CEEPR Center for Energy and Environmental Policy Research

Merchant Transmission Investment

by
Paul Joskow and Jean Tirole

03-004 WP February 2003

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

Merchant Transmission Investment

Paul Joskow* and Jean Tirole†

February 10, 2003

Abstract

We examine the performance attributes of a merchant transmission investment framework that relies on "market driven" transmission investment to provide the infrastructure to support competitive wholesale markets for electricity. Under a stringent set of assumptions, the merchant investment model has a remarkable set of attributes that appear to solve the natural monopoly problem and the associated need for regulating transmission companies traditionally associated with electric transmission networks. We expand the merchant model upon which these conclusions are based to incorporate imperfections in wholesale electricity markets, lumpiness in transmission investment opportunities, stochastic attributes of transmission networks and associated property rights definition issues, the effects of the behavior of system operators and transmission owners on transmission capacity and reliability, coordination and bargaining considerations, forward contract, commitment and asset specificity issues. Incorporating these more realistic attributes of transmission networks and the behavior of transmission owners and system operators significantly undermines the attractive properties of the merchant investment model. Relying primarily on a market driven investment framework to govern investment in electric transmission networks is likely to lead to inefficient investment decisions and undermine the performance of competitive markets for electricity. A significant research challenge is to design regulatory mechanisms for system operators and incumbent transmission owners and a better framework for defining transmission property rights that will stimulate efficient investments by regulated incumbent transmission owners and by merchant entrants responding to market opportunities when they are the most efficient suppliers.

*MIT. Research support from the MIT Center for Energy and Environmental Policy Research and the Cambridge-MIT Institute is gratefully acknowledged.

†IDEI and GREMAQ (UMR 5604 CNRS), Toulouse, CERAS (URA 2036 CNRS), Paris, and MIT.

1 Introduction

Many countries are in the process of restructuring their electric power sectors to promote the developing of competitive wholesale and retail electricity markets. These programs involve unbundling the generation of electricity and the retail marketing of electricity from the transmission and distribution of electricity, creating competitive wholesale and retail markets for these services, and providing access to the transmission and distribution networks which provide the infrastructure platforms that support these competitive markets. Economic research supporting these developments has mostly focused on the organization and functioning of spot markets for energy and other generation services (e.g. operating reserves and frequency regulation). This theoretical and empirical research has analyzed, among other things, the organization of day-ahead and real time (balancing) energy markets and associated auction rules, the role of bilateral contracts, congestion management, nodal pricing, physical and financial transmission contracts, and associated market power issues. This work typically takes the transmission network as given, assumes that there is a fixed non-stochastic amount of transmission capacity available on the network, that the available capacity is unaffected by decisions made by the transmission owner and system operator, and that this capacity is common knowledge to all market participants, transmission owners and the system operator.

In reality, even in the short run, the capacity of a transmission network is stochastic as a consequence of facility outages and variations in external conditions such as weather. Moreover, the actual capacity of the transmission network under any particular set of supply and demand conditions depends on decisions made by the transmission owner (TO) (e.g. maintenance) and the system operator (SO) (e.g. actions designed to achieve target risks of system failures), which may (as in England and Wales) or may not (as in California and PJM) be the same entity. In the medium and long term as demand grows and new generating capacity is added to replace older less efficient capacity or to meet growing demand efficiently, investments in transmission capacity are likely to be necessary to minimize the overall costs of wholesale electricity supplies, to maintain reliability, to mitigate locational market power, and to improve the performance of competitive wholesale and retail markets. Indeed, most new investments in generation of any significant size must be accompanied by expansions of the transmission network.

Decisions regarding investments in new generation (including location) and transmission facilities are inherently interdependent. A new generator requires at least some supporting investment to connect it to the network. More interestingly, additional investments to expand generating capacity may be inefficient if the increased power flows from the new generator increase network congestion costs, constrain the operation of low-cost generating plants at particular locations, or reduce reliability. In addition, the locations chosen by new generators, and retirement decisions by existing generators at particular locations, will depend, in part, on forecasts of network congestion that may affect prices for generation service at different locations over many years into the future.[1] Finally, when there is transmission congestion, market power may be enhanced at particular locations where competition is limited by import constraints into the area. Locational market power leads to inefficiencies from dead-weight losses resulting from deviations of prices from marginal costs, from inefficient entry and other rent-seeking behavior, and from alternative imperfect market power regulatory mechanisms such as price caps.

When the electric power industry was made up of regulated vertically integrated monopolies, decisions about investments in generation and transmission and associated locational decisions were typically made jointly by the same firm, arguably internalising these interdependencies. In addition, potential market power problems that can arise when the

[1] Generator location decisions depend on many variables include the availability and price of land, the availability of cooling water, the costs of transporting fuel, the costs of connecting to the network, and the costs of congestion on the network at different locations. Generators' decisions to continue operating once investments have been sunk are likely to be more sensitive to locational prices for energy and operating reserves.

prices generators can charge for power are deregulated, were not an issue for regulated vertically integrated firms and their investment decisions did not take such market power considerations into account (Joskow 2002). Accordingly, in restructured electricity sectors where generation and transmission investment decisions are made independently and power prices deregulated, some governance framework must be found to facilitate efficient coordination of generation and transmission investments and to account for the short run and long run social costs of congestion, changes in reliability and market power.

Despite the importance of developing such a governance structure, and growing problems associated with stimulating transmission investment in many restructured electricity markets, there has been surprisingly little research on the institutions governing transmission investment in restructured competitive wholesale electricity markets. Early formulations of the structure for competitive wholesale markets envisioned the creation of independent regulated regional transmission and system operating entities (Transcos) that would be responsible for building, owning and operating transmission facilities and would be subject to economic regulation (Joskow and Schmalensee, 1983). More recent research has explored the attributes of incentive regulatory mechanisms that could be applied to such regulated transmission monopolies (e.g. Celebi, Nasser 1997, Léautier 2000, Vogelsang 2001) to integrate energy price (congestion) signals with transmission investment. We refer to this approach as a *regulated Transco* (or regulated Transmission Company) model. The institutional arrangements governing transmission operation and investment in England and Wales reflect this basic institutional approach. The regulated Transco model is necessarily subject to the classical challenges of regulated monopoly. How to specify and apply regulatory mechanisms that provide good performance incentives to the regulated firm while minimizing the economic rents that the regulated firm can earn given the asymmetry of information between the regulator and the regulated firm.

An alternative (or complement) to the regulated Transco model relies on decentralized property-rights based institutions to govern transmission investment (Hogan 1992; Bushnell and Stoft 1996,1997; Chao and Peck 1996). The hope is that by relying on competitive "market driven" transmission investment, the imperfections associated with the institution of regulated monopoly can be avoided. The market-based approaches envision new transmission investment creating transmission rights for the merchant investor (either physical or financial as described in Joskow and Tirole 2000)[2] based on the increase of the capacity of the network to transfer power from points of injection to points of consumption. The value of these transmission rights, which are typically equated to the expected congestion charges either avoided (physical rights) or rebated by the system operator (financial rights) over the life of the transmission investment, then provides the financial incentive for incumbent suppliers or new entrants to invest in new transmission capacity. We call this the *merchant transmission* model and its attributes are the focus of this paper.

Research on this model has focused almost entirely on simple cases where transmission investments are characterized by no increasing returns to scale, there are no sunk cost or asset specificity issues, nodal energy prices fully reflect consumers' willingness to pay for energy and reliability, all network externalities are internalised in nodal prices, transmission network constraints and associated point-to-point capacity is non-stochastic, there is no market power, markets are always cleared by prices, there is a full set of futures markets, and the TO/SO has no discretion to affect the effective transmission capacity and nodal prices over time. Under these assumptions it can be demonstrated (a) that efficient transmission investments that create transmission rights satisfying certain simultaneous feasibility constraints will be profitable and (b) that inefficient transmission investments will not be profitable (Hogan 1992; Bushnell and Stoft 1996,1997). These two results are the primary economic foundation for relying on a merchant transmission model. While

[2]As we will discuss, however, the nature of the rights that are typically assumed in the literature appear to be poorly adapted to important physical attributes of transmission networks.

there has been some recognition that relaxing these assumptions undermines key results regarding the optimality of merchant investment (e.g, Hogan 1992, pp.228-230, Bushnell and Stoft 1996, 1997; Oren et al 1995), little analysis of more realistic cases has been forthcoming (Perez-Arriaga et al, 1995 is an exception).

In this paper we examine the performance attributes of this new *merchant transmission* investment and ownership framework when assumptions that better reflect the physical and economic attributes of real transmission network are introduced. We show that a variety of potentially significant performance problems then arise. While the focus of our analysis is electric transmission networks, several of the issues that we address arise in connection with decentralized market driven investments in network infrastructure capacity in other sectors, including railroads, highways, and gas pipeline networks. The paper proceeds in the following way. The next section provides additional background regarding competitive wholesale electricity market institutions, the allocation and pricing of scarce transmission capacity, transmission rights, and the nature of investments in transmission capacity. Section 3 then outlines the attributes of the basic merchant transmission model that has appeared in the literature and the case for its attractive performance atributes. In the sections that follow, we examine how these results are affected by imperfections in wholesale energy markets, lumpy transmission investments, the stochastic properties of transmission capacity and the associated definitions of property rights, network operator behavior, coordination issues, and extensions to account for loop flow. We find that the attractive properties of the merchant transmission model are seriously undermined when more realistic characterizations of transmission networks are introduced.

2 Merchant transmission investment: background

No restructured electric power industry has adopted a pure merchant transmission model of the type described above, though Australia has adopted a mixed merchant and regulated transmission model.[3] However, recent academic proposals,[4] as well as FERC's July 2002 Standard Market Design (SMD) proposals, call for relying primarily on "market driven" transmission investments, while recognizing that at least some regulated transmission investments may be necessary.[5] The extreme version of market driven investment relies entirely on free entry of investors into the activity of constructing transmission lines and no regulation of the prices that they can charge. The owners of these transmission lines are rewarded through the congestion rents associated with these lines.

Transmission investment institutions cannot be considered independently of the institutions that govern the determination of energy prices, operating reserves, contingency constraints, congestion management, and the specification of transmission capacity and increments to it. No single paradigm has emerged from the liberalization effort of the last decade for these attributes of the design and operation of wholesale markets, system operations, and congestion management. So, to evaluate the properties of alternative transmission investment frameworks we need to be more precise about the organization of the wholesale market, congestion management and price determination to understand and evaluate alternative institutional frameworks to govern transmission investment. In what follows we will assume that a nodal or locational marginal pricing (LMP) system is in place with attributes similar to those being proposed by the U.S. Federal Energy Regulatory Commission (FERC) in its SMD proposals and to what is in operation in New

[3]Two merchant lines supported by differences in spot prices in the two market areas they connect have been placed in operation under this arrangement in Australia. Directlink is a 180 MW, 40 mile merchant DC link connecting Queensland and New South Wales and began operating in 2000. Murraylink is a 220 MW, 108 mile merchant DC link connecting South Australia and Victoria which began operating in October 2002. On October 18, 2002, Murraylink applied to the regulatory authorities in Australia to change its status from a merchant line to a regulated line that would be compensated based on traditional cost of service principles combined with a performance incentive mechanism. Neither merchant link appears to be profitable. As far as we can tell, these are the only two merchant transmission lines operating anywhere in the world that have been built in anticipation of recovering their costs entirely from congestion rents arising from the difference in nodal prices.

[4]e.g., Hogan (2002).

[5]*Remedying Undue Discrimination through Open Access Transmission Service and Standard Electricity Market Design*, Federal Energy Regulatory Commission, Notice of Proposed Rulemaking, July 31, 2002. See in particular ¶335–351.

York and PJM in the U.S. This is the most conducive framework for merchant investment because nodal prices provide a measure of locational scarcity.

Under this model, an independent system operator (ISO)[6] operates a real-time balancing market and allocates scarce transmission capacity using bids from generators and consumers to increase or decrease generation or demand at each node. That is, the ISO takes all of the bids (generation and demand) and finds the "least cost" set of uniform market-clearing price bids to balance supply and demand at each generation and consumption node on the network using a security constrained dispatch model that takes transmission constraints into account. This establishes day-ahead quantity commitments and nodal prices that vary by location when there is congestion. The resulting nodal prices reflect both congestion and marginal losses. Generators may enter into bilateral contracts with marketers or load serving entities (LSEs) and schedule supplies with the ISO separately from the organized day-ahead market. However, they still have to pay any congestion charges associated with their schedules based on the difference in nodal prices between the injection and receipt points. The day-ahead schedules, nodal prices, and congestion charges are "commitments." They can be adjusted in real time by submitting adjustment bids to the real time balancing markets (which again rely on bids, a security constrained dispatch and nodal prices) to allow these schedules to be changed based on real time physical and economic conditions.

The FERC SMD model recognizes that there are incumbent regulated transmission owners (TO) that own the existing transmission assets and requires that the SO and TO be separate entities and operate independently.[7] The incumbent TO receives some cost-of-service compensation for the usage of a grid that it no longer controls to compensate it for legacy investments and ongoing maintenance costs. New investments in transmis-

[6]Renamed Independent Transmission Provider (ITP) in the FERC proposal.

[7]Exactly what functions are assigned to the TO and what functions to the SO is a subject of continuing debate and depends in part on whether the TO is "independent" of generators and marketers that use its facilities to participate in the wholesale market.

sion are anticipated to be made by competing merchant investors whose compensation is based on the value of Congestion Revenue Rights (CRRs)[8] created by their investments.[9] These financial rights represent the right to receive congestion revenues defined as the difference between the locational prices between the two nodes (point-to-point) covered by the relevant CRR times the quantity of CRRs held. In Joskow and Tirole (2000) we defined these rights as representing a *share* of the congestion revenues (or merchandizing surplus) earned by the system operator. This formulation implies that the obligations to pay rights holders is always the same as the congestion revenues earned by the system operator. Under FERC's proposed formulation, however, the quantity of point to point financial rights is fixed ex ante and allocated to holders to reflect estimates of the capacity of the network to accommodate schedules that fully utilize these rights under "normal operating conditions." In this case, deviations between actual transmission capacity and the number of allocated rights results either in the congestion revenues earned by the SO being too little to fully cover the associated financial obligations to rights holders or in congestion revenues in excess of what is owed to rights holders. For example, if K rights are issued to inject power at node 1 and receive it at node 2, the rights holders are owed $K(p_2 - p_1)$, where p_1 and p_2 are the prices at the 2 nodes. If the actual capacity of the network turns out to be K_a then the system operator will have a congestion revenue deficit or surplus equal to $(K - K_a)(p_2 - p_1)$.

