NHH

Norges Handelshøyskole

Norwegian School of Economics and Business Administration

Topics on Electricity Transmission Pricing

by Mette Bjørndal

A dissertation submitted for the degree of dr. oecon.

ISBN : 82- 405-0040-4

Acknowledgements

I would like to express my deepest respect and gratitude to my advisor, professor Kurt Jörnsten. Without his involvement with creative ideas, insightful comments, encouragement and support, this work would have been different, and certainly less exciting.

I am also thankful to professor Einar Hope, Balbir Singh, Christian Andersen, Linda Rud, and Endre Bjørndal for introducing me to the area of electricity transmission, and for contributing to the many inspiring discussions that this thesis has benefited from. In particular my thanks go to Linda and Endre who have read and commented on my thesis.

I would also like to thank the other members of my committee, professor Sverre Storøy, for being partly responsible for me attending Kurt's operations research class in the first place, and professor Geir Dahl, for recommending the economist's perspective on network problems the first time that we met.

This research has been financed by the Foundation for Research in Economics and Business Administration, through a research project sponsored by Telenor. This is very much appreciated, as is the hospitality and friendly atmosphere at the Department of Finance and Management Science that has provided me office space and the opportunity to meet many interesting people and great friends during these years.

Finally, I would like to thank my family for all their love and support. Especially, Jan-Erling who has taken more than his fair share of the responsibilities at home, our children Marie and Ivar who have felt the "sufferings" of a doctorate candidate at an early age, and my mother and father who have "filled in" for us whenever asked. I am looking forward to returning your kind favors.

NHH, Bergen, February 2000
Mette Bjørndal

Contents

1. INTRODUCTION........1

2. POWER FLOW MODELS........5

2.1. SINGLE-PHASE AC CIRCUITS........5
2.2. NETWORK EQUATIONS........9

3. OPTIMAL DISPATCH AND OPTIMAL PRICES........21

3.1. NODAL PRICES........23
3.2. TRANSMISSION CONGESTION CONTRACTS (TCCS)........29
3.3. CHAO-PECK PRICES........32
3.4. CP+HUB PRICES........35
3.5. ZONAL PRICING........35
3.6. COORDINATED MULTILATERAL TRADES........37
3.7. COMPARISON OF THE DIFFERENT COORDINATION MECHANISMS IN AN EXAMPLE........40
3.8. MARKET STRUCTURE AND THE ROLE OF THE SYSTEM OPERATOR........49
3.9. REACTIVE POWER AND SECURITY CONSTRAINTS........50
3.10. LONG RUN CONSIDERATIONS........52

4. AN OVERVIEW OF THE NORWEGIAN TRANSMISSION SYSTEM........55

5. A NOTE ON LOOP FLOW AND ECONOMIC MODELING........61

6. IMPLEMENTATION OF TARIFFS FOR MARGINAL LOSSES........67

6.1. COMPUTING MARGINAL LOSS FACTORS........68
6.2. DO LOSS FACTORS REFLECT MARGINAL LOSSES?........74
6.3. ALTERNATIVE LOSS FACTORS........79
6.4. CONCLUDING REMARKS........83

7. ZONAL PRICING........85

7.1. COMPUTING ZONAL PRICES........87
7.2. EXAMPLES........88
7.3. THE CORRECT BASIS FOR ZONE ALLOCATIONS........98
7.4. CONCLUSIONS AND FUTURE RESEARCH........108

8. PARADOXES IN ELECTRICITY NETWORKS........111

8.1. BRAESS' PARADOX AND GENERALIZATIONS........111
8.2. GRID INVESTMENTS IN ELECTRICITY NETWORKS........113
8.3. MARKET INTEGRATION........125
8.4. SUGGESTED CURES........127
8.5. CONCLUSIONS........128

9. COMPETITIVE EFFECTS OF CONGESTION........129

9.1. EXAMPLE WITH THREE IDENTICAL LOCAL MARKETS........130
9.2. CONCLUSIONS AND FUTURE RESEARCH........137

10. SUGGESTIONS FOR FUTURE RESEARCH........139

11. REFERENCES........141

1. Introduction

Within the last decade we have experienced deregulation of several industries, such as airlines, telecommunications and the electric utility industry, the last-mentioned being the focus of this work. Both the telecommunications and the electricity sector depend on network facilities, some of which are still considered as natural monopolies. In these industries, open network access is regarded as crucial in order to achieve the gains from increased competition, and transmission tariffs are important in implementing this.

Based on the Energy Act that was introduced in 1991, Norway was among the first countries to restructure its electricity sector. On the supply side there are a large number of competing firms, almost exclusively hydro plants, with a combined capacity of about 23000 MW, producing 105-125 TWh per year, depending on the availability of water. Hydro plants are characterized by low variable costs of operation, however since water may be stored in dams, water has an opportunity cost, generally known as the water value, which is the shadow price of water when solving the generator's intertemporal profit maximization problem. Water values are the main factor of the producers' short run marginal cost.

Total consumption amounts to 112-117 TWh a year, and consumers, even households, may choose their electricity supplier independent of the local distributor to which the customer is connected. In fact, approximately 10% of the households have actually changed supplier. The web-site www.konkurransetilsynet.no indicates available contracts, and www.dinside.no provides an "energy-calculator" where one can check whether it is profitable to switch supplier. If a customer buys energy from a remote supplier, the local distributor only provides transportation facilities for the energy and is compensated accordingly.

Transmission and distribution have remained monopolized and regulated by the Norwegian Water Resources and Energy Directorate (NVE). To prevent cross-subsidization, the largest state-owned firm was split to form Statkraft SF, Norway's largest generation company, and Statnett SF, the major transmission company. Statnett is also the system operator of the entire Norwegian power system. Many distribution companies, usually publicly owned, are still part of vertically integrated companies, but there is strict separation of the financial accounts for transmission/distribution and generation/marketing.

Trading is accomplished through several channels. The largest organized market is the common Norwegian-Swedish pool, NordPool. NordPool ASA is owned by Statnett Marked (which is a subsidiary of Statnett) and the Swedish grid-company, Svenska Kraftnät. NordPool also supports the Finnish and Danish markets, thus through cooperation with the grid-companies of Denmark (Eltra) and Finland (Fingrid), locational markets are provided in these countries as well. Unlike England and Wales, only part of actual production is traded on this market. In addition, independent brokers offer various energy-contracts and facilitate bilateral contracts. In 1998 22% of the energy was traded on the hourly day-ahead (spot) market of NordPool. NordPool also organizes a forward and futures market and settles contracts from the regulation market organized by Statnett. The regulation market is a real-time market used to settle imbalances in real time. In total, close to 30% of the energy traded is organized through NordPool, and this share is increasing year by year.

The efficiency of the market is heavily affected by the operation and pricing of the transmission system, and the topic of this thesis concerns the interaction of the transmission network and the energy markets. Throughout, the findings are illustrated by means of simple examples. This is to enhance readability and intuition. To start with, chapter 2 is devoted to power flow models, whereas chapter 3 provides an overview of models developed to efficiently coordinate the allocation of transmission resources. The focus is mainly on short-term efficiency, and the survey is only partial, but provides an integrated overview of some of the theoretical models most frequently cited in the literature. Chapter 4 gives an overview of the Norwegian transmission system, especially the central high voltage grid, and we describe the tariff structure, which applies in this part of the network.

In chapter 5 we comment on loop flow and economic modeling. Chapters 6 and 7 concern the implementation of short run marginal cost pricing in the Norwegian transmission system, marginal losses are treated in chapter 6, and in chapter 7 zonal pricing is examined. Loop flow induces seemingly paradoxical situations in power transmission, and in chapter 8 we show that a new line may reduce social surplus, whereas in chapter 9 the competitive effects of a new line are studied. The main contributions of the thesis are found in chapters 5-9. In chapter 10 suggestions for future research are indicated.

2. Power Flow Models

In this chapter we will give a (strongly) simplified overview of power flow analysis. The purpose is to give a flavor of this intricate matter and to provide a background for the models described in later chapters. More details may be found in, for instance, Cuthbert [16], Wood and Wollenberg [88] and Young and Freedman [93].

2.1. *Single-Phase AC Circuits*

In an *alternating current* (AC) circuit operating under steady-state conditions the instantaneous current and voltage are functions of time. More specifically, they are sinusoidals. A general AC circuit involves any combination of resistors, inductors and capacitors, as well as the AC source. Resistors are characterized in terms of resistance, inductors and capacitors in terms of inductive and capacitive reactances, respectively. For a resistor, voltage and current will be in phase. In the case of inductors, voltage leads current by 90°, while in a capacitor voltage *lags* current by 90°. When combining elements in a circuit, voltages are added in the manner of vector addition, and the voltage of the entire circuit comes out with a *phase angle* of some δ with respect to the current. This is our starting point.

Instantaneous Power

Instantaneous current, $i(t)$, and voltage, $v(t)$, at time t can then be written as

$$i(t) = \mathrm{I}_{max}\cos(\omega t)$$
$$v(t) = \mathrm{V}_{max}\cos(\omega t + \delta)$$

where I_{max} and V_{max} are the maximal values of current and voltage (i.e. the amplitudes of the sinusoidals), ω is the (constant) frequency and δ is the phase angle of voltage with respect to current. Instantaneous power, $p(t)$, equals the product of voltage and current, i.e.

(2-1) $$p(t) = v(t) \cdot i(t) = \mathrm{V}_{max}\mathrm{I}_{max} \cos(\omega t + \delta)\cos(\omega t).$$

Instantaneous current and voltage have a pulsating nature, taking on both positive and negative values. Currents and voltages in distribution systems are usually described in terms of their *root-mean-square* values or rms-values. As indicated by the name, the instantaneous function is squared, one takes the average, and finds the square root of this. It is easily shown that the rms-values of current and voltage are equal to

(2-2) $$\mathrm{I} = \frac{\mathrm{I}_{max}}{\sqrt{2}}$$

(2-3) $$\mathrm{V} = \frac{\mathrm{V}_{max}}{\sqrt{2}}$$

where I and V denote the rms-values of current and voltage, respectively.

Real and Reactive Power

Instantaneous power may be expressed in terms of rms-values. By employing trigonometric identities together with (2-2) and (2-3), (2-1) can be transformed into

(2-4) $$p(t) = \underbrace{\mathrm{VI}\cos\delta \cdot \{1 + \cos[2(\omega t + \delta)]\}}_{p_R(t)} + \underbrace{\mathrm{VI}\sin\delta \cdot \sin[2(\omega t + \delta)]}_{p_X(t)}.$$

We have divided the expression for $p(t)$ into two components where $p_R(t)$ can be interpreted as the power absorbed by the resistive elements of the circuit, whereas $p_X(t)$ is the reactive part of the load, and is associated with the inductive and capacitive elements[1].

The average value of $p_R(t)$ equals

$$P = \mathrm{VI}\cos\delta\ .$$

[1] Strictly speaking, this pictures the circuit as a parallel R-X load.

P is the *real power* or *active power*, and it can be given the following interpretation: Total *energy* absorbed by a load during a time interval consisting of n cycles of the sinusoidal voltage, each of time T, equals PnT. The factor $\cos\delta$, with an ideal value of 1 ($\delta = 0$), is called the *power factor*. A low power factor (resulting from a large angle δ of lag or lead) is usually undesirable in power circuits because for a given voltage, a large current is needed to supply a given amount of power. Since losses are approximately equal to resistance times the square of the current, this will result in large losses in the transmission lines.

The average value of $p_X(t)$ is zero, while the maximal value, or amplitude, equals

$$Q = \mathrm{VI}\sin\delta .$$

The term Q is referred to as *reactive power*. Even though the average is zero, implying zero energy, reactive power is necessary to completely describe the operation of power systems. In particular, reactive power is useful for controlling voltage. For instance, when connected to a circuit, inductors and capacitors may shift the phase angle δ. This affects the power factor and it may be shifted towards its ideal value of 1, thus reducing losses. Inductors and capacitors themselves absorb no energy. This however, does not mean that reactive power is a free good. It does require fixed equipment, which is costly and limits the supply.

Complex Power

Power in AC systems is often described by means of complex numbers. The phasor-representation[2], $V = \mathrm{V}e^{j\delta} = \mathrm{V}\cos\delta + j\mathrm{V}\sin\delta$, of the rms-voltage is used, and one of the basic rules of circuit theory is that power is equal to the product of voltage, V, and the *conjugate* of the current, I^*. Since δ denotes the phase angle with respect to current, S, frequently referred to as *apparent power*, is equal to

$$S = VI^* = \mathrm{V}e^{j\delta} \cdot \mathrm{I}e^{j\cdot 0} = \mathrm{VI}\cos\delta + j\mathrm{VI}\sin\delta = P + jQ .$$

[2] By using Euler's identity, $e^{j\varphi} = \cos\varphi + j\sin\varphi$, we can express a sinusoid in terms of a phasor. For the voltage, $v(t) = \mathrm{Re}[\mathrm{V}_{max}e^{j(\omega t+\delta)}] = \mathrm{Re}[\sqrt{2}(\mathrm{V}e^{j\delta})e^{j\omega t}]$, where $j = \sqrt{-1}$ and Re denotes the 'real part of'. The rms phasor-representation of the voltage is given by $V = \mathrm{V}e^{j\delta} = \mathrm{V}\cos\delta + j\mathrm{V}\sin\delta$.

The real part of S is the real or active power (which is the time average of instantaneous power), whereas the imaginary part is the reactive power (which is the amplitude of the reactive part of the load). The relationship between the quantities may best be pictured by the power triangle of Figure 2-1.

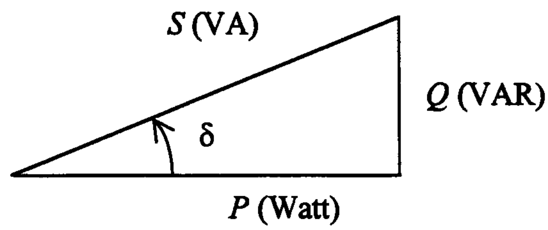


Figure 2-1 The Power Triangle

In Figure 2-1 units are shown in parentheses, VAR is short for VoltAmperesReactive and VA stands for VoltAmpere. Even if all the power terms are products of voltage (Volt) and current (Ampere), the notational differences of the units are maintained to distinguish real and reactive components.

Poly-Phase Circuits

In a single-phase system, instantaneous power has a pulsating nature, which is regarded undesirable for medium and large loads. In real world applications, this is avoided by using poly-phase generators, usually three-phase systems where voltage and current of the second and third phases are shifted by 120° and 240° with respect to the first. The equation for every single phase looks like (2-4), but when taking the sum of the three phases, it has the effect that $p(t) = P$, i.e. power is not a function of time, but a constant. This has several advantages, including reduced capital and operating costs of transmission and distribution, and better voltage control. However, the use of poly-phase systems does not alter the basics of this section, where the intention has been to show the relation of real and reactive power to the sinusoidal voltage and current.

2.2. Network Equations

In this section we consider a network of electrical buses, or nodes, and the equations governing the power flows of the grid. More specifically, we want to show the relationships between power injections and withdrawals in the nodes and the power flowing over transmission links. We want to characterize these power flows in terms of network characteristics like line resistances and reactances and phase angle differences. In this section, all the electrical quantities are represented by complex numbers.

Kirchhoff's Laws and Conservation of Energy

Together with the law of conservation of energy, Kirchhoff's laws are essential when modeling electricity networks. The laws are used a number of times when developing the flow equations and are consequently built into the power flow equations as stated below.

1. *Kirchhoff's current law (junction rule)*

The current flow into any vertex is equal to the current flowing out of it. If there are n nodes in the network, there will be $n-1$ independent equations[3]. It follows immediately from Kirchhoff's junction rule that $P_i = \sum_{k \neq i} P_{ik}$ (definitions follow).

2. *Kirchhoff's voltage law (loop rule)*

The algebraic sum of the potential differences across all components around any circuit or cycle is zero. If there are n nodes in the network and m edges, there will be $m-n+1$ independent loops. A procedure to identify the independent loops is given by Dolan and Aldous [19]. Kirchhoff's loop rule implies that we can calculate the voltage across a line as the nodal voltage difference, i.e. the voltage drop of line ik, V_{ik}, is equal to $V_i - V_k$.

3. *The law of conservation of energy*

Total generation is equal to total consumption plus losses.

[3] Dolan and Aldous [19] also state the junction rule in terms of current over cutsets.

Given any two of these three laws, it is possible to derive the third. Versions of rules 1 and 3 are familiar from ordinary transportation networks, for instance commodity flows, while rule 2 is special for electricity networks[4] and limits the possibility of routing.

Power Flow Equations

Consider a power network consisting of n electrical buses, operating in sinusoidal steady-state. Let $\mathbf{I} = (I_1, \ldots, I_n)$ be the vector of complex currents, $\mathbf{V} = (V_1, \ldots, V_n)$ the vector of complex voltages and $\mathbf{Y} = [Y_{ik}]$ the *admittance* matrix. The admittance matrix is constructed from the complex *impedances* of the transmission lines. The impedance of line ik is denoted z_{ik}, and

$$z_{ik} = r_{ik} + j \cdot x_{ik} = z_{ki}$$

where r_{ik} is the *resistance* of the line, x_{ik} the *reactance* (inductive – capacitive), and $j^2 = -1$. The admittance of a line between two different nodes is defined as

$$\underset{i \neq k}{Y_{ik}} = -\frac{1}{z_{ik}} = -\frac{1}{r_{ik} + j \cdot x_{ik}} = -\frac{r_{ik}}{r_{ik}^2 + x_{ik}^2} + j \cdot \frac{x_{ik}}{r_{ik}^2 + x_{ik}^2} = G_{ik} + j \cdot B_{ik}$$

where G_{ik} is called the *conductance* of line ik and B_{ik} the *susceptance.* If there is no direct line between i and k, $Y_{ik} = 0$. The entries of the diagonal of the matrix are defined by[5]

$$Y_{ii} = \sum_{k \neq i} \frac{1}{z_{ik}} = \sum_{k \neq i} -Y_{ik} = \sum_{k \neq i} \frac{r_{ik}}{r_{ik}^2 + x_{ik}^2} + j \cdot \sum_{k \neq i} \frac{-x_{ik}}{r_{ik}^2 + x_{ik}^2} = G_{ii} + j \cdot B_{ii}$$

The admittance matrix is symmetric and every row and every column sum to zero.

Current, voltage and admittance are related by *Ohm's law,* $\mathbf{I} = \mathbf{YV}$, implying that current at bus i is equal to

[4] The rule applies to some other physical networks too, for instance pipelines.

$$I_i = \sum_{k=1}^{n} Y_{ik} V_k \,.$$

Let P_i and Q_i be the real and reactive powers injected at bus i, and I_i^* the conjugate of the current. Then apparent power is given by

$$\text{(2-5)} \qquad S_i = P_i + j \cdot Q_i = V_i \cdot I_i^* = V_i \sum_{k=1}^{n} Y_{ik}^* V_k^* \,.$$

Inserting $V_i = \mathrm{V}_i e^{j\delta_i} = \mathrm{V}_i(\cos\delta_i + j \cdot \sin\delta_i)$ and $Y_{ik} = G_{ik} + j \cdot B_{ik}$ into (2-5) gives real and reactive power (after some manipulations) as

$$\text{(2-6)} \qquad P_i = \sum_{k=1}^{n} \mathrm{V}_i \mathrm{V}_k \left[G_{ik} \cos(\delta_i - \delta_k) + B_{ik} \sin(\delta_i - \delta_k) \right]$$

$$\text{(2-7)} \qquad Q_i = \sum_{k=1}^{n} \mathrm{V}_i \mathrm{V}_k \left[G_{ik} \sin(\delta_i - \delta_k) - B_{ik} \cos(\delta_i - \delta_k) \right]$$

showing the power flows as functions of the phase angles $\delta_i, i = 1,\ldots,n$ and rms-voltages $\mathrm{V}_i, i = 1,\ldots,n$. P_i and Q_i may be positive (representing *injections*) or negative (representing *withdrawals*). The AC power flows are now given by a system of $2n$ nonlinear equations. The equations are, however, not independent. If the rms-value of voltage at each node and the admittance-matrix are given, only phase angle *differences* matter. We can fix a single δ - value, say δ_1, and once the $n-1$ angle differences $(\delta_2 - \delta_1, \ldots, \delta_n - \delta_1)$ are assigned, so are all the other angle differences, and so are all the power flows.

