toujours centralisé (par le gestionnaire du réseau) car le temps disponible pour trouver un équilibre production-consommation n'est pas suffisant pour envisager un design décentralisé. Mais, les différents designs de marché électrique ont abordé le problème de l'équilibre du système en temps réel (consommation=production) avec divers points de vue.

Pour notre part, nous différencions ici deux types de designs : marché du temps réel (proprement dit) et « mécanisme d'ajustement »⁶. Notre critère principal pour les différencier est la possibilité d'accès au marché pour tous les participants du marché ou l'absence de barrières d'entrée. Une des formes de barrières d'entrée la plus explicite est l'application de pénalités sur le marché du temps réel. Par exemple, pour nous un « marché du temps réel » avec pénalités est en fait la même chose qu'un mécanisme d'ajustement. Bien qu'il existe

⁶ Cette classification stricte permet de simplifier la liste des designs réels dont les caractéristiques sont beaucoup plus nuancées. Notre nomenclature ne correspond pas exactement aux nomenclatures utilisées par les marchés eux-mêmes.

d'autres barrières d'entrée qui transformeraient un marché du temps réel en un mécanisme d'ajustement, nous nous centrerons principalement sur l'utilisation de pénalités. Néanmoins nous commenterons sommairement les facteurs qui peuvent déplacer un design plus du côté d'un marché du temps réel ou plus du côté mécanisme d'ajustement.

Marché du temps réel vs. Mécanisme d'ajustement

En principe, sur un **marché du temps réel**, tous les participants peuvent acheter ou vendre l'électricité en temps réel au prix du temps réel. Ce design permet de définir un prix du temps réel qui devrait correspondre au coût marginal (ou à la « valeur marginale ») en temps réel (si l'hypothèse de concurrence parfaite est vérifiée). Les participants peuvent vendre ou acheter de l'énergie librement en se mettant en déséquilibre (production-consommation) avant la *gate closure* (dernier délai pour présenter les programmes de production et de consommation). Concrètement, il n'y pas de pénalités pour l'achat et la vente d'électricité en temps réel. Le point clé ici, pour différencier le modèle du marché de celui du mécanisme d'ajustement, est qu'en temps réel il existe un seul prix de l'énergie et, que ce prix est déterminé sur un marché (enchère). Ce prix sera utilisé pour régler toutes les productions (ou consommations) propres au temps réel.

En revanche un **mécanisme d'ajustement** comme un design qui pénalise les achats « passifs » et les ventes « passives »⁷ d'énergie en temps réel. Cependant, les participations actives, à la demande du gestionnaire du réseau ne sont pas pénalisées. Dans ce design, les agents qui n'ont pas une certaine flexibilité de commande sur leurs moyens de productions ou sur leurs consommations en temps réel sont pénalisés quand ils soutirent ou injectent de l'énergie en temps réel en écart par rapport à leurs programmes antérieurs. Les pénalisations peuvent apparaître comme une formule explicite directe où la valeur des écarts est calculée en multipliant (ou en divisant) les prix (calculés sur la base des offres présentées sur ce mécanisme) par un facteur (par exemple, $1+K$). Elles peuvent apparaître implicitement, comme une manière particulière de calculer les prix des écarts (par exemple, cas de NETA où les prix sont calculés comme une moyenne pondérée en volume pour chaque demi-heure). Nous pourrions ajouter ici les designs de marchés où la valeur de l'énergie achetée ou vendue en temps réel provient d'une formule ou tarif défini administrativement (par exemple, cas de la Belgique). Cette valeur ne correspond pas aux conditions réelles du moment où cette énergie a été consommée ou produite et de plus les tarifs sont choisis pour diminuer le volume de déséquilibres. La différence principale entre mécanisme d'ajustement et marché du temps réel est l'existence d'au moins deux prix de l'énergie en temps réel (*dual imbalance pricing*): un prix pour ceux qui ont un écart positif (vendeurs « passifs » d'énergie en temps réel) et un autre prix pour ceux qui ont un écart négatif (acheteurs « passifs » d'énergie en temps réel). Les acheteurs actifs et les vendeurs actifs (ceux qui présentent des offres qui sont acceptées par le gestionnaire du réseau) peuvent être payés le prix de l'offre (*pay-as-bid*) ou le prix marginal.

⁷ Le terme « passif » est utilisé ici pour caractériser les achats et les ventes pour des « écarts ». En revanche, le terme « actif » sera utilisé pour les achats et les ventes suite à l'acceptation des offres sur le mécanisme d'ajustement.

Il est important de remarquer la différence entre les pénalités pour non-exécution des ordres du Gestionnaire du réseau sortant des offres acceptées sur le marché du temps réel ou des déploiement des réserves qui ont été acheté antérieurement et les pénalités sur des achats et des ventes en temps réel (présentes dans les mécanismes d'ajustement).

Effets des pénalités sur le marché du temps réel

Un mécanisme d'ajustement avec des pénalités (ou géré de manière administrée) provoque des distorsions sur les prix des marchés *forward* d'énergie car ces prix sont construits sur l'espérance mathématique des prix du temps réel (qui, dans ce cas se basent sur des pénalités administrées) ([Boucher-Smeers 2002] et [Boucher-Smeers 2003]). Les pénalités en temps réel, quand elles n'ont pas un fondement économique, limitent (créent des barrières) ou distordent la possibilité d'arbitrage temporel entre les marchés d'énergie. Or cet arbitrage est une des forces principales qui mènent un marché bilatéral ou décentralisé vers un résultat optimal global (un *dispatch* économiquement efficace) ([Chandley-Harvey-Hogan 2000], [Chandley-et al. 2003] et [Wolak et al. 2002]).

Un des principaux problèmes des pénalités est qu'elles ne sont pas symétriques pour tous les types d'acteurs. Les pénalités n'affecteront pas un générateur qui peut participer activement au mécanisme d'ajustement. Il peut y présenter des offres calculées de manière optimale pour être appelé à modifier la production seulement si les conditions économiques du système en temps réel lui conviennent. Le calcul de l'offre optimale dépendra du système de paiement du mécanisme (*pay-as-bid*⁸ ou prix de l'offre marginale⁹ [Green-McDaniel 1999]). En revanche les acteurs qui n'ont pas de contrôle de la production (ou la consommation) en temps réel seront toujours pénalisés s'ils souhaitent vendre ou acheter en temps réel (cas des consommateurs, cas des traders, cas des générateurs lents, cas des producteurs éoliens). De plus, normalement les pénalités ne sont pas symétriques pour les ventes et pour les achats en temps réel. Cela vient du fait que le gestionnaire du réseau craint plus d'être « court » en énergie et préfère donc pénaliser les écarts négatifs.