The separation between transmission ownership (TO) and network dispatch (system

[8]Congestion Revenue Rights is the name FERC has now given to financial rights that have been referred to in the literature as Transmission Congestion Contracts (TCCs) or Financial Transmission Rights (FTRs).

[9]While the FERC SMD reflects a preference to rely on "market driven" transmission investments, it recognizes that because market forces "may not" yield efficient transmission investment and provides backstop investment policies. The primary backup policy proposed was some type of (largely undefined) competitive RFP process, with regulated investments by incumbent transmission owners considered to be a "last resort." Responding the comments received since the SMD rules were proposed, as this is written, the competitive RFP proposal appears to be off the table, and some combination of merchant investment and regulated incumbent TO investment mediated through a regional planning process appears to be gaining favor.

operating or SO) functions in this model is motivated by two considerations. First, a market driven transmission system leads to multiple owners of parts of the grid; while the owners can form a cooperative to operate the grid, their goals are in general antagonistic,[10] and it is well-known that cooperatives of members with heterogeneous interests face complex governance problems.[11] Second, and quite crucially, grid owners face a serious potential conflict of interest when operating a transmission grid *if* their compensation varies directly with the level of congestion rents. In practice, due to the lack of market-based penalties for outages, dispatching does not quite correspond to the least-cost optimization used in economic and engineering models; rather, grid operators have substantial discretion over how much outage they are willing to take while dispatching.[12] This discretion in turn potentially provides incentives for system operators to manipulate the congestion rents received by the owners (Glachant and Pignon, 2002). By conservatively "withdrawing" transmission capacity (under the cover of a safe management of the network), the system operator can substantially raise the congestion rents. Third, in many countries there continues to be vertical integration between generation power marketing and transmission. The creation of an independent SO is thought to be a way to mitigate the potential problems that may arise from common ownership and control of transmission and generating assets and associated power marketing activities. However, the separation of ownership (TO) and operations (SO) carries other potential costs caused by inefficient coordination between the SO and the TO. Accordingly, there is a tradeoff between integration of TO/SO functions and separation of these functions that has largely been ignored in the literature and by policymakers. We examine the resulting

[10] Incumbent owners of transmission lines that are being compensated based on congestion rents will have incentives to oppose investments by others in generation or transmission that reduce these congestion rents. Generators located in congested areas will have incentives to oppose transmission enhancements that would reduce or eliminate the congestion.

[11] See, e.g., Hansmann (1996).

[12] For example, the so-called (N-1) and (N-2) constraints are crude self-imposed constraints and are subjective responses to the perceived risk.

"moral hazard in teams" issues further below.

Finally, the restructuring of electric power systems to rely on competitive wholesale markets does not start with a blank slate. There generally exists an extensive legacy transmission network and an associated fleet of generating plants. The configuration of these assets may not be "optimal" in the ex ante sense for at least two reasons. First, supply and demand conditions are likely to have changed from what was assumed when these investments were made. Second, the investments were not made to be optimally configured to accommodate a decentralized competitive wholesale market. For example, vertically integrated firms would not have taken local market power problems into account since they would have had no incentive to exercise market power against themselves (Joskow 2002). The configuration of the legacy network must be taken into account in the evaluation of alternative institutional arrangements to govern its operation and investments to expand its capabilities.

For these reasons, we have found it useful to consider two types of transmission investments that can increase the capacity of the network (or, alternatively reduce congestion) to accept injections of energy at a particular location A on the network for consumption at another point B on the network.[13]

Network deepening investments: These are investments that involve physical upgrades of the facilities on the incumbent's existing network (e.g. adding capacitor banks, phase shifters, reconductoring existing transmission links, upgrading transformers and substations, installing new communications and relay equipment spread around the network to increase the speed with which the SO can respond to sudden equipment outages and re-

[13] We focus on transmission investments that affect congestion on the high voltage network. Regulators in the U.S. often break transmission investments down into additional categories. First, there are local transmission investments "inside" the demand node. These investments are sometimes called "reliability" investments. Second, interconnection investments are investments that must be made by an incumbent grid owner to connect new generators with the rest of the network. These are often treated like radial links and are typically paid for by the generators seeking interconnection. However, it is hard to draw bright lines between reliability investments, interconnection investments, and investments that affect network congestion.

lax contingency constraints, etc.). These are investments that are physically intertwined with and inseparable from the incumbent TO's facilities. These investments are specific investments (as described by Williamson 1983) that we assume can be undertaken most efficiently by the incumbent network owner. Similar to network deepening investments are *network maintenance* decisions. Like network deepening, maintenance is most efficiently performed by the owner of the link or grid.

Independent network expansion investments: These are investments that involve the construction of separate new links (including parallel links) that are not physically intertwined with the incumbent network except at the point at either end where they are interconnected. These investments can (in principle) be made either by incumbent transmission owners, by stakeholders (generators, load-serving entities), or by a third-party merchant investor. The two operating DC merchant links in Australia appear to fall into this category. However, as in Australia, these links may have effects on power flows on the rest of the network, including on parallel lines, but are physically separable projects from a construction and maintenance perspective.

Remark: The merchant investment paradigm requires that there effectively be free entry into the development of new transmission capacity. One can think of at least two situations in which free entry is not a good assumption. First, *network deepening* investments can, as a practical matter, only be implemented efficiently by the owner of the existing lines. Defining an efficient "competitive access to deepening investments" policy is likely to be extremely difficult for several reasons. First, adding third-party facilities that are fully integrated with the existing network from a physical and maintenance perspective creates significant incentive problems with decentralized ownership. The problems of defining a good set of rules for investing in and maintaining facilities of this type with decentralized ownership is further exacerbated by the heterogeneous nature of transmission facilities. While it is theoretically possible to devise contractual arrangements that will solve the incentive problems, including opportunistic behavior of one or more parties, investments with these attributes are most likely to be governed efficiently through ownership by a single firm ; second and relatedly, one would need to carefully allocate the new capacity of the line between the initial design and maintenance choice of the original owner and the actions of the renters who make deepening investments. This "moral hazard in teams" problem is a substantial obstacle to the design of an effective third party access policy for this type of transmission investment.

This raises the question of how incumbent transmission owners are to participate in a "market driven" transmission investment framework. On the one hand, precluding them from participating would mean that potentially low-cost network deepening investments will be lost. On the other hand, allowing them to make unregulated merchant investments for network deepening enhancements to which they have unique access would allow them to exercise market power, restrict supplies and capture rents that might otherwise go to consumers under a regulated investment regime. It is natural then to think about allowing incumbents to make regulated investments and new entrants make merchant investments. However, to the extent that the regulated and merchant investments involve parallel lines whether or not the most efficient investments are made will depend heavily on the regulatory mechanisms adopted. Mixing regulated and unregulated activities that are (effectively) in competition with one another is always a very challenging problem. In Australia, this mixture of competition and regulation has led to extensive litigation between proponents of regulated and merchant transmission links, delaying investments in both.[14]

[14] Comments of Transgrid on Standard Market Design, FERC NOPR Proceeding, January 10, 2003.

3 The case for merchant transmission investment

Let us start from the theoretical case for market driven or "merchant" transmission investment (this rationale has been developed, inter alia, by Hogan 1992 and Wu et al 1996, Chao and Peck 1996 and examined further by Bushnell and Stoft 1996, 1997 for simple cases.) The basic argument is conveyed in the two-node framework of figure 1.

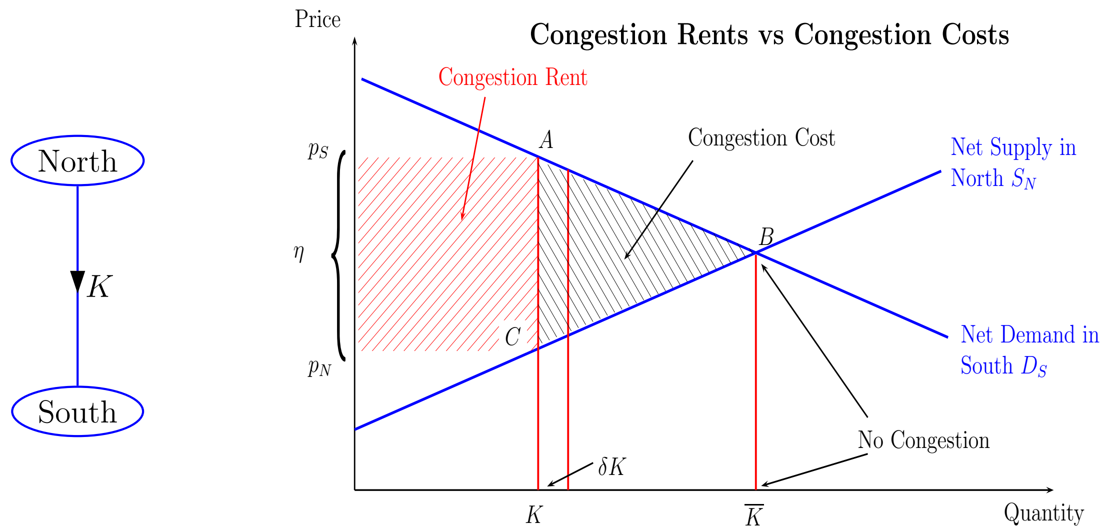

Figure 1

Figure 1 depicts a simple situation in which load serving entities (distribution companies or marketers buying on behalf of retail consumers, or large industrial customers buying directly) in the South (say, a large city) buy their power from cheap generation sources in the North and, possibly, more expensive sources in the South. Alas, the capacity of the line from North to South is limited to K, and faced with net demand /supply curves in the North and the South, the system operator is forced to dispatch "out of merit". For example, the system operator calls on expensive generators in the South while generators in the North would be willing to supply this amount at a lower price if more transmission capacity were available. The rationing of the scarce North-South capacity is implemented by setting two nodal prices, p_S and p_N that clear the markets in the South and the North, respectively. The difference, $\eta = p_S - p_N$ is the shadow price of the transmission capacity constraint. The area ηK is the *congestion rent* and the triangle ABC in the *congestion* or *redispatch cost*. The latter represents the cost of running more costly generation in the South because less costly imports from the North are limited by transmission congestion. The congestion *rent* and the congestion *cost* are sometimes confused and it is important to recognize their appropriate definitions and meaning.

Now consider a *marginal* (unit) increase in transmission capacity (δK). This unit increase allows one more KWh to flow from North to South, replacing a marginal generator in the South with cost p_S by a cheaper generator in the North producing at cost p_N. That is, the social value of the investment is given by the reduction in the area ABC in Figure 1.

Assume that the builder of this marginal capacity, whether it is a new entrant or the incumbent TO, is rewarded through a financial transmission right that pays a dividend equal to the shadow price of the transmission constraint. A non-incumbent merchant company will enter to build this extra capacity as long as η exceeds the cost of building it. By contrast, *if* an incumbent grid owner is compensated through the payment of congestion rents, it may not want to make this marginal investment as it must compare the extra revenue η net of the cost of expanding the capacity with the reduction in the congestion rent on its inframarginal transmission units ($-K d\eta/dK$). It is only when the incumbent grid owner's capacity has been rated at some level K^* not too different from actual capacity, and that the corresponding rights, with value ηK^*, have been auctioned off, that the monopoly distortion vanishes. The incremental capacity then yields $\eta + (K - K^*)\,\dfrac{d\eta}{dK}$ close to η. As in the case of Contracts for Differences,[15] forward sales restore proper incentives for a player with market power.

Hogan (1992) and Bushnell and Stoft (1996, 1997) show that under certain condi-

[15]See Green (1992).

tions (e.g. no increasing returns to scale, simultaneous feasibility constraints bind when awarding congestion rights, efficient nodal prices clear all markets, no market power in the wholesale market, well defined property rights, a complete set of competitive liquid forward markets to provide sufficient statistics for long run demand and supply conditions and risk management, etc.), all *efficient* transmission investments will at least recover their costs from congestion revenues and that *inefficient* investments will not be profitable. These are potentially powerful results that appear to transform the transmission investment problem from one that appears to be almost intractable to one that requires a simple implementation of a property-rights based market system.

Accordingly, merchant investment's appeal is that it allows unfettered competition to govern investment in new transmission capacity, placing the risks of investment inefficiencies and cost overruns on investors rather than consumers, and bypassing planning and regulatory issues associated with a structure that relies on regulated monopoly transmission companies. In addition, in theory, it allows investment in new generating capacity in the constrained area to "compete" with new transmission investment that reduces the import constraint. In this way, market driven transmission investment is an economist's dream, solving the problems associated with imperfect regulation of a "natural monopoly" transmission company and aligning competitive transmission investments with the newly developed competition in the generation segment. Unfortunately, the optimality of the market driven approach depends on a number of strong assumptions and conditions that are likely to be inconsistent with the actual attributes of transmission investments and the operation of wholesale markets in practice. (some of the critiques will apply to alternative frameworks as well).

We turn now to a discussion of what we view as the most important assumptions underlying the case for the merchant model and the implications of relaxing these assumptions. We will assume that wholesale markets are organized around the "nodal pricing" model utilized in PJM, New York and proposed in FERC's SMD. We will ignore issues associated with common ownership of generation and transmission and, following the FERC SMD, assume that the TO and SO are independent. However, unlike much of the analysis underlying this model we will recognize that the TO and SO have objective functions, can make discretionary decisions that can affect market performance, including reliability risks, respond to incentives they face (including political pressures in the case of non-profit SOs), and that TO and SO decisions may be interdependent leading to potential costs of imperfect coordination between them.