A closer examination of the expression for real power at node i gives

$$\begin{aligned} P_i &= \mathrm{V}_i^2 \cdot G_{ii} + \sum_{k \neq i} \mathrm{V}_i \mathrm{V}_k \left[G_{ik} \cos(\delta_i - \delta_k) + B_{ik} \sin(\delta_i - \delta_k) \right] \\ &= \mathrm{V}_i^2 \cdot \sum_{k \neq i} \tfrac{r_{ik}}{r_{ik}^2 + x_{ik}^2} + \sum_{k \neq i} \mathrm{V}_i \mathrm{V}_k \left[G_{ik} \cos(\delta_i - \delta_k) + B_{ik} \sin(\delta_i - \delta_k) \right] \end{aligned}$$

[5] We ignore any impedances from buses to ground.

$$= -\sum_{k\neq i} G_{ik}V_i^2 + \sum_{k\neq i} G_{ik}V_iV_k\cos(\delta_i - \delta_k) + \sum_{k\neq i} B_{ik}V_iV_k\sin(\delta_i - \delta_k),$$

which is the expression used as the starting point of the analyses of for instance, Schweppe et al. [64] and Wu and Varaiya [89]. Define P_{ik} as power leaving node i and flowing towards node k. A positive P_{ik} indicates outflow, while a negative P_{ik} indicates inflow. Calculating power using the voltage difference across a line[6] gives P_{ik} equal to

(2-8) $$P_{ik} = -G_{ik}V_i^2 + G_{ik}V_iV_k\cos(\delta_i - \delta_k) + B_{ik}V_iV_k\sin(\delta_i - \delta_k)$$

and it follows that

$$P_i = \sum_{k\neq i} P_{ik}.$$

Similar expressions can be found for reactive power. Reactive power leaving node i in direction node k equals

$$Q_{ik} = B_{ik}V_i^2 + G_{ik}V_iV_k\sin(\delta_i - \delta_k) - B_{ik}V_iV_k\cos(\delta_i - \delta_k)$$

and

$$Q_i = \sum_{k\neq i} Q_{ik}.$$

Transmission Losses and Capacity Constraints

Consider a transmission line ik and assume that power flows from i to k, i.e. $P_{ik} > 0$ and $P_{ki} < 0$ (Figure 2-2).

Figure 2-2 Transmission Losses

Generally, $|P_{ik}| \neq |P_{ki}|$ because some power is "lost" on its way, heating up the line. The real power loss of line ik equals $P_{ik} + P_{ki}$, and is approximately proportional to the square of the power flowing over the line. In power systems, total losses typically constitute 2-5% of total generation, and are found by adding (positive) injections and (negative) withdrawals $\sum_{i=1}^{n} P_i$.

If P_{ik} is very large, the heating of the line can be of such a magnitude that it causes physical damage to the line. Depending on the characteristics of the wire, there will be a thermal limit $C_{ik} = C_{ki} > 0$ associated with the line, and it is required that[7]

(2-9) $\quad P_{ik} \leq C_{ik}$.

There are also a number of other constraints concerning the reliability and security of operation. This may be voltage-limits, or it may be constraints insuring that the system will operate under different fault-situations (contingency constraints). Contingency constraints may be expressed as constraints on power flow, similar to (2-9).

Computing Power Flows

A set of consistent real power injections is a vector $\mathbf{P} = (P_1, \ldots, P_n)$ that satisfies the power flow equations (2-6). Because of the interdependencies of power flows and losses, it is not trivial to find a vector of consistent injections. Positive node injections must account for total withdrawals plus losses, but at the same time the distribution of nodal injections is decisive

---

[6] $S_{ik} = P_{ik} + jQ_{ik} = V_i \cdot I_{ik}^*$, and it follows from Kirchhoff's loop rule that $I_{ik} = \frac{V_{ik}}{z_{ik}} = \frac{V_i - V_k}{z_{ik}} = Y_{ik} \cdot (V_k - V_i)$.

for the line flows, which again determine the size of the losses. Phrased differently, assuming voltage magnitudes fixed, given a set of injections $\mathbf{P}$, it is either not allowed or it determines completely the voltages and line flows. This means that routing is given whenever injections are. In contrast to, for instance, transportation problems, this implies that it is not possible to route around a capacity constraint.

Consider the following simple example, Figure 2-3 showing a fragment of a network, i.e. node i and incident edges 1, 2 and 3.

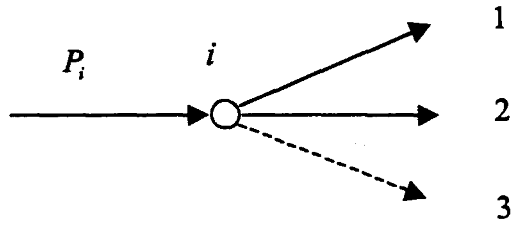


Figure 2-3 Congestion

Node i produces a positive injection of P_i resulting in outflows over links 1, 2 and 3. We assume that link 3 is congested. Increasing P_i will increase outflow on lines 1, 2 and 3 in given proportions, depending on the loads of the entire network. This means that even if lines 1 and 2 are uncongested, generation cannot increase in node i because it would increase flow over line 3, which is already at its limit. For generation in node i to increase, there must be a change in injections in other nodes relieving the congested line 3. The only way to control this system is through nodal injections. We have only n of those, and only $n-1$ are independent.

From the power flow equations (2-6) and (2-7), it is evident that an AC power flow generally works with four basic types of variables, that is real and reactive power, voltage magnitudes and phase angles. To find a vector of consistent injections, some variables must be specified because there are $4n$ unknowns and only $2n$ equations. Because of the interdependencies in

[7] In practice, transmission capacity constraints are often imposed as limits on apparent power. For fixed rms-voltages and small angle differences however, apparent power and real power differ by (approximately) a scaling factor (Chao and Peck [12]).

the system of equations, it is not arbitrary which variables to specify. For instance one might specify power injections (real and reactive) in $n-1$ buses and voltage (rms-magnitude and phase angle) in a single bus. We can then solve for the $n-1$ unknown voltages and find the resulting power injections at the last bus, which we call the swingbus. This procedure might be carried out using numerical methods like Gauss-Seidel or Newton-Raphson (for more details see for instance Cuthbert [16] or Wood and Wollenberg [88]).

Approximations and Simplifications

Often there is a need to simplify the flow-calculations. Power system engineers have observed that there are, on the one hand, strong couplings between real power and voltage angle, and between reactive power and voltage magnitude. On the other hand, there are much weaker couplings between voltage angle and reactive power, and between voltage magnitude and real power. McGuire [51] classifies different approximations in a very informative manner, showing the relation to the full AC power flow equations. Using our notation, a model is coded GB-PQ if the conductance matrix **G**, the susceptance matrix **B**, and the cost and value parameters for real and reactive power are all non-zero. If one of the items is zero, the corresponding letter is dropped from the type code.

The Decoupled Power Flow

A simplification known as "the decoupled AC power flow" uses the observation of the weak $P-\mathrm{V}$ and $Q-\delta$ relations to decouple the AC power flow equations into two subsystems: the $P-\delta$ equations and the $Q-\mathrm{V}$ equations, which then iterate with each other (Wood and Wollenberg [88]). This is a full GB-PQ model that introduces a number of approximations to ease computation. The procedure may not converge (find a solution of the full AC power flow) if some of the underlying approximations are not valid.

Direct Current (DC) Power Flows

A G-P model refers to a strictly direct current (DC) network containing resistors only. In relation to the AC power flow, the DC power flow can be regarded as a special instance, leaving out all imaginary parts. The entries of the admittance matrix are

$$Y_{ik} = -\frac{1}{r_{ik}} \text{ for } i \neq k \text{ and } ik \text{ being an edge in the network}$$

$$Y_{ik} = 0 \text{ if there is no edge between } i \text{ and } k$$

$$Y_{ii} = \sum_{k \neq i} \frac{1}{r_{ik}}.$$

Since currents and voltages are real numbers, power injected at node i can be written as

$$P_i = V_i I_i = V_i \cdot \sum_{k=1}^{n} Y_{ik} V_k = V_i \left(Y_{ii} V_i + \sum_{k \neq i} Y_{ik} V_k \right)$$

$$= V_i \left(\sum_{k \neq i} \frac{1}{r_{ik}} \cdot V_i + \sum_{k \neq i} -\frac{1}{r_{ik}} \cdot V_k \right) = V_i \left(\sum_{k \neq i} \frac{V_i - V_k}{r_{ik}} \right)$$

and

(2-10) $$P_{ik} = \frac{V_i (V_i - V_k)}{r_{ik}},$$

exhibiting the (maybe) more familiar way of computing currents from voltage differences. We will use this simplification when analyzing pricing of marginal losses in the Norwegian transmission network in chapter 6.

The "DC" Model

When analyzing congestion constraints, models of type B-P are frequently used. Considering real power only and letting $\mathbf{G} = \mathbf{0}$, which implies $r_{ik} = 0 \; \forall ik$

$$P_{ik} = B_{ik} \mathrm{V}_i \mathrm{V}_k \sin(\delta_i - \delta_k) = \frac{\mathrm{V}_i \mathrm{V}_k}{x_{ik}} \sin(\delta_i - \delta_k) = -\frac{\mathrm{V}_i \mathrm{V}_k}{x_{ik}} \sin(\delta_k - \delta_i) = -P_{ki}.$$

Since $P_{ik} + P_{ki} = 0$ this is a lossless system. Assuming voltage magnitudes equal to 1 for all nodes

$$P_{ik} = \frac{1}{x_{ik}}\sin(\delta_i - \delta_k) \text{ and } P_i = \sum_{k \neq i} \frac{1}{x_{ik}}\sin(\delta_i - \delta_k),$$

leaving us with a system of $n-1$ independent nonlinear equations. For $|\delta_i - \delta_k|$ small, we have that $\sin(\delta_i - \delta_k) \approx (\delta_i - \delta_k)$, giving a linear approximation of the power flows, namely

(2-11) $$P_{ik} = \frac{\delta_i - \delta_k}{x_{ik}} \text{ and } P_i = \sum_{k \neq i} \frac{\delta_i - \delta_k}{x_{ik}}.$$

Models of this kind are called "DC", pseudo-DC or even DC. The term "DC" arose in this context because the linear relationship (2-11) between P and δ is analogous to the relationship (2-10) between current and voltage in a direct current network containing resistors only. Losses are incorporated in these models by, for instance, associating a quadratic (in power flow) loss function to the lines.

The linear "DC" approximation allows us to work with power variables directly, i.e. we do not need to consider phase angles. Assuming all line reactances equal to 1, the power flow equations can be stated as

(2-12) $$P_i = \sum_{k \neq i} P_{ik}, \quad i = 1, \ldots, n-1$$

(2-13) $$\sum_{ik \in L_l} P_{ik} = 0, \quad l = 1, \ldots, m-n+1$$

where $L = (L_1, \ldots, L_{m-n+1})$ is a set of independent loops and L_l is the set of directed arcs in a path going through loop l. (2-12) follow from Kirchhoff's junction rule and (2-13) follow from the system of equations that result when letting one of the phase angles equal to zero in (2-11). In a lossless network the law of conservation of energy simply implies that

$$\sum_i P_i = 0;$$

Example

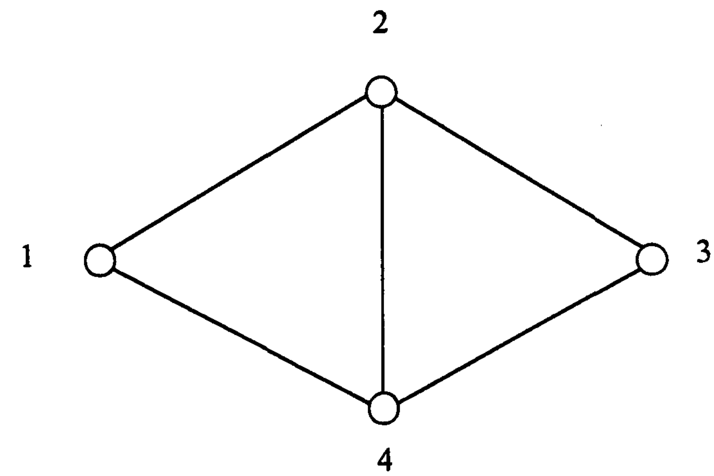


Figure 2-4 Four-Node Network

Using (2-11) with all impedances equal to 1, and letting $\delta_2 = 0$, we find from the P_{ik} expressions the two loop equations

(2-14) $$P_{14} = P_{12} + P_{24}$$

(2-15) $$P_{24} = P_{23} + P_{34}$$

constraining flows in loops 1-2-4-1 and 2-3-4-2, respectively. (Choosing $\delta_1 = 0$ instead, results in similar expressions for loops 1-2-4-1 and 1-2-3-4-1.) In addition, we need conservation of flow in 3 nodes, for instance

(2-16) $$P_1 = P_{12} + P_{14}$$

(2-17) $$P_2 = -P_{12} + P_{23} + P_{24}$$

(2-18) $$P_3 = -P_{23} + P_{34}$$

and finally

(2-19) $$P_1 + P_2 + P_3 + P_4 = 0.$$

The example can also serve as an illustration of the "routing-problem" in electricity networks. Consider increasing injection in node 1, withdrawing it in node 2. Flow will change on all the lines of the network in given proportions, according to line impedances. These proportions are referred to as distribution factors or load factors. Generally, the distribution factors depend not only on the grid's geometry, but also on the level of power flows. In the "DC" approximation, however, distribution factors are constant, and for trades between nodes 1 and 2, they can be found by setting $P_1 = 1$, $P_2 = -1$, $P_3 = P_4 = 0$ and solving for the line flow variables in the system of equations (2-14)-(2-19). The resulting factors are shown in Figure 2-5.

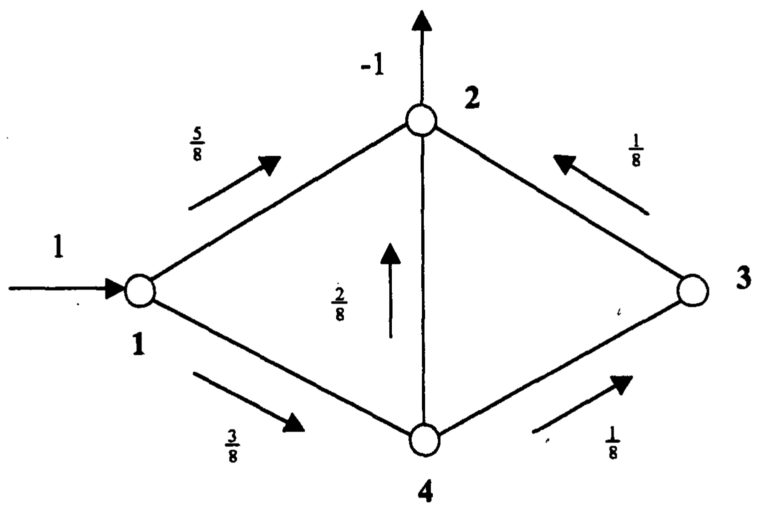


Figure 2-5 Distribution Factors

We observe that injecting a unit in node 1, withdrawing it in node 2, affects all line flows, and the impact is completely determined by the power flow equations. We cannot decide which lines to use for the transmission, and therefore it is not possible to route around capacity constraints. Congestion in electricity networks is therefore a *network* problem rather than a *link* problem.

3. Optimal Dispatch and Optimal Prices

The goal of deregulating the market has been to achieve overall short run and long run efficiency through competition on the supply and demand side and through efficient pricing of transmission. This provides the basis for decentralized decision making, also when it comes to the usage and development of the network.

In the short run, demand functions are given and we want to optimize the use of the existing facilities, both generating facilities and network capacity. In electric power transmission there are mainly three short-term cost components that must be taken into account, i.e. congestion cost, resistive losses, and the cost of ancillary services like the provision of reactive power and spinning reserve. Congestion cost is the opportunity cost that results from out-of-merit order dispatch, i.e. not being able to dispatch the cheapest generators first. Green [26] shows that by applying uniform prices, inferring that location means nothing, welfare is reduced even if transmission constraints are handled through efficient re-dispatch. In addition, this is likely to give the wrong long-term signals. Marginal losses can easily lead to a 5% or 10% premium or penalty for specific locations, while even small deviations from the unconstrained cases can produce even larger differences in the marginal prices at the buses in the system (Hogan [38]). Such differences could have significant effect on the evaluation of the relative economics of location decisions.

In the long run, the most important requirement of transmission pricing is to provide the right incentives for the location and construction of new generating and consuming facilities. Also the network must be expanded optimally, and this requires addressing the question of how to compensate network owners. This should however not be confused with revenue requirements, which concern the compensation of invested capital, which is mostly sunk cost. Long run efficiency is clearly the most difficult objective, but also maybe the most rewarding task. As phrased by Hogan [35]: "By comparison with the costs of poor choices on these major plant investment decisions, there would likely be small operating inefficiencies from any failure to adopt a perfect short-run transmission pricing model."

Short-term pricing decisions should at least not be detrimental to long run efficiency, but the opposite is also valid. The pursuit of long run efficiency should not interfere with short run optimal use of the network. Ideally, one would hope for the possibility of attaining long run efficiency through a sequence of optimal short run pricing decisions. According to Hogan [35] however, this requires constant or decreasing returns to scale in the transmission system, which is not likely to apply in practice. The question then is whether long-term efficiency is attainable at all in a decentralized market-based system. We may instead have to rely on different regulatory mechanisms with central investment decisions.

In the literature, much attention has been given to short-term efficiency. When discussing efficient pricing in the scheduled power market, the benchmark has been the so-called optimal (economic) dispatch, which in words may be stated roughly as the solution of

max social welfare (consumer benefit – production costs)
s.t. power flow equations
thermal constraints
reliability constraints

Assuming a utilitarian welfare function, social welfare is found by adding consumer benefits from real and reactive power consumption (represented by the area under the demand functions in every node) and subtracting the cost of generation (corresponding to the area under the supply functions of the nodes)[8].

In general, the power flow equations are highly nonlinear and the social welfare maximization problem is non-convex. However, in the normal operation of a power system, focusing on real power, voltage magnitudes are kept close to rated levels and phase angle differences between nodes are small, implying that the power flow equations can be approximated by convex functions. Assuming a well-behaving objective function, the resulting social welfare maximization problem is then (at least locally) convex, which is a

[8] The related OPF (optimal power flow) problem corresponds to optimal dispatch with inelastic demand (i.e. demand is fixed, and generation costs are minimized subject to network constraints).

prerequisite for the existence of an efficient market mechanism to replicate social optimum. For a discussion, see Chao and Peck [12].

Assume for the moment that we consider real power only, and that we ignore any security constraints like contingency constraints or constraints on nodal voltages. We are then left with the power flow equations and the thermal limits constraining the dispatch.

3.1. Nodal Prices

The concept of nodal prices is discussed by Schweppe et al. [64]. Optimal nodal prices are produced by the solution of the welfare maximization problem as the dual prices of the power flow equations, and are interpreted as the value of power in each node. A mechanism enforcing optimal nodal prices together with generators and consumers adapting to their local (nodal) market price when deciding on output will ensure social optimum.

Marginal Losses

To see how nodal prices work in the case of losses, assume that the price of power at node i is λ_i, and that there are no binding capacity or security constraints. Any difference in nodal prices is then due to marginal losses, and $\lambda_k - \lambda_i$ is the marginal cost (value of marginal losses) of transporting power between nodes i and k. Nodal prices of marginal losses can be constructed by introducing a system price s, letting $t_i^{inj} = s - \lambda_i$ be the price of injection in node i, and $t_k^{con} = \lambda_k - s$ the price of withdrawals in node k. The tariff paid when injecting a unit in node i and withdrawing it in node k is $t_i^{inj} + t_k^{con} = s - \lambda_i + \lambda_k - s = \lambda_k - \lambda_i = \Delta\text{cost}$. This construction also implies $t_i^{inj} = -t_i^{con}$.

The cost of marginal losses corresponding to a trade depends on the operating point, i.e. the distribution of generation and load in the entire network. It is therefore not trivial to calculate the correct prices of losses. It should also be noted that marginal losses typically are about two times the size of average losses. Because of this, a tariff covering marginal losses will cover more than total losses in the system. This may be a problem if the agents are allowed to pay in kind, i.e. to increase power injection (or reduce withdrawals) according to marginal losses. If all agents do so, it will result in imbalances in the system.

Congestion

Capacity constraints due to thermal limits are extensively studied in the literature because of the externalities created by loop flow (i.e. "the routing problem" as discussed at the end of section 2.2, demonstrating that a single trade between two nodes affects every line flow of the network). Wu et al. [90] point to a number of counter-intuitive and possibly troublesome characteristics of nodal congestion prices. Their analysis is based on an optimal dispatch problem considering real power only and using the lossless linear "DC" model (i.e. a linear type B-P model). In every node, only net power injections are considered, and the benefit functions are represented by negative cost, i.e. $C_i(P_i) > 0$ for $P_i > 0$ (net injections) and $C_i(P_i) < 0$ for $P_i < 0$ (net withdrawals). In addition, it is assumed that $C_i(0) = 0$ and $C_i(P_i)$ is non-decreasing and convex[9].