La position de la *gate closure* est elle aussi importante et elle est liée au fonctionnement en temps réel [Smeers 2004a]. Pour les mécanismes d'ajustement la position temporelle de la *gate closure* déterminera la fermeture des marchés *forward* et la présentation des derniers programmes donnant la base pour calculer l'énergie achetée ou vendue en temps réel. Plus la *gate closure* est proche du temps réel, moins il y aura d'incertitudes. C'est-à-dire que la variance est réduite et la position présentée dans les programmes finaux sera plus proche de la valeur du temps réel. Et les participants seront moins exposés aux pénalités. En théorie, avec une *gate closure* proche du temps réel, et si les marchés *forward* qui fonctionnent juste avant la *gate closure* étaient suffisamment liquides et transparents, les agents ne paieraient presque pas de pénalités. Pour les marchés du temps réel, la *gate closure* est importante car

⁸ Ces générateurs pourraient être affectés s'ils ont une composante stochastique de défaillance par exemple ou s'ils n'arrivent pas à prévoir et calculer correctement les offres optimales dans les enchères discriminatoires (*pay-as-bid*).

⁹ Pour un système de paiement au prix de l'offre marginale, l'offre optimale pour un générateur sera son coût marginal (sous l'hypothèse de concurrence parfaite).

la position de production présentée doit être respectée à un certain degré (bien qu'elle puisse être déséquilibrée, si la production programmée est différente de la consommation du même agent). Les déviations de production après la *gate closure* sont traitées par un autre type de pénalités, les pénalités pour programmes physiques non accomplis (*uninstructed deviations* pour le marché texan).

Une autre source d'inefficacité importante avec l'utilisation de pénalités sur le marché du temps réel est que les participants sont incités à s'auto équilibrer (en gardant des moyens d'équilibrage « individuels »), et le bénéfice d'un équilibrage global, où les déséquilibres individuels peuvent être annulés, est diminué [Newbery-McDaniel 2003], [Roques-Newbery-Nutall 2004].

L'utilisation des pénalités a été justifiée principalement pour trois raisons : La première est la sécurité du réseau. La seconde a été de créer une forte incitation à réaliser des contrats *forward*. La troisième est d'éviter des phénomènes de *gaming*.

L'utilisation de pénalités comme une manière d'améliorer la sécurité du réseau serait valide si le calcul des « pénalités » se faisait sur la base des coûts correspondants à la diminution de la sécurité. Mais ces coûts sont très difficiles à évaluer. De plus les gestionnaires du réseau peuvent avoir l'incitation à préférer des grandes pénalités car elles feraient augmenter la sécurité (artificiellement ?) et diminuer la charge de travail en temps réel et tout cela gratuitement. Si les gestionnaires du réseau seraient incités par rapport à l'efficacité du marché en générale (et non pas seulement à la sécurité du réseau) ils préféreraient des pénalités plus basses ou nulles. De plus les incitations données aux gestionnaires (après l'utilisation des pénalités) sont encore plus compliquées car étant donné qu'il y a deux prix du temps réel, le gestionnaire du réseau a un revenu additionnel (*beer fund*). Si le montant ramassé par cette différence de prix est donné au gestionnaire du réseau il sera encore plus incité à préférer des grandes pénalités. Si ce montant doit être réalloué aux participants proportionnellement aux quantités mesurées, cela pourrait donner des mauvaises incitations aux grands participants [Cornwall 2002].

Donner une forte incitation pour réaliser des contrats *forward* serait une solution de compromis pour forcer une plus grande liquidité pour des marchés qui ne sont naturellement très liquides, marchés bilatéraux non organisés (Smeers 2004b). Mais il faudrait évaluer les pertes d'efficacité pour comparer.

D'autres facteurs autres que les pénalités peuvent limiter l'accès sur le marché du temps réel. Entre ces facteurs nous pouvons citer la façon de calculer les prix de déséquilibre et la longueur de l'intervalle de calcul. La façon de calculer les prix de déséquilibre a une influence directe sur la prédictibilité des prix. Les mécanismes d'ajustement utilisent des façons de calculer les prix généralement obscures qui rendent la prédiction très difficile et cela provoque un effet très similaire à une pénalité.. Un autre point important est la période de calcul du prix de déséquilibre (elle peut aller d'entre 5 minutes et 1 heure). Plus la période est longue , plus il est difficile d'avoir une vision précise pour traduire les coûts en prix de déséquilibre. Par exemple sur 30 minutes il y a des générateurs qui peuvent être en équilibre global sur l'intégrale de l'énergie de la période mais pas en équilibre à tout moment de la période. Ils sont par exemple en déséquilibre négatif pour les 15 premières minutes et en

déséquilibre positif pour le reste. Cela impose donc sur le système des coûts d'équilibrage supplémentaires. Etant donné que les coûts supplémentaires sont repartis seulement sur ceux qui sont en déséquilibre sur l'intégrale de la période, ces derniers doivent supporter les coûts provoqués par les autres [Cornwall 2002].