4 Imperfections in energy markets

The reasoning above assumes that the prices that clear the markets in the North and the South reflect the marginal costs of production (and the marginal willingnesses to pay[16]) at each location, so that the congestion rent perceived by merchant investors does reflect the social savings brought about by the investment. That is, potential investors in new transmission capacity see the correct locational price signals in the wholesale markets. There are a number of reasons why this is unlikely to be the case. Market power may distort nodal prices; regulatory interventions like price caps may distort prices; the absence of a complete representation of consumer demand in the wholesale market may distort prices; discretionary behavior by system operators may distort prices under "extreme" conditions when the network is constrained.

Suppose for example that there is a generator with market power in the South, and that the latter region is import constrained. The generator exercises market power by withdrawing capacity and driving the price in the South up. Hence

$$p_S > c_S,$$

[16]We will present the argument in terms of cost savings; because what matters is net supply at each node, the same argument would apply to the demand side.

where c_S denotes the marginal cost of production in the South. The measured congestion rent then overestimates the cost savings associated with the replacement of one unit of power generated in the South by one unit of power generated in the North, suggesting an *over-incentive* to reinforce the link, ignoring the potential impacts of other market imperfections.[17] On the other hand, the increased transmission capacity does not replace production in the South one-for-one; it also leads to an increase in total energy consumption in the South, which yields a social benefit equal to $(p_S - c_S)$ times the expansion in consumption. The box below shows that, under a very weak assumption,[18] the first effect dominates, and therefore market power results in an over-incentive to invest in merchant transmission. Similarly, and to the extent that reinforcing the line is akin to adding production capacity in the South, this suggests that entrants in generation have too much of an incentive to invest in the South as well. The box verifies that this is indeed the case. Thus, market power in the importing area does not reduce incentives to invest in additional transmission capacity between the exporting and importing areas. Other things equal, market power in the importing area produces enhanced incentives for transmission investment.

The impact of locational market power on merchant investment incentives

• Consider a monopoly supplier in the South producing at marginal cost c_S and facing demand function $D(p_S)$. Provided that the transmission capacity K between North and South is fully utilized, the monopolist solves:

$$\max_{p_S} \left\{ (p_S - c_S) \left[D\left(p_S\right) - K \right] \right\};$$

equivalently, this monopolist selects a consumption q_S in the South so as to solve :

[17] For example, as we discuss below, lumpiness in transmission investment leads to *under-investment*.

[18] The assumption is that the Southern monopolist's reaction curve be downward sloping in a Cournot game. Intuitively, the transmission line creates a Cournot "duopoly" in the South, in which the Southern firm faces a fixed output from its (transmission) rival. A downward sloping reaction curve means that the Southern firm curtails its output as the transmission capacity expands. This implies that the energy consumption increase effect is smaller than the inflated signal effect (the two effects would cancel if the output in the South were invariant to an increase in imports from the North).

$$\max_{q_S} \left\{ \left[P\left(q_S\right) - c_S \right] \left(p_S - K\right) \right\},$$

where $P(\cdot)$ denotes the inverse demand function and $S^g(\cdot)$ the gross surplus. Neglecting consumption in the North, social surplus is

$$W = S^g\left(q_S\right) - c_N K - c_S \left[q_S - K\right].$$

The marginal gross surplus, dS^g/dq_S is equal to price p_S in the South, and so when the line's capacity is increased by dK, resulting in a consumption change dq_S, welfare changes by

$$dW = \left(p_S - c_S\right) dq_S + \left(c_S - c_N\right) dK.$$

Note that, with *perfect competition*, $p_S = c_S$ (and $p_N = c_N$) and so $dW = \eta dK$. With monopoly power in the South, though,

$$p_S - c_S > 0$$

and

$$c_S - c_N < \eta.$$

There is an over-incentive to invest if and only if

$$dW < \eta dK,$$

and

$$\left(p_S - c_S\right) dq_S + \left(c_S - c_N\right) dK < \left(p_S - c_S\right) dK,$$

that is if and only if

$$dq_S < dK,$$

For there to be an over-incentive to invest, the monopolist must "absorb" some of the increase in imports from the North. To know whether this is the case, differentiate the first-order condition for profit maximization:

$$\frac{dq_S}{dK} = \frac{P'}{2P' + \left(q_S - K\right) P''}.$$

Thus, there is an over-incentive to invest under merchant investment if and only if

$$P' + \left(q_S - K\right) P'' < 0.$$

A sufficient condition for this is that the demand curve be concave. More generally, this condition is the standard condition for quantities to be strategic substitutes.

• The same reasoning can be applied to generation investments in the South. Indeed, K could alternatively denote the amount of power produced by a competitive fringe in the South in the profit maximization exercise. And so

$$dq_S < dK$$

as long as

$$P' + (q_S - K) P'' < 0.$$

There is an over-incentive to invest if and only if

$$dW - (p_S - c_S) dK < 0$$

or

$$d [S^n (q_S) - c_S q_S] < (p_S - c_S) dK$$

$$\iff \quad (p_S - c_S)(dK - dq_S) > 0.$$

Hence, there is in general an over-incentive to invest in generation in the South as well.

Conversely, a generator with market power in the North may (while still making full use of the link) be able to raise price p_N by withdrawing production capacity — perhaps to the level of p_S if it faces no competition in the North (Oren, 1997; Stoft, 1999; Joskow and Tirole 2000). In this case, the congestion rent underestimates the gain from expanding the line's capacity, resulting in an *under-investment* by merchant transmission investors. At the same time, it could lead to inefficient entry of generating capacity in the North in response to the short run monopoly rents created there.

As a second example, in the case of market power in the South (this is the situation that will generate very high prices for consumers in the South), the regulator may be tempted to impose a price cap.[19] While the price cap improves economic efficiency if it really is about constraining market power, it may also distort price signals if high prices are at least sometimes due to tight competitive supply and demand conditions rather than market power. A cap $p_S \leq \bar{p}_S$ then reduces the congestion rents during those hours that are very important because they produce the bulk of the rents to support investment, yielding *under-investment* in transmission.

Third, prices may not clear supply and demand in real time because market clearing processes are not fast enough to respond to rapid changes in supply and demand condi-

[19] Or a de facto price cap as when the system operator curtails load administratively when prices don't clear the market.

tions while maintaining physical requirements for frequency, voltage, and stability on the network. To maintain physical network parameters, administrative rationing is then substituted for prices to balance supply and demand as a consequence of what is effectively a problem of incomplete markets (Wilson 2002). Moreover, the system operator has discretion in determining the exact nature of the responses to operating reserve deficiencies or "scarcity." For example, Patton (2002) shows that during tight supply conditions the system operator takes actions that tend to depress market-clearing prices. Whether it is administrative rationing in response to incomplete markets or price controls motivated by efforts to constraint market power or price distortions caused by market power or discretionary decisions by the system operator, actual prices will depart from the efficient prices required to give the efficient signals for new investment. These imperfections are potentially important with regard to transmission (and generation) investment because the prices that create significant congestion rents tend to occur in a relatively small number of hours and these hours also happen to be the hours when these types of price distortions are most likely to occur.

5 Lumpy transmission investments

The analysis of the effects of market power in investment incentives in the previous section, as well as most of the literature upon which the merchant investment model relies, assumes that transmission investment is not characterized by economies of scale or "lumpiness". However, *network expansion investments* are likely to be lumpy. That is, the average cost of a new link declines as its capacity increases, other things equal (Baldick and Kahn 1992, Perez-Arriaga et al, 1999). (Many *network deepening investments* may be less lumpy, but as we have already discussed, these investments are most conducive to investment by the incumbent network owner rather than a merchant entrant. We discuss deepening investments further below.) The impact of lumpiness is illustrated in figure 2. The initial capacity is K_0 and economically can be brought to a level K_1.

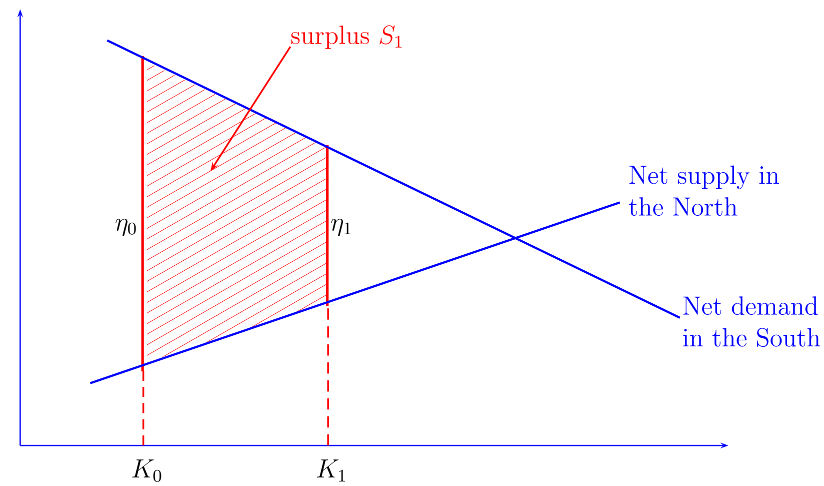

Figure 2

Assuming that the energy market participants are perfectly competitive, so that net demand/supply curves represent the true marginal costs/willingnesses to pay, and the market is cleared by efficient nodal prices, the surplus created (or congestion cost reduced) by the expansion of capacity from K_0 to K_1 is depicted by the shaded area in figure 2. The value, $\eta_1 (K_1 - K_0)$ of the transmission rights (equal to the ex post congestion *rents*) granted to the merchant investors building this capacity expansion understate the social surplus S_1 it creates by reducing congestion *costs*. Lumpiness thus results in an *underincentive* to reinforce the system for the same reason that an incumbent grid owner rewarded by congestion rents has suboptimal incentives to remove these congestion rents.

Another source of lumpiness for *network expansion* investments arises because there may be a *scarcity of rights of way*, for example a unique corridor between a cheap and an expensive area. The difficulties that new transmission corridors face in obtaining siting authority suggests that the available corridors for new lines through many areas will be limited in the sense that (for example) one additional corridor may be available through the Pyrenees between France and Spain, and it may accommodate *one* new link that could be of any size between 100 MW and 1000 MW. This scarcity is particularly problematic as demand grows. Merchant investment is then likely to end up in a "preemption and monopoly" situation. A merchant will install a small capacity on the corridor to gain a toehold and will later expand this capacity (presumably, the merchant will underinvest in this expansion as we have seen), to the extent that the expansions are now deepening investments. To be certain, under perfect competition, rents will be dissipated through a very early entry into the scarce corridor, or, if the corridor is put up for auction, through high bids. But the outcome is then similar to a monopoly outcome. Moreover, scarce corridors are typically not allocated through auctions but rather through a regulatory process that places a premium on being first in line.

Besides generating too little investment, lumpiness also may make merchant investment occur too early when it takes place in order to pre-empt additional entry. In a system with growing demand, pre-emption leads to an investment at the first date at which the discounted value of the financial rights on the additional capacity is equal to the investment cost. It could also lead to the investment being "undersized". For example, if the optimal investment is 600 MW, a merchant developer may find it most profitable to invest in a 300 MW enhancement, pre-empting additional investment. The box below contains an analysis of the incentive to get a toehold by sinking a small investment to pre-empt additional entry and produce monopoly rents for the merchant transmission investor and under-investment in transmission capacity, generally.

Lumpy investments: preemption and toeholds

• Suppose that at some future date T, the net demand in the South jumps up to a new level (the dotted line in figure 3); the post-reinforcement shadow price jumps from η_1 to $\eta_2 > \eta_1$. Letting r denote the interest rate and I the investment cost, suppose that

$$\eta_1 (K_1 - K_0) < rI < \eta_2 (K_1 - K_0) .$$

Then, under free entry into merchant transmission investment, investment occurs at date $\tau < T$ such that

$$\left[\frac{1-e^{-r(T-\tau)}}{r}\eta_1+\frac{e^{-r(T-\tau)}}{r}\eta_2\right](K_1-K_0)=I.$$

Note that this preemption is actually socially beneficial if the surplus S_1 brought about by the expansion before the increase in demand exceeds the interest on the investment cost, i.e.:

$$S_1>rI.$$

Otherwise, preemption is socially wasteful.

And the point about underinvestment remains: Letting S_2 denote the surplus after demand has grown, if

$$\eta_2\,(K_1-K_0)<rI<S_2,$$

then no merchant investment ever takes place even though it is socially desirable.

• Similarly, we can show that preemption may encourage inefficiently small investments. Suppose that capacity K_1 can either be reached in one stage, at cost I, as discussed above, or in two stages. We assume that the second-stage upgrade can be performed only by the first-stage merchant investor: see the final remark below. The first stage costs I' and yields capacity K', $K_0<K'<K_1$, which can then be upgraded at cost I'' to K_1.
Let $\eta'\in(\eta_1\,,\,\eta_2)$ denote the congestion cost for capacity K' before demand in the South jumps up. Similarly, let η_2' denote the shadow price after T when transmission capacity is K'. See figure 3. The first-stage merchant investor has an incentive to upgrade the facility at date T to yield total capacity K_1 if and only if

$$\eta_2\frac{(K_1-K_0)}{r}-I''>\eta_2'\frac{(K'-K_0)}{r},$$

which we will assume. Let us look for an equilibrium in which a merchant investor preempts at date $\tau<T$ by investing a little (I') and then upgrades the line at time T :

$$I'=\frac{1-e^{-r(T-\tau)}}{r}\eta'\,(K'-K_0)+e^{-r(T-\tau)}\left[\frac{\eta_2\,(K_1-K_0)}{r}-I''\right].$$

For this, it must be the case that preemption at $(\tau-\varepsilon)$ with the full investment does not pay off:

$$I\geq\left[\frac{1-e^{-r(T-\tau)}}{r}\eta_1+\frac{e^{-r(T-\tau)}}{r}\eta_2\right](K_1-K_0)\,,$$

or

$$I-\left[I'+e^{-r(T-\tau)}I''\right]\geq\frac{1-e^{-r(T-\tau)}}{r}\left[\eta_1\,(K_1-K_0)-\eta'\,(K'-K_0)\right].$$

The right-hand side of this inequality is negative if the total value of the rights (the total congestion rent) decreases with the capacity of the link.