Social welfare is then maximized by solving the following convex minimization problem

(3-1) $$\min_{\delta,P} \sum_{i=1}^{n} C_i(P_i)$$

(3-2) s.t. $$P_i = \sum_{k=1}^{n} B_{ik}(\delta_i - \delta_k) \qquad i = 1,\ldots,n$$

(3-3) $$B_{ik}(\delta_i - \delta_k) \le C_{ik} \qquad 1 \le i,k \le n$$

where susceptance $B_{ik} = 1/x_{ik}$, and x_{ik} is the reactance of line ik. C_{ik} is the thermal limit of line ik. The shadow prices of equations (3-2) are denoted λ_i, while shadow prices associated with inequalities (3-3) are represented by $\mu_{ik} \ge 0$. The Lagrangian is

[9] In this way, simultaneous generation and consumption at the same node are "netted" out. An alternative would be to consider both production and consumption in every node explicitly (as in Chao and Peck [12]) maximizing $\sum_i [B_i(P_i^d) - G_i(P_i^s)]$, where $B_i(P_i^d)$ is the benefit of consuming P_i^d, and $G_i(P_i^s)$ is the cost of generating P_i^s. Net injection in node i is equal to $P_i = P_i^s - P_i^d$.

$$\Phi = \sum_{i=1}^{n} C_i(P_i) + \sum_{i=1}^{n} \lambda_i [\sum_{k=1}^{n} B_{ik}(\delta_i - \delta_k) - P_i] + \sum_{i=1}^{n}\sum_{k=1}^{n} \mu_{ik}[B_{ik}(\delta_i - \delta_k) - C_{ik}]$$

In addition to (3-2) and (3-3), the first order conditions of the minimization problem are given by

(3-4) $$\frac{\partial C_i}{\partial P_i} = \lambda_i \qquad i = 1,\ldots,n$$

(3-5) $$\sum_{k=1}^{n} B_{ik}(\lambda_i - \lambda_k + \mu_{ik} - \mu_{ki}) = 0 \qquad i = 1,\ldots,n$$

(3-6) $$\mu_{ik}[B_{ik}(\delta_i - \delta_k) - C_{ik}] = 0 \qquad 1 \le i,k \le n$$

(3-4) follows from $\partial\Phi / \partial P_i = 0$, and since $C_i(P_i)$ is non-decreasing, $\lambda_i \ge 0$. (3-5) follows from $\partial\Phi / \partial\delta_i = 0$, noting that the admittance matrix is symmetric. (3-6) is the complementary slackness condition due to $\mu_{ik} \ge 0$.

Optimal Dispatch

(P, δ) is an optimal dispatch if it solves (3-1)-(3-3). Since the optimization problem is a convex program, (P, δ) is an optimal dispatch if and only if there exist λ_i and $\mu_{ik} \ge 0$ such that (3-2)-(3-6) hold.

Market Equilibrium

Assuming price λ_i is interpreted as the node i market price and that transmission from i to k is charged the price difference $\lambda_k - \lambda_i$, (P, δ, λ) is a market equilibrium if (3-2), (3-3) and (3-4) hold. In a market equilibrium, every market clears (3-2), i.e. supply equals demand in every nodal market. Also transmission constraints are met (3-3), and there is consumer and supplier equilibrium at every node (3-4). There is no opportunity for profit from buying power at one node and selling at another, since buying at node i and selling at node k would cost $\lambda_i + (\lambda_k - \lambda_i) = \lambda_k$.

The merchandizing surplus (MS) at a market equilibrium (P, δ, λ), is defined as

$$\text{(3-7)} \qquad \mathrm{MS} = -\sum_{i=1}^{n} \lambda_i P_i = \tfrac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} (\lambda_k - \lambda_i) \cdot P_{ik}$$

($\frac{1}{2}$ because flow over line ik is counted twice, since $P_{ik} = -P_{ki}$), and can be interpreted as the profit accruing to a "market maker" who is party to every purchase and every sale of power. For instance, if a grid company is to accommodate every trade that needs transmission over the network, this would be the revenue of the grid company resulting from market equilibrium (P, δ, λ).

A market equilibrium is non-unique, and does not necessarily correspond to optimal dispatch, as shown by the two-node example of Wu et al. [90], which is displayed in Figure 3-1. In fact, any set of nodal prices leading to a feasible flow will be a market equilibrium given that the consumers and generators adapt competitively to their local prices. In the example, a single line connects production in node i with consumption in node j, with S_i and D_j denoting the corresponding supply and demand functions.

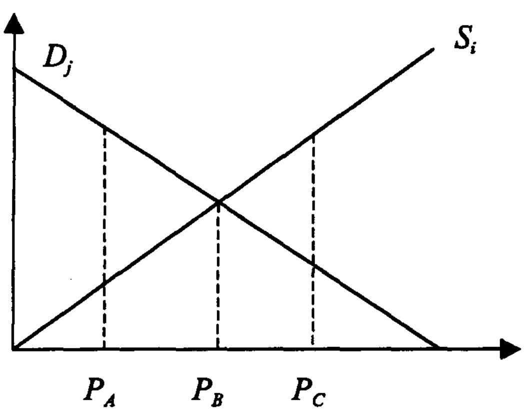


Figure 3-1 Three Different Market Equilibria

Assuming that the capacity-limit of the line is not binding, letting $\lambda_i^A = S_i(P_A)$ and $\lambda_j^A = D_j(P_A)$, $(P_A, \lambda_i^A, \lambda_j^A)$ corresponds to a market equilibrium. Similar equilibria are found for quantities P_B and P_C. In the case of output P_A, $\lambda_i^A < \lambda_j^A$ and MS is positive, whereas for

P_C, $\lambda_i^C > \lambda_j^C$ and MS is negative, however only P_B is efficient (with MS equal to zero due to the line not being constrained). Thus the example shows that a market equilibrium need not be efficient and MS can be positive or negative. To attain optimal dispatch then requires central coordination, for instance provided by a system operator.

Moreover, in Wu et al. [90] it is proven that MS is non-negative in optimal dispatch, but as the example shows it is not necessarily maximal. This means that a system operator may have a wide range of alternatives when choosing the system operating point, and if the system operator also receives the revenue from the grid, the income potential from choosing a non-optimal dispatch may be huge. This calls for tight regulation, and one possibility considered is that of *optimal price-regulation*, with the purpose of making the system operator fix nodal prices at the Lagrangian multipliers corresponding to optimal dispatch. For this to work, suppliers and consumers must truthfully reveal cost and demand functions, and the system operator must calculate correctly the optimal dispatch and implement it.

Unfortunately, both producers and consumers may not be willing to give such strategic information. Among others, Hogan [37], Stoft [75] [79] and Younes and Ilic [92] show how congestion affects competition and that suppliers and consumers may have strong incentives not to be truthful about cost and demand data. Even if truthful information comes forward, the system operator, as the central authority, is very well informed and has the power to determine prices. This power could be abused by favoring some suppliers, customers or even the system operator itself. To prevent this, regulation must be strong and audit procedures must be implemented that can check whether the operating point is the optimal dispatch. This is indeed a difficult task, and it is not simplified by the fact that the selection of thermal and contingency constraints involves a considerable degree of judgement on the part of the system operator.

In optimal dispatch, the μ_{ik} 's are the (shadow) congestion prices for the thermal constraints, with the standard interpretation as the marginal values of increased thermal capacity. We can define the (shadow) congestion rent as

$$\sum_{i=1}^{n}\sum_{k=1}^{n}\mu_{ik}C_{ik}$$

and it can be shown that in optimal dispatch the congestion rent is equal to the merchandizing surplus (for proof see Wu et al. [90]).

From the analysis of an n-node network containing exactly one congested link $1n$, i.e. $\mu_{1n} > 0$, while $\mu_{ij} = 0$ for $ij \neq 1n$, Wu et al. [90] have shown that the relationship between nodal prices is given by the equation

$$\textbf{(3-8)} \qquad \lambda_i = \lambda_n - \gamma_i \mu_{1n} \qquad i = 1,\ldots,n-1$$

where γ_i depends on network characteristics only and not on the power flows. If $\gamma_i \neq \gamma_k$, then $\lambda_i \neq \lambda_k$, even if line ik is not congested and $\mu_{ik} = \mu_{ki} = 0$. This implies that a single congested line may affect nodal prices throughout the network, and that nodal price differences generally are not equal to congestion prices.

From (3-8) it is evident that if $\gamma_i > \gamma_k$, then $\lambda_i < \lambda_k$. This relationship holds whether the line flow is from node i to node k, or from k to i. In any network that is not radial, it is easy to construct an optimal dispatch involving flow from a higher priced node to a lower priced node. This results in a negative merchandizing surplus over the line[10] and implies that price differentials are not indicative of merchandizing opportunities. This may also imply that optimal dispatch is not sustained by (profitable) bilateral contracts only. Instead, a coordination mechanism must facilitate multilateral contracts in which some trades are subsidized in order to reach optimal dispatch. For a more detailed discussion and examples, see for instance Wu et al. [90], Wu and Varaiya [89], and Jörnsten and Singh [40]. We also provide an example in section 3.7.

In a long-term perspective, the system operator may have incentives not to invest in the network to reduce congestion, because it reduces the revenue from the grid. Loop flow may

[10] I.e. the line owner must pay if merchandizing surplus is used to allocate revenue to line owners.

also result in the counter-intuitive outcome that strengthening a line may reduce capacity and social surplus. This means that lines cannot be strengthened arbitrarily (which is rather obvious) however, since the flow changes constantly, it is not evident which line should be strengthened. This somewhat paradoxical characteristic of power networks is analyzed in chapter 8, where it is also shown that the merchandizing surplus may increase as a consequence of a detrimental investment. Conditional on the distribution of the merchandizing surplus, i.e. compensation to network owners, this is one case in which optimal short run pricing may negatively affect long run efficiency.

3.2. Transmission Congestion Contracts (TCCs)

Loop flow characteristics are recognized by Hogan [35], introducing the concept of a contract *network* instead of a contract *path* to reflect the influence of a trade on the network. Hogan discusses the importance of open network access and suggests distributing network capacity to the traders as transmission capacity rights in order for them to secure long-term access to the network. A transmission capacity right is defined as the right to inject power in a node and take out the same amount of power at another bus. It is assumed that the simultaneous use of all allocated rights is feasible. Moreover, one allows for either the specific performance of the capacity right or the receipt of an equivalent rental payment. Open access is also facilitated by the existence of efficient secondary markets for transmission rights.

In a system of optimal nodal prices, transmission capacity rights, often referred to in the literature as transmission congestion contracts or TCCs, constitute a sort of transmission price hedge because it is the owners of the transmission rights that receive the transmission revenue accruing from nodal price differences. By acquiring TCCs the parties of a long-term contract can avoid the risk of transmission prices rising in the future because of increased congestion leading to large differences in nodal prices.

Consider a long-term contract of delivering T_{ik} units of power at node k. A TCC of T_{ik} units between nodes i and k would entitle the owner to a compensation of $T_{ik}(\lambda_k - \lambda_i)$. This compensation makes the right holder indifferent between i) purchasing power at node i at the price λ_i, transporting it over the network for a cost $T_{ik}(\lambda_k - \lambda_i)$, and receiving the compensation $T_{ik}(\lambda_k - \lambda_i)$, and ii) purchasing power at node k at price λ_k and receiving the compensation $T_{ik}(\lambda_k - \lambda_i)$. I.e. if the right holder uses his full capacity right, the congestion charge paid by the right holder is just offset by the revenue received. Whenever the right holder is precluded from using the full capacity, the compensation received is just the amount needed to make the right holder indifferent between actually delivering the power at the destination point, and fulfilling the long-term contract by buying power at the destination point and receiving the compensation[11].

Oren et al. [56] and Wu et al. [90] have discussed different aspects of transmission capacity rights and pointed to the fact that physical transmission rights (i.e. requiring actual deliveries) may constrain the optimal dispatch and thus be detrimental to short run efficiency. Harvey et al. [34] also assert that even if TCCs are allocated optimally at every time interval[12], the distributed rights should not interfere with the complex actual dispatch procedures of power systems, in which there are many considerations that are difficult to incorporate in the economic optimization problem. Financial contracts like Hogan's TCCs can sustain optimal dispatch, however, Oren et al. [56] demonstrate that the income-stream from a transmission right between nodes i and k can be replicated by positions in a nodal forward market. More specifically, a short forward at node i, a long forward at node k and a fixed annuity equal to the difference in the forward prices at k and i at the issuing time would match this TCC. Thus it is argued that from a hedging consideration, TCCs are redundant.

Bushnell and Stoft [8] [10] investigate the effect of TCCs on investment decisions. A rule for awarding new transmission rights to investors, based on feasible dispatch, is formalized. The *Feasibility Allocation Rule* works as follows. Let T be the set of allocated TCCs before grid modification. The modifier of the grid is granted any set of contracts ΔT such that the

[11] In this discussion, we have not considered losses. In the case of losses, the transmission charge would consist of a congestion charge and a charge covering marginal losses. The compensation of the TCC would equal the congestion charge, so that if the right holder uses the full capacity right, the net cost of transmission is just the charge of marginal losses. Bushnell and Stoft [8] suggest alternatively to include losses in the definition of TCCs or to define a TCC with respect to a single node instead of a pair of nodes. In that case one would have to acquire a bundle of TCCs to hedge a trade, but this may make it easier to include losses in economic decisions.

dispatch corresponding to $T+\Delta T$ is feasible under the new grid configuration. This allocation is contrasted with "link-based rights" in which the line owner receives a rent equal to the merchandizing surplus of the line, i.e. $P_{ik}(\lambda_k - \lambda_i)$. In this case, an investment detrimental to social benefit could give positive revenue. A rule allocating TCCs to investors based on the change in flow capacity of the lines as a result of the investment would effectively be equivalent to a system of "link-based rights" and inherit its problems. With the simple "link-based" or capacity-based TCC allocation rule, the new TCCs, ΔT, together with the old ones may not be feasible in the new network configuration and would not be allowed under the Feasibility allocation rule.

It is shown that if transmission rights replicate actual dispatch for all agents (i.e. everyone has hedged their spot positions perfectly) no one would benefit from detrimental investments under the Feasibility allocation rule, because the negative external effects would be accounted for. This is so because any increase in network revenue for the investor would be (more than) offset by purchase *costs* of TCCs. The authors however point to a difficulty with beneficial expansions, namely that the full benefit will most likely not be captured by the investor who makes the expansion. In that case "free-riders" will also benefit, possibly with under-investment as a result, and this possibility is not eliminated by the presence of TCCs. A weaker result states that if the combined TCCs replicate actual dispatch *on aggregate*, detrimental investments will not be profitable for any *new* entrants, since TCCs on their own cannot generate profits from detrimental investments.

The results rest on the assumption of TCCs matching actual dispatch at different degrees. In the strongest form the match is extremely unlikely, one reason being that actual dispatch varies constantly, whereas contract positions will probably not. However, as argued by Bushnell and Stoft [8], since the highest potential revenue from a set of contracts is reached when those contracts match the dispatch, the agents have at least no incentives to *mismatch* dispatch with regards to the transmission contracts held.

Hogan's system of transmission capacity rights is based on a nodal pricing system, and offers no solution to the problem of finding optimal nodal prices without actually solving the

---

[12] For instance through the trading of rights.

optimal dispatch problem. Also, since the transmission capacity rights effectively are claims on proportions of the merchandizing surplus, there is the problem of making the initial distribution of capacity rights. This may lead to conflicts between relatively new users and users claiming "grand-fathered" rights to a network they have paid the fixed costs of. As shown by Wu et al. [90], for some right holders an allocation of rights may produce a negative income-stream. An alternative would be to permit the right holder not to exercise the right, thus receiving a minimum of 0 rent[13]. This would require strengthening the feasibility condition, demanding that any set of injections corresponding to a possible way of exercising the rights should be feasible.

3.3. Chao-Peck Prices

The price system suggested by Chao and Peck [12], which we will call CP-prices, represents a system for "explicit congestion pricing". Conceptually, one prices the use of scarce transmission resources instead of the energy as the nodal prices do. To explain how this system works, we continue using the lossless and linear "DC" approximation. CP-prices are based on the definition of transmission capacity rights, which intent is to simulate the power carrying capability of individual links. A transmission capacity right entitles the owner the right to send a unit of power through *a specific transmission line* in *a specific direction* (note the difference compared to Hogan's TCCs which are defined between two arbitrary nodes not necessarily linked by a direct connection). A set of transmission capacity rights consistent with thermal limits is issued for each (directed) link, and these rights are tradable.

A trading rule is then defined to govern the exchange of rights by specifying the transmission capacity rights that traders must acquire in order to complete an electricity transaction. The trading rule is therefore determined by the distribution factors (or load factors) $\beta_{ik}^{lm}, 1 \le i,k,l,m \le n$, where β_{ik}^{lm} is interpreted as the units of transmission capacity rights on the directed link ik that are required to transfer one unit of power between nodes l and m. The factor β_{ik}^{lm} can be positive or negative (representing counter-flows). If π_{ik} is the price of

[13] A TCC as originally defined, is a kind of "property" that entails both rights and obligations (therefore the

the transmission capacity right on link ik (from i to k), the transmission cost of a trade of one unit between nodes l and m will be

$$\sum_{i=1}^{n}\sum_{k=1}^{n}\pi_{ik}\beta_{ik}^{lm}.$$

Chao and Peck define a *competitive equilibrium* as the vector-triple (P,λ,π) that satisfies three conditions:

A) The price of electricity must be equal to the marginal cost/benefit at each node, i.e. $\frac{\partial C_i}{\partial P_i}=\lambda_i$ for $1\leq i\leq n$.

B) There should be no profit from transferring power from one node to another, i.e. $\lambda_m=\lambda_l+\sum_{i=1}^{n}\sum_{k=1}^{n}\pi_{ik}\beta_{ik}^{lm}$ for $1\leq l,m\leq n$.

C) The price of a transmission capacity right is zero when there is excess supply, i.e. $\pi_{ik}[\sum_{l=1}^{n}\sum_{m=1}^{n}\beta_{ik}^{lm}q_{lm}-C_{ik}]=0$ for $1\leq i,k\leq n$ where $q_{lm}\geq 0$ is the power that is to be transferred from node l to node m.

If $\pi_{ik}=\mu_{ik}$ (the shadow price of the capacity constraint in the optimal power flow problem), a competitive equilibrium is an optimal economic dispatch. This highlights three appealing properties of CP-prices that are not valid for nodal prices:

1) Lines that are not used to their capacity limit have a zero congestion price[14].
2) Congested lines have always a positive congestion price in the direction of the flow.
3) The number of positive prices is equal to the number of congested links, and this number is usually far less than the number of nodes in the system[15].

possibly negative value). In the alternative definition, a TCC takes the form of an option.

[14] This is obvious from point C) in the definition of competitive equilibrium, but follows also from the complementary slackness conditions of the optimal dispatch problem.

[15] Still, in a fully meshed network with K congested lines, K+1 nodal prices may be enough to characterize the economic information completely. All other prices are just weighted averages of these K+1 prices, and the weights depend only on network characteristics (Stoft [76]).

CP-pricing can be interpreted as using the congestion rent instead of merchandizing surplus to charge transmission. In optimum the total revenues are the same. Furthermore, CP-prices must be consistent with optimal nodal prices for each trade as well, which is guaranteed by A) and B)[16]. The requirement of CP-prices being consistent with optimal nodal prices implies that CP-prices could be calculated from optimal nodal prices and distribution factors. However, the point is to find a market mechanism that quickly and reliably finds the congestion prices without a central authority having to gather all the needed information and solve the optimal dispatch problem.

Chao and Peck themselves suggest a decentralized mechanism where the congestion prices emerge as a result of agents maximizing congestion rents while allowing generators at least to break even. In principle, a trade affects power flows on every link of the network so that transmission rights must be acquired for virtually every link of the grid, possibly involving many right owners. This may be a burdensome procedure. In addition, these transmission right owners may have tremendous market power, noting that if any one person owns all of a transmission line, that person has monopoly power over all users of the grid.

Stoft [78] suggests three remedies. First, one could designate a maximum percentage that each owner is allowed to hold of a line, thus reducing market power. Secondly, the line owners could be required to sell the entire capacity of the lines. This would supposedly drive the price of a line with unused capacity to zero. Finally, Stoft suggests adopting a centralized initial auction and the establishment of a successive bid-ask market in which transmission capacity rights are traded continuously up until "real-time" (the closing of the scheduled power market) to determine prices. This may reduce trading costs and provide a more liquid and competitive market. However, the specific design of a mechanism determining CP-prices, including algorithms for evaluating bids quickly, is still regarded an unresolved problem.

So far we have discussed CP-prices in the context of the lossless linear "DC" approximation. Real power markets have several complications that must be accounted for. First, losses must be included. This is readily available as shown by Chao and Peck [12], resulting in a slightly

[16] As phrased by Stoft [78]: "These nodal prices are completely uncontroversial, for in spite of the dispute between the "bilateralists" and the "poolco" advocates, all agree that these are the only energy prices that induce a least-cost, efficient dispatch."

more complicated trading rule having more or less the same structure as the one above[17]. Second, the non-linear nature of the AC power flows implies that distribution factors are not constants. The results above are however valid if we exchange the distribution factors with *marginal* distribution factors. These depend on the load flow of the system, so in addition to solving non-linear systems, in principle we have to recalculate distribution factors whenever the load changes.