3.2 Design de l'internalisation d'externalités (provoquées par les congestions)

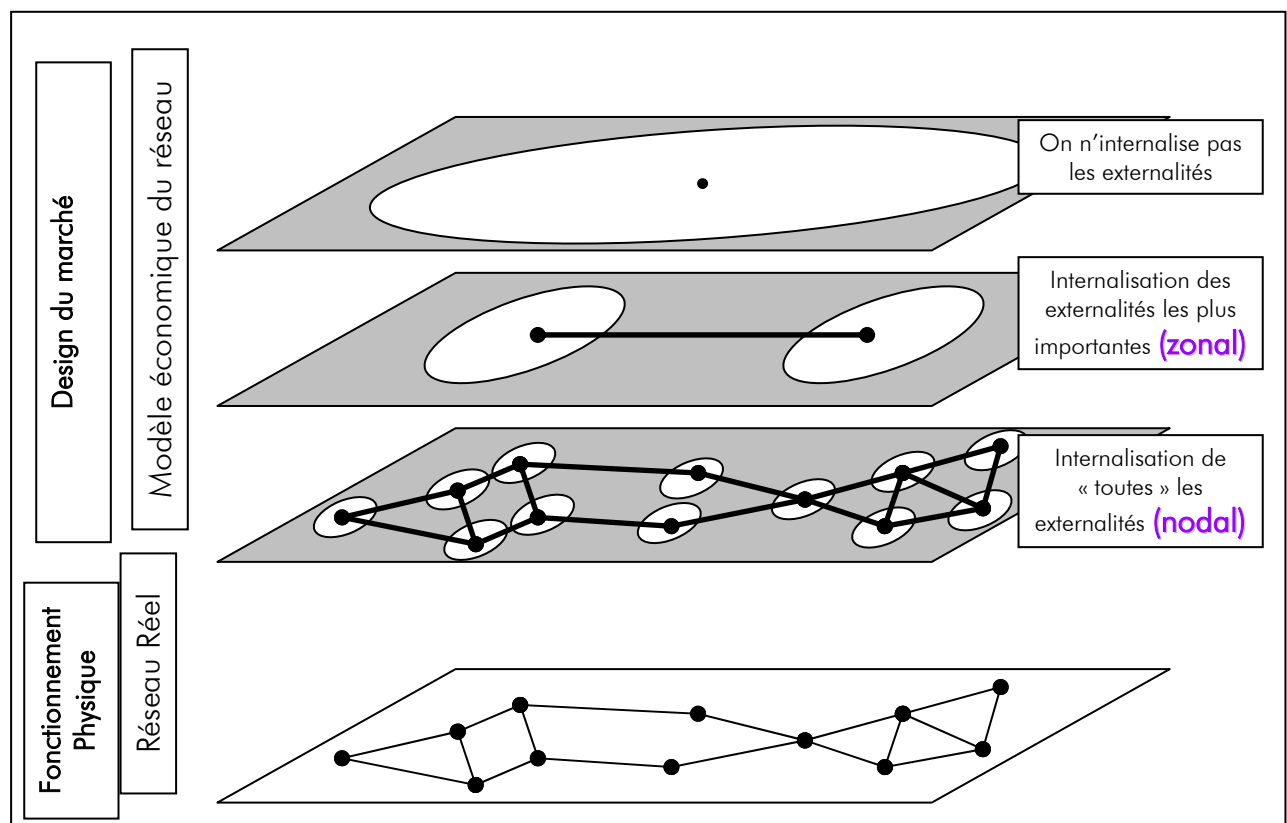
Les congestions sur le réseau de transport sont une des principales causes de la présence d'externalités entre les participants d'un marché électrique. Si les externalités ne sont pas internalisées, le marché est incomplet [Smeers 2002] et des problèmes peuvent apparaître : inefficacités, gaming/mauvaises incitations, besoin des mécanismes de contrôles, multiples équilibres, etc. Une des manières pour internaliser les externalités est d'introduire dans le marché d'énergie (enchère) un « modèle économique » du réseau qui permettrait de différencier géographiquement l'énergie. Le choix de ce modèle économique du réseau va déterminer le compromis entre les externalités qui seront gérées par des marchés et celles qui seront gérées par des contrôles. Si le choix du « modèle économique » du réseau correspond très clairement aux caractéristiques physiques du réseau on pourrait internaliser toutes les externalités, c'est-à-dire elles seraient gérées par le marché. Mais, internaliser toutes les externalités peut s'avérer impraticable car le produit « électricité » serait très différencié géographiquement et donc cela correspondrait à beaucoup de produits/marchés élémentaires avec peu de liquidité. Il y a donc, un compromis entre le choix des externalités que l'on internalise (le degré de précision d'internalisation) et la fonctionnalité des marchés résultants. De plus pour pouvoir internaliser les externalités il faut d'abord pouvoir les mesurer. Nous avons vu dans la section 2 qu'on est capable de mesurer plus précisément les externalités lorsqu'on s'approche du temps réel. Il y a donc aussi un autre facteur important qui est la position temporelle de l'internalisation. C'est pourquoi, ici nous analyserons deux causes de distorsion des signaux de prix (par rapport à la localisation) qui peuvent avoir une influence sur l'efficacité. La première vient (de ce que) du fait que le degré de précision d'internalisation d'externalités (son modèle économique) ne correspond pas exactement avec la réalité physique du réseau. La seconde vient de ce que l'internalisation des externalités se fait en J-1 et qu'il n'y a pas de « correction » en J (temps réel). Cette cause est importante car l'internalisation des externalités peut être plus précise lorsqu'on se rapproche du temps réel.

Pour simplifier, nous considérerons par la suite que tous les designs de marché fonctionnent avec un modèle intégré¹⁰. C'est-à-dire que l'énergie est intégrée au transport, et que la valeur de la capacité de transport est implicite dans les prix de l'énergie pour chaque position « géographique » (nœud ou zone) où l'énergie est injectée ou soutirée. Ceci vaut aussi pour un design de marché « séparé » mais parfaitement arbitré (géographiquement) pour ses différents marchés d'énergie. Si cette hypothèse n'était pas retenue, on devrait ajouter une incertitude supplémentaire.

¹⁰ Le terme « intégré » sera utilisé avec deux acceptions. Dans l'une, il correspond à l'intégration entre le marché d'énergie et de capacité de transport. Mais le terme « intégré » sera aussi utilisé dans le sens d'intégration des marchés régionaux.

3.2.1 Degré de précision d'internalisation (Zonal/Nodal)

Le degré de précision d'internalisation peut être sélectionné en choisissant un « modèle économique » du réseau (présent dans le modèle mathématique du calcul des prix de l'énergie) qui représente plus ou moins le vrai réseau de transport. Simplement, un modèle complet (celui qui représente tous les éléments -lignes, transformateurs, etc- du réseau)¹¹ peut différencier des prix d'énergie pour chaque nœud du réseaux (modèle nodal). Un modèle simplifié du réseau peut agréger plusieurs nœuds en zones et donc fournir des prix d'énergie différenciés par zones (modèle zonal). Si un seul prix de l'énergie est défini (pas de différenciation géographique) correspond au cas particulier du modèle zonal avec zone unique.



L'idée de base, en principe, serait de faire correspondre au maximum les caractéristiques des réseaux économiques avec la réalité physique. Mais rappelons tout d'abord un concept concernant la caractéristique physique des réseaux. Nous avons considéré deux types de réseaux : les réseaux « zonables » et les réseaux « non-zonables ». La définition des réseaux « zonables » est que, si nous appliquons un modèle nodal sur un réseau « zonable », dans tous les cas les prix calculés pour les nœuds composant la zone hypothétique sont les mêmes¹². Pour les réseaux non-zonables la seule manière de respecter cette propriété serait d'utiliser une zone pour chaque nœud.

¹¹ Le modèle nodal est aussi une approximation car il modélise jusqu'à un niveau de tension déterminé.

¹² Une des propriétés physiques nécessaires pour les réseaux « zonables » serait que tous les nœuds à l'intérieur de la zone hypothétique aient le même PTFD par rapport à toutes les lignes susceptibles d'être congestionnées.

A première vue, on aurait le degré de précision d'internalisation optimale quand le modèle « économique » du réseau (définition zonale/nodale des prix) représente correctement sa caractéristique physique : une définition nodale des prix pour les réseaux non-zonables et une définition zonale des prix pour les réseaux zonables. En revanche, l'utilisation d'un modèle « économique » du réseau qui n'est pas exactement en accord avec la réalité physique du réseau pourrait provoquer des inefficacités. Par exemple, cela apparaît dans les designs qui utilisent des modèles « économiques » zonaux sur des réseaux non-zonables ou dans les designs qui se comportent comme s'il n'y avait pas de congestions alors qu'il y en a (zone unique). Mais un modèle économique du réseau plus détaillé pourrait créer beaucoup des produits différenciés. Les marchés sont donc nombreux et en conséquence peu liquides. Le manque de liquidité a deux conséquences : il implique des coûts spécifiques et incite à la manipulation. Les agents ont tendance à désertir ce type de marché avec, pour conséquence une nouvelle perte de liquidité (Smeers 2004b).