Aside from the timing considerations discussed above, note that given an entry at τ, a social planner might want to jump to capacity K_1 directly, as the social surplus is larger under capacity K_1 than under capacity K'. Note also that if the “upgrade” from K' to K_2 can be made by another investor than the first-stage merchant investor, the latter will need to perform the upgrade *before* T (that is, preemption occurs at both stages). The reason for this is that the upgrade reduces the first-stage investor's inframarginal rents (the rent on the $(K' - K_0)$ units of capacity), while entrants have no such rent).

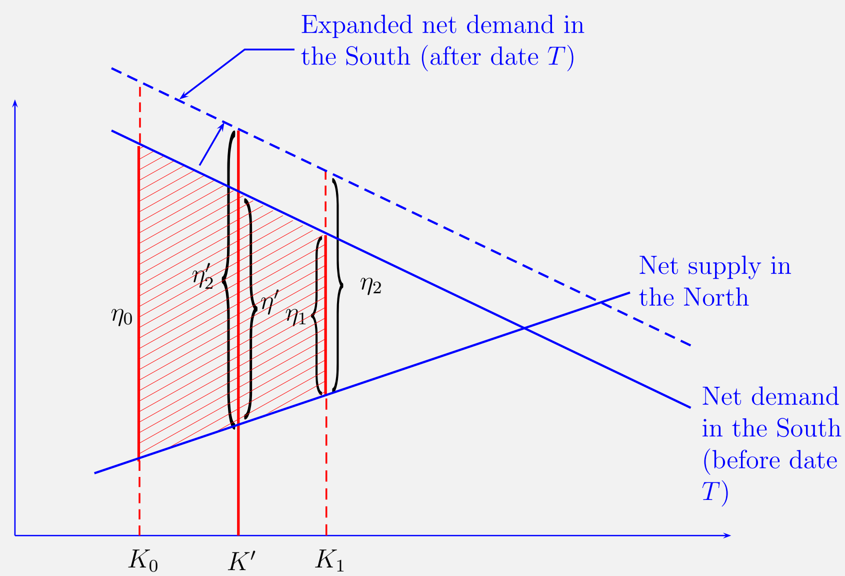

Figure 3

6 State-contingent rights and diversification

The analysis of merchant investment assumes that the capacities K_0 and K_1 are well-defined and non-stochastic. This abstracts from some important issues that arise even in the two-node case, but are especially problematic in more complex networks with loop flow, which we discuss in a separate section below. In practice, even in the two-node model, the actual capacity of the North/South link depends on exogenous environmental parameters; furthermore, system operators have substantial discretion on defining and implementing security constraints, affecting the actual power flows on the link in real time. For example, the physical capability of transmission lines depends on temperature and other exogenous contingencies.[20] And, of course, even a well-maintained system will have some random outages that cause the available capacity of the link to be reduced. Moreover, the *ex ante* estimated physical capabilities of a transmission network are defined by relatively crude administrative risk criteria (N-1, N-2) and a set of assumed system "study" conditions that are discretionary decisions of the system operator that could change in the future. In addition, the effective transmission capacity of the network varies dynamically from the study condition assumptions as system conditions change.

This all raises the issue of the number of financial rights to be allocated for the existing system and as a consequence of new investments, how congestion revenue deficiencies or surpluses arising from deviations between the number of rights allocated ex ante and the actual capacity of the network ex post are handled, and how these allocation and compensation decisions affect investment and the ultimate performance of the system. We consider this issue in the simple two-node case and explore the issue further when we consider loop flow below.

Suppose that K is stochastic: $K = K(\theta)$, $K'(\theta) > 0$ and θ is distributed between θ^- and θ^+. Let's say that the line is congested for all values of θ, but the value of η will vary with $K(\theta)$. For which value of θ should one compute the number of financial rights? One could be conservative and set the number of financial rights equal to $K(\theta^-)$. One would issue $K(\theta^-)$ financial rights and owe the holders $\eta K(\theta^-)$ in congestion payments. When the realized θ is θ^-, one satisfies the feasibility and revenue adequacy condition. But what happens when $\theta > \theta^-$? The merchandising surplus will exceed what is owed to the rights

[20] The rated capacity of Path 15, connecting Northern and Southern California falls by about 600 MW as the ambient temperature rises, other things equal. The rated capacity of Path 15 varies by about 1300 MW depending on the availability of various remedial action schemes to respond to transmission and certain generation outages. California ISO, Operating Procedure T-122A, November 6, 2002. It is also important to recognize that in the U.S. and Europe there is not a single SO controlling the network, but multiple SOs controlling independent segments of the network. To maintain reliability and avoid free riding less flexible contingency criteria must be defined than might be the case if there were a single SO operating the network in real time. For example, the simultaneous import transmission capacity into Southern California varies by 700MW depending on the operating status of the three units of a nuclear generating plant in Arizona. California ISO, Operating Procedure T-103, November 6, 2002.

holders. What does one do with the excess and how does the distribution affect investment incentives? At the other extreme, one could set the number of financial rights to reflect the maximum capacity $K(\theta^+)$. There would be revenue adequacy when $\theta = \theta^+$ but not when $\theta < \theta^+$, which would be most of the time since the system operator would owe $\eta K(\theta^+)$ regardless of the actual realization of θ. Where does the shortfall come from and how does it affect investment incentives? The answers to these questions necessarily affect the incentives merchant investors will have to make transmission investments. Realistically, especially at this stage of the development of a competitive wholesale electricity markets, SO discretion, as it affects the number and value of transmission rights and uncertain rules for implementing feasibility standards and defining the number of rights introduce uncertainties and potential opportunism problems that are not present for typical property rights.

The impact of a generous ($K(\theta^+)$, say) or conservative ($K(\theta^-)$, say) distribution of rights on investment incentives depends on the way the resulting shortfall or surplus is financed or redistributed. Suppose, first, that one appeals to the taxpayer. Even if taxation were lump-sum, there would still be distortions in investment behavior; generous distributions over-incentivize merchants, while conservative ones under-incentivize them. By contrast, appropriate taxes on users of the transmission network[21] make biased distributions of rights neutral in this radial network, provided that the dispatch is efficient: An efficient dispatch implies that in each state of nature, the cum-tax (or subsidy) price at a given node fully utilizes the link. And so end-users are unaffected by a generous or conservative distribution of rights on the line. Because there is no source or sink of money outside of the industry, rights owners receive the same overall dividend (for example, through a smaller per-unit dividend in the case of a generous distribution). This

[21] We here have in mind proportional taxes on electricity in the South (so the price is $p_S + \tau_S$, where p_S is the nodal price in the South) and a tax on exports from the North (where generators receive $p_N + \tau_N$). When actual capacity (K_a) is larger than the number of rights K, this effectively reduces the net dividend paid to rights holders so that there is sufficient revenue to compensate them.

taxation solution is however complex, since the taxes have to be state contingent. One might therefore as well have state contingent rights.

The basic problem here is that the non-contingent property rights that have been used in the literature are not well matched to the stochastic and dynamic physical attributes of transmission networks. At the very least, a merchant framework requires different types of property rights.

For example, rather than dealing with the payment shortfall or surplus problems associated with the kinds of property rights that have been proposed, it seems more natural simply to divide the merchandizing surplus proportionately among the rights' owners based on their relative ownership shares. The next box analyzes the optimal relative allocation rule (in the same way percentage ownership, but not the total number of shares, matters to determine one's proceeds from the distribution of a firm's dividend, the exact number of rights does not matter as long as the merchandizing surplus is distributed proportionately among rights' owners). It shows that the optimal allocation rule derives from standard asset pricing (CAPM) principles in finance. An addition to an existing link is particularly valuable if its actual capacity remains high when the primary link is very congested. Its construction then creates a diversification benefit. These diversification benefits are ignored both in the traditional definition of fixed MW point-to-point property rights payment obligations and with a simple proportional allocation rule.

For example, suppose that the primary North-South AC line exhibits reduced capacity (or breakdowns) during very hot weather. Then an addition along the same path has less social value if it is an aerial AC line than if it is an underground DC line not subject to the same climatic shocks even if average line availability is the same for both. Allocations of rights based only on expected link capacity therefore miss the point that, for a given capacity, some lines may provide better insurance than others. More generally, an allocation of rights solely in proportion to (or equal to) expected capacities provides insufficient incentives to build lines whose availability covaries with the shadow price less (in absolute value) than that of the existing lines. (This point applies to the case of networks with loop flow discussed further below. That is, if a transmission investment has a diversification benefit, it must be reflected in the allocation of rights if the rights are relied upon to stimulate efficient transmission investments.)

Non-contingent financial rights under state-contingent capacities.

Consider the North-South network described at the beginning of the section (see figure 1). The initial expected capacity of the link is K (the actual capacity will in a moment be assumed to be state-contingent). A merchant investor contemplates adding a small amount δK of expected capacity to the line.

Actual dispatching depends on the realization of the state of the world ω. The state of the world encompasses the uncertainty about net demand in the South, $D_S(p_S, \omega)$, that about net supply in the North, $S_N(p_N, \omega)$, and the actual capacity of the lines:

$$[1 + \theta(\omega)]\, K \quad \text{for the existing link(s)}$$

and

$$[1 + \mu(\omega)]\, \delta K \quad \text{for the new facility,}$$

where we can normalize the noises to have zero means

$$E[\theta(\omega)] = E[\mu(\omega)] = 0.$$

Let us assume that the SO dispatches optimally given the state of nature; and so, in states of nature in which the North-South link is congested:

$$D_S(p_S(\omega), \omega) = S_N(p_N(\omega), \omega) = [1 + \theta(\omega)]\, K + [1 + \mu(\omega)]\, \delta K.$$

Let $\eta(\omega) \equiv p_S(\omega) - p_N(\omega)$ denote the (state-contingent) shadow price of the link. Suppose further that K^* rights are distributed among all rights owners, including the merchant investor, and that the distribution is proportional to average capacities; and so the merchant investor receives

$$\frac{\delta K}{K + \delta K} K^* \text{ rights.}$$

The merchandizing surplus is distributed to rights owners. Needless to say, distributions of rights that would not reflect expected capacities could by themselves introduce a bias in merchant investors' incentives. For example, suppose that the incumbents in the past received a very generous rating of the existing lines form North to South, while rating standards are strengthened for new comers. The latter then receive a disproportionately small share of total rights, which penalizes them when the merchandizing surplus is distributed among rights' owners, and thereby gives little incentive to sink new investment. To avoid such obvious biases, we assume an allocation of rights proportional to average capacity. Even so, merchant incentives may be inappropriate, as we will see shortly.

The congestion dividend, $d(\omega)$ paid to the owner of a right is therefore:

$$K^* d(\omega) = [[1+\theta(\omega)]\, K + [1+\mu(\omega)]\, \delta K]\, \eta(\omega).$$

The merchant investor's expected revenue, R, is therefore

$$\begin{aligned} R &= E\left[\left(\frac{\delta K}{K+\delta K} K^*\right) d(\omega)\right] \\ &= (\delta K)\, E_\omega\left[\frac{[1+\theta(\omega)]\, K + [1+\mu(\omega)]\, \delta K}{K+\delta K} \eta(\omega)\right] \\ &\simeq (\delta K)\, E_\omega\left[[1+\theta(\omega)]\, \eta(\omega)\right] \end{aligned}$$

for δK small.

By contrast, the increase in social welfare is

$$\delta W \simeq (\delta K)\, E_\omega\left[[1+\mu(\omega)]\, \eta(\omega)\right],$$

Since in state ω, the merchant investor's expansion delivers $[1+\mu(\omega)]\, \delta K$ units of transmission capacity which each have value $\eta(\omega)$.
Hence,

$$R \leq \delta W \quad \Longleftrightarrow \quad \operatorname{cov}(\mu, \eta) \geq \operatorname{cov}(\theta, \eta).$$

Let us draw the implications of this simple characterization in specific environments:

Example 1 (diversification effect). Suppose that all uncertainty results from line availability. Existing line(s) may exhibit reduced capacity due to harsh weather (extreme heat) conditions. The merchant investor's line, by contrast is not (or at least less) subject to these harsh weather conditions (or is better protected against them). For example, the new line could be underground, or cross a climatically distinct area. Then $\theta(\omega)$ and $\eta(\omega)$ are (in a first approximation) perfectly negatively correlated, while η is not perfectly correlated with θ:

$$\mu = k\theta + \varepsilon$$

with

$$k < 1 \quad \text{and} \quad E(\varepsilon \mid \theta) = 0.$$

Hence:

$$\operatorname{cov}(\mu, \eta) \geq \operatorname{cov}(\theta, \eta),$$

implying

$$R < \delta W.$$

Non-contingent rights create an under-incentive to invest. Intuitively, the new line supplies a disproportionately high share of the transmission capacity in those states of nature in which transmission capacity is scarce and therefore very valuable. This contribution however is not reflected in the distribution of dividends which is based on fixed (non state-contingent) shares. It is only when the availabilities of the lines (old and new) are perfectly correlated that the private and social incentives coincide.

The analysis can be generalized to encompass uncertainty about energy market participants's demand and supply curves. Suppose that

$$\begin{aligned} S_N &= a_N + b_N p_N + \varepsilon_N \\ D_S &= a_S - b_S p_S + \varepsilon_S. \end{aligned}$$

So $\omega = (\theta, \mu, \varepsilon_N, \varepsilon_S)$. Under efficient dispatching

$$\eta(\omega) \simeq \left[\frac{a_S}{b_S} + \frac{a_N}{b_N}\right] + \left[\frac{\varepsilon_S}{b_S} + \frac{\varepsilon_N}{b_N}\right] - \left[\frac{1}{b_S} + \frac{1}{b_N}\right] K(1+\theta).$$

This implies that the analysis above generalizes when line availabilities are independent of demand and supply shocks (cov $(\varepsilon_i, \theta) =$ cov $(\varepsilon_i\,, \mu) = 0$ for $i = N, S$).