3.4. CP+Hub Prices

A slight modification of CP-prices is suggested by Stoft [78]. By allowing energy bids at any given bus in the network, a "hub" price could be determined. This would simply be the standard nodal price of energy at the selected bus. The number of prices that a market mechanism would have to discover, is equal to the number of congested links plus one, and this is still usually far less than the number of nodes in the network. Adding the "hub" price to the CP prices makes it easy to calculate the price of energy at every node using point B) in section 3.3. Introducing a price of energy may also ease the correct treatment of losses. For every trade it can be calculated how much power to inject (or withdraw) at the "hub" to cover marginal losses, resulting from the trade. The trade could then simply be charged for the added injection (or compensated for the withdrawal) by the energy price at the "hub".

3.5. Zonal Pricing

Zonal pricing is an approximation of nodal pricing, dividing the network into zones, each with a uniform price. Zonal pricing is used as the basis for several practical implementations to account for congestion, for instance in the US and Norway. Generally, optimal nodal pricing would induce a unique price for every node in the network, and except for the case of zonable networks, optimal dispatch is not achieved by zonal pricing. The combination of a

[17] One of the reasons for the more complicated appearance is that marginal losses are accounted for by two different components, one accounting for average losses and the second accounting for the inherent rent when

network and a list of congested lines is zonable if optimal nodal prices are uniform within zones. If for instance all buses belonging to a zone have the same distribution factors on all congested lines, the buses of the zone will have the same price (refer to point B) in section 3.3), and the network is zonable. In a zonable network, intra-zonal lines must have flow below their limit, only *inter*-zonal lines can be congested.

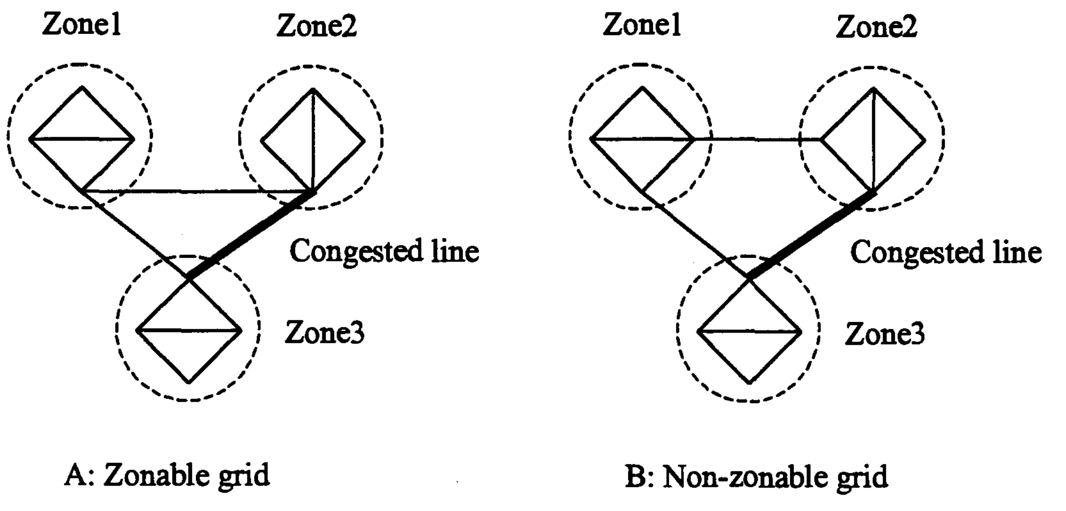


Figure 3-2 Zonable Networks

Radial networks are zonable, but Stoft [76] also shows examples of zonable grids with more complicated meshed structures, like grid A in Figure 3-2. The congested loop only touches every zone at a single node. The sub-networks of the zones can be of any complexity. In grid B, where the congested loop passes through zones 1 and 2, there will be a price gradient within the zones and the network is no longer zonable. Generally, networks are not zonable and defining zone-boundaries[18] is a difficult task even if we knew the optimal nodal prices. Zones should be defined based on price differences, but prices change gradually throughout the network, and it is hard to know where to draw the line. We will study zonal pricing in chapter 7.

paying for marginal losses. This structure facilitates payment in kind.

[18] As long as we are not allowed to have a zone for every node!

3.6. *Coordinated Multilateral Trades*

The coordinated multilateral trade (CMT) model suggested by Wu and Varaiya [89] is intended to attain optimal dispatch without requiring the explicit description of cost/benefit functions, i.e. the model proposed intends to avoid a system operator with all private information. To achieve this, the decision mechanisms regarding economics and the reliability of system operation are separated. Economic decisions are carried out by private multilateral trades among generators and consumers, while the function of reliability is coordinated through the system operator who provides publicly accessible data, based upon which generators and consumers can determine profitable trades that meet the secure transmission loading limits.

More specifically, the model requires traded quantities to be reported to the system operator. Based on network characteristics and net injections in every node, the system operator performs a power flow calculation. Assuming this leads to a single congested link ik, in direction i to k, trades are curtailed to meet the flow limit, and a loading vector is made public to guide traders searching for additional feasible trades. The loading vector β_{ik} consists of load factors for each node when referenced to a specific bus (the "swingbus"), for instance bus n. This means that $\beta_{ik} = (\beta_{ik}^1, \beta_{ik}^2, \ldots, \beta_{ik}^n) = (\beta_{ik}^{1n}, \beta_{ik}^{2n}, \ldots, \beta_{ik}^{nn})$ and the loading vector is equivalent to Chao-Peck's trading rule since a trade of one unit between nodes l (injection) and m (withdrawal) increases flow on line ik by $\beta_{ik}^l - \beta_{ik}^m = \beta_{ik}^{lm}$.

Starting with the set of curtailed trades, for this simple bilateral trade between nodes l and m to be feasible, we must have $\beta_{ik}^l - \beta_{ik}^m = \beta_{ik}^{lm} \leq 0$, since the trade cannot increase congestion on the line (which is already at its limit). More generally, to reach optimal dispatch multilateral trades might be necessary. A multilateral trade may be represented by a vector $q = (q_1, q_2, \ldots, q_n)$ where q_l is positive (negative) if the trade involves an injection (withdrawal) in node l, and zero otherwise. Ignoring losses, a profitable feasible trade must satisfy

(3-9) $$\sum_{l=1}^{n} q_l = 0$$

(3-10) $$\sum_{l=1}^{n} MC_l q_l \leq 0$$

(3-11) $$\sum_{l=1}^{n} \beta_{ik}^{l} q_l \leq 0$$

where $MC_l \geq 0$ denotes the marginal cost/benefit of production/consumption in node l. (3-10) ensures profitability while (3-11) guarantees feasibility regarding congested link ik.

If there are several congested lines (or more generally, congested cuts), the system operator must publish a loading vector for every transfer limit. During the process of additional trading new constraints may become binding, thus central coordination is achieved through an iterative process that can be described by a step by step algorithm as the following:

Step 1: *Initialization*

Brokers arrange trades, resulting in an initial vector of power injections and withdrawals, P^0.

Step 2: *Curtailment*

If P^0 is not feasible, the system operator curtails the trades to a point where the resulting injection vector P is feasible.

Step 3: *Announcement*

If lines are congested at P, the system operator announces a loading vector β_{ik} for each congested line ik.

Step 4: *Trading*

Using the loading vectors to evaluate if a trade is feasible, a profitable feasible trade is arranged in the market. If no profitable feasible trade is found, go to step 6.

Step 5: *Feasibility*

If the trade is infeasible, the system operator curtails it and we return to step 3. If the trade is feasible, go to step 4.

Step 6: *Termination*

Stop.

It is shown that this coordinated multilateral trading process converges to optimal dispatch under the assumption that all participants make reasonable decisions, i.e. whenever there is a worthwhile profitable trade, the participants will carry it out.

In the CMT process brokers (which may be generators, consumers or others) arrange profitable and feasible trades between generators and consumers. Checking profitability and feasibility of a given trade is simple. Finding the *optimal* trade between a group of generators and consumers is far more complicated and involves maximizing net profit, i.e. minimizing the left-hand side of (3-10), subject to (3-9) and feasibility constraints of type (3-11). This problem is a version of the optimal dispatch problem, considering only a group of market participants and a subset of the transmission constraints. Solving this problem requires knowledge of marginal cost and willingness to pay.

As pointed to by Allen et al. [2], if there are K congested lines, one will generally need to combine $K+1$ bilateral markets (which at least involve $K+2$ nodes) in a multilateral trading process to arrive at the social optimum. However, the number of congested lines is usually far less than the number of nodes in the grid. Allen et al. [2] also point to the fact that different curtailment procedures may be used. If curtailed trades are traded at the original price agreed upon before curtailment, then the specific curtailment procedure chosen will affect the final allocation of social surplus among the participants.

Wu and Varaiya [89] also describe different refinements of the CMT model, including losses and adjusting for nonlinear flows. Regarding losses, it is searched for an explicit expression for losses in order to determine losses due to an arbitrary trade. The complexity of the power flow equations precludes a closed form expression. Therefore, a Taylor expansion is performed to provide a quadratic approximation of the power flow equations, and losses become functions of the Jacobian and Hessian matrices of the power flow equations. From this, a loss allocation scheme is developed, allocating total losses to individual trades. Central in the allocation scheme is a loss vector, or alternatively a loss matrix that also demonstrates the interactions of trades in terms of their effect on losses. Loss vectors or matrices are calculated and published by the system operator so that brokers can estimate and account for losses caused by each trade[19].

3.7. *Comparison of the Different Coordination Mechanisms in an Example*

To arrive at social optimum, central coordination is necessary. However, as shown in the preceding sections it may be possible to reach the same level of coordination through different information and decision-making structures. As phrased by Wu and Varaiya [89]: "an examination of coordination should focus on ways to distribute information (who knows what) and control (who does what) for all parties: generators, consumers, system operator, regulator, etc., to achieve the goals."

The CMT model relies very much on the strength of the competitive forces governing the arrangement of profitable and feasible trades between generators and consumers, and the system operator is reduced to a body providing shared and public information that is based on obvious and reproducible calculations. The stated goals of this arrangement are to 1) achieve economic efficiency and 2) encourage the search for alternatives and innovations for any function that requires a centralized authorized. When using nodal pricing or Chao-Peck prices, coordination is obtained through prices imposing transmission charges and thus favoring or disfavoring certain trades. This is contrasted with the CMT model, which coordinating mechanism consists of a loading vector used to determine feasibility[20].

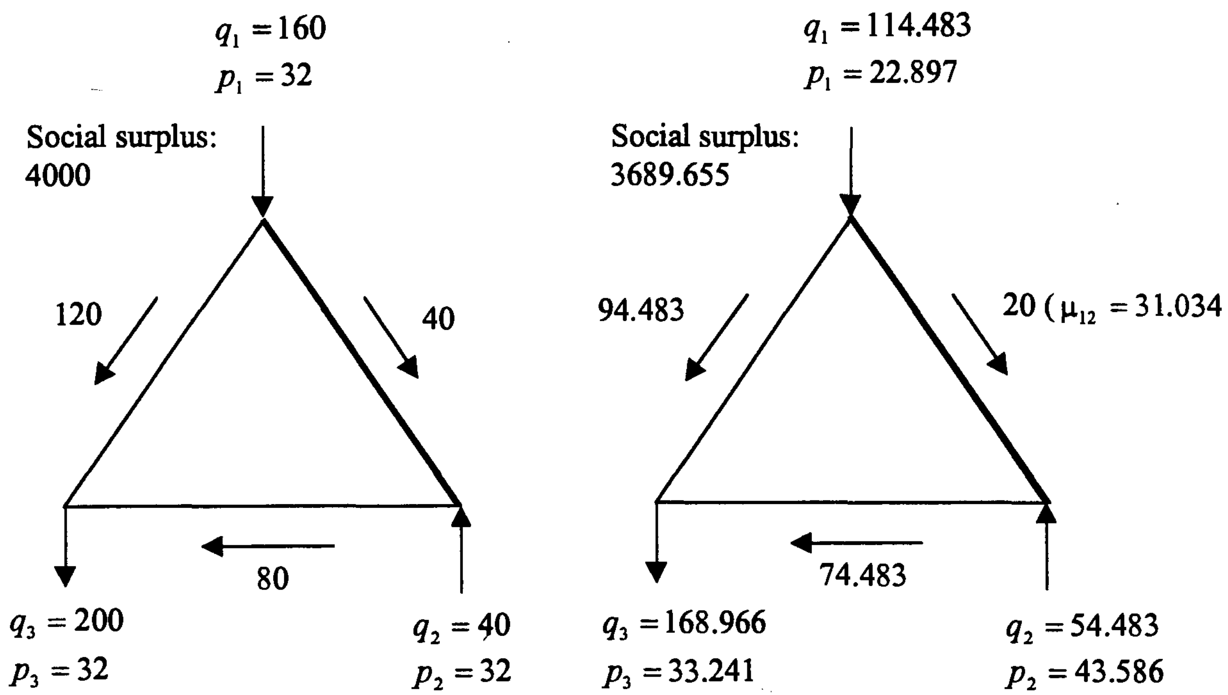
To illustrate how the different models work, consider a triangle network, using the lossless linear "DC" approximation with reactance equal to 1 on every link. We assume that nodes 1 and 2 are net injection nodes with net supply functions $p_1 = 0.2q_1$ and $p_2 = 0.8q_2$,

[19] From the description of the process it is somewhat unclear if a trade is to account for average or incremental losses. In the latter case, or more generally, to account for external effects pertaining to losses, a loss *matrix* is necessary. The share of losses attributed to a trade when using the loss *vector* is an average quantity.
[20] In this part we ignore losses.

respectively. Node 3 is a net withdrawal point with demand function $p_3 = 40 - 0.04q_3$[21]. Note the change of notation letting p_i be the price of region i, while $q_i \geq 0$ is the quantity of real power that is supplied or demanded. $q_1 + q_2 = q_3$, and it is assumed that line 1-2 has a flow limit of 20 units.

Optimal Dispatch

In Figure 3-3 part A, the unconstrained optimal dispatch is displayed. The total traded quantity is equal to 200 at a uniform price of 32. Flow over line 1-2 is 40, which is well over the limit of 20. The solution of the constrained optimal dispatch problem is exhibited in part B of the figure. Total quantity traded is reduced to 168.966, and the reduction is greater than the "overflow" of line 1-2 of 20 units. Now there is a unique price prevailing in each region. In the following we will relate the different coordination mechanisms to the example.



A: Unconstrained Dispatch

B: Constrained Dispatch

Figure 3-3 Optimal Dispatch of a Three-Node Network

[21] The cost and benefit functions are $C_1 = 0.1q_1^2$, $C_2 = 0.4q_2^2$ and $B_3 = 40q_3 - 0.02q_3^2$.

Nodal Pricing

The system operator collects supply and demand schedules from generators and loads and solves the optimal dispatch problem. Optimal nodal prices result from the optimization, and the system operator communicates the nodal prices to the market. Agents adapt to the local prices, increasing output until marginal cost/benefit is equal to the nodal price. A trade between nodes i and k is charged $p_k - p_i$ for transmission, and there can be no arbitrage from trading. Assuming that the merchandizing surplus (MS) is attributed to the system operator/grid, we have the following allocation of social surplus

REGION 1	REGION 2	REGION 3	GRID	TOTAL
1310.630	1187.348	570.987	620.690	3689.655

Note that line 1-3 is not congested ($\mu_{13} = \mu_{31} = 0$), but $p_1 \neq p_3$. Also we have flow from a high priced node to a lower priced node on link 2-3, resulting in a negative contribution of $74.483 \cdot (p_3 - p_2) = -770.527$ to MS from this line. With the capacity of line 1-2 equal to 20, the system operator has no incentive to announce a lower capacity, because MS will then be reduced. However, as shown by Jörnsten et al. [39], it is not difficult to find examples where this is the case. For instance, if the capacity of line 1-2 was 30 units, MS in optimal dispatch is equal to 465.517 with a total social surplus of 3922.414. Reducing capacity to 29 units will increase MS to 495.000, while total social surplus is reduced to 3906.121[22].

Consider a trade of 10 units between nodes 1 and 3 in this market, and assume that the producer has matched the trade by a TCC of 10 units between nodes 1 and 3. If the transmission right is exercised, the cost of fulfilling the obligation for the producer is equal to

$$10 \cdot p_1 + 10 \cdot (p_3 - p_1) - 10 \cdot (p_3 - p_1) = 10 \cdot p_1,$$

where the first part is the (alternative) cost of generation at node 1, the second part is the transmission charge and the third part is the compensation of the transmission right. If the

[22] In fact, in this example the system operator maximizes income (with respect to capacity) when acting as if the capacity is equal to 20.

producer is not dispatched, and must fulfill the contract with the customer by buying power at node 3, the cost would be

$$10 \cdot p_3 - 10 \cdot (p_3 - p_1) = 10 \cdot p_1,$$

i.e. the producer is indifferent between the two.

The example also shows the difficulty of zoning. Assuming we want to divide the network into two zones, how exactly should this be carried out? The zone boundary should intersect the congested line so that nodes 1 and 2 belong to different zones. It is however not clear which zone node 3 should belong to, the price differentials being considerable both in relation to node 1 and 2. We will investigate the zonal approach in chapter 7.

Chao-Peck Pricing

The system operator calculates the load factors that constitute the trading rule. In the example the complete set of load factors is shown in Table 3-1.

Table 3-1 Load Factors

LINKS/TRADES	1-2	2-1	2-3	3-2	1-3	3-1
1-2	2/3	-2/3	-1/3	1/3	1/3	-1/3
2-1	-2/3	2/3	1/3	-1/3	-1/3	1/3
2-3	-1/3	1/3	2/3	-2/3	1/3	-1/3
3-2	1/3	-1/3	-2/3	2/3	-1/3	1/3
1-3	1/3	-1/3	1/3	-1/3	2/3	-2/3
3-1	-1/3	1/3	-1/3	1/3	-2/3	2/3

The entries of Table 3-1 are interpreted in the following way: The column labeled 1-3 shows the transmission rights used (positive numbers) and produced (negative numbers) by a trade of one unit from node 1 to node 3. For instance $\beta_{12}^{13} = \frac{1}{3}$ and $\beta_{21}^{13} = -\frac{1}{3}$, meaning that the trade uses $\frac{1}{3}$ units of capacity on line 1-2 in direction 1 to 2 and *produces* a unit of capacity in the opposite direction. Since $\beta_{ik}^{lm} = -\beta_{ik}^{ml}$, $\beta_{ik}^{lm} = -\beta_{ki}^{lm}$ and $\beta_{ik}^{ll} = 0$ (not shown in the table), the load factors can be summarized by a loading vector for each line by introducing a reference point, for instance node 3. The loading vectors are

$$\beta_{12} = (\beta_{12}^1, \beta_{12}^2, \beta_{12}^3) = (\beta_{12}^{13}, \beta_{12}^{23}, \beta_{12}^{33}) = (\tfrac{1}{3}, -\tfrac{1}{3}, 0)$$

$$\beta_{23} = (\beta_{23}^1, \beta_{23}^2, \beta_{23}^3) = (\beta_{23}^{13}, \beta_{23}^{23}, \beta_{23}^{33}) = (\tfrac{1}{3}, \tfrac{2}{3}, 0)$$

$$\beta_{13} = (\beta_{13}^1, \beta_{13}^2, \beta_{13}^3) = (\beta_{13}^{13}, \beta_{13}^{23}, \beta_{13}^{33}) = (\tfrac{2}{3}, \tfrac{1}{3}, 0)$$

and β_{ik}^{lm} is found by $\beta_{ik}^{l} - \beta_{ik}^{m}$.

The Chao-Peck mechanism rests upon a market bringing forward the prices of transmission rights on the links. In the example, there should be only one positive price, namely on link 1-2 in direction from 1 to 2. In a competitive market for transmission rights, the price is equal to the shadow price $\mu_{12} = 31.034$ of the capacity constraint. Transmission charges for a trade are found by multiplying prices with the appropriate load factors. Here, a trade between nodes 1 and 3 will be charged $\tfrac{1}{3} \cdot 31.034$ per unit, while a trade between nodes 2 and 3 will be charged $-\tfrac{1}{3} \cdot 31.034$, i.e. the trade is compensated. As the markets adapt to the transmission charges, output will increase until

$$p_3 = MB_3 = MC_1 + \tfrac{1}{3} \cdot \mu_{12} = MC_2 - \tfrac{1}{3} \cdot \mu_{12} \tag{3-12}$$

It is easy to check that this holds true in optimal dispatch. The revenue from the grid is equal to $\mu_{12} \cdot 20 = 620.680$, which is equal to the merchandizing surplus. The CP+Hub price model would provide a price of energy in a node (the "hub"), for instance the price $p_3 = 33.241$, which together with load factors and CP-prices makes it easy to calculate the nodal price of any node from (3-12). The distribution of social surplus depends on who owns the transmission rights and the development of the trading process.