Quand on est sur un réseau « zonable » et que le nombre des zones est faible, le choix est plus clair. En revanche pour les réseaux « non-zonables », trancher entre un modèle nodal et zonal (ou calculer le nombre de zones nécessaires optimal) n'est pas une tâche facile¹³. Le choix d'un modèle économique du réseau résultera d'un compromis entre les externalités qui seront internalisées par des marchés d'un côté et la fonctionnalité de ces marchés (par exemple, liquidité) d'un autre côté.

Lorsque l'on sort du choix « optimal » du modèle économique du réseau, trois problèmes vont apparaître : design des mécanismes de contrôle de congestions (et design des incitations pour le gestionnaire du réseau), des inefficacités et des problèmes de Gaming (mauvaises incitations pour les participants).

Les congestions qui ne sont pas gérées directement pour le marché doivent être traitées d'une manière ou d'une autre par le gestionnaire du réseau. Une congestion qui n'a pas été traitée et résolue peut provoquer la panne des certains éléments et même de tout le réseau. Cela exige un design d'un mécanisme de contrôle pour gérer les congestions « hors marché ». Les gestionnaires du réseau peuvent utiliser divers mécanismes à cette fin. Ces mécanismes aident à rendre réalisable l'ensemble des programmes de production car les acteurs pensent que les capacités de transport sont infinies quand ils prennent leurs décisions mais les capacités de transport ne sont pas infinies pour le gestionnaire du réseau. Des mécanismes pour résoudre des congestions locales liés au marché du temps réel ou mécanisme d'ajustement sont les plus communs (normalement appelées *counter-trading* ou *redispatching*). Aussi des contrats bilatéraux peuvent être utilisés pour résoudre des congestions plus systématiques. Mais étant donné que les coûts des congestions sont socialisés (par le biais du tarif de transport ou par le biais d'une charge complémentaire - *uplift*) les participants ne sont pas incités à prendre les décisions optimales. De même, étant donné que le coût pour résoudre les congestions n'est pas affecté directement sur le gestionnaire du réseau, il n'est pas directement incité à le faire de la manière la plus efficace.

¹³ Il ne faut pas oublier que les designs de marchés qui ont choisi des modèles nodales utilisent aussi des approximations et non des modèles de réseau complet (par exemple, il ne modélisent pas les éléments de basse tension généralement radiaux ?).

Dans un environnement avec deux technologies hypothétiques de production des inefficacités apparaissent principalement quand les congestions sont résolues avec des moyens de production chers (ou d'une façon sous-optimale). En théorie, les décisions de production que doivent être prises avant le temps réel (par exemple, les générateurs LENTS) se font sur la base de l'espérance mathématique des prix du temps réel et du prix *forward*. Ces deux prix devraient coïncider si l'arbitrage est parfait et en absence de pénalités et aversion aux risques des acteurs. Si la définition du prix en temps réel ne prend pas en compte les phénomènes de congestions (n'internalise pas les coûts des congestions), le prix *forward* peut présenter des distorsions et cela peut conduire les générateurs à prendre de mauvaises décisions physiques. Quand ces mauvaises décisions ne peuvent pas être corrigées sans coût, elles ont une influence sur l'efficacité. Par exemple, si un générateur LENT, placé en un nœud particulier du réseau, prend une mauvaise décision de production, cela peut provoquer des surcoûts pour résoudre les congestions en temps réel (avec les moyens rapides de production plus chers) [Benintendi-Boccard 2002].

Quand il y a des externalités qui ne sont pas internalisées (c'est-à-dire que les acteurs ne prennent pas en compte dans leurs fonctions objectif les effets qu'ils provoquent sur le réseau) des mauvaises incitations apparaissent. Ces mauvaises incitations peuvent provoquer des phénomènes de *GAMING* (« *DEC game* » pour les congestions). Étant donné que les coûts pour résoudre les congestions ne sont pas répercutés directement sur ceux qui provoquent les congestions, les acteurs peuvent créer des congestions volontairement et se faire payer pour les résoudre, c'est-à-dire qu'ils se font payer pour ne rien faire. Les phénomènes de *Gaming* sont apparus dans plusieurs marchés (nous le verrons dans la section 4) et parfois avec une intensité si forte qui ont provoqué le dysfonctionnement complet du marché.

Ces deux sources d'inefficacités pourraient être contre-balançées par le gestionnaire du réseau. Mais d'abord, le gestionnaire du réseau devrait avoir la compétence, le pouvoir et les incitations correctes. Le problème principal ici est de donner les incitations nécessaires et le pouvoir nécessaire au gestionnaire du réseau pour pouvoir réduire les coûts de congestions. On pourrait penser que la compétence nécessaire pour diminuer les coûts de congestions doit être accordé à la complexité du réseau (zonable ou non-zonable). Le pouvoir donné au gestionnaire du réseau correspond aux actions qu'il est autorisé à réaliser pour accomplir son objectif (par exemple, un gestionnaire du réseau fort serait autorisé à vendre et acheter de l'énergie afin de pouvoir améliorer l'efficacité du système). Finalement et la plus importante est la manière comment les incitations sont données au gestionnaire du réseau pour veiller à la fois sur l'efficacité et la sécurité.

3.2.2 Position temporelle de l'internalisation (J-1 et/ou J)

Étant donné que les « vraies limites de capacité » et les injections et soutirages finales ne sont connues qu'en temps réel, l'internalisation des externalités devrait se faire au moins en temps réel (c'est-à-dire que le marché du temps réel devrait avoir une définition des prix nodal ou

zonal). L'internalisation définitive d'externalités avant le temps réel devrait avoir des conséquences sur l'efficacité du marché. En principe, si l'on internalise les externalités en temps réel et si les participants peuvent prévoir correctement ces prix nodaux ou zonaux, les décisions de production (même pour les générateurs LENTS) pourraient se faire optimalement. Mais, ces prix différenciés géographiquement seront très variables et les participants peuvent vouloir se couvrir contre le risque des prix très variables. Deux solutions de design pour cela : soit on le fait de manière décentralisée (on laisse les acteurs choisir la forme et niveau de couverture) soit on le fait de manière organisée (ou semi centralisé) en créant des marchés (financiers) en J-1 où on utilise un réseau économique de transport au moins si détaillé que celui utilisé en J. Si nous supposons qu'il existe un arbitrage efficace entre J-1 et J (avec absence de pénalités et sans aversion au risque, par exemple), les prix nodaux (zonaux) en J-1 devront représenter les espérances mathématiques des prix nodaux (ou zonaux) en temps réel.