On the other hand, line availability may be related to demand and supply shocks. For example, it may be that a line (old or new) is subject to the same climatic shock as the demand node. Hot weather may simultaneously increase demand and limit the capacity of the line bringing electricity from a cheaper node (precisely when the line is most needed). Such a line obviously has a lower social worth than one whose availability is less negatively correlated with increases in demand at the expensive node.

Example 2 (uncertainty about energy market players only). Suppose that there is no uncertainty about the actual capacities of the lines:

$$\theta(\omega) = \eta(\omega) = 0 \quad \text{for all} \quad \omega.$$

Hence all uncertainty comes from generation and consumption. In this case, the private and social incentives coincide:

$$R = \delta W.$$

To sum up, the analysis in this section leads to several conclusions. The type of non-contingent transmission rights that have been contemplated in the literature are not well adapted to the physical attributes of transmission investment. Both the quantities of non-contingent rights awarded and the distributions of congestion revenue surpluses and shortfalls are in some sense arbitrary as a result. But these decisions also necessarily affect investment incentives. Rights whose capacities are contingent on exogenous factors that change the capacity of the network are conceptually a superior approach to defining transmission rights. On a two-node network, transmission rights that give the holders the rights to a proportionate share of the congestion revenues in each (exogenous) state of nature replicate the type of contingent right that we have in mind and do not create market liquidity problems. As we shall see below, however, there is no equivalent to simple proportional rights for creating the desired contingent rights on a network with loop flow. Finally, regardless of the types of rights offered, the quantity of rights given to investors in new transmission capacity should reflect the insurance attributes of the new links based on the covariance between the actual varying of these links over time, reflecting outages and deratings, and the actual varying capacity of the existing links.

7 Nominal and actual transmission capacity

The difficulty in putting a number in front of a line's "capacity" (in the two-node case) raises other issues. In a nutshell, the actual capacity of the line depends on discretionary choices, and under a merchant transmission model control is likely to be separated from ownership due to the conflict of interest associated with the measurement of congestion rents. This section considers two such discretionary actions: dispatching and maintenance.

7.1 Dispatching

As we already noted, rewarding merchant investment through congestion rents requires separating ownership and dispatch in order to obtain an unbiased measure of this rent.

But this separation of ownership and dispatch raises a moral-hazard-in-teams problem. The electric system's state-contingent output (to simplify, the intensity of power in the absence of outage and the probability and duration of an outage) depends on both the care and the forecasts of the owner (the quality of the line, its maintenance, and the adequacy to consumers' needs) and the quality of the management of the grid by the system operator, as the latter must use her acumen to get lots of power through without creating a high risk of outage.

In other words, the transmission owner's measure of performance is conditioned by the system operator's behavior and therefore incentive scheme. This raises two points: First, one cannot consider incentives given to merchant investors without also specifying those of the system operator. Second, moral hazard in teams reduces accountability. An outage can be claimed to result from poor line maintenance or from imprudent dispatching. Conversely, high power prices may be due to a proper dispatching motivated by low line quality or to an undue conservatism of the system operator.

There is also a potential moral-hazard-in-teams problem among line owners. Recall that merchant investment incentives are better aligned with the public interest when merchants don't have inframarginal units whose congestion rent is to be preserved. The total North-South capacity may then belong to different owners. The same value of a given actual capacity K selected by the independent system operator may correspond to different quality configurations of the various components of the network with multiple owners. The question is then one of allocation of total capacity and congestion rents among the different owners.

Moral hazard in teams : transmission owners and system operator

Consider the North-South network. Let K denote the nominal capacity of the line. In a first step, we assume that this capacity is known to the system operator (for example, the line's maintenance is perfectly observed by the SO). The system operator choose to allow an amount $\widehat{K}$ to flow through the link. The greater is $\widehat{K}$, the higher is the probability that the link breaks down. We assume that with probability $x\left(\widehat{K}-K\right)$ the link breaks down and no power flows through it. With probability $1-x\left(\widehat{K}-K\right)$, $\widehat{K}$ flows through. The function x is increasing in $\widehat{K}$. Let $L\left(\widehat{K}\right)$ denote the out-of-merit dispatch cost when the realized capacity of the line is $\widehat{K}$. [For example, in figure 1, L was equal to the area of the triangle ABC (for capacity K)]. We assume that there is no market power at either mode and so L represents the social loss attached to the inability to import power without constraint from the North. Note that

$$L' = -\eta.$$

The socially optimal dispatch solves, for a given K,

$$\min_{\widehat{K}} \left\{ x\left(\widehat{K}-K\right)L(0) + \left[1 - x\left(\widehat{K}-K\right)\right] L\left(\widehat{K}\right)\right\}.$$

And so $\widehat{K} = \widehat{K}^*$ is given by

$$x'\left(\widehat{K}^*-K\right)\left[L(0) - L\left(\widehat{K}^*\right)\right] + \left[1 - x\left(\widehat{K}^*-K\right)\right] L'\left(\widehat{K}^*\right) = 0.$$

The marginal social gain from capacity expansion is then (using the envelope theorem):

$$\frac{dW}{dK} = x'\left[L(0) - L(K)\right] = (1-x)\eta.$$

And so, if the marginal investment is rewarded by the congestion cost in the absence of outages, merchant investors face the proper signal for investment.

• *Dispatcher with conservative incentives.*
Turn now to the system operator's incentives. Suppose that the SO is penalized more for outages than she is rewarded for increases in the amount of power flowing through the network ; that is, she solves:

$$\min_{\widehat{K}} \left\{ x\left(\widehat{K}-K\right)\theta L(0) + \left[1 - x\left(\widehat{K}-K\right)\right] L\left(\widehat{K}\right)\right\},$$

where $\theta > 1$. This yields first-order condition:

$$x'\left[\theta L(0) - L\left(\widehat{K}\right)\right] + [1-x]\, L'\left(\widehat{K}\right) = 0.$$

The marginal social gain from capacity expansion is

$$\frac{dW}{dK} = x'\left[L(0) - L\left(\widehat{K}\right)\right] - x'(1-\theta)L(0)\frac{d\widehat{K}}{dK}.$$

And so

$$\frac{dW}{dK} - (1-x)\eta \quad = \quad x'(1-\theta)L(0)\left[1 - \frac{d\widehat{K}}{dK}\right],$$

using the SO's first-order condition.

Next, rewrite the SO's optimization program as the choice of a risk factor $\Delta \equiv \widehat{K} - K$:

$$\min_{\widehat{K}} \left\{ x\left(\Delta\right)\theta L(0) + \left[1 - x\left(\Delta\right)\right] L\left(K + \Delta\right)\right\}.$$

The cross-partial derivative of the minimand with respect to K and Δ is positive as $L'' > 0$, and so, by a revealed preference argument, Δ is decreasing in K, or:

$$\frac{d\widehat{K}}{dK} < 1.$$

In words, the system operator takes less risk as K increases, because the marginal gain from increased throughflow decreases. We therefore conclude that

$$\frac{dW}{dK} < (1 - x)\eta,$$

and so congestion rent payments over-incentivize merchant investors. In a sense, the SO's conservative behavior implies that insufficient use will be made of the added capacity and so the shadow price of the link overstates the value of additional capacity. This result shows that one cannot properly analyze merchant investment (or, for that matter, the incentives of a Tranco company not responsible for dispatching) without considering the system operator's incentives.

Moral hazard in teams : general considerations

At an abstract level, one can view transmission owners and the SO as a team (in the sense of Holmström 1982) jointly delivering an output — state-contingent power— to the final consumers. A general principle is that proper incentives require that each member of the team be made the residual claimant for the team performance. So for example each member of a n-member team should receive 1 when the team's profit increases by 1 (third parties must act as "budget breakers" to bring the missing \$$(n-1)$). Here, the performance of the team is not a profit, but rather (minus) the social loss

$$xL(0) + [1 - x]\, L\left(\widehat{K}\right),$$

and the members of the team are the SO and the transmission owners. Making each residual claimant is however very costly for two reasons:

- Adverse selection : fortuitous improvements in performance give rise to n rents.
- Collusion : relatedly, the members of the team have an incentive to collude. Suppose for example that a merchant investor has a marginal project that costs as much as the marginal reduction of redispatching cost it brings about. While this merchant investor is indifferent as to whether to implement the project, the other participants (SO, other transmission owners) each costlessly receive the value of this reduction in redispatching cost if he implements it. They therefore have incentives to bribe him into investing. More generally, collusion will induce investments whose cost vastly exceed their social benefit.

To avoid or alleviate these problems, one can make each member accountable for only a fraction of the social benefit. But this policy creates moral hazard. For example, the SO has reduced incentives to dispatch properly (for example, x increases for a given) and transmission owners have reduced incentives for maintenance.

7.2 Maintenance

It is now well understood that an owner of *physical* transmission rights may have an incentive to withdraw some of the physical capacity in order to create artificial scarcity and thereby raise the value of the rights. Overt withdrawals can of course be prevented through a use-it-or-lose-it rule, but withdrawals can take more subtle forms. For instance, maintenance can be strategically planned so as to occur during tight supply periods.

Similar concerns arise in the case of *financial* rights. Consider a merchant investor who is in charge of the maintenance (to create more accountability at the building stage) and keeps the financial rights. Then, planning the maintenance in high-congestion periods raises the price of the rights in these periods and may raise the owner's revenue. This suggests that it might be desirable to require merchant investors to sell their rights to third parties along with a performance-based contract that provides incentives for the transmission owner to make efficient maintenance decisions. Namely, the merchant investor will be held liable for revenue shortfalls that his choice of maintenance schedule and technique imposes on the owners of acquired rights.

The box below analyzes the merchant investor's maintenance choices in the absence and presence of such liability assuming that capacity availability depends only on the transmission owner's maintenance practices. (Of course in reality other factors that affect transmission capacity that are not under the control of the owner would have to be controlled for.) The performance contract induces the merchant investor to "withdraw" less capacity for maintenance purposes and to use more expensive methods (such as helicopter-based live maintenance) during peak periods, but it induces excess line availability and expenditure during these periods. This overshooting comes from the fact that the accountability for revenue shortfalls makes the merchant investor a net *buyer* of rights.

Financial rights and incentives for maintenance

Suppose that there are $t = 1, \cdots, T$ subperiods, with varying degrees of congestion. Let E_t denote the maintenance expenditure and k_t the capacity made available in period t. The merchant owner chooses expenditures E_t and availabilities $k_t \leq K$, the rated capacity, so as to maximize

$$\sum_{t=1}^{T} \left[k_t \eta_t - E_t\right].$$

Letting $w_t \equiv K - k_t$ denote the capacity withdrawn for maintenance, the period-t contribution to maintenance is

$$m_t = f\left(w_t, E_t\right) \tag{1}$$

and is increasing in its two arguments. In a sense, money and unavailability are substitutes in maintenance production. For example, the owner can use helicopters to maintain while letting lines operate ("live maintenance"). An adequate level of maintenance is obtained when

$$M\left(m_1, \cdots, m_T\right) \geq \overline{M}, \tag{2}$$

where M is increasing in its arguments. The *socially optimal* levels of expenditure and maintenance satisfy, for all t:

$$\eta_t = \mu \frac{\partial M}{\partial m_t} \frac{\partial f}{\partial w_t}, \tag{3}$$

$$1 = \mu \frac{\partial M}{\partial m_t} \frac{\partial f}{\partial E_t}, \tag{4}$$

where μ is the multiplier of the maintenance adequacy constraint (2).

A *profit-maximizing owner* yields (4) and

$$\eta_t - k_t \frac{\partial \eta_t}{\partial w_t} = \mu \frac{\partial M}{\partial m_t} \frac{\partial f}{\partial w_t}. \tag{5}$$

The marginal revenue loss of withdrawing capacity is (unless k_t is small) smaller than is socially optimal, since this creates additional scarcity ($\partial \eta_t / \partial w_t > 0$) when the line is congested. Note that

$$\frac{\dfrac{\partial f}{\partial w_t}}{\dfrac{\partial f}{\partial E_t}} < \eta_t$$

in congested periods. That is, the merchant investor overconsumes line unavailability relative to expenditures during peak periods.

Suppose now that,

- the merchant investor sells the K financial rights, and
- he operates the link's maintenance subject to a performance-based scheme making him accountable for the revenue shortfall.

The merchant investor then receives a fixed revenue for the financial rights and then selects his maintenance strategy so as to solve:

$$\min_{\{w_t, E_t\}} \left\{ \sum_{t=1}^{T} \left[(K - k_t)\, \eta_t + E_t \right] \right\}$$

subject to (1) and (2). The first-order conditions are (4) and

$$\eta_t + (K - k_t)\,\frac{\partial \eta_t}{\partial w_t} = \mu \frac{\partial M}{\partial m_t}\,\frac{\partial f}{\partial w_t}. \tag{6}$$

And so

$$\frac{\dfrac{\partial f}{\partial w_t}}{\dfrac{\partial f}{\partial E_t}} > \eta_t$$

in congested periods.

The merchant investor, who has sold K rights forward, has become a net *buyer* of rights. He therefore has an incentive to excessively enhance availability (underschedule maintenance) during congested periods so as to reduce the penalty paid for unavailability. The merchant investor also tends to substitute expenditures (live maintenance for instance) for withdrawals of capacity during peak time. While this clearly is a good thing qualitatively, there is some overshooting as the merchant investor now benefits from lowering the value of the rights.

Policymakers are sometimes attracted to the merchant investment model because they think that it relieves them of the difficult task of designing and implementing good incentive regulation mechanisms. This discussion should make it clear that this view is misguided. The system operator remains a monopoly and providing it with appropriate incentives continues to be an important regulatory policy challenge. Allowing (or requiring) merchant investors to sell financial transmission rights without performance contracts can distort maintenance behavior. Regulatory rules may be necessary to obtain the best design for such performance contracts. And, of course, if the incumbent transmission owners continue to be regulated, incentive issues regarding both maintenance and investment continue to be of great importance.