Coordinated Multilateral Trades

Following the algorithm of section 3.6, we will show a possible outcome of the CMT process. Initial trading reaches the unconstrained dispatch, and we assume this is achieved by bilateral contracts as depicted by the arrows inside the triangle of Figure 3-4 part A. There are two trades, 160 units from 1 to 3 and 40 units from 2 to 3. Since the unconstrained dispatch violates the thermal limit, trade 1-3 is curtailed to the feasible level of 100 as shown in Figure 3-4 part B. The result of the initial trading/curtailment is trades 1 and 2 of Table 3-2. Marginal values in the nodes are $MC_1 = 20$, $MC_2 = 32$ and $MB_3 = 34.4$. The flow on line 1-2 is at its limit, and the system operator announces the loading vector $\beta_{12} = (\frac{1}{3}, -\frac{1}{3}, 0)$ to the market.

Considering first additional bilateral trades, both trades between 1 and 3 and 2 and 3 are profitable. However, only trades between 2 and 3 are feasible since a trade between nodes 1 and 3 will increase congestion ($\beta_{12}^1 - \beta_{12}^3 = \frac{1}{3} - 0 = \frac{1}{3}$). Assuming that a broker arranges a trade between nodes 2 and 3 that equalizes the marginal values in the nodes, produces trade 3 of Table 3-2. The trade also relieves the transmission constraint and profitable trades can be arranged between nodes 1 and 3. Equalizing marginal values in nodes 1 and 3 produces a trade of 59.524 at price 31.905, and the trade is curtailed due to congestion, resulting in trade 4 of Table 3-2. Also the system operator announces the loading vector $\beta_{12} = (\frac{1}{3}, -\frac{1}{3}, 0)$.

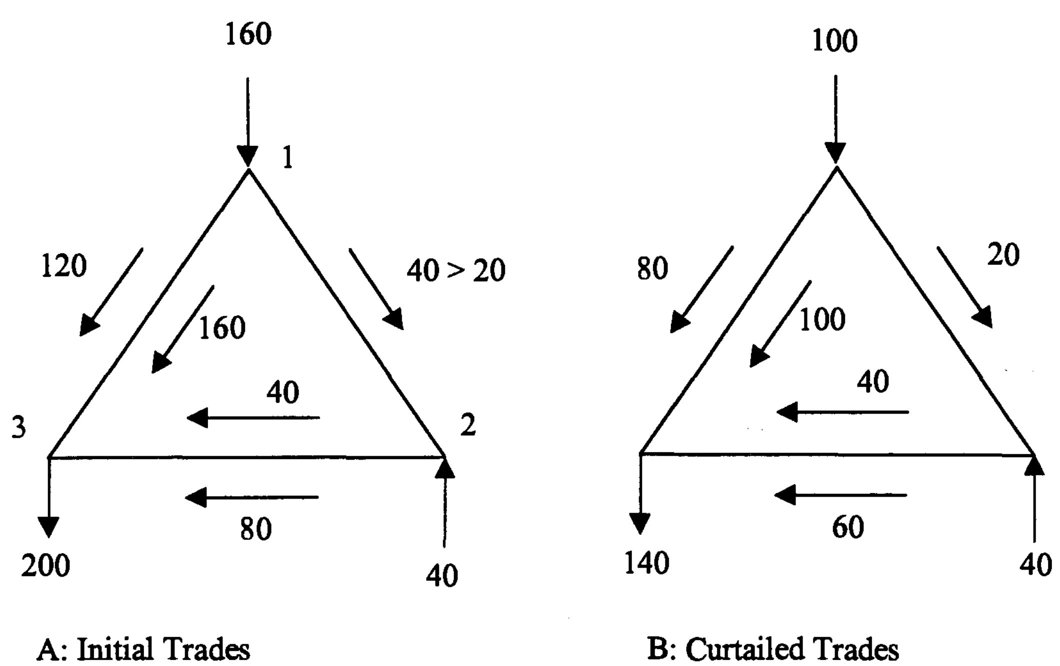


Figure 3-4 Coordinated Multilateral Trades

At this point marginal values are $MC_1 = 20.571$, $MC_2 = 34.286$ and $MB_3 = 34.171$ with a total traded quantity of 145.714 units, i.e. we have not yet reached optimal dispatch. This illustrates that optimal dispatch cannot be achieved by profitable and feasible bilateral contracts only. We must search for profitable and feasible *multilateral* contracts, in this case, trilateral contracts with production in nodes 1 *and* 2 and consumption in node 3. Since $\beta_{12}^1 = -\beta_{12}^2$, a unit produced in node 2 cancels the extra flow over line 1-2 resulting from producing a unit in node 1. Marginal cost in the trilateral contract is an average of the costs of producers 1 and 2, and the contract can be interpreted as producer 1 paying producer 2 to generate to relieve congestion, permitting producer 1 to generate at low cost. The additional trilateral trade is found by equalizing the average marginal cost and the marginal benefit of the consumers[23], and the result is trade 5 of Table 3-2 with an average price of 33.241.

Table 3-2 Trades

TRADE	FROM-TO	QUANTITY	PRICE
1	1-3	100.000	32.000
2	2-3	40.000	32.000
3	2-3	2.857	34.286
4	1-3	2.857	31.905
5	2-3	11.626	43.586
	1-3	11.626	22.897
Total		168.966	

The traded quantities correspond to optimal dispatch and marginal values in the nodes are now equal to the optimal nodal prices $(p_1, p_2, p_3) = (22.897, 43.586, 33.241)$. Allocation of social surplus generally depends on how the trading process develops, at which prices the trades take place and how the curtailment procedure is carried out (in the example, the only possibility is to curtail producer 1). The trades of Table 3-2 bring about the following allocation:

	REGION 1	REGION 2	REGION 3	GRID	TOTAL
CMT	2246.711	697.320	745.624	0	3689.655
Nodal	1310.630	1187.348	570.987	620.690	3689.655

[23] The size of the additional trade is found by solving $(20.571 + 0.2x + 34.286 + 0.8x) \cdot \frac{1}{2} = 34.171 - 0.04 \cdot 2x$. A general framework for finding "optimal" trilateral trades can be found in Wu and Varayia [89].

This is contrasted with the allocation under nodal pricing in row 2. Most notably, the grid or system operator receives no congestion rent. This rent is divided between the regions, and the grid/system operator must collect its income from other sources. Also the surplus of node 2 is reduced, the reason being that producer 2 in the CMT process trades most of his quantity at prices lower than the optimal nodal price p_2.

As is illustrated by the numerical example, the CMT model and CP-price model are quite similar since both rely on the communication of load factors as elements of the loading vector and trading rules respectively. In the CP model the load factors are combined with the prices in the market for (link based) transmission rights to determine transmission charges. In the CMT model load factors determine feasibility of trades, and the valuation of favorable load factors are implicit in the market for multilateral trades. Market power is a problem in the CP-price model, and one might ask whether a favorable β -value can result in market power for traders also in the CMT-model. The loading vector shows the "power" of all players. Will a player with a favorable β -value trade at marginal cost or will he try to bid up the price? And what is the final effect of this? Will quantities and marginal values be distorted or will it just reallocate social surplus? In the example, assuming no strategic bids, the producer with the favorable β -value is shown to be *worse* off compared to the outcome under nodal pricing.

The coordination models can be interpreted as different relaxation schemes, with competitive players in generation and consumption and the system operator solving different subproblems, and information is exchanged back and forth. The decompositions corresponding to nodal pricing and Chao-Peck prices are price driven. In the case of nodal prices the system operator hands out the optimal nodal prices of energy obtained after solving the optimal dispatch problem, and optimal dispatch is achieved as producers and consumers adapt to their local prices. For Chao-Peck prices a market is supposed to bring forth the competitive prices of transmission rights while the system operator provides information of how trades affect every single link. When traders adapt to the transmission charges of the links imposed by the prices of transmission rights, the overall problem is solved. The CMT model can be interpreted as a Benders decomposition where the market players maximize net profit and quantities are communicated to the system operator, which checks feasibility and generates constraints. The new constraints must be taken into consideration when additional trades are placed and the process continues.

The coordinating mechanisms also differ when it comes to compensating network owners and the system operator. In the nodal pricing model, the merchandizing surplus (which is non-negative in optimum) is distributed as link revenue, either directly to line owners or as a reward to holders of TCCs. The use of CP-prices also results in a transmission charge (equal to the merchandizing surplus in optimum) that is distributed through the trade of (link-based) transmission rights. In the CMT model there are no rents accruing to the system operator or grid, neither from marginal losses nor from congestion. When dealing with congestion, only feasibility is considered, and feasibility is attained when imposing constraints on the dispatch, not by pricing out the transfer limits.

In the discussion of marginal losses, we have stated that traders pay for their share of total losses, and though marginal or incremental information may be used when one is considering a new trade and its impact on losses, only total losses are covered in the CMT model. This is necessarily so since losses may be included in the trades. The difference between nodal prices, where each trader pays as if it were the last trade superimposed on the grid, and the payment according to the CMT model, can be illustrated by Figure 3-5.

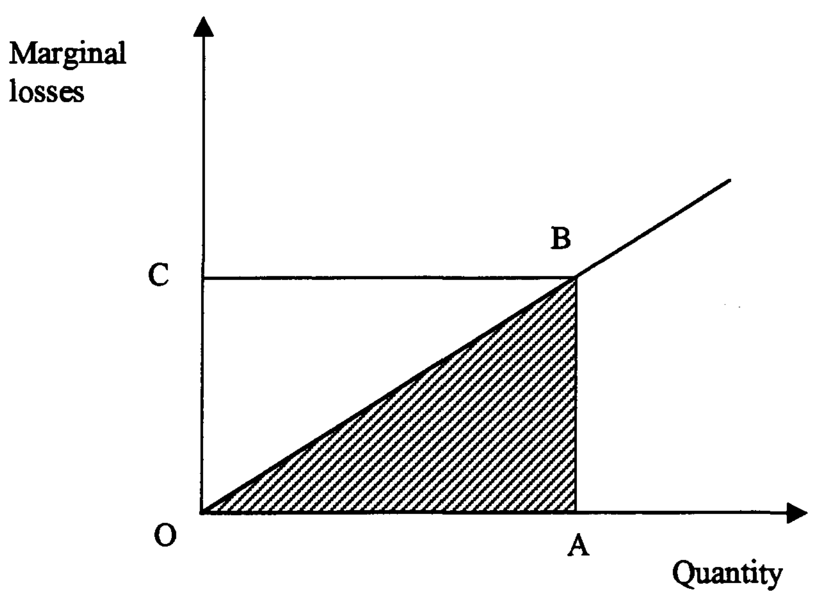


Figure 3-5 Charging for Losses

Total losses correspond to the area of triangle OAB, and this is covered by the traders in the CMT model. The income from a tariff in which all traders pay marginal losses (like the nodal pricing model) equals the area of rectangle OABC. Area OBC then represents the rent associated with marginal losses.

3.8. Market Structure and the Role of the System Operator

Until now we have not considered the dispute over market organization. We will differentiate between i) a mandatory pool (like UK and Wales), ii) bilateral trading (including multilateral trading), and iii) a voluntary pool combined with bilateral trading (US, Norway, Sweden). The transmission strategy must be discussed in the context of a specific market structure, and the responsibilities of the system operator differ fundamentally, depending on the type of energy market it is supposed to serve.

A nodal pricing system where the system operator solves optimal dispatch and acts as a market maker is best suited with a mandatory pool. The problem of bilateral trades in this system is that there are no cost/benefit data available from these markets, only quantities are reported. The system operator must then solve the optimal dispatch problem with only partial information. Chao-Peck prices and the CMT model on the other hand, are very well suited for bilateral trading. Bilateral traders then take into account transmission charges or feasibility based on the information provided by the system operator. This information is of technical character and must be completed with prices from the markets for transmission rights or trade arrangements supported by brokers. The role of the system operator concerning optimal dispatch is to calculate power flows based on the information received from the market and communicate load factors and relevant information concerning losses.

Stoft [78] argues in favor of bilateral trading and nodal markets (like a pool) to work simultaneously. The reason being that one should let the market choose which way it prefers to trade. This is also one reason for suggesting the CP+Hub price-system. When CP-prices are extended to CP+Hub prices, it provides a complete nodal energy spot market that can operate side by side with bilateral trading that relies on the CP-prices. Also Wu and Varaiya

[89] are concerned about allowing for non-discriminatory treatment of bilateral, multilateral and voluntary pool trading. The CMT model facilitates this since the pool is effectively a multilateral contract. A final comment on the CMT model is appropriate. Since it involves an iterative procedure, it requires some lead-time[24], and is therefore more suitable in scheduled power markets, and less appropriate in for instance the UK and Wales, which run a real-time market only.

3.9. *Reactive Power and Security Constraints*

In the previous discussion we have focused on real power. The pricing of reactive power is left out, assuming sufficient amounts are provided by the system operator. Hogan [36] discusses the validity of this approach in the presence of tight voltage constraints[25], claiming the necessity of both real and reactive nodal prices. By way of examples in which prices calculated in a "DC" model are compared to prices calculated by employing the full AC power flow equations, Hogan shows that reactive power affects the price of real power both through the effect on losses and voltage constraints. Voltage constraints may require a spatial reallocation of production of real and reactive power, and this may lead to an out-of-merit order generation of real power.

The examples show that if there are only thermal limits constraining the flow, the "DC" approximation provides a good model for calculating locational prices of real power. However, "translating" a voltage constraint into a flow limit does not help much, even if the "translation" is consistent with the underlying physics. Hogan's examples also show that voltage constraints can lead to prices of reactive power of the same magnitude as real power prices.

On the other hand, Kahn and Baldick [42] maintain that reactive power is really a cheap constraint as long as dispatch is optimal. The high prices of reactive power in Hogan's examples result from an uneconomic and unnecessarily constrained dispatch in which there is

[24] Although this can be as short as half an hour (Wu and Varayia [89]).

production of reactive power in the node with the low cost producer only. Generally, generators produce a bundled product of both real and reactive power, and it is argued that even though there is a substitution in the sense that more reactive power means less real power, this is an unnatural constraint. Letting the high cost producer generate reactive power, reduces the price of reactive power to a fraction of its original value. Moreover, reactive power can also be supplied locally by relatively cheap compensation equipment such as capacitors, thus reducing (net) demand and the need for generation and transmission.

The examples illustrate that quite subtle changes to dispatch have great consequences for transmission prices, i.e. the group of participants controlling the dispatch can easily manipulate prices. In complex real world networks there is great information asymmetry between transmission owners / system operator and any other party. According to Kahn and Baldick [42], the analysis emphasizes the need for monitoring and audit functions to detect potential abuses in an open access regime using marginal cost pricing.

Focusing on reliability, power balance must be maintained to guarantee the continuity of supply, even after a disturbance in the system such as the failure of a generator or the outage of a transmission line. In the previous discussion we have neither considered contingency constraints, which concern the ability to handle cascading outages as a result of disturbances in the power system. Stoft [78] argues that regarding Chao-Peck prices, this may call for two prices for each congested line since a constraint can be relieved either by reducing flow on the constrained line or the contingent line[26]. In addition to increasing the number of prices that must be found, an auction program also needs to find contingent lines. Generally, a proper framework for dealing with contingency constraints may require differentiation of power by reliability along the lines of for instance Chao and Wilson [14] and Woo et al. [87]. Agnagic et al. [1] take a broader view, including reserves also when formulating the economic dispatch to allocating both power and reserve generation resources.

[25] "The second major source of congestion in a power network arises from voltage magnitude constraints at buses."

[26] Stoft considers first-contingency constraints that take into account the power flow of the contingent line that would be most damaging to the constrained line.

We have discussed pricing/coordination of planned dispatch. Generally, we also need settlement procedures to handle failures and outages in generation and transmission or surges in demand that cause actual dispatch to deviate from planned dispatch. Chao and Peck [13] suggest a procedure offering all electricity traders priority insurance against interruptions. By offering insurance contracts with varying coverage and holding the system operator financially responsible, the system operator will have incentives to operate this part of the system efficiently and reliably.

3.10. Long Run Considerations

In the long run, equipment for production, consumption and transmission of electric power is not fixed, and the optimal development of facilities in a long run perspective is a main challenge in a deregulated electricity industry. As already mentioned, developing a transmission price mechanism to induce optimal long-term decisions is a difficult (maybe impossible) task. As described in the literature, and demonstrated in chapter 8, there are incentives to over-invest or under-invest in the network, and there are no immediate cures.

For instance, as shown by Bushnell and Stoft [8] [10], TCCs only partly eliminate the distortions of investment incentives in the case of negative externalities due to loop flow. Under-investments may result because of:

1) TCCs do not capture the full benefits of network expansions, "free-riders" may also benefit, and
2) The transmission capacity decision contains an element of monopoly power, yielding a capacity that is lower than the socially efficient level, and this effect is exacerbated by the large scale economies that are typical of transmission investments.

This means that there is probably still an important role to be played by the regulator in the lightly regulated approach to organize the transmission function. This is also pointed to by Marangon Lima and de Oliveira [49] who focus on the coordination problems of investment-decisions regarding generation and transmission in a deregulated electricity market.

In the Norwegian system, transmission tariffs are designed according to a set of criteria, namely

1) Economic efficiency
2) Cost recovery
3) "Fairness"

Our focus has been on the efficient usage of the existing grid through marginal cost pricing, which is intended to charge marginal losses and the cost of out-of-merit order dispatch caused by congestion. According to reasons explained in Pérez-Arriga et al. [59] marginal cost tariffs are not likely to recover cost.

In the pursuit of solving the revenue reconciliation problem, Kim and Baughman [43] analyze the effects of methods for adjusting marginal cost. Generally, and quite expectedly, among the linear tariffs, Ramsey pricing performs the best, generating the highest surpluses. An alternative is a non-linear tariff with variable charges to reflect short run marginal cost and a fixed charge for revenue reconciliation. A variant of this scheme has been implemented in the Norwegian system since the beginning of the 1990s. This also corresponds to the suggestions of Tabors [81] and Yu and David [94]. To recover revenue shortfalls, Yu and David distinguish between embedded cost methods and long run incremental cost approaches. Reliability is also considered, and approaches to differentiate the fixed tariff are mentioned, including capacity usage (for instance based on load factors as in Rudnick et al. [61]) and game theoretic approaches.

Differentiating the fixed charges is necessary to induce economic efficiency (for instance we do not want to exclude marginal generation and loads), but it is also closely connected to the "fairness" criterion. In this setting it seems natural to consider the use of game theoretic concepts, which can also provide guidelines on how to allocate cost not only to different locations, but also to producers and consumers, a topic that is addressed also by Rudnick et al. [61].

Except for describing some long-term effects of the fixed charges of the Norwegian transmission tariff in the next chapter, we will not consider long-term aspects any further, although it is a topic that clearly will receive our attention in future research. A fairly recent overview of contributions on transmission access issues is given by McCalley et al. [50].

4. An Overview of the Norwegian Transmission System

In the Norwegian electricity market several local and regional network monopolies are connected by a national high voltage grid. The result is a three-level network monopoly serving competing generators and customers as depicted in Figure 4-1. The Central Network, or transmission network, consists of all 300 and 420 kV lines and some 132 kV lines, and has a meshed structure. The network spans 166 connection points under which there may be production only, consumption only or both. The central grid accounts for approximately 15% of total network cost, whereas the share of cost in the regional networks (R) is about 35%. The rest is due to the lower level, which consists of a number of distribution networks (D) with voltages mainly below 22 kV and a radial structure. Production (P) is present at all levels.

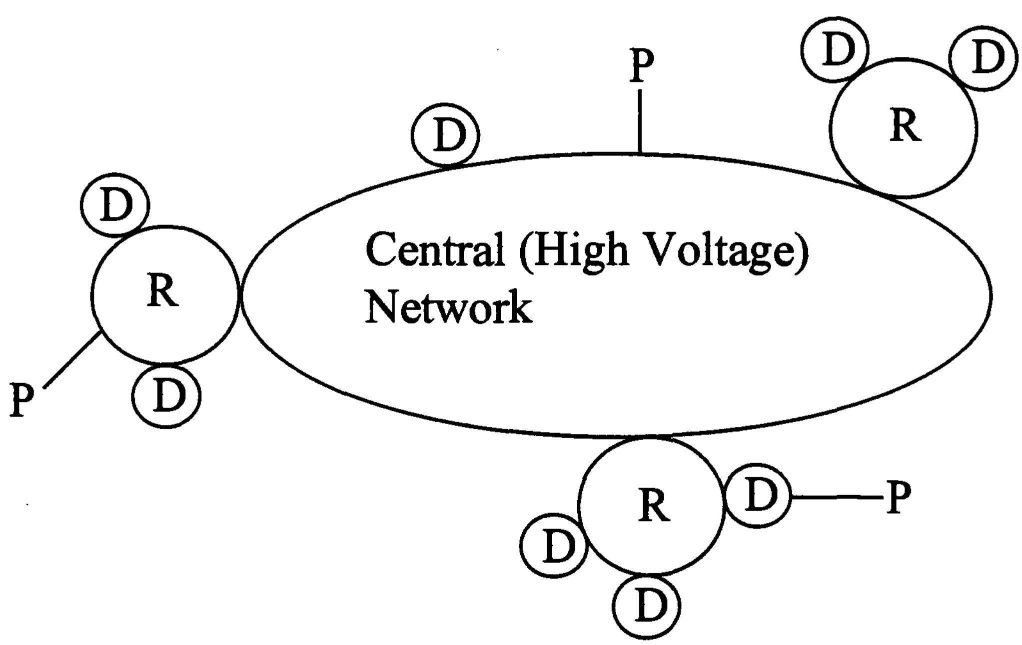


Figure 4-1 The Norwegian Transmission System

The Norwegian Power Grid Company, Statnett, is a state enterprise founded in 1991 as a result of the Energy Act of 1990, which provided the framework for increased competition. Statnett supervises and coordinates the operation of the entire Norwegian power system. This includes ensuring that the balance between production and consumption is maintained at all times, and that the power grid's capacity is not exceeded. Moreover, Statnett is responsible for the development and operation (including charging) of the central high voltage grid.