Les cas où nous pouvons trouver des inefficacités apparaissent quand les externalités sont internalisées avant le temps réel et quand elles ne sont pas « corrigées » en temps réel. Pour voir clairement ce problème il faut penser à un design de marché où l'internalisation d'externalités se fasse seulement en J-1. L'exemple le plus parlant est la proposition de marché intégré pour l'Europe Continentale. Dans la proposition de ce design, des capacités de transport entre les zones et des facteurs d'utilisations (par exemple, PTFD) sont définies et utilisées pour arbitrer les marchés régionaux en J-1. Il n'y a pas des mécanismes de coordination ou d'internalisation prévus pour le temps réel. Deux types d'inefficacités pourraient être présents ici : Premièrement, si la capacité exacte en J entre les zones est incertaine en J-1, la capacité mise à disposition pour le marché *forward* ne devrait pas être supérieure à la capacité minimale possible dans tous les états de la nature. Comme il n'y a pas de recoordination en temps réel, tous les états de la nature doivent être physiquement réalisables. Deuxièmement, quand le marché *forward* intégré « alloue » implicitement cette capacité en J-1, et donc les flux programmés entre les zones seraient optimisés pour des conditions « moyennes » dans chacune des zones. En temps réel, après la réalisation d'un état de la nature pour chaque zone (par exemple, demande totale), le flux fixé en J-1 peut n'être plus optimal. Une preuve de l'existence de ce type d'inefficacité serait donnée par la comparaison entre le signe (+ ou -) de la différence des prix en temps réel dans les deux zones par rapport à la direction du flux physique en temps réel.

3.3 Marché *forward* centralisé (organisé) ou décentralisé

Pour finir nous nous intéressons au degré de centralisation ou organisation des marchés *forward*. La plupart des designs de marché *forward* organisés sont des marchés volontaires. Mais différentes règles contraignent ou favorisent plus ou moins la participation des acteurs sur le marché *forward* « centralisé ». Le résultat de l'application de ces règles est que le marché *forward* « centralisé » capte un certain pourcentage du volume total des transactions d'énergie. Par exemple, quand il y a une obligation absolue de participer au marché *forward*, on trouve un cas extrême de centralisation. En revanche, quand il n'y a aucun marché *forward* organisé, on a un cas extrême de décentralisation. La plupart des marchés

en fonctionnement se trouvent entre ces deux cas extrêmes, c'est-à-dire, qu'ils sont en quelque sorte semi centralisés. Pour simplifier, nous définirons ici un modèle comme « centralisé » quand le pourcentage du volume de transactions traité par le marché *forward* organisé est important (disons, supérieur à 15 %) et qu'un modèle est « décentralisé » quand le pourcentage du volume de transactions est moins important (inférieur à 15 %).

Cette caractéristique de design est important pour trois points : l'information disponible pour les acteurs, la liquidité des marchés et la vitesse pour trouver un équilibre.

3.3.1 Information disponible

Dans cette perspective de décision sous incertitude nous pouvons étudier le design de marché *forward* « centralisé » ou « décentralisé » par rapport à l'information disponible pour les agents à un moment donné. Un point clé pour l'efficacité de la prise de décisions sous incertitude pour les acteurs est une capacité de prédiction assez large. Si l'on considère des hypothèses d'information complète, les participants du marché peuvent connaître parfaitement les différents états de la nature possibles et leur associer une probabilité et le profit correspondant. Simplement, tout le monde peut estimer les prix en temps réel avec sa distribution de probabilités. Cette propriété (que tous les agents ont la même information à un moment donné) est très importante. Du point de vue de l'information disponible, si tous les agents possédaient la même information au moment de prendre leurs décisions, le type du design du marché *forward* (centralisé ou décentralisé) ne jouerait pas un rôle important. En revanche, si l'on suppose que chaque participant possède une information limitée sur l'état de la nature possible pour le futur, le choix du design du marché *forward*, centralisé ou décentralisé, peut devenir important. A première vue, un modèle plus centralisé permettrait de corriger les erreurs de prédiction sur les états de la nature et les conséquences possibles pour le lendemain. Ici encore, on trouverait des différences selon le type du générateur. Car une erreur de prédiction du générateur LENT aurait des conséquences plus graves qu'une erreur de prédiction d'un générateur RAPIDE étant donné que ce dernier pourrait toujours modifier ses décisions de production. En cas de congestions, la prédiction des prix nodaux ou zonaux peut être beaucoup plus difficile que dans les cas sans congestions. La liaison entre les prix aux nœuds ne dépend pas seulement des conditions de l'offre et la demande en chaque nœud, mais aussi des conditions du réseau. Une centralisation du marché *forward* d'énergie (intégré au marché de capacité de transport) aiderait à ajuster les estimations des prix pour l'ensemble des participants et donc serait positive pour l'efficacité.

3.3.2 Liquidité

Nous avons déjà évoqué que à cause de la grande différenciation du produit électricité dans le temps et dans l'espace, le nombre de marchés élémentaires est assez important et donc souvent des marchés *forward* pour ces nombreux sous-jacents sont très peu liquides. Si le volume des ces marchés est déjà trop bas et si de plus on doit le distribuer entre plusieurs formes de marché, le problème de liquidité devient plus grave. De ceci on pourrait supposer que les marchés plus centralisés auraient une liquidité plus importante que les marchés plus décentralisés.

3.3.3 Vitesse pour trouver un équilibre

Un marché organisé peut trouver un équilibre offre demande plus rapidement qu'un marché décentralisé. Le temps est un facteur clé pour la production de l'électricité car lorsqu'on se rapproche du temps réel les moyens de production deviennent moins flexibles mais les incertitudes sont moins importantes. D'autre coté, les marchés organisés fonctionnent dans des guichet déterminé du temps et ne fonctionnent pas continuellement comme les marchés décentralisés.

Il faut voir aussi la flexibilité des arrangements, les coûts de transaction. **Non-convexités !!!**

4 Analyses de cas

En prenant en compte les caractéristiques décrites précédemment nous allons analyser les options choisies pour 4 designs de marché fonctionnant actuellement.

Dans le tableau suivant nous présentons les options pour chaque caractéristique.