8 Defining and allocating rights with loop flows

Loop flow introduces additional practical complications. First, how does one define the "capacity" created by new investment and the associated financial rights that go along with the new capacity on a network with three or more nodes and associated loop flows? Second, full or partial outages of one link may affect the effective capacity and nodal prices on other links and at other nodes in less straightforward ways than in the two-node case, especially when there are multiple owners.[22] And as is now well-known, an addition of capacity may have negative social value and even in the absence of system operator discretion, the increase in a link's capacity is unrelated to the system's increased capacity. Finally, and more crucially, small investments may no longer be "marginal".

With a two-node network or a radial network with multiple generation nodes but without loop flow, transmission rights (whether physical or financial) are naturally conceptualised as "link-based" rights reflecting the capacity of each link. When there are more than two nodes and loop flow, then there are at least two ways of introducing financial rights (Joskow and Tirole 2000, pp. 478–479). One approach is to use "link-based" rights (Oren et. al. 1995) which are rights associated with each transmission line on the network and, in the case of financial rights, paying a dividend equal to the shadow price of the congestion on each line. The other approach proposed by Hogan (1992) is to specify point-to-point financial rights from each injection node to each receipt node on

[22]For example, there are simultaneous import limitations into California that depend on the availability of links from the Southwest to Southern California, the Northwest to California, and the operating generating capacity inside California. These limits are presently managed administratively with "nomograms" that define the curtailments that are triggered when the constraints are binding.

the network, with the rights paying a dividend (which could be negative) equal to the difference in nodal prices at the two nodes due to congestion. We will focus on point-to- point financial rights here since they are being used in several areas of the U.S. and appear to be the favored approach in the FERC's SMD rules. Moreover, in theory the values of point-to-point rights internalise network externalities associated with loop flow since they reflect the shadow prices on all lines affected by an injection at one node and an equivalent withdrawal at another node. The shadow price on a particular transmission link, however, does not reflect the social value of the link to the network overall.

For the general case of a multi-node network with loop flow, Hogan (1992) envisions that point-to-point transmission rights will be defined and allocated through a process in which a set of all feasible (i.e., consistent with the transmission network) *physical* combinations of bilateral contracts between injection and receipt points is first calculated. The process of defining the feasible set must be conducted by the SO by performing a large set of simulations of the use of the network under various supply and demand conditions and contingencies (e.g. line outages) using load flow models. The process envisioned for defining the feasible set appears to be purely physical in the sense that the SO does not rely on prices or other valuation procedures to define the set of feasible rights. A second process (e.g. grandfathered allocations to incumbents, auctions, bilateral trading) is then used to define the specific combination of rights/capacities from within (or on the frontier of) the feasible set that will be allocated initially to generators, marketers and/or load serving entities. Once a specific combination of feasible transmission rights is defined and allocated they become the property of the holders. These rights may then be traded in secondary markets.

Investments in new transmission capacity are translated into *incremental* transmission rights through the effects these investments have on *shifting* the initial frontier of the set of all feasible bilateral transactions and the associated configurations of point-to-point capacities/rights. The literature generally assumes (a) that the initial feasible set and shifts in its frontier are well defined in the sense that there is no uncertainty about the relevant parameters of the feasible set, (b) that the feasible set does not itself vary with exogenous random variables, and (c) that the shifts in the frontier of the feasible set do not make any rights/capacity combinations that were previously in the feasible set infeasible, post investment, or else that efficient trading arrangements are in place that ensure reallocations to ensure feasibility.

There are both practical and theoretical issues that may undermine these assumptions. As we have already discussed, the feasible set of bilateral schedules that can be accommodated without causing congestion and the associated transmission capacities and accompanying rights depend on exogenous environmental parameters. While this fact literally contradicts assumption b) above, the existing theory can straightforwardly be extended in the usual manner by allocating *state-contingent rights*, as long as the contingencies can be described (temperature, output of specific generators that affect contingency limits, conditions in interconnected control areas, etc.). The drawback of this extension is that the large number of potential contingencies that are relevant for defining and implementing the feasibility requirement call for a large number of state-contingent rights, with the concomitant problems that these create: large transaction costs, thinness and market power in the secondary markets for these rights.

Assumption a) can also be questioned. In particular, system operators have substantial discretion in defining and implementing security constraints, affecting the actual power flows on the network in real time, and random line outages. Moreover, for complex networks the physical feasibility evaluation necessary to define the numerous potential configurations of transmission rights that are simultaneously feasible and the incremental configurations of transmission rights created by new investments involves many discretionary assumptions and is likely to be based on load flow models that are approximations to real networks, are subject to SO discretion and may not be especially good approximations under stressed conditions when losses are significant and contingency constraints binding. These are the conditions when transmission rights are likely to be especially valuable. What is modeled as being feasible and what is feasible in actual operations can differ, especially when reactive power and voltage constraints are important.

It should be clear as well that in practice the merchant transmission model cannot operate "as if by an invisible hand," since some *de facto* regulatory authority must have the ability accurately to simulate load flows on the network, apply contingency criteria, define feasible sets and changes in feasible sets associated with transmission investments, and ensure that rights allocations are consistent with feasibility under numerous contingencies.

Let us finally come to assumption c). It is clear that, except in radial networks, the expansion of the network both creates new feasible allocations and makes some initially feasible allocations infeasible. So, in general, the expansion may infringe on existing property rights. This problem has been recognized in the academic literature, though not very clearly in the policy arena, and it has been proposed that the merchant investor building a new line leave existing property rights intact, which in general requires the merchant investor to compensate for the loss of property rights by buying existing ones and turning them back to initial owners who were expropriated. We now proceed to elaborate on these issues for the case of a 3-node network with loop flow.

The most elegant explanation of how the contract network framework can be applied in practice to a simple network with loop flow is provided by Bushnell and Stoft (1997), and we follow their presentation very closely here. Figure 4 depicts the standard simple three-node network. There is generation in the North, generation in the South and demand in the East. The transmission lines connecting these nodes have capacities K_{NE}, K_{SE} and K_{NS} respectively as depicted in Figure 4.

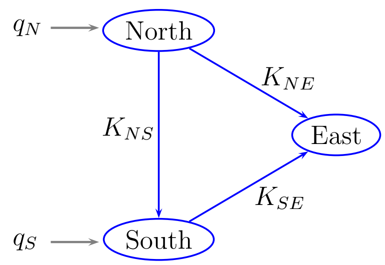

Figure 4

Assuming that the transmission links connecting the three nodes are of equal length (resistance) and ignoring losses, the physical laws of electricity (Kirchoff's) determine the flows through the three transmission links associated with alternative configurations of generation (q_N and q_S) in the North and South and consumption (q_E) in the East. The relevant constraints applicable to the definition of the feasible set of bilateral transactions are:

$$\begin{aligned}
\frac{2}{3}q_N + \frac{1}{3}q_S &\leq K_{NE}\,, \\
\frac{1}{3}q_N + \frac{2}{3}q_S &\leq K_{SE}\,, \\
\left| \frac{1}{3}q_N - \frac{1}{3}q_S \right| &\leq K_{NS}\,, \\
q_S + q_N = q_E.
\end{aligned}$$

The feasible combinations of q_N and q_S associated with each constraint are depicted by lines in Figure 5 and the intersection of these sets defines the feasible set of bilateral transactions.[23] The feasible set is depicted as the hatched area in Figure 5 (equivalent to Figure 2 is Bushnell and Stoft 1997). The dispatch of the system and the allocation of point-to-point transmission rights must lie within this feasible set.

[23] In what follows, K_{NS} is assumed to be small relative to K_{NE} and K_{SE}. The capacities K_{NE} and K_{SE} do not have to be equal, but the examples in Bushnell and Stoft (1997) assume that they are and we will follow that assumption in the graphical presentation here.

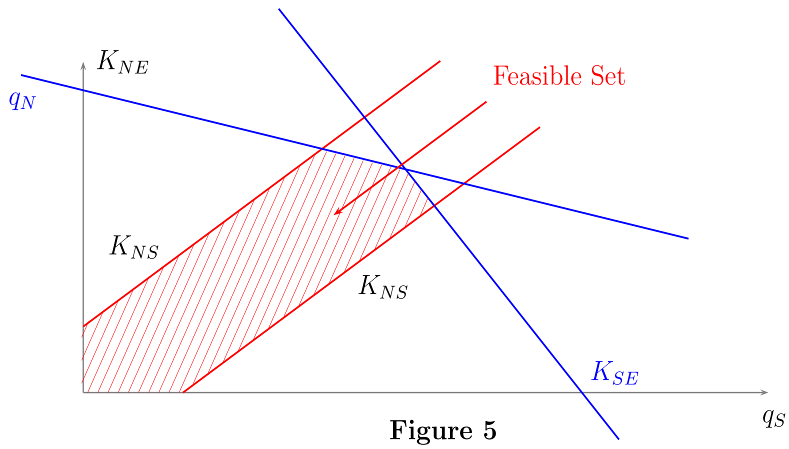

Figure 5

Accommodating investment into this framework is tricky because grid expansions can both make combinations of q_N and q_S and associated point-to-point rights feasible that were not previously feasible and make some combinations of q_N and q_S and associated point-to-point rights that were feasible pre-investment, infeasible post-investment. To see this, it is again useful to follow Bushnell and Stoft (1997) and start with a radial network that does not have a link connecting the North and the South, and, accordingly, no loop flow. A radial network of this type is depicted in Figure 6.

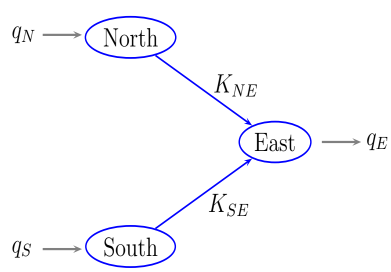

Figure 6

For this network, the feasible set is very simple to define. It satisfies:

$$q_N \leq K_{NE}$$
$$q_S \leq K_{SE}$$
$$q_N + q_S = q_E.$$

Generation at each node is limited only by the capacity of the link connecting it to the demand node. The feasible set of bilateral transactions for this radial network is depicted as the hatched area in Figure 7. Figure 7 also depicts a hypothetical optimal dispatch consistent with generation in the North being less costly than generation in the South, but limited transmission capacity from North to East requires some more expensive generation from the South to be dispatched to clear the market.[24] The marginal cost of the expensive generation that clears the market would also determine the market clearing price in the East.

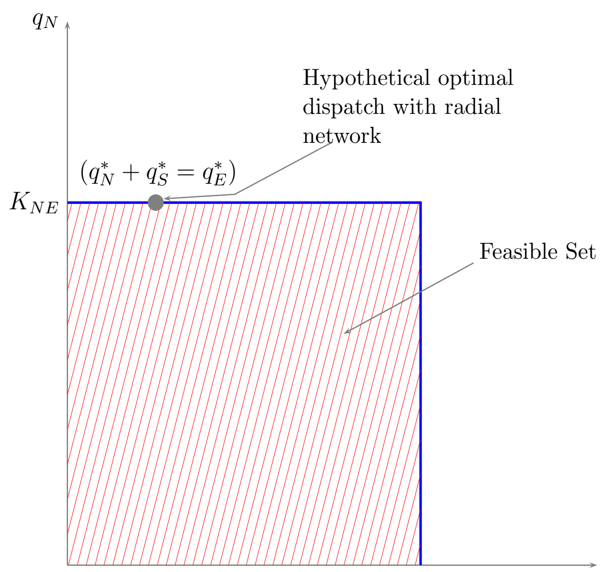

Figure 7

Now let's consider adding to this radial network a third link between North and South to create the three-node network with loop flow depicted in Figure 4. The changes in the

[24]The optimal dispatch does not have to be on the frontier of the feasible set in general, but can lie inside of the frontier. This would be the case here if the marginal cost of generation at both nodes is increasing in output and demand is at a level that does not require that generators operate at full capacity to clear the market.

feasible set resulting from this investment are displayed in Figure 8.

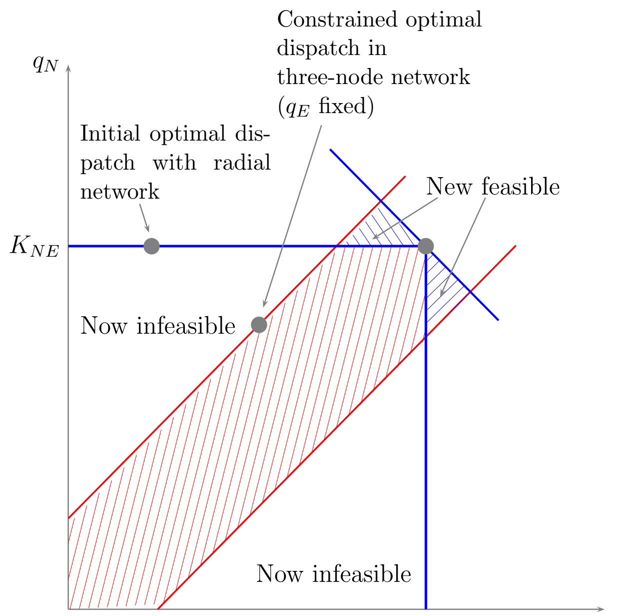

Figure 8

(This figure is equivalent to Figure 4 in Bushnell and Stoft 1997.) Some allocations that were previously feasible are now infeasible and some allocations that were not previously feasible are now feasible. In particular, the initial optimal dispatch is no longer feasible. In order to go forward with the new line, the investor would (effectively) have to buy back sufficient rights from those who hold them initially to restore feasibility (or as in Bushnell and Stoft require the investor to take rights that have negative values and require payments rather than receiving dividends to restore feasibility). An efficient economic transmission rights reallocation process must complement any physical analysis of the effect of a transmission investment on the feasibility of the existing allocation of rights. Otherwise, if the SO were to take the allocation of existing rights as fixed when performing a feasibility test, the set of investments that satisfy the constraint that no existing right will be made infeasible will lead to a set of allowable investments that is much smaller than the set of investments that increase social welfare.