Since the entire network is still monopolized, it is subjected to public regulation and control. As a means to obtaining efficient prices, the Norwegian Water Resources and Energy Directorate (NVE) has developed guidelines for calculating transmission tariffs, applying to central, regional and distribution networks. The tariff guidelines imply a system of nodal prices, a mechanism that guarantees that transmission tariffs are determined independent of power purchase agreements, and that a single agreement at the connection point suffices to gain access to the entire coordinated network, and thus to the power market. More specifically, a variant of a nonlinear pricing scheme is adopted, consisting of a two-part tariff with variable parts reflecting short-term marginal costs associated with transmission, and fixed charges to recover any revenue shortfalls[27]. In the following we will concentrate on the tariffs of the central grid.

The variable part of the tariff consists of a loss component and a component intended to ration capacity. The capacity component is only applicable whenever transmission requirements exceed network capacity, and is then fixed at a level aiming to balance these figures. Roughly, when the scheduled market clears so that capacity limits would be violated, the network is divided into zones separated by congested cuts. The price is then lowered in net supply areas and increased in net demand areas so that markets clear in each zone without violating the congested cuts. Zonal pricing as implemented in the Norwegian scheduled market will be studied in detail in chapter 7. To settle real-time imbalances, counter purchases are used. I.e. one of the reasons for Statnett to engage in the regulation market and purchase increments and/or decrements is to relieve congestion.

The loss component is based on the value of marginal losses. Marginal losses represent changes in losses within the system due to a customer's input or takeout of energy from the network. The value of marginal losses is calculated as the product of exchanged energy, the system price of the spot market of NordPool, and the marginal loss-percentage in the

[27] At efficient operation, transmission tariffs shall provide the network owner with a reasonable return on invested capital. This determines the *permitted income* of the network owner, and this income is meant to cover the cost within his own network as well as the network owner's share of cost in superadjacent networks.

connection point (which is in the range of –10% to +10%[28]). The loss component will be differentiated over time because of variations in spot price and energy exchanged. Due to the nonlinear behavior of losses, loss components should also be differentiated in relation to total load variations in the network, requiring time-dependent or, more precisely, load-dependent loss-percentages in every connection point. In chapter 6, we will investigate the pricing of marginal losses.

Also fixed charges have two parts with different quantity bases. The "access part" is based on a measure of total production and consumption under the connection point, whereas the "net load part", which is the residual part of the tariff, is based on a measure related to the power exchanged in the node. Although the fixed (also called "usage independent") charges are related to various load measures, they are not load prices, as the load measures are associated with the load during a specific hour, namely the hour with maximal load in the price area that the connection point belongs to. The specific hour used is determined ex post and cannot be anticipated with certainty[29]. For production, the load measure is *available winter capacity*, which is the installed capacity in kW, possibly adjusted for the availability of water if this limits production during maximal winter load.

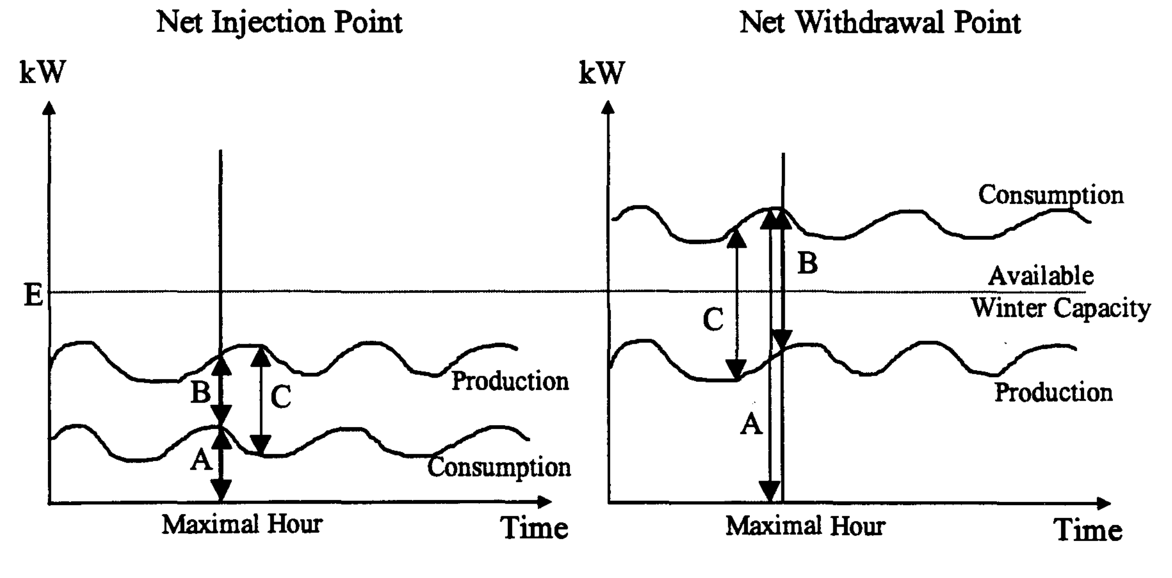


Figure 4-2 Bases for Fixed Charges

[28] These are administratively determined minimum and maximum charges.
[29] Usually, it is the coldest day of the year.

In Figure 4-2 we show the quantity bases of the fixed charges of a net injection and a net withdrawal point with the same available production capacity, E, and the same production pattern. A is total consumption and B is power exchanged in the point during the maximal hour. The maximal power exchanged in the connection point (measured by C) does not necessarily occur during the maximal hour. The quantity bases of the fixed charges of the tariff are:

"Access part":

Production: E

Consumption: A

"Net load part":

Net injection point:

	Power exchanged	B
+	Idle capacity	[E - (A + B)]
=	Basis[30]	E – A

Net withdrawal point:

	Power exchanged	B
-	Idle capacity	[E - (A - B)]
=	Basis	A - E

Thus, the "access part" charges every kW in the entire transmission system, whereas the quantity that is being charged through the "net load part" in a net injection (withdrawal) point corresponds to the maximal (minimal) power that can be exchanged given the consumption in the maximal hour.

For year 2000 the annual rates are equal to:

"Access part"	Production:	11 NOK/kW
	Consumption:	14 NOK/kW

"Net load part"	Net injection point:	45 NOK/kW
	Net withdrawal point:	63 NOK/kW

Also reactive consumption pays a fixed charge of 20 NOK/kVAR per year, and special charges apply to consumption that can be closed down at short notice.

Historically, fixed charges have accounted for approximately 70-80% of the total tariff in the central (high voltage) network. However, since the revenue from the variable parts varies over the years, and the fixed parts are to cover residual cost, the exact share vary. Lately, this figure has increased due to smaller revenues from marginal losses. We will come back to the reasons for this in chapter 6.

In Jörnsten et al. [41] we have investigated the incentive effects of the "usage independent" charges of the central grid. In the short run, fixed charges have little or no effect on output decisions. In the long run however, the tariff structure may negatively influence reserve capacity, a consequence that has received attention also in the economic press lately. According to NVE's guidelines, "The central grid's input tariffs shall serve as a reference for the usage independent tariffs in the regional- and distribution networks." This means that a producer should pay fixed charges to his connection point as if connected directly to the central grid in a "pure" injection point (i.e. a connection point with production only). In that case, every kW of available winter capacity (which is most often the installed capacity) will be charged $11+45=56$ NOK/kW per year. Considering a 40-year period with an interest rate of 7%, the present value of the tariff is approximately 750 NOK/kW. Compared to the cost of new capacity, this is a considerable cost. Based on 12 specific expansion projects described in Wangensteen et al. [85], the cost of new capacity is found to vary from 1000 to 3000 NOK/kW, i.e. the present value of the tariff constitutes 25-75% of the cost of new capacity.

Why is this a problem? In Norway, power generation is almost exclusively based on hydroelectric plants, and traditionally there have been huge capacity reserves, while energy (or water) is scarce. Lately however, load capacities have shown to become "almost" scarce on several occasions, and this is a potentially dangerous situation. Because of the fixed

[30] Special rules exist regarding minimum charges/quantities, i.e. if E-A is close to zero.

charges, investments in new capacity are discouraged. In addition, the charges give incentives to close down existing capacity, which also contributes to avoiding considerable maintenance cost. This development has already been identified in the Norwegian power supply system.

Since the establishment of a common Norwegian-Swedish power market, there has been concern about the necessity of harmonizing the tariffs of Statnett and the Swedish counterpart, Svenska Kraftnät. Both systems adhere to the nodal pricing principles, but the details of the tariffs vary. At present, capacity problems are handled by counter purchases in the Swedish system, both in the scheduled market and for settling real-time imbalances. Fixed charges are also different, as the administrator of a connection point subscribes to an exchange capacity in the connection point. Thus, the fixed tariff consists of a sort of "net load" charge only, and the rates are geographically differentiated. For the moment, the two systems exist side by side, and in spite of several reports written on the subject and convergence in some parts, there is still no agreement on a common tariff structure.

5. A Note on Loop Flow and Economic Modeling

Wu et al. [90] give counter-examples to a number of propositions regarding the characteristics of optimal nodal prices, which at first sight, without any specific knowledge of power networks, seem quite intuitive. Among the "folk theorems" that are proven false are

1) Uncongested lines do not receive congestion rents (defined through node price differences)
2) In an efficient allocation power can only flow from nodes with lower prices to nodes with higher prices, and
3) Strengthening transmission lines or building additional lines increases transmission capacity

It is argued that these assertions stem from the incorrect analogy between power transmission and the transportation of goods. Economic analyses of the transportation of goods can be found already in the classical works on *spatial price equilibrium* by Enke [21] and Samuleson [63][31]. While appealing to economic intuition, this note intends to give one possible explanation of the foundation for the difference between markets that are based on power transmission networks and spatial markets based on simpler models for transportation of goods, like commodity flows or transportation problems.

Let $B_i(S_i^d)$ be the benefit from consuming complex power $S_i^d = P_i^d + jQ_i^d$, and $C_i(S_i^s)$ the cost of producing $S_i^s = P_i^s + jQ_i^s$ in node *i*. A general formulation of the optimal dispatch problem, taking into account thermal capacity limits, is then given by problem (5-1)-(5-7) (ref. Wangensteen et al. [86]). (5-1) is the objective function, maximizing social surplus while summing benefits and withdrawing cost over all the nodes. (5-2) defines net injection $S_i = P_i + jQ_i$ in every node, and (5-3) and (5-4) relate complex power to complex voltage V_i and the conjugates of complex node and line currents I_i and I_{ik}. Inequalities (5-5) represent the thermal capacity constraints, which are stated in terms of limits C_{ik} on the magnitude of

[31] The spatial price equilibrium model can be phrased as follows. Buyers and sellers of a commodity are located at the nodes of a transportation network and the issue is to determine simultaneously the quantities supplied and demanded at each node, the local (nodal) prices at which the commodity is bought and sold, and the commodity flows between pairs of nodes.

apparent power, $|S_{ik}| = \sqrt{P_{ik}^2 + Q_{ik}^2}$. Equations (5-6) represent Kirchhoff's junction rule and (5-7) Ohm's law with Kirchhoff's loop rule incorporated, Y_{ik} being the admittance of line *ik*.

(5-1) $$\max \sum_i [B_i(S_i^d) - C_i(S_i^s)]$$

(5-2) $$\text{s.t.} \quad S_i = S_i^s - S_i^d \qquad \forall i$$

(5-3) $$S_i = V_i \cdot I_i^* \qquad \forall i$$

(5-4) $$S_{ik} = V_i \cdot I_{ik}^* \qquad \forall ik$$

(5-5) $$|S_{ik}| \leq C_{ik} \qquad \forall ik$$

(5-6) $$I_i = \sum_{k \neq i} I_{ik} \qquad \forall i$$

(5-7) $$I_{ik} = Y_{ik}(V_i - V_k) \qquad \forall ik$$

It is well known (since the work of Kirchhoff and Maxwell in the 19th century) that the physical equilibrium of electric networks can be described in terms of minimization of total power-losses, i.e. the electric current follows the path of least resistance. To simplify, consider now a direct current (DC) model, where all power flows, voltages and currents of problem (5-1)-(5-7) are real numbers. Given node currents I_i, optimal line currents I_{ik} are obtained by solving the following convex flow problem (see for instance Dembo et al. [18]):

(5-8) $$\min \tfrac{1}{2} \sum r_{ik} I_{ik}^2$$

(5-9) $$\text{s.t.} \quad I_i = \sum_{k \neq i} I_{ik} \qquad \forall i$$

where r_{ik} is the resistance of line *ik*.

Introducing dual variables V_i of equations (5-9), the Lagrangian can be written

$$\Phi = \tfrac{1}{2}\sum_{ik} r_{ik} I_{ik}^2 + \sum_i (I_i - \sum_{k \neq i} I_{ik}) \cdot V_i$$

with first order conditions

(5-10) $$\frac{\partial \Phi}{\partial I_{ik}} = r_{ik} I_{ik} - V_i + V_k = 0 \qquad \forall ik$$

and

$$\frac{\partial \Phi}{\partial V_i} = I_i - \sum_{k \neq i} I_{ik} = 0 \qquad \forall i.$$

Condition (5-10) implies

$$I_{ik} = \frac{V_i - V_k}{r_{ik}} = Y_{ik}(V_i - V_k) \quad \forall ik$$

since admittance $Y_{ik} = 1/r_{ik}$ in a DC network. I.e. the first order conditions of problem (5-8)-(5-9) correspond to equations (5-6) and (5-7). This means that we can reformulate the optimal dispatch problem (assuming a DC network with real power only, i.e. $S_i = P_i$) to:

P1 $$\max_{P_i^s, P_i^d, I_i} \sum_i [B_i(P_i^d) - C_i(P_i^s)]$$

$$\begin{aligned} \text{s.t.} \quad & P_i = P_i^s - P_i^d && \forall i \\ & P_i = V_i I_i && \forall i \\ & P_{ik} = V_i I_{ik} && \forall ik \\ & P_{ik} \leq C_{ik} && \forall ik \end{aligned}$$

and given I_i $\forall i$, I_{ik} is implicitly defined by,

$$\text{P2} \qquad \min \ \tfrac{1}{2}\sum r_{ik} I_{ik}^2$$

$$\text{s.t.} \quad I_i = \sum_{k \neq i} I_{ik} \qquad \forall i$$

which provides also the dual variables V_i.

Problem P1-P2 fits into the framework of *bilevel* programs that are discussed in Kolstad [45]. Thus, the optimal dispatch problem can be seen as a bilevel program consisting of an *upper level* program, which is the social maximization problem P1, and a *lower level* program or *behavioral* problem P2, which determines line currents and, as a byproduct, voltages. The intention of this bilevel construction is to reveal the structure of the problem, not to indicate how it should be solved. In general, the problem is highly nonlinear and non-convex with interdependencies between the variables. However, according to the classification of Kolstad [45], formulation (5-1)-(5-7) can be understood to arise after applying a *Kuhn-Tucker-Karush*-method to P1-P2, transforming the behavioral problem P2 into Kuhn-Tucker-Karush necessary conditions for optimality, and solving (5-1)-(5-7) is equivalent to solving P1-P2.

A number of economic problems can be interpreted as bilevel programs. For instance, a *Stackelberg leader-follower* game can be viewed as a bilevel program with the leader's problem corresponding to P1 and the follower's problem corresponding to P2 (Kolstad [45], Migdalas and Pardalos [53], and Vicente and Calamai [82]). In this type of model, the follower chooses his strategy in full knowledge of the leader's decision, a fact that the leader takes into consideration when determining his own actions. Similarly, *principal-agent* problems can be interpreted in the same manner, as the principal takes into account the behavior of the agent acting in his own self interest (modeled through P2) when solving the upper level program P1.

Returning to the optimal dispatch problem of electrical networks, and the discussion of Wu et al. [90] concerning the incorrect analogy between power transmission and transportation of goods, constraint (5-6), which is Kirchhoff's junction rule, is normally accounted for in most transportation models. However, if one is to disregard Kirchhoff's loop rule in the analysis, thus assuming power is routable, the error made may be of the same order as ignoring the behavior of the followers in a Stackelberg leader-follower game or the behavior of the agents in a principal-agent setting. Despite obvious similarities between the operation of the power market and spatial price equilibrium models, focusing on the physical equilibrium of a power network leads to the awareness that one should rather have in mind something similar to *traffic equilibrium* problems as the underlying network model when investigating power markets. This is highlighted by the contents of chapter 8. Also the same *non-cooperative* phenomenon is recognized in communication networks, as is evident from the works of for instance MacKie-Mason and Varian [48], Shenker [65], Shenker et al. [66], Korilis et al. [46] and Gupta et al. [28].

Viewing the optimal dispatch problem as a bilevel mathematical program with interacting physical and economic equilibria may lead to new ideas regarding optimal pricing in a decentralized electricity market. For instance, instead of (or additional to) checking if a market equilibrium is physically feasible, one could check whether a physical equilibrium is economically viable. Whether this is an interesting approach, and how it could be used in a practical procedure, is a topic for future investigation.

6. Implementation of Tariffs for Marginal Losses

As described in the overview of the Norwegian transmission system in chapter 4, the charge for marginal losses at node i in a given hour is equal to

$$\begin{pmatrix}\text{Marginal}\\ \text{loss percentage}\\ \text{of node } i\end{pmatrix} \times \begin{pmatrix}\text{Energy}\\ \text{exchanged}\\ \text{over the hour}\end{pmatrix} \times \begin{pmatrix}\text{System}\\ \text{price of}\\ \text{energy}\end{pmatrix}.$$

The charge varies from hour to hour since the spot price of NordPool (the system price of energy) changes on an hourly basis, and so does the energy exchanged at the connection point. The marginal loss percentage or loss factor of a given point gives the change in total system losses relative to a change in the quantity exchanged in that point. In general, the marginal loss percentage can be positive or negative, and depends on network characteristics, as well as the power flows in the network. In practice, the marginal loss percentages of the 166 connection points of the central grid are calculated beforehand and communicated to the market[32]. The loss factors are based on load patterns that are "typical" for the period they apply to, and there are different loss percentages for daytime and nights/weekends. In the present system loss percentages are recalculated every 8-10 weeks, and more often if required because of changes in load.

In this chapter we will take a closer look at how the loss factors are computed. Loss factors can be interpreted as loss "prices" because they show how much total system losses change due to increasing injections/withdrawals in a node by one unit. We will start by describing the method that was used until 1997, and continue with the new procedure that was adopted in January 1998. The procedures and the characteristics of the resulting charges will be illustrated by means of simple examples.

[32] For instance, they can be found at www.statnett.no.

6.1. Computing Marginal Loss Factors

Former practice (before 1998)

Based on a given load flow (the base load pattern), every node is identified as a net injection or net withdrawal point. Further, extra load is added and the incremental losses generated by this are computed in a load flow program. In accordance with the nodal pricing system, a destination-independent charge for injections is established in every node, as well as an origin-independent charge for withdrawals.