Marché	Marché du temps réel ou mécanisme d'ajustement ?	Design d'internalisation d'externalités ¹⁴ en temps réel ou en J-1 ? zonale ou nodale ?	Marché <i>forward</i> centralisé ou décentralisé ?
ERCOT (Texas)	Marché	Temps réel/ Zonal	Décentralisé
PJM (Pennsylvanie)	Marché	J-1 et temps réel/ Nodal	Centralisé
NETA (Angleterre)	Mécanisme	Temps réel/ Zone unique	Décentralisé
Nord Pool (Scandinavie)	Marchés et mécanismes	J-1/ Zonal	Centralisé

Tableau N° 4.1 : Caractéristiques du design de marché

4.1 Cas texan

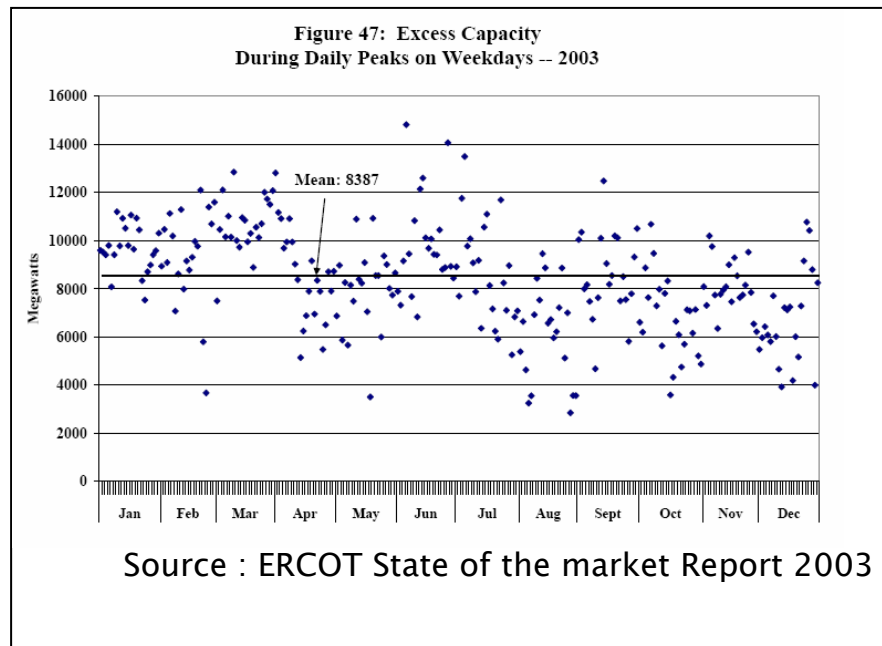
En temps réel, le design d'ERCOT est un marché du temps réel. Il existe un seul prix pour l'électricité en temps réel. Il est important de remarquer qu'ERCOT au début (2001) fonctionnait avec un « mécanisme d'ajustement ». La restriction à participer au marché du temps réel ne se faisait pas en utilisant des pénalités. ERCOT exigeait des programmes de production équilibrés correspondants quasi exactement à la prédiction de leurs consommations. Postérieurement ERCOT est passé à des programmes « relaxés » ou déséquilibrés (possibilité d'acheter ou de vendre de l'électricité sur le marché du temps réel au prix du temps réel¹⁵). Avant la relaxation par ERCOT de la règle des programmes équilibrés, la possibilité d'arbitrage temporel entre les marchés était très réduite. La règle de programmes équilibrés fonctionnait comme une barrière pour l'arbitrage [PUCT 2004]. Après la relaxation de la règle, bien que cela n'ait pas produit une grande augmentation du volume des transactions sur le temps réel (phénomène qui aurait pu avoir des conséquences pour la sécurité) les prix du temps réel et les prix *forward* se sont rapprochés. Cela signifie que les agents peuvent maintenant arbitrer entre les deux marchés. Ces arbitrages sont bénéfiques pour l'efficacité du système.

¹⁴ Pour NETA, OMEL et UK Pool, le terme correct serait ici « gestion de congestion » (et non « internalisation d'externalités ») car les coûts des congestions pour les trois marchés sont socialisés.

¹⁵ Néanmoins ces deux marchés appliquent aussi des charges sur les écarts. Dans ERCOT, ceux qui sont en écart doivent payer aussi un pourcentage des coûts d'appel de ressources supplémentaires en J-1, quand la capacité on-line prévue pour le lendemain n'est pas suffisante. Normalement ces coûts ne sont pas importants.

ERCOT a commencé à fonctionner avec un modèle économique de réseau zonal (4 ou 5 zones). L'internalisation d'externalités a été réalisée en temps réel. Un mécanisme pour résoudre des congestions locales a été conçu pour fonctionner juste après le marché du temps réel. Mais ce modèle a présenté plusieurs phénomènes de gaming et des inefficacités. Le coût pour résoudre les congestions locales a augmenté énormément entre 2002 et 2003 [PUCT 2004]. ERCOT pourrait être un exemple typique du choix incorrect du modèle économique du réseau. Ce design a utilisé une définition zonale des prix bien que le réseau ERCOT n'est pas tout à fait « zonables » [Baldick 2003] (en tout cas, pas zonables avec 5 zones). ERCOT va passer dans un modèle nodal bientôt. Des études de coût-bénéfice comparant les modèles zonal et nodal pour le réseau ERCOT ont été favorables au modèle nodal [Tabors-KEMA 2004].

Sur le marché ERCOT, il n'y a pas de marché *forward* organisé (centralisé). Le marché *forward* est complètement décentralisé. Des analyses faites sur ce marché ont conclu que le manque de marché organisé (et l'absence des prix de référence en J-1) a rendu difficile la tâche des générateurs pour prendre des décisions optimales de production. Une des preuves qui ont permis d'arriver à ces conclusions est le graphique suivant :



Dans ce graphique nous pouvons observer l'excès de capacité on-line pour les heures de pointe et les jours hors week end. Cet excès est calculé comme la différence de la puissance totale disponible à l'heure de pointe moins la capacité nécessaire pour fournir la demande à la pointe et les réserves. Simplement, cet excès représente des machines qui tournent sans produire ni énergie ni réserves. L'excès est relativement important, coûteux et peut représenter un manque de coordination (ou des mauvaises décisions de production). La solution proposée par les analystes est la création d'un marché *forward* centralisé.

4.2 Cas PJM

En temps réel le design à PJM est un marché du temps réel. Il existe un seul prix pour l'électricité en temps réel¹⁶. Les prix sont calculés tous les 5 minutes mais les règlements se réalisent avec la moyenne horaire de ces prix.

A PJM, le modèle économique du réseau a une définition nodale. L'internalisation se réalise tant en temps réel qu'en J-1. Le réseau de transport correspondant à PJM est assez congestionné et maillé donc on pourrait le classer comme un réseau « non-zonable ». En 1997, avant d'utiliser un modèle économique nodal, PJM a introduit un modèle zonal avec 8 zones (approx), le modèle zonal a failli complètement. La cause principale a été le phénomène de Gaming qui été provoqué. Les inefficacités n'ont pas été mesurées mais le fort gaming a provoqué le changement de design (nodal). Actuellement le marché fonctionne correctement.

PJM organise un marché en J-1. Ce marché n'est pas obligatoire (les participants peuvent présenter des programmes de production et consommation) mais des règles particulières font qu'une part conséquente (importante) de la capacité passe par ces marchés centralisés. La majeure partie des participants qui ont négocié des contrats bilatéraux, présentent des programmes de production dans le marché J-1 organisé pour se couvrir des coûts de transport. Tous les participants doivent payer la différence des prix du temps réel entre le nœud de soutirage et injection comme coût de transport. Ceux qui ont présenté des programmes en J-1 paieront la différence de prix nodaux en J-1 si en temps réel ils respectent ces programmes.