With such an efficient reallocation mechanism in place, Bushnell and Stoft show that

that it will be most profitable for the investor to acquire rights that lead to an allocation equal to the most efficient dispatch given the constraints associated with the new investment and associated network topology. Using their numerical assumptions, the new efficient dispatch and allocation of point-to-point financial rights would be the point depicted in Figure 8 that involves less (cheap) generation in the North and more (expensive) generation in the South than was the case without the new link. The new link is therefore inefficient and should not be built and, indeed, Bushnell and Stoft show that the obligation of the investor to restore feasibility will make this investment unprofitable for a merchant investor under these assumptions.

This naturally raises the question of why transmission links such as the one between North and South that cause loop flow are so common. One reason might be that the post-investment optimal dispatch lies in the small new feasible region in Figure 8, allowing increased production from the cheap generator in the North. This could be the case, for example, when demand is high and the optimal dispatch for the radial network is further to the right on the K_{NE} constraint in Figure 8, involving more generation from (expensive) South. The new link would then have the effect of increasing the feasible (cheap) supplies from North and reducing the (expensive) supplies from South to balance supply and (higher) demand, by effectively increasing the capacity from North to East via South. But, this situation requires that it is less costly to invest in transmission capacity to increase supplies from the North over an indirect path (North to South to East) than simply to increase the capacity of the direct link from North to East.

If the new link does not move the efficient dispatch into the new areas of the feasible set, the third link between generation nodes in the standard three-node network reduces social welfare because it is a binding constraint on low-cost generation schedules which would otherwise be accommodated without congestion on the direct links between each generation node and the demand node. This link only makes sense when we recognize that one of the other links may fail and the third link provides an alternative path for delivering supplies to satisfy demand (Joskow and Tirole 2000, p. 477). So, this link will have negative value under some contingencies and positive value under others. Adding the link will reduce the feasible capacity from at least one node to another under some contingencies and increase it under others.[25]

We can see this by extending the Bushnell and Stoft's (1997) examples to take account of line outages. Let's go back to the radial network depicted in Figure 6. Assume now that that the capacity of the link from North to East is reduced as a consequence of equipment failures or other contingencies; assume that the capacity is cut by 1/3. The new feasible set is depicted in Figure 9. The new optimal dispatch involves less cheap generation and more expensive generation and increases total generation costs.

Now we consider the effects of adding a link between North and South under the contingency that the capacity of the NE link is reduced due to a line outage. The new feasible set (under the condition that capacity on the link between North and East has

[25]This is a potential problem with link-based rights. The shadow price on the link between North and South is positive in both states of nature and could look profitable to an investor if she could ignore the impacts on the rest of the network.

been cut by a third) is depicted in Figure 9.

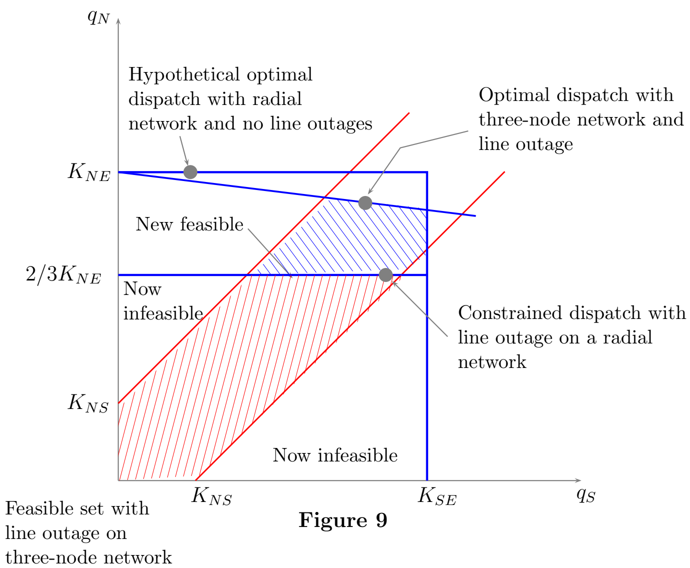

Figure 9

As before, adding a line that causes loop flow makes some allocations that were previously feasible now infeasible and others that were previously infeasible are now feasible. As portrayed in Figure 9, the link in this case makes it possible to increase generation at (cheap) North and reducing generation and (expensive) South when the capacity of the NE link is reduced significantly. Accordingly, the third link will reduce generation costs (and perhaps reduce the probability that demand will have to be shed to balance supply and demand) when there are transmission line outages of the type examined here. Whether it is an efficient investment will depend on the benefits of the link during contingencies like these (and others when it is valuable), the costs of the link during conditions when it is not "needed" and leads to an inefficient dispatch, and the cost of the investment. It is clear that *non-contingent* transmission rights cannot be defined properly to capture the varying valuations of a transmission investment under the many contingencies that characterize real electric power networks and provide the right incentives to support efficient investments. Only *contingent rights* provide the proper incentives. Moreover, for any mechanism like this to work well a liquid competitive secondary market for rights would have to exist to make it possible for investors to easily buy and sell rights at their competitive market values to restore feasibility and to allow welfare enhancing investments to go forward.

The success of any property-based system in attracting efficient levels of investment depends on the ability to define and enforce clear and consistent property rights. This appears to be an especially challenging problem on an electric power network with loop flow where the feasible set of property rights and their efficient allocation (i.e. not just their value) are contingent on changing supply and demand conditions, the application of contingency constraints by the system operator,[26] and their interaction with new investments.

9 Coordination issues

9.1 General considerations

As noted earlier, the optimality of merchant investment requires that the net demand and supply curves in the wholesale market represent the true demands and supplies of energy market participants. Most of the literature supporting transmission investment is static in the sense (a) there is no uncertainty about supply, demand or prices, and (b) all investments in generation and transmission occur simultaneously. However, investments in transmission are long-lived sunk investments and their value depends on changing and uncertain supply and demand conditions over many future years. The economic calculus necessarily involves forecasting future supply and demand conditions which are uncertain, including changes in locational supply and demand conditions resulting from future investments in generating and transmission capacity, and the associated uncertain nodal prices. As a result, the presentation of the supply and demand functions in the

[26]As we earlier noted, a further complication may be that the contingency constraints applied to define transmission capacity and the network topology used to manage congestion depend in part on the location, size and operating status of specific generating units on the network, effectively making transmission capacity endogenous.

previous figures, and the standard formulations of these problems must stand for the *long-term* demand and supply curves. The latter of course reflect the possibility of investments in generation and (for consumers) bypass via self-generation. Existing and new electricity producers formulate investment plans, whose implementation depends on the expectations of market conditions; similarly large and small users may adopt equipment that allow them to switch to alternative sources of energy. For example, investments in the North will be unprofitable if they are not accompanied by a strengthening of the North-South line. Conversely, the reinforcement of the line won't be profitable if the congestion rent is too small, that is if no investment occurs in the North.

In principle, this coordination can be achieved through a planning procedure, in which all interested parties announce their (price-contingent) investment plans. Such coordination however becomes more involved if either some party (or coalition of parties) have market power or an incentive to block investments to create it or if investments are lumpy. Mechanisms designed to aggregate stakeholder preferences to make choices about major transmission investments have not been particularly successful.[27]

For example, the owner(s) of the existing capacity K_0 of the line may announce a substantial reinforcement in the hope of attracting investment in the North, and later not implement this capacity building. The price collapse in the North brought about by "excessive" investment in generation there increases the congestion rent and benefits the transmission owner. Similarly, investments in generation in the South might be announced, that are meant to preempt a reinforcement of the transmission line and will never be implemented.

In general, proper incentives must be put in place in order to prevent such manipulations of other parties' investments. For example, proposed transmission projects that are recognized and approved by the system operator might be required to post a bond to

[27]See Chisari et. al. (2001) for a discussion of the experience in Argentina.

secure their commitment that would be forfeited if the project is not completed (subject to *force majeur* exceptions).

9.2 Does the Coase theorem apply?

It is sometimes argued that the problems created by lumpy investments can be resolved through negotiations between the various market participants who will benefit from the investment. That is, that the "Coase theorem" applies. There are many reasons, exposited in the following box, to believe that negotiations among the affected market participants is unlikely to solve the problems.

Lumpy projects: Will winners get on board and losers be compensated?

As was discussed in section 5, the price system does not supply the proper signals when the transmission investment has a substantial impact on nodal prices (according to this definition, a very small transmission investment creating a new link that causes loop flow — as in section 8 — is a lumpy investment). Merchant investment, to be efficient, then must involve some stakeholders process. Its proponents build on the Coase Theorem to argue that winners will participate in the funding of socially desirable projects, while losers will pay enough so to prevent socially undesirable ones. There however are several issues with this argument, some general and well-known, and some more specific to the context:

- *Transaction costs*: Coasian bargaining involves transaction costs especially when the number of stakeholders is large.

- *Asymmetric information*: In order to reach an efficient agreement, parties must make reasonable demands. When imperfectly informed as to what others will accept, participants in a bargaining process may end up being too greedy, resulting in an inefficient bargaining breakdown.

- *Absence of future players*: efficiency requires that all parties be present to negotiate a deal. Future newcomers in generation, transmission and consumption by definition do not sit at the bargaining table and so their interests will not be accounted for by the existing parties.

- *Non-excludability of winners and free riding*: The design of restructured electricity markets imposes that prices be uniform at a given node: this is mechanistically so in the case of pools, but it applies as well to bilateral-transactions-based systems. Thus, a consumer in the South benefits from the reinforcement of the North-South line and has no incentive to join other consumers in the South and other winners (producers in the North) to contribute to defray the transmission investment if such free-riding does not prevent the investment from being made. The current market design does not call for exclusive rights for the financing parties, and thus encourages free riding in a merchant investment context.

• *Hold-up of potential losers*: To reach an efficient agreement when a socially inefficient transmission investment is being considered, it must be the case that losers be able to "bribe" the winners not to make this investment. While this bribe enables an efficient outcome "ex post", when anticipated "ex ante", it may provide parties with the wrong incentives.

To see this, consider a fixed demand q_S in the South (corresponding to gross surplus vq_S — we assume an inelastic demand for computational simplicity). The unit cost of building generation capacity in the South is $I_G \equiv c_0 q_S$, where c_0 is the per-unit cost of investment. This capacity can be built by a monopoly generator (the analysis can be altered to accommodate the case of a competitive set of generators in the South). The variable (ex post) cost of producing q_S is cq_S.

Suppose now that, at cost

$$I_T > I_G,$$

a link with capacity exceeding q_S from another region (North) to South can be built. There is already plenty of competitive generation in the North at some variable cost c. Thus, the building of the transmission line involves a social loss of $I_T - I_G$ if the generation in the South is not yet built and of I_T if this generation is already in place.

Now suppose that there is investment in generation in the South (the efficient solution). The monopolist charges v per unit, resulting in profit $(v - c)\, q_S$ (gross of the investment cost I_T). Suppose next that the consumers form a coalition and threaten to build the transmission line. If they do so, Bertrand competition brings the price down to c, and so the threat is credible if

$$(v - c)\, q_S - I_T > 0,$$

which we will assume. Coasian bargaining implies that the monopoly generator will bribe the consumer coalition not to implement the project, where the bribe b lies in an interval:

$$b \in \left[(v - c)\, q_S - I_T \ , \ (v - c)\, q_S\right].$$

The lower bound of the interval corresponds to the coalition surplus from bypass, and the upper bound the quasi-rent enjoyed by the monopolist in the absence of bypass. So, if β denotes the coalition's bargaining power:

$$\begin{aligned} b &= \beta\left[(v - c)\, q_S\right] + (1 - \beta)\left[(v - c)\, q_S - I_T\right] \\ &= (v - c)\, q_S - (1 - \beta) I_T. \end{aligned}$$

Knowing that it exposes itself to paying a ransom, the generator in the South invests if and only if

$$(1 - \beta) I_T \geq I_G,$$

a condition that is violated unless β is small. So, if the coalition does not hold, the efficient investment in generation does not occur.

Two comments are in order. First, the inefficiency can be prevented by moving Coasian bargaining to an ex ante stage — that is, in effect by having the generator and the consumers engage in long-term contacting. [This solution of course raises other issues, such as the absence ex ante of relevant parties at the bargaining table; or the barrier-to-entry potential of long-term contracts as studied by Aghion and Bolton (1987).]

Alternatively, an ex ante agreement that does not lock in the consumers must be akin to greenmail in that it must involve a commitment by consumers and/ or transmission enabler (in the case of a scarce corridor) not to play this bypass-and-holdup game. Second, the same logic applies to the situation described in figure 1. Consider an inefficient reinforcement of the line that fully eliminates congestion and whose investment cost lies between the congestion cost and the sum of the congestion cost and the congestion rent. Then, a merchant investor can hold up consumers, producers and rights owners collectively, a behavior that would have an impact on the ex ante incentives for consumption-, production-, and transmission-related investments.

9.3 Gaming between merchant investment projects

Last, we conclude this section with an example of possible gaming between merchant projects rather than between a merchant project and producers or consumers.

Gaming and coordination of merchant investment projects

An important benefit of market mechanisms is the freedom economic agents enjoy in their investment decisions. They don't need to coordinate with other agents. In the case of electric networks, coordination will be needed not only between transmission and generation investments, as we just discussed, but also between transmission investments. To illustrate this, consider the pair of transmission investment projects depicted in figure 10. Complementary investments from North to Middle and from Middle to South will allow cheap power to flow from North to South. While this pair is really a single investment from an economic viewpoint, the investments may be undertaken by different entities for technological reasons (one is an AC line and the other a DC line and the two companies have different expertise) or other reasons (for example, through separate ownerships of rights of way or different political jurisdictions enabling siting).