More specifically, let the network be represented by a graph, $G = G(V,E)$, V being the set of nodes, and E the set of edges. With respect to the base load pattern, define $D \subset V$ as the set of net withdrawal points, $S = V \setminus D$ is then the set of net injection points. Letting P_i denote load in node *i*, $P_i \geq 0$ if $i \in S$, and $P_i \leq 0$ if $i \in D$, $\mathbf{P} = (P_1, P_2, ..., P_n)$ is the load vector, showing injections and withdrawals in all the *n* nodes. A change in load, positive or negative, in node *i*, is represented by ΔP_i, while $\Delta\mathbf{P} = (\Delta P_1, \Delta P_2, ..., \Delta P_n)$ indicates a change in the load *vector*. Total losses associated with transmitting load $\mathbf{P}$ are measured by $L(\mathbf{P})$. The calculation of loss factors for a particular node *j* is then undertaken in the following manner[33] (Statnett [68]):

Injections

Some $\Delta P > 0$ is injected into node *j* and withdrawals are increased in every net withdrawal point $i \in D$, relative to total net withdrawals, i.e.

$$\Delta P_i = -\Delta P \cdot \frac{P_i}{\sum_{l \in D} P_l} \leq 0 \qquad \forall i \in D.$$

Net change of load in node *j* is $\Delta P > 0$ if $j \in S$ (increasing net generation) and $\Delta P + \Delta P_j \geq 0$ if $j \in D$ (reducing net consumption). This gives a new generation/consumption pattern, $\mathbf{P} + \Delta\mathbf{P}$, and the losses of this load pattern are calculated using the load flow program. The

[33] We assume that the consistency of a load vector is accomplished by adjusting the load of the swing bus.

change in total system losses, $\Delta L = L(\mathbf{P}+\Delta\mathbf{P}) - L(\mathbf{P})$, is then used to establish the loss factor, ρ_j^{inj}, for injections in node j

$$\rho_j^{inj} = \frac{\Delta L}{\Delta P}.$$

Withdrawals

Some $\Delta P < 0$ is withdrawn in node j and injections are increased in every net generation node i relative to total injections to the system:

$$\Delta P_i = -\Delta P \cdot \frac{P_i}{\sum_{l \in S} P_l} \geq 0 \qquad \forall i \in S.$$

Net change in node j equals $\Delta P + \Delta P_j \leq 0$ if $j \in S$ (reducing net generation) and $\Delta P < 0$ if $j \in D$ (increasing net consumption). Losses for the new load vector, $\mathbf{P}+\Delta\mathbf{P}$, are calculated, and the change in total system losses, ΔL, is used to establish loss factor, ρ_j^{con}, for consumption in node j

$$\rho_j^{con} = -\frac{\Delta L}{\Delta P}.$$

In principle, loss factors ρ_j^{inj} and ρ_j^{con} were used to charge for marginal losses, as loss factors were multiplied by the system price of energy and by the actual energy that was delivered/withdrawn during the hour in question. However, in the actual implementation, after having computed loss factors for every connection point, nodes were aggregated into price-areas with a common (average) loss percentage. When proceeding, we will not consider this approximation.

Example

To illustrate the general procedure and the loss factors that result, we consider an example involving 5 nodes and 5 edges (Figure 6-1). We assume that there are no binding capacity or security constraints, and to simplify computations we use a direct current (DC) network with resistances only. To find the power flows we employ a simple Gauss-Seidel numerical procedure (Wood and Wollenberg [88]). In the Gauss-Seidel procedure we specify loads in all but one node (the swing bus) and the voltage of a single node (in our case, the swing bus). Assume node n is the swing bus. From chapter 2, we know that in a DC network $P_i = V_i I_i$ and by using Ohm's law

$$I_i = \sum_{k=1}^{n} Y_{ik} V_k = V_i \cdot \sum_{k \neq i} \frac{1}{r_{ik}} - \sum_{k \neq i} \frac{1}{r_{ik}} \cdot V_k = \frac{P_i}{V_i}.$$

Considering the last equality sign, we solve for V_i on the left-hand side to obtain

$$V_i = \frac{1}{\sum_{k \neq i} \frac{1}{r_{ik}}} \left(\frac{P_i}{V_i} + \sum_{k \neq i} \frac{1}{r_{ik}} \cdot V_k \right).$$

Given estimates of the voltages $V_k, k < n$, and the voltage of the swing bus V_n, a new estimate of $V_i, i < n$ is obtained by

$$V_i^{\text{new}} = \frac{1}{\sum_{k \neq i} \frac{1}{r_{ik}}} \left(\frac{P_i}{V_i^{\text{old}}} + \sum_{k \neq i} \frac{1}{r_{ik}} \cdot V_k^{\text{old}} \right).$$

When $\left| V_i^{\text{new}} - V_i^{\text{old}} \right| < \varepsilon \;\; \forall i < n$, the procedure is terminated. The load P_n can then be found by

$$P_n = V_n I_n = V_n \cdot \sum_{k=1}^{n} Y_{nk} V_k,$$

and it includes losses.

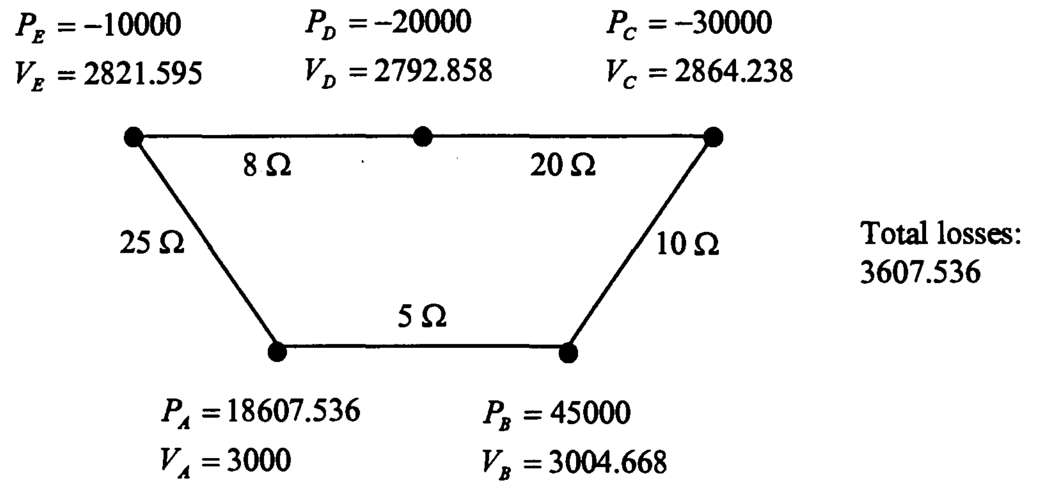


Figure 6-1 Base Load Flow

In Figure 6-1, a base load pattern used to define loss factors is exhibited. Node A is assumed to be the swing bus, such that loads are specified for nodes B, C, D and E, together with voltage V_A. P_A, V_B, V_C, V_D and V_E are then obtained by the Gauss-Seidel procedure. In the following we will show how the loss factors of node A are computed.

Injections in point A

Injection increases with $\Delta P > 0$ and is matched by increased consumption in nodes C, D and E relative to total consumption that is transmitted in base load flow (Figure 6-2).

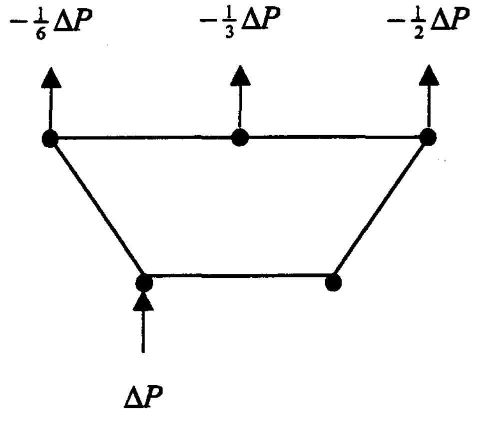


Figure 6-2 Loss Factor of Injections in Node A

$\mathbf{\Delta P} = (\Delta P, 0, -\frac{1}{2}\Delta P, -\frac{1}{3}\Delta P, -\frac{1}{6}\Delta P)$ is added to base load flow $\mathbf{P}$ and gives a new set of injections and withdrawals. We calculate the new losses and compare them with the losses of $\mathbf{P}$. With $\Delta P = 1$, we find that the loss factor of injections in point A is equal to

$$\rho_A^{inj} = \frac{\Delta L}{\Delta P} = 0.12774356.$$

Withdrawals from point A

Consumption increases with $\Delta P < 0$ and is matched by increased injection in nodes A and B relative to total net injection in base load flow (Figure 6-3).

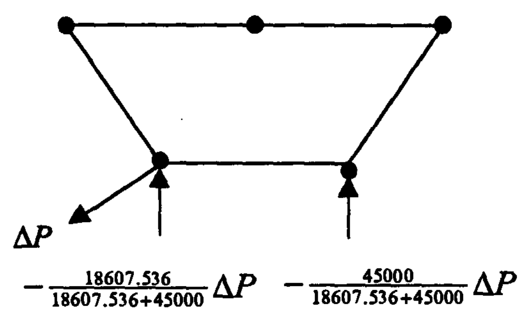


Figure 6-3 Loss Factor of Withdrawals from Node A

The change of flow is equal to $\mathbf{\Delta P} = (\Delta P - \frac{18607.536}{18607.536+45000}\Delta P, -\frac{45000}{18607.536+45000}\Delta P, 0, 0, 0)$ and new losses are calculated for $\mathbf{P} + \mathbf{\Delta P}$. With $\Delta P = -1$ the loss factor for withdrawals in point A becomes

$$\rho_A^{con} = -\frac{\Delta L}{\Delta \mathrm{P}} = 0.00033489.$$

From the examples of Figure 6-2 and Figure 6-3 it is obvious that injections and withdrawals are not treated symmetrically, i.e. they are assumed to influence power flows differently, and we cannot expect that $\rho_i^{con} = -\rho_i^{inj}$.

Present practice (since 1998)

Loss factors ρ_j^{inj} and ρ_j^{con} are the basis for charging marginal losses in the present system as well. However, the factors are now adjusted according to a specific formula, so that charges for generation and consumption in a node become equal in absolute value, but have opposite signs. This is achieved by letting

$$\nu_j^{inj} = \frac{\rho_j^{inj} - \rho_j^{con}}{2}$$

for injections, and

$$\nu_j^{con} = \frac{\rho_j^{con} - \rho_j^{inj}}{2}$$

for withdrawals.

For the given example, with $\Delta P = +/-1$, the resulting loss factors according to the old and new methods are given by columns $\rho(\text{inj})/\rho(\text{con})$ and ν, respectively, in Table 6-1 below. $\nu(\text{con}) = -\nu(\text{inj})$ and is not displayed.

Table 6-1 Loss Factors for DC Example

	ρ(inj)	ρ(con)	ν(inj)
A	0.12774356	0.00033489	0.06370434
B	0.12821622	-0.00013833	0.06417727
C	0.02467858	0.10339975	-0.03936059
D	-0.03262769	0.16070751	-0.09666760
E	-0.00877664	0.13685651	-0.07281657

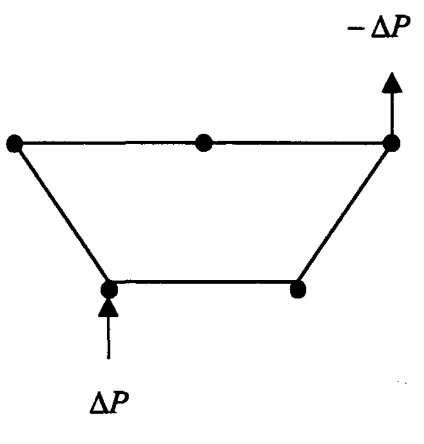
Why is it desirable to have loss factors, f, where $f_i^{con} = -f_i^{inj}$? In the first place, as is demonstrated in section 3.1, optimal dispatch indicates that this is a central feature of an optimal tariff. Secondly, it is claimed that it is easier to carry over tariffs to underlying networks (Statnett [70]). Finally, it has been an essential feature of the Swedish system, and in the process of integrating the Nordic power markets there has been a desire to harmonize the tariff-systems of the different countries. At first sight, the chosen method seems to be merely a "trick" to attain injection- and withdrawal-charges that are equal in absolute values, but have opposite signs, and it has been questioned by people in the industry whether the new set of loss factors reflects marginal cost, i.e. marginal losses.

6.2. *Do Loss Factors Reflect Marginal Losses?*

In this section we show 4 different $\Delta\mathbf{P}$ s and check how they affect marginal losses. This is compared to the losses that are charged according to loss factors ρ and ν. Let ΔL be the change in marginal losses due to $\Delta\mathbf{P}$, i.e. ΔL is the marginal cost of $\Delta\mathbf{P}$ in units of energy/power. $\Delta Q(f)$ is the change in the quantity that is charged using loss factors given by f. Thus, ΔL and $\Delta Q(f)$ are comparable.

Example 1

Increasing net injection in node A, withdrawing it in node C is a change that increases the total quantity that is transmitted over the network.



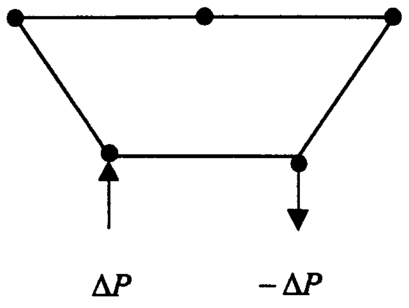
$\Delta P = 1$ gives

- $\Delta L = 0.10306560$
- $\Delta Q(\rho) = \rho_A^{inj} + \rho_C^{con} = 0.12774356 + 0.10339975 = 0.23114331 \neq \Delta L$
- $\Delta Q(\nu) = \nu_A^{inj} + \nu_C^{con} = 0.06370434 + 0.03936059 = 0.10306493 \approx \Delta L$

I.e. using ρ as the basis for charging marginal losses does not reflect marginal losses due to $\Delta\mathbf{P} = (1,0,-1,0,0)$. On the other hand, the ν -based charge gives a very good approximation.

Example 2

Increasing net injection in node A, withdrawing it in B represents a redistribution of injections.



$\Delta P = 1$ gives

- ΔL = -0.00047249
- $\Delta Q(\rho) = \rho_A^{inj} - \rho_B^{inj} = 0.12774356 - 0.12821622 = -0.00047266 \approx \Delta L$

  (However if B becomes a net consumption node[34], the change of quantity charged per unit of withdrawal is $\rho_A^{inj} + \rho_B^{con} = 0.12774356 - 0.00013833 = 0.12760523 \neq \Delta L$)
- $\Delta Q(\nu) = \nu_A^{inj} - \nu_B^{inj} = 0.06370434 - 0.06417727 = -0.00047293 \approx \Delta L$

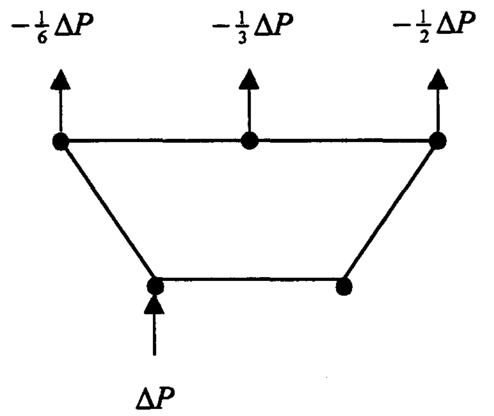
I.e. using ρ as the basis for charging marginal losses reflects marginal losses due to $\Delta\mathbf{P} = (1,-1,0,0,0)$ as long as the change in flow still keeps point B a net injection point, i.e. ρ_A^{inj} and $-\rho_B^{inj}$ applies. The ν -based charge also gives a very good approximation of marginal losses in this case. By checking every change in load involving two nodes, it is seen that ν reflects marginal losses consistently.

Example 3

The $\Delta\mathbf{P}$s that are considered in examples 1 and 2 are different from the $\Delta\mathbf{P}$s used to establish loss factors because they involve a single origin and a single destination point.

[34] For this to happen in our example would require a large change in loads, which would bring about an entirely different operating point, where ΔL does not apply. However, in a larger network, ΔL may still be a good approximation of the change in losses even if the change of load in a single node is extensive.

However, the loss factors ρ_i^{inj} and ρ_i^{con} can be viewed as average values since they assume a specific distribution of withdrawals and injections when found. In this example we consider the $\mathbf{\Delta P}$ used to calculate the loss factor of injections to node A.



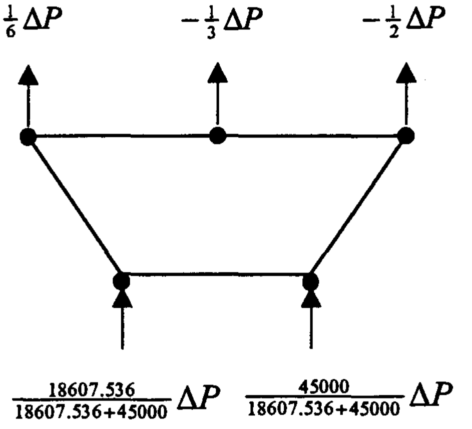
$\Delta P = 1$ gives

- $\Delta L = 0.12774356$
- $\Delta Q(\rho) = \rho_A^{inj} + \frac{1}{2}\rho_C^{con} + \frac{1}{3}\rho_D^{con} + \frac{1}{6}\rho_E^{con} = 0.25582202 \neq \Delta L$
- $\Delta Q(\nu) = \nu_A^{inj} + \frac{1}{2}\nu_C^{con} + \frac{1}{3}\nu_D^{con} + \frac{1}{6}\nu_E^{con} = 0.12774326 \approx \Delta L$

Even in this case, using ρ as the basis for charging marginal losses does not reflect marginal losses due to $\mathbf{\Delta P}$. Again, the ν-based charge gives a very good approximation.

Example 4

In the final example we consider a vector $\mathbf{\Delta P}$ that is proportional to base load.



$\Delta P = 1$ gives

- $\Delta L = 0.12807784$
- $\Delta Q(\rho) = 0.12807846 + 0.12807795 = 0.25615641 \approx 2 \times \Delta \mathrm{L}$
- $\Delta Q(\nu) = 0.06403892 + 0.06403892 = 0.12807784 \approx \Delta \mathrm{L}$

When using ρ on this $\Delta \mathbf{P}$, marginal losses are charged twice, as they are fully paid by *both* injections *and* withdrawals. This is not the case for the ν -based charges, which make a 50-50 split of marginal losses on injection- and withdrawal points.

This is further illustrated if we consider the total quantities that are charged in base load under the different loss factor structures. In the case of ν this is equal to 7915.693, while it is 15831.431 (≈2×7915.693) when combining ρ^{inj} and ρ^{con}. Both quantities are considerably greater than the total losses of 3607.536, and this is natural since total transmission losses on a line vary quadratically with currents. However, the difference between ν and ρ is very large, and both cannot be correct.

When using ρ^{inj} and ρ^{con}, injections and withdrawals are treated asymmetrically when changing loads to calculate loss factors. A common analogy is to refer to the network as a "market place"[35]. In this analogy, computing ρ^{inj} and ρ^{con} corresponds to choosing different locations for the "market place" for injections (a weighted average of the withdrawal points) and withdrawals (a weighted average of the injection points). The effect is that $\Delta \mathbf{P}$ s that use loss factors from both the $\rho(\text{inj})$-column and $\rho(\text{con})$-column of Table 6-1, are mis-priced. This applies to other transportation networks such as commodity flows and transportation problems as well. To illustrate the error that is made, consider a single-line network with production in node A and consumption in node B, and a transportation cost of c per unit on the line between the nodes (Figure 6-4).

[35] I.e. a "market place" in the sense of a point of delivery.

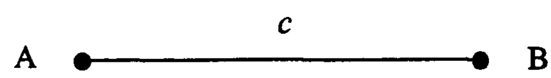


Figure 6-4 A Single-Line Transportation Network

ρ_A^{inj} is defined as the per unit cost of transporting from A to B, i.e. c. ρ_B^{con} is also defined as the per unit cost of transporting from A to B, implying that a unit that uses the line between A and B is charged $\rho_A^{inj} + \rho_B^{con} = 2c$. When computing ρ_A^{inj} the "market place" is assumed to be at point B, while it is assumed to be located in point A when ρ_B^{con} is determined. In this way, both producers *and* consumers pay the full cost of transporting a unit between A and B. This characteristic of loss-factors ρ^{inj} and ρ^{con} is recognized by Statnett[36] (Statnett [68]). It is also claimed that this practice is correct![37]

In the cited report, we also find suggestions of how to reflect "total losses" in the system[38]. Whether this refers to total marginal losses or total losses is not clear. The suggested Method 1 corresponds to ν (possibly combined with an administratively determined adjustment factor). Method 2 on the other hand, consists of adjusting ρ^{inj} and ρ^{con} by a factor k, i.e. $p^{inj} = k \cdot \rho^{inj}$ and $p^{con} = k \cdot \rho^{con}$, where k is defined as "physical losses in the network considered divided by marginal losses"[39]. It is not straightforward whether Method 2 is intended to cover total losses (corresponding to 3607.536 in the example) or marginal losses. In any case, Method 2 will generally not have ρ^{inj} equal to $-\rho^{con}$.

[36] "En økning i uttaket et sted i nettet forutsetter en tilsvarende økning i innleveringen til nettet. Motsatt gjelder at en økning i innleveringen forutsetter en samtidig økning i uttaket. Ved å bruke de beregnede tapsfaktorene ved fastsettelse av energileddet, fører det til at både den som tar ut kraft og den som leverer inn kraft dekker de fulle marginale tap som følge av en endring i uttak/innlevering. En slik beregning av energileddet gir en tariff som i sum overstiger de marginale kostnadene som aktørene påfører systemet med sine handlinger."

[37] "Prinsipielt kan det hevdes at det er riktig at hver enkelt aktør fullt ut møter de marginale tapene som systemet påføres. Tapsfaktorene kan da anvendes direkte."

[38] "Hvis man imidlertid bare ønsker å avbilde de totale kostnadene i systemet, kan følgende to alternative metoder anvendes for å tilpasse tapsfaktorene:"

[39] "Faktoren k beregnes som fysiske tap i aktuelt nett dividert på marginale tap."

From the income statements of Statnett[40] we find that marginal losses have accounted for 17-34% of the total tariff, as shown in Table 6-2.