4.3 Cas anglais

En temps réel, le design sur NETA est un mécanisme d'ajustement. Dans ce design il n'y a pas de pénalités explicites. Mais il y a deux prix en temps réel pour ceux qui sont en écart : un prix pour acheter de l'énergie en temps réel (SBP-System Buying Price) et un prix pour vendre de l'énergie en temps réel (SSP-System Selling Price)¹⁷. Normalement le SBP est plus grand que le SSP, donc ces deux prix pénalisent les achats passifs et les ventes passives en temps réel. De plus, ces prix (surtout le SBP) sont très difficiles à prédire. Des problèmes liés aux « pénalités » ont été trouvés pour le NETA (Angleterre) [Cornwall 2002], [Newbery 2004]. En revanche, la « gate closure » a été placée très proche du temps réel et donc elle permet aux acteurs de modifier leurs positions contractuelles jusqu'à très proche du temps réel et donc de cette manière d'éviter des grandes expositions aux prix des écarts (positions contractuelles plutôt équilibrées). Un autre point important pour le NETA est la période de calcul du prix de déséquilibre, qui est de 30 minutes. Pendant cette période assez longue, il est difficile d'avoir une vision précise pour traduire les coûts en prix de déséquilibre. Sur 30 minutes il y a des générateurs qui peuvent être en équilibre global sur l'intégrale de l'énergie de la période mais pas en équilibre à tout moment de la période. Ils sont par exemple en

¹⁶ Dans PJM il y a aussi des charges pour les déviations par rapport aux programmes *day ahead*. Nous ne les considérons pas comme des pénalités parce qu'elles ne sont pas importantes et de plus parce qu'elles sont basées sur des coûts réels supportés par le système à cause de déviations.

¹⁷ Sur NETA il y a aussi une différence pour les écarts qui sont dans le même sens (augmentent l'écart global) que le système et ceux qui sont dans le sens contraire (aident à diminuer l'écart global).

déséquilibre négatif pour les 15 premières minutes et en déséquilibre positif pour le reste. Cela impose donc sur le système des coûts d'équilibrage supplémentaires. Étant donné que les coûts supplémentaires sont repartis seulement sur ceux qui sont en déséquilibre sur l'intégrale de la période, ces derniers doivent supporter les coûts provoqués par les autres [Cornwall 2002].

Avant NETA (UK POOL) et après NETA, le design concernant l'internalisation d'externalité utilisait une zone unique. Il n'y a donc pas d'internalisation d'externalités. Il y a eu des phénomènes de gaming et d'inefficacité (surtout avant NETA). Mais ces problèmes se sont solutionnés lorsque le gestionnaire du réseau a été incité correctement à minimiser ces coûts de congestions locales. Sur NETA des pouvoirs spéciaux ont été donnés aussi, le gestionnaire du réseau anglais NGC peut agir comme agent de marché, vendre ou acheter de l'énergie pour améliorer le fonctionnement du réseau. De plus, le réseau a été renforcé et donc crée un système sans congestions donc les congestions ont diminué [Smeers 2004b].

Il n'y a pas de marché organisé sur NETA. Il y a deux marchés privés mais le volume traité par ces deux marchés est très faible. En revanche, le marché des contrats bilatéraux est assez liquide. Deux raisons ont contribué à améliorer la liquidité du marché bilatéral : le renforcement des capacités de transport et les pénalités sur les mécanisme d'ajustement [Smeers 2004b]. Étant donné que les fortes pénalités présentes sur le mécanisme d'ajustement anglais, les acteurs sont incités à réduire les écarts au maximum. Pour réduire les écarts les acteurs doivent négocier des contrats jusqu'à très proche le temps réel (jusqu'à la gate closure). Le gain de liquidité est augmenté grâce aux pénalités. Aucune étude n'a été réalisée pour évaluer les bénéfices d'une plus grande liquidité contre les inefficacités provoquées par les pénalités.

4.4 Cas Nordpool

En temps réel le design du Nordpool est mixte. Nordpool est composé par plusieurs pays (et gestionnaire du réseau). Bien que depuis un certain moment l'intégration des designs a été commencée, chaque gestionnaire du réseau national gère son propre marché du temps réel ou mécanisme d'ajustement, surtout pour le règlement des écarts. Par exemple le design de marché en Norvège est un marché du temps réel. En revanche, en Suède, le design est un mécanisme d'ajustement. L'importance de la production hydraulique dans le système nordique le fait moins sensible aux incertitudes présentes entre J-1 et J, alors le choix du design est moins fondamental.

L'internalisation d'externalités se réalise en utilisant un modèle économique du réseau zonal et en J-1. Le réseau de transport en Scandinavie a été développé pour permettre le transport des grandes quantités d'énergie entre les zones de production à bas coûts et des zones de consommation. Le réseau est donc facilement « zonable ». Ce modèle d'internalisation zonal a donné des résultats corrects. Le fait que l'internalisation soit organisée principalement en J-1 n'a pas provoqué des grandes inefficacités. La haute disponibilité de moyens flexibles et de

bas coûts (hydraulique) permet de modifier les programmes facilement après le changement entre J-1 et J.

Sur Nordpool, il y a un marché forward en J-1 organisé ainsi qu'un très grand nombre des autres produits financiers avec des horizons temporels très grands. Nordpool est le marché où les différents produits financiers se sont mieux développés. Les produits financiers ont tous comme sous-jacent les produits vendus dans le marché forward J-1. La liquidité du marché forward J-1 est assez haute et cela vient principalement du fait de la grande capacité de production hydraulique présente en Scandinavie qui fait de l'électricité un produit un peu plus stockable et plus similaires aux autres commodités.

5 Conclusions et travaux futurs

En prenant en compte les caractéristiques « physiques » d'un système électrique (typologie des moyens de production, incertitudes et type du réseau de transport), nous avons choisi trois caractéristiques du design des marchés électriques. Pour chacune de ces caractéristiques, nous avons présenté des options et nous avons analysé les conséquences du choix de ces options.

Les caractéristiques choisies sont : le design en temps réel, le design d'internalisation d'externalités (provoquées par les congestions) et le design du marché *forward* en J-1. Les options pour le design en temps réel considèrent l'utilisation ou pas de pénalités. Les options pour le design d'internalisation d'externalités considèrent différents degrés d'internalisation (zonal/nodal) et différentes positions temporelles (J-1 ou/et J). Finalement, pour le design du marché *forward* en J-1 les options correspondent au degré de centralisation ou organisation de ce marché.

Dans un environnement incertain et avec les contraintes imposées par le système électrique (moyens de production, capacités des lignes), les différents designs (choix des options) donnent différents niveaux d'efficacité.