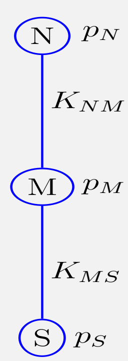

Figure 10

Suppose the two complementary projects are built (that is, the standard "coordination failure" does not arise) and that companies choose their capacities K_{NM} and K_{MS}. The value of the point-to-point rights are

$$K_{NM}\left(p_M - p_N\right),$$

and

$$K_{MS}\left(p_S - p_M\right),$$

respectively. The incentive for gaming comes from the fact that the lower-capacity line grabs the entire rights' value: Suppose, for instance, that $K_{MS} < K_{NM}$. Then, $p_M = p_N$ while $p_S > p_M$. This gives rise to a game in which each would like to have a capacity slightly lower than the other. Hence, none dares to move first as the other will then make sure to collect the entire rent.

To be certain, this example is extreme, but it illustrates a general point: Merchant investment is conducive to preemption (see section 5) and war-of-attrition strategies (Tirole 1988, pp.311-314). These detrimental behaviors can be avoided through a centralized process involving various forms of incentives to promote incentive compatibility. But, by so doing, the market moves closer to a centralized process.

10 Forward markets and commitment

As we discussed earlier, merchant investment is likely to be most appropriate for major new links that expand the geographic "footprint" of the transmission network, rather than for network deepening investments that involve enhancements to existing facilities. Constructing a major new line, however, involves both a long lead time and substantial uncertainty as to the availability of a crucial input, namely the various authorizations needed to build the line, and as to the nodal prices of electricity in the distant future. This gives rise to three concerns:

- *Availability of financing.*

Merchant investment is a high-powered-incentives activity. Merchants thus bear a substantial long-term risk. To obtain financing, they probably will want to offload a good part of this risk.[28] One technique for doing involves entering into financial arrangements with generators and load-serving entities. The latter then still face energy price risk as

[28] For theoretical foundations for the desirability of this offloading, see Holmström-Tirole (2000).

well as (if this transmission project cannot be brought to completion) counterparty risk. In principle, some insurance should also be supplied by non-stakeholders. For reasons that have received insufficient attention in economic theory, such long-term forward markets are usually poorly developed, though. This fact can make it hard for merchant investors to raise financing.

- *Credibility vis-á-vis projects with shorter lead times.*

Merchant transmission projects that increase capacity between an import constrained area with high nodal prices and an export constrained area with low nodal prices are, in a sense, substitutes for generation projects of equivalent capacity inside the import constrained area. While the merchant transmission project does not compete directly with this kind of generation project, it does make it possible for generators outside the constrained area to compete with existing and new generators inside the import constrained area. Suppose for instance that a merchant investor plans to invest in a new North-South line, whose acquisition of siting permits plus actual construction will take say 10 years.[29] If the transmission project is not built assume that a generator with equivalent capacity would be built in the South and that such a project would take only 3 years to obtain siting approvals and to be constructed. There is room for only one of these two alternative ways of reducing the price wedge between North and South. The merchant transmission investor is at a strategic disadvantage even if his project is socially more valuable. If most of the costs involved in building a new line are sunk after the first two years, then the merchant investor is likely to cancel his project if the new generation plant in the South is built. Knowing this, the generator may well try to use his short-term investment period to preempt the transmission project, and this even if the merchant investor has

[29]Transmission lines do not take very long to build once they have obtained siting permits. However, for major new transmission corridors, the permitting process can be very lengthy.

announced his intention and has started work on this project.[30]

- *Regulatory uncertainty and opportunism.*

Government and regulators have substantial discretion over the profitability of energy projects. In the case of the construction of a new line, they will first affect the probability that the company receives the authorizations needed to build it. And, once it is built, the choice of rating paradigm (which determines the number of rights allocated to the merchant), the imposition of energy price caps, the definition of incentives for the System Operator (see section 7 above), the build-up of parallel lines under different incentives (e.g., by a Transco regulated under cost-of-service and aiming at reducing nodal price differences or market power in the South) all impact the merchant investor's long-term return.

While this commitment problem exists for all investments, it is partially mitigated on the short end by institutional factors (short-term stability of the regulatory environment, 5-year regulatory commitments,) and by the current regulators' reputation concerns. But long-term commitments are less desirable and administrations change. This is why long term investments whose payoffs are heavily dependent on government policies are often performed either by a State-owned enterprise or by a utility under some cost-of-service scheme, but not by a private company under a high-powered incentive scheme.[31]

11 Conclusion

We have examined the performance attributes of a "merchant transmission" model in which investments in electric transmission capacity rely upon competition and free entry to exploit profitable transmission investment opportunities rather than on regulated

[30] Such timing issues are of course not specific to transmission investments. But the latter are particular vulnerable to preemption strategies due to their long lead time.

[31] Unless legal protection against expropriation may be supplied by the court system, which requires that expropriation take blatant, rather than subtle forms.

monopoly transmission companies. In return for investment in additional transmission capacity, merchant investors receive property rights that allow them to collect congestion revenues equal to the difference in nodal energy prices associated with the incremental point-to-point transmission capacity their investments create. The value of these rights to receive congestion revenues represent the revenues merchant investors receive to cover the capital and operating costs of these investments. Previous theoretical research has demonstrated that under a fairly stringent set of assumptions (a) all efficient investments in transmission capacity will be profitable and (b) all inefficient investments will be unprofitable. These results undermine the traditional assumption that transmission networks are "natural monopolies" that must be subject to regulation of their maintenance and operating decisions as well as their investment decisions. If they are correct in practice, they would lead to the remarkable conclusion that both the generation of electricity and the transmission of electricity can be largely deregulated.

Our work examines how these results are affected by introducing assumptions that more accurately reflect the physical and economic attributes of transmission networks. We argue first that there are many investment opportunities that involve enhancements to the existing transmission network and that are physically and operationally inseparable from the incumbent transmission owner's network. These "network deepening" investments can be undertaken most efficiently by the incumbent transmission owner. If incumbents were able to pursue these investments on a merchant basis they would be able to extract significant monopoly rents and would underinvest in these types of transmission capacity expansion opportunities. Accordingly, these types of investments are not good candidates for governance by the merchant investment model. Investments that expand the footprint of the existing transmission network by building major new transmission lines, what we call "network expansion investments," are the most likely candidates for successful merchant investment. However, we show how various attributes of transmission investments lead to market imperfections and inefficiencies with the merchant model. When there are imperfections in the competitive wholesale electricity markets that lead nodal spot electricity prices to depart from their efficient levels, investment incentives will be distorted. For example, when unregulated generators have market power, nodal energy prices will be distorted from their efficient levels. These distortions may lead to over-investment or under-investment depending upon where on the network electricity generators have market power. Imperfect government interventions to control market power in competitive wholesale electricity markets may also distort investment incentives. The absence of a good set of liquid forward markets for trading wholesale electricity at all points on the network is a further deterrent to efficient investment.

Network expansion investments that are most conducive to supply by competitive entrants are also likely to be characterized by economies of scale or "lumpiness." We show how economies of scale will lead to under-investment, to monopoly pre-emption of competitive generation or transmission investments, and distort the timing of investments. We argue that these problems are unlikely to be resolved by relying on bilateral or multilateral negotiations among the market participants who gain and lose from these investments. Indeed, opportunities for multilateral bargaining may further distort investment incentives.

We also show that the types of non-contingent transmission property rights that have been assumed to be awarded to investors in transmission capacity are poorly matched to the stochastic characteristics of transmission networks and investments in them. Instead, transmission property rights that are contingent on exogenous variations in transmission capacity and reflect the diversification attributes of new investments would be necessary to properly align investment incentives with the stochastic attributes of transmission networks. We explore these property rights definition and allocation issues for both radial networks and networks with loop flow. Unfortunately, defining and allocating contingent property rights that provide efficient investment incentives is also likely to be inconsistent with the development of liquid competitive markets for these rights or derivatives on them.

Under the merchant transmission model the ownership and maintenance of transmission facilities are to be separated from decisions regarding security-constrained bid-based dispatching of generators and price-responsive demand on the network and managing reliability criteria and constraints. The latter decisions are to be made by independent system operators that are unaffiliated with generators, energy marketers or transmission owners. We show that maintenance and operating decisions made by transmission owners are interdependent with dispatch and network reliability decisions made by monopoly system operators. Separating these decisions potentially leads to moral hazard problems and associated inefficiencies. Transmission owners and system operators must have a compatible set of incentives to avoid these inefficiencies, but designing these incentives is challenging. Vertical integration between transmission ownership and system operations is likely to reduce these incentive problems and, if transmission owners are also independent of generator and marketing of power, may be a superior organizational structure.

As a practical matter it is appears to be unlikely that we can rely primarily on competitive merchant investment to provide efficient investments in transmission infrastructure necessary to support efficient competitive wholesale power markets. The challenge for future research is to develop regulatory mechanisms that facilitate efficient investment and operating decisions by incumbent regulated transmission network owners, stimulate merchant investment when it is more efficient, and convey the net benefits of efficient investment and operating decisions made by both regulated and merchant transmission owners to consumers. Developing good performance based regulatory mechanism to govern both the behavior of incumbent transmission owners and to stimulate efficient investment decisions by incumbents and merchant developers is an essential feature of a system that relies on competitive wholesale power markets whose participants depend on the transmission network to support wholesale electricity market competition.

References

[1] Aghion, P., and P. Bolton (1987) "Contracts as a Barrier to Entry," *American Economic Review*, 77(3): 388-401.

[2] Baldick, R., and E. Kahn (1992) "Transmission Planning in the Era of Integrated Resource Planning: A Survey of Recent Cases," Lawrence Berkeley Laboratory, LBL-32231.

[3] Bohn, R.E., Caramanis, M.C., and F.C. Schweppe (1984) "Optimal Pricing in Electrical Networks over Space and Time," *Rand Journal of Economics*, 15 (3): 360–376.

[4] Bushnell, J. (1999) "Transmission Rights and Market Power," *Electricity Journal*, 12(8): 77–85.

[5] Bushnell, J., and S. Stoft (1996) "Electric Grid Investment Under a Contract Network Regime," *Journal of Regulatory Economics*, 10: 61–79.

[6] —— (1997) "Improving Private Incentives for Electric Grid Investment," *Resource and Energy Economics*, 19: 85–108.

[7] Celebi, M. (undated) "An Analysis of Incentives to Provide Line Capacity and Reliability in Deregulated Power Networks," mimeo, Brattle Group.

[8] Chao, H. P., and S. Peck (1996) "A Market Mechanism for Electric Power Transmission," *Journal of Regulatory Economics*, 10(1): 25–59.

[9] Chisari, O. et al (2001) "High-Tension Electricity Network Expansion in Argentina: Decision Mechanisms and Willingness-to-pay Revelation," *Energy Economics*, 23: 697-715.

[10] Glachant, J-M, and V. Pignon (2002) "Nordic Electricity Congestion's Arrangement as a Model for Europe: Physical Constraints and Operators' Opportunism," mimeo.

[11] Green, R. (1992) "Contracts and the Pool: The British Electricity Market," mimeo, Department of Economics, University of Cambridge.

[12] —— (1997) "Transmission Pricing in England and Wales," *Utilities Policy*, 6(3): 185–93.

[13] Hansmann, H. (1996) *The Ownership of Enterprise*, Belknap Harvard.

[14] Hogan, W. (1992) "Contract Networks for Electric Power Transmission," *Journal of Regulatory Economics*, 4: 211–242.

[15] —— (2002) "Financial Transmission Rights Formulations," mimeo, Harvard University.

[16] Holmström, B. (1982) "Moral Hazard in Teams," *Bell Journal of Economics*, 13: 324–340.

[17] Holmström B., and J. Tirole (2000) "Liquidity and Risk Management," *Journal of Money, Credit and Banking*, 32(3): 295–319.

[18] Joskow, P. (2002) "Electricity Sector Restructuring and Competition: A Transaction Cost Perspective," in Eric Brousseau and Jean-Michel Glachant (eds.) *The Economics of Contracts: Theories and Applications*. Cambridge: Cambridge University Press.

[19] Joskow, P., and R. Schmalensee (1983) *Markets for Power*. MIT Press.

[20] Joskow, P., and J. Tirole (2000) "Transmission Rights and Market Power on Electric Power Networks," *Rand Journal of Economics*, 31(3): 450–487.

[21] Laffont, J-J., and J. Tirole (1993) *A Theory of Incentives in Procurement and Regulation*, MIT Press.

[22] Léautier, T. (2000) "Regulation of An Electric Power Transmission Company," *The Energy Journal*, 24(1): 61–92.

[23] Nasser, T.O. (1997) "Imperfect Markets for Power: Competition and Residual Regulation in the Electricity Industry," PhD Dissertation, MIT Department of Economics.

[24] Oren, S. (1997) "Economic Inefficiency and Passive Transmission Rights in Congested Electric Systems with Competitive Generation," *The Energy Journal*, 18: 63–83.

[25] Oren, S., P. Spiller, P. Varaiya, and F. Wu (1995) "Nodal Prices and Transmission Rights: A Critical Appraisal," *The Electricity Journal*, 8(3): 24–35.

[26] Patton, D. (2002) "Summer 2002 Review of New York Electricity Markets," October 15, 2002, pp.25, 42-64.

[27] Perez-Arriaga, I. J. et al (1995) "Marginal Pricing of Transmission Service: An Analysis of Cost Recovery," *IEEE Transactions on Power Systems*, 10(1): 546–553.

[28] Stoft, S. (1999) "Financial Transmission Rights Meet Cournot: How TCCs Curb Market Power," *The Energy Journal*, 20: 1–23.

[29] Tirole, J. (1988) *The Theory of Industrial Organization*, MIT Press.

[30] Vogelsang, I. (2001) "Price Regulation for Independent Transmission Companies," *Journal of Regulatory Economics*, 20(2): 141–165.

[31] Williamson, O. (1983) "Credible Commitments: Using Hostages to Support Exchange," *American Economic Review*, 73: 519–540.

[32] Wilson, R. (2002) "Architecture of Power Markets," *Econometrica*, 70(4): 1299–1340.

[33] Wu, F.P., Varaiya, P., Spiller, P., and S. Oren (1996) "Folk Theorems on Transmission Access : Proofs and Counterexamples," *Journal of Regulatory Economics*, 10(1): 5–24.