Après l'analyse de différents cas des marchés en fonctionnement nous avons trouvé un lien très clair entre les résultats de l'application de certains designs et les caractéristiques « physiques » propres aux systèmes électriques correspondants. Les designs de marchés ne sont pas directement transposables, un bon fonctionnement d'un design dépend de son adaptation à la réalité physique correspondante.

Certaines configurations physiques laissent plus la place pour des options de design. C'est le cas quand les réseaux ne sont pas trop congestionnés ou quand ils sont congestionnés mais zonables. Une capacité importante de production hydraulique avec réservoir facilite aussi le design du marché. Un gestionnaire du réseau fort avec des incitations correctes permettrait aussi de choisir plus facilement les designs, c'est-à-dire de choisir un compromis entre marchés et contrôles (NETA). En revanche, d'autres configurations physiques plus « défavorisées » n'ont pas beaucoup de flexibilité dans le choix d'options. C'est le cas pour le réseau non-zonables. PJM et ERCOT fonctionnent basiquement sur des réseaux non-zonables. Ils ont commencé à fonctionner avec deux designs complètement différents. Après quelques années de fonctionnement, ils convergent vers un design commun (marché du temps réel, modèle économique nodal, marché *forward* ?).

Le choix des différentes options a des conséquences sur l'efficacité. Ce choix d'options n'est pas ni tout à fait noir, ni tout à fait blanc, mais des décisions de compromis. Mais normalement ces décisions sont faites avec un certain degré d'ignorance sur les résultats. Nous avons essayé de montrer où et pourquoi ces inefficacités peuvent se présenter mais nous n'avons pas encore vérifié la taille de ces inefficacités. Il est très difficile de vérifier l'efficacité des marchés sur un marché qui fonctionne « correctement ». Des études

approfondies pour ceux qui ont l'information (gestionnaires du réseau) sont réalisés seulement quand quelque chose fonctionne vraiment mal (écroulement des marchés). Mais des pertes d'efficacité importantes provoquées par des défauts de designs peuvent être présentes même quand les marchés fonctionnent « correctement ». Un marché sans black-out n'est pas nécessairement un marché efficace. Des travaux futurs seront orientés à modéliser les options et comparer quantitativement dans la mesure du possible afin d'avoir une idée estimative des pertes (ou gains) d'efficacité après le choix des différentes options.

6 Bibliographie

- [Baldick 2003] Baldick R., « Shift Factors in ERCOT Congestion Pricing », WP disponible sur www.ece.utexas.edu/~baldick/papers/shiffactors.pdf, Mars 2003;
- [Benintendi-Boccard 2002] Benintendi D., Boccard N., « An Efficient Mechanism for Cross-Border Congestion Relief », Conference IDEI, Toulouse novembre 2002;
- [Boucher-Smeers 2002] Boucher J., Smeers Y., « Towards a common European Electricity Market – Paths 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 standard market design ? », disponible sur www.ksg.harvard.edu/hepg/Lessons_from_Abroad.htm, Mars 2003;
- [Chandley-et al. 2003] Chandley J., Danner C., Groves C., et al. « California's Electricity markets : Structure, Crisis, and Needed Reforms », LECG, disponible sur www.lecg.com, Janvier 2003;
- [Chandley-Harvey-Hogan 2000] Chandley J., Harvey S., Hogan W., « Electricity Market Reform in California », LECG, disponible sur www.lecg.com, Novembre 2000;
- [Cornwall 2002] Cornwall N. « The Need for Balance in NETA Imbalance Pricing », disponible sur www.ksg.harvard.edu/hepg, Mars 2002;
- [ETSO 2004a] ETSO (European Transmission System Operators), « Cross-border electricity exchanges on meshed AC power systems », disponible sur www.ets-net.org, Avril 2004;
- [ETSO 2004b] ETSO (European Transmission System Operators), « Flow-based Market Coupling », disponible sur www.ets-net.org, Septembre 2004;
- [Green-McDaniel 1999] Green R., McDaniel T., « Modelling Reta : A model of forward trading and the balancing mechanism », Departement of Applied Economics, Cambridge, Avril 1999;
- [Hirst 2001] Hirst E., « Real-time Balancing Operation and Markets : Key to Competitive Wholesale Electricity Markets », Edison Electric Institute, Avril 2001 ;
- [Hogan 2000] Hogan, W., « Flowgate Rights and Wrongs », disponible sur www.ksg.harvard.edu/hepg, Août 2000;

- [Hogan-et al. 1999] Hogan, W., Cadwalader M., Harvey S., Pope S., « Coordinating congestion relief across multiple regions », disponible sur www.ksg.harvard.edu/hepg, Octobre 1999;
- [Newbery 2002] Newbery D., « England's experience with NETA », présentation CAE, Oviedo, Espagne, Juillet 2002 ;
- [Newbery 2004] Newbery D., « Electricity Liberalisation in Britain : the quest for a satisfactory market design », présentation en conférence SESSA : « Refining Market Design », Cambridge, Juillet 2004;
- [Newbery-McDaniel 2003] Newbery D., « Auctions and Trading In Energy Markets – An Economic Analysis », Ch 10, CRI Regulatory Review 2002/2003;
- [PUCT 2004] Public Utility Commission of Texas, « 2003 State of the Market Report for the ERCOT Wholesale Electricity Markets », Potomac Economics, LTD., août 2004;
- [Roques-Newbery-Nuttall 2004] Roques F., Newbery D., Nuttall W., « Generation Adequacy and Investment Incentives in Britain : from the Pool to NETA », Cambridge Working Papers in Economics CWPE 0459, Octobre 2004;
- [Ruff 2000] Ruff L., « Flowgates, Contingency-Constrained Dispatch and FTRs », disponible sur www.ksg.harvard.edu/hepg, Octobre 2000;
- [Smeers 2003a] Smeers Y., « Market Incompleteness in Regional Electricity Transmission Part II: The Forward and Real Time Markets », Networks and Spatial Economics 3 (2): 175-196, Juin 2003;
- [Smeers 2003b] Smeers Y., « International Congestion Management », présentation en CMI Electricity Project Transmission Workshop, Cambridge, Juillet 2003 ;
- [Smeers 2004a] Smeers Y., « TSO, electricity markets and market power », présentation en « European Regulatory TSO benchmarking », Den Haag, Mai 2004 ;
- [Smeers 2004b] Smeers Y., « Marchés organisés et marchés de gré à gré en électricité », Revue de l'Institut d'Economie Publique no 14 - 2004/1, Décembre 2004;
- [Tabors-KEMA 2004] Tabors, Caramanis & Associates, KEMA Consulting, « Market Restructuring Cost-Benefit Analysis », Rapport final pour ERCOT, Novembre 2004 ;
- [Wolak-et al. 2002] Wolak F., Barber B., Bushnell J., Hobbs B., « Opinion on Oversight and Investigation Review », Market Surveillance Committee of the California ISO, Juillet 2002;