Table 6-2 The Share of Income from Marginal Losses

1993	1994	1995	1996	1997
17%	29%	24%	34%	23%

When Method 1 was adopted, the practice of aggregating nodes into price-areas with common (average) loss factors was abandoned, and loss factors are recomputed more often than before. As a consequence, the charge for marginal losses was budgeted to 20% of the total tariff for 1998. Having observed the actual incomes in 1998, the share was further reduced to 9% in the 1999-budget. In our opinion, this is not surprising since the new method for computing loss factors in itself will halve the income from marginal losses.

6.3. Alternative Loss Factors

Stoft [77] uses an alternative computation of loss factors, involving a "hub". The "hub" is a reference node that can also be interpreted as corresponding to a "market place". The loss factor of injections in node *i* is the per unit incremental loss of transmitting power from node *i* to "hub". Similarly, for withdrawals the loss factor of withdrawing a unit in node *i* is the change in losses due to transmitting a unit from "hub" to node *i*. Choosing node A as "hub" in our example gives the loss factors of Table 6-3.

Table 6-3 Loss Factors with "Hub" in Node A

	η(inj)	η(con)
A	0.0	0.0
B	0.00047351	-0.00047249
C	-0.10306209	0.10306560
D	-0.16036735	0.16037314
E	-0.13651697	0.13652201

[40] http://www.statnett.no/aarsrapport/kapitler/html/kap15.html and "Sentralnettstariffen, Sammenstilling av data 1994 (uke 1-52)", Statnett, Kraftsystemdivisjonen, Avdeling for Sentralnett, September 1995.

Like ν, the loss factors given by η correctly reflect marginal losses for any trade (as long as it is small relative to the base load flow).

Consider now computing loss factors by $\Delta\mathbf{P}$ s that involve every node in the system, i.e. letting increased injections (withdrawals) in a point be matched by increased (reduced) consumption in net withdrawal points *and* reduced (increased) injections in net generation nodes. This corresponds to seeing the "market place" as a weighted average of *all* the nodes of the grid. If we make a 50-50 split on injections and withdrawals[41], and the load changes are distributed relative to net injections/withdrawals in base load flow, we obtain loss factors ν.

As long as we retain our definition of the "market place" when computing loss factors for injections and withdrawals, the charges of a trade will reflect marginal losses. This is so whether the "market place" is an arbitrary "hub" or any combination of nodes. This means that loss factors given by ρ^{inj} and ρ^{con} can be used, only if ρ^{inj} is used for charging injections, $-\rho^{inj}$ must be used to charge withdrawals (refer example 2). Similarly, if ρ^{con} is used to charge withdrawals, then $-\rho^{con}$ must be used to charge injections.

Thus we have already computed 4 different sets of loss factors, corresponding to different "market places". These are given in Table 6-4, where only f_i^{inj} is shown since $f_i^{con} = -f_i^{inj}$. Compared to the numbers in Table 6-1, the signs of the $\rho(\text{con})$-numbers are reversed since this column in Table 6-1 consists of loss factors for *withdrawals*.

Table 6-4 Loss Factors Corresponding to Different "Market Places"

	"MARKET PLACE"			
	A [η]	A,B [ρ (inj)]	C,D,E [ρ (con)]	A,B,C,D,E [ν]
A	0.0	0.12774356	-0.00033489	0.06370434
B	0.00047351	0.12821622	0.00013833	0.06417727
C	-0.10306209	0.02467858	-0.10339975	-0.03936059
D	-0.16036735	-0.03262769	-0.16070751	-0.09666760
E	-0.13651697	-0.00877664	-0.13685651	-0.07281657

[41] I.e. half of the increase in injections (withdrawals) of point i is covered by reducing (increasing) injections in net generation points, and half of it stems from increasing (reducing) withdrawals in net consumption nodes.

The loss factors of a specific node vary considerably from column to column, however the differences between nodes are the same. For instance, injections in node B are approximately 0.00047 more expensive than injections in node A, and injections in A that are withdrawn in C, cost approximately 0.10306 for any of the sets of loss factors in Table 6-4. Thus the relative cost differences prevail.

This means that the choice of "market place" determines the level of the loss factors in the nodes, but it does not affect the market outcome and the final distribution of revenue. As phrased by Stoft [77]: "The surprise about loss factors is that it does not matter what bus is chosen as the hub. Any bus will yield a efficient system of loss charges and the same final distribution of revenue. The basic principle behind this phenomenon is most easily understood by considering a one-line network with generation at one end and load at the other. If the hub is placed at the generation end, the generator pay no loss charges. If the hub is placed at the load end, the generators pay all of the loss charges. But of course, if the generators pay for losses, they raise their prices by exactly this amount and so the loads end up no better off than if they had paid for the losses directly. Complex networks change nothing."

Which agents that finally bear the transmission cost, depends on the elasticities of supply and demand. This is illustrated in Figure 6-5.

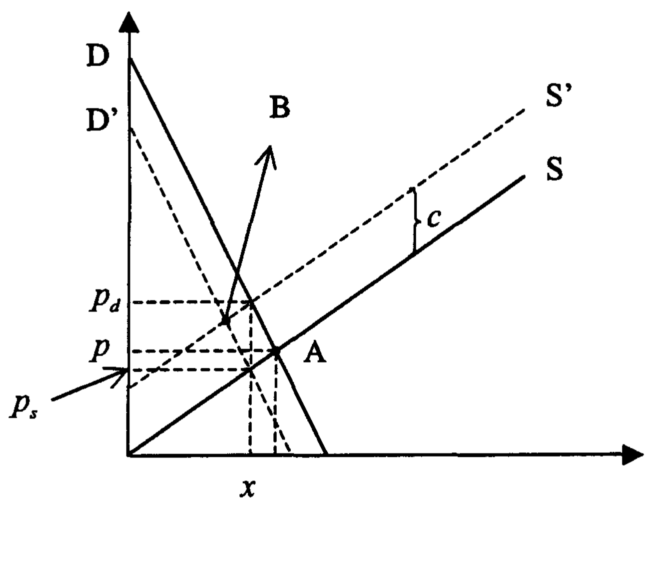


Figure 6-5 Final Distribution of Charges

Without loss charges the market clears in point A. In S' the marginal cost c is paid by producers, and the supply curve is shifted to the left (at a given price, less is offered). In D' marginal cost is collected by consumers and the demand curve faced by suppliers is shifted to the left compared to D (at a given price, less is consumed). In any case, the market outcome is x, consumers pay p_d, and producers receive p_s, and the difference $p_d - p_s = c$. In the example, demand is the more inelastic and pays the gross part of the charge $(p_d - p > p - p_s)$. If both producers and consumers are charged the full cost of transportation, the market clears in point B, at a higher price and a lower quantity than optimal.

In Stoft [77] the focus is on methods that cover total losses only, i.e. loss factors are to cover average losses. In the US system, generators are allowed to pay in kind. If every generator delivers its marginal losses, it would result in huge imbalances since marginal losses exceed average losses. If it is essential to allow paying in kind, Stoft suggests that loss factors should not be adjusted by a scaling factor as in Method 2 of section 6.2. Instead, loss factors should be *shifted* by subtracting a constant from every single loss factor. This is to correctly reflect the cost-differences between generation (consumption) in different nodes.

In the US market there is also a second limitation on the design of the charge for marginal losses, namely that generators are to pay for all losses. In a nodal pricing system (where the "prices" of injections are destination-independent), (electrically) distant loads will consume too much, as they are subsidized by loads that are (electrically) close to the site of generation. By charging withdrawals as well, the cost of any trade can in principle be correctly charged, as in the present Norwegian system. This is also applicable to transportation networks in general. In Samet et al. [62] and Hallefjord et al. [29], cost allocation on the destination points of a transportation problem is considered. The Aumann-Shapley price mechanism is the unique mechanism fulfilling a set of five axioms, which seem very general and reasonable for a "fair" cost allocation mechanism. In general, the Aumann-Shapley prices do not correspond to marginal cost. If one is allowed to charge *both* supply and destination points, the problem considered is in some sense trivial, as any set of dual prices will do the job, due to strong duality of linear programming.

6.4. Concluding Remarks

Even if loss factors are used consistently[42], the way marginal losses are charged for still implies a number of approximations. The most obvious simplification is that a few operating points are chosen as the basis for calculating loss factors, and these operating points do not necessarily correspond to optimal dispatch or the actual points of operation. In principle, loads and charges for losses should be determined simultaneously, but this would be computationally and informationally complex. Instead, loss factors are announced 14 days prior to the period they apply to, based on expected load patterns. When producers and consumers adapt to these charges, dispatch will not be strictly optimal.

In addition, there are a number of simplifications of minor importance. Compared to the prices found in the optimal dispatch problem of section 3.1, the loss factors constitute a secant instead of a tangent to the marginal cost function. When ΔP is small, the error is negligible, for instance in our example, letting $\Delta P = +/-1$, the difference for injections and withdrawals occur in the fifth decimal as can be seen in Table 6-3 showing the "hub"-based loss factors. These loss factors also illustrate the nonlinear quality of losses, since the absolute values of the positive prices are a little bit higher than the absolute values of the negative prices.

When computing loss factors using some $\mathbf{\Delta P}$, $\mathbf{\Delta P}$ is generally not based on the elasticities of supply and demand as the charges according to optimal dispatch would be. Contrary to transportation networks in which total supply and demand balance, different choices of "market place" will affect the total charges collected from the tariff. In our example, the quantities that are charged in base load by the 4 different loss factor sets of Table 6-4 are given in Table 6-5.

Table 6-5 Quantities Charged under Different Loss Factors

Loss factors	η	ρ(inj)	ρ(con)	ν
Quantity	7685.687	8146.686	7684.701	7915.693

[42] And disregarding administratively determined maximum and minimum factors.

The reason for the differences is that we assume losses to be produced by the swing bus. Thus losses are injections that have no counterpart on the demand side, and are charged according to some nodal loss factor that varies over the different loss factor sets. For instance, $\rho_A^{inj} \neq -\rho_A^{con} \neq \nu_A^{inj} \neq \eta_A^{inj}$, and the difference between the 8146.696 of ρ(inj) and the 7684.701 of ρ(con) is approximately equal to the losses of 3607.536 times the difference in injection "prices" in node A (0.12774356-(-0.00033489)). In addition, the choice of swing bus will in itself affect the size of the losses (Wood and Wollenberg [88]). Thus, losses are generally not produced in the most cost-effective manner, as it would be in optimal dispatch, and this will to a certain extent affect the loss factors that are computed.

In this chapter we have tried to illustrate that even if the principle of marginal cost pricing is well known and agreed upon, implementation counts. The differences between the old and new loss factors did not seem to be very well understood in the market. However, the procedure adopted in 1998 for computing loss factors is far better than the old method when it comes to reflecting marginal losses.

7. Zonal Pricing

As already mentioned in previous chapters, a zonal approach to managing congestion has been adopted in the Norwegian scheduled power market. In this chapter we will illustrate some of the problems that the zonal pricing system, as implemented in Norway, has. With the use of small network examples we illustrate the difficulties involved in defining the zones, the redistribution effects of the surplus that a zonal pricing system has, as well as the conflicting interests concerning zone boundaries that are present among the various market participants. We also show that a zone allocation mechanism based on optimal *nodal* prices does not necessarily lead to a zone system with maximal social surplus. Finally, we formulate an optimization model that when solved yields the zone system that maximizes social surplus given a pre-specification of the number of zones to be used.

The trading process of the Norwegian system works approximately as follows:

1) Based on the supply and demand schedule bids given by the market participants, the market is cleared while ignoring any grid limitations. This produces a system price p of energy.
2) If the resulting flows induce capacity problems, the nodes of the grid are partitioned into zones.
3) Considering the case with two zones defined, the zone with net supply is defined as the low-price area, whereas the net demand zone is determined the high-price area.
4) Net transmission over the zone-boundary is fixed when curtailed to meet the violated capacity limit.
5) The zonal markets are now cleared separately giving one price for each zone, p_L being the low price and p_H the high price. If the flow resulting from this equilibrium still violates the capacity limit, the process is repeated from step 4). If any new limits are violated the process would be repeated from step 2), possibly generating additional zones.
6) The revenue of the grid-company (from capacity charges) is equal to the price difference times the transmission across the zone-boundary.

An assumption made in the six steps given above, is that a zone boundary should cut the link with the capacity problem. In a large network this still leaves the grid-company, Statnett, with a huge flexibility when defining the zone-boundaries. According to Statnett [69] the system can be interpreted as inflicting a positive capacity charge $p - p_L$ in the low price area and a negative charge $p - p_H$ in the high price area (relative to the system price of energy). This means that withdrawals are charged in the high price area and compensated in the low price area. For net injections the opposite is valid.

As pointed out above it is not exactly clear how the number of zones and zone-boundaries are to be determined. Stoft [74] [76] shows that the partition of the network into zones generally is not obvious[43], but states that it should be based on price differences, the reason being that the dead-weight loss resulting from erroneous prices is generally proportional to the square of the pricing error. Walton and Tabors [84] also focus on price differentials and suggest that it might be possible to use statistical methods using the standard deviation of nodal price distributions as a criterion to determine the number of zones and which nodes belong to/do not belong to the different zones.

In this chapter we will show the multitude of possible cuts, representing zone-boundaries, that exists even in a small example and study the resulting welfare effects. Different zone allocations will affect both the overall efficiency and the allocation of social surplus. We will also illustrate that the partition of the network into zones based on absolute values of optimal nodal price differences does not necessarily lead to a zone system with maximal social surplus. Gaming is an aspect that we will not consider since we assume nodal markets to be competitive when calculating the market outcomes.

[43] Networks are generally not zonable.

7.1. Computing Zonal Prices

We consider real power in a lossless and linear "DC"-model with all line-reactances equal to 1. When the number of zones K $(\leq n)$ and the allocation of nodes to zones $Z_1,\dots,Z_K$ are determined, the optimal zonal prices can be found by solving the following problem:

(7-1) $$\max \sum_{i=1}^{n}\left(\int_0^{q_i^d} p_i^d(q)dq - \int_0^{q_i^s} p_i^s(q)dq\right)$$

(7-2) $$\text{s.t.} \quad q_i^s - q_i^d = \sum_{j\neq i} q_{ij} \qquad i=1,\dots,n-1$$

(7-3) $$\sum_{ij\in L_l} q_{ij} = 0 \qquad l=1,\dots,m-n+1$$

(7-4) $$\sum_{i=1}^{n}(q_i^s - q_i^d) = 0$$

(7-5) $$q_{ij} \leq C_{ij} \qquad 1\leq i,j\leq n$$

(7-6) $$\begin{cases} p_i^s(q_i^s) = p_{Z_k} \\ p_i^d(q_i^d) = p_{Z_k} \end{cases} \qquad i\in Z_k,\ k=1,\dots,K.$$

Here $p_i^d(q_i^d)$ is the demand function of node i and q_i^d is the quantity of real power consumed. $p_i^s(q_i^s)$ is the supply function of node i, while q_i^s is the quantity of real power produced. C_{ij} is the capacity of link ij, q_{ij} is the power flow over the link from i to j, and p_{Z_k} is the price in zone Z_k.

The objective function (7-1) expresses the difference between consumer benefit (the area under the demand curve) and the cost of production (the area under the supply curve). Equations (7-2) correspond to Kirchhoff's junction rule, while equations (7-3) represent the loop rules (equivalent to (2-12) and (2-13) respectively). Equation (7-4) is the conservation of energy, while inequalities (7-5) are the capacity constraints. Equations (7-6) guarantee that prices are uniform over nodes belonging to the same zone.

Solving (7-1) (or alternatively (7-1)-(7-4) to obtain line flows) gives the unconstrained dispatch and the system price. Problem (7-1)-(7-5) corresponds to the optimal dispatch problem, and solving (7-1)-(7-6) provides us with the optimal zonal prices. It is obvious that the social surplus of the optimal dispatch is less than or equal to the unconstrained social surplus and greater than or equal to the social surplus of the zonal solution. Moreover, it is obvious that a finer partition of the grid (dividing a zone into two or more "sub-zones" by allowing more prices) will increase social surplus or leave it unchanged.

In practice we would not solve problem (7-1)-(7-6) to find the zonal solution, because this would be equally complicated as solving the optimal dispatch problem. A practical algorithm based on the described procedure of the Norwegian system could be based on curtailment of the unconstrained dispatch. When capacity limits are violated, the grid is partitioned and trades between zones are curtailed until limits are restored. Zonal markets are then cleared separately and new flows are calculated. If these flows still violate the constraints, the flows are curtailed further and we repeat the process. Following the description of the Norwegian system, defining high price and low price areas, we could alternatively lower the price in the low price area and increase the price in the high price area until balance is restored. We will discuss possible problems pertaining to these procedures in relation to the examples of the next section.

7.2. Examples

The network considered contains 5 nodes connected by 8 edges like the grid of Figure 7-1. In every node there is both production and consumption, and we assume quadratic cost and benefit functions implying linear supply and demand curves. Demand in node i is given by $p_i = a_i - b_i q_i^d$, where p_i is the price in node i and a_i and b_i are positive constants. Supply is given by $p_i = c_i q_i^s$ where c_i is a positive constant. In the specific example considered, we assume identical demand curves in every node, while the cost functions vary as shown in Table 7-1.

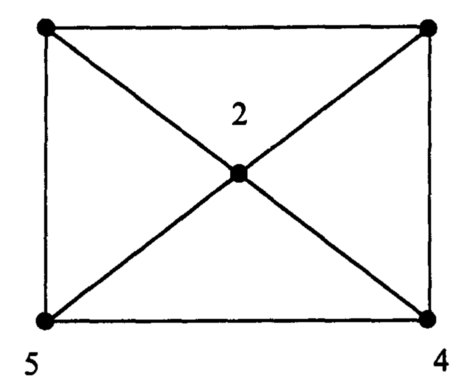
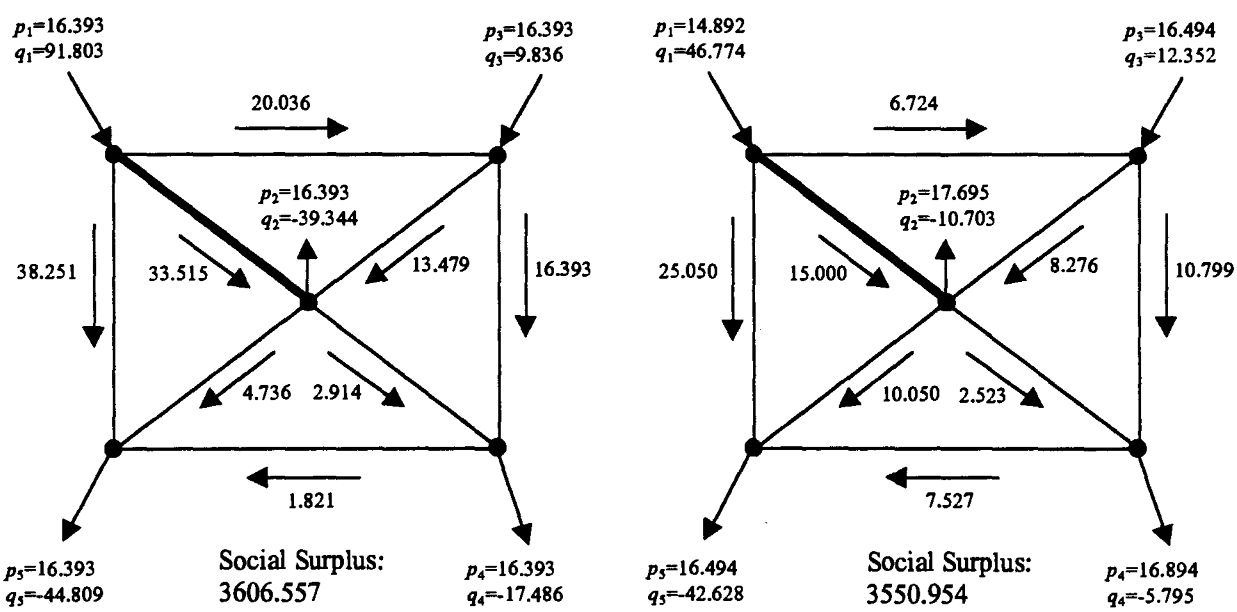


Figure 7-1 Grid Topology

Table 7-1 Supply and Demand Data

NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.5
3	20	0.05	0.2
4	20	0.05	0.3
5	20	0.05	0.6

In the unconstrained dispatch we get a uniform nodal price of 16.393 (the system price of energy). Net injections, $q_i = q_i^s - q_i^d$, and line flows are shown in Figure 7-2 part A. Line 1-2 is assumed to have a capacity of 15 units and is overloaded in the unconstrained dispatch. Taking into account the flow limit, we get the optimal dispatch shown in Figure 7-2 part B.



Part A:
Unconstrained Dispatch

Part B:
(Constrained) Optimal Dispatch

Figure 7-2 Optimal Dispatch

In the following we will examine zonal pricing. Even if we are restricted to use only two zones in the example, several allocations are possible. In practical zonal implementations[44], the nodes at the endpoints of the congested line would typically be allocated to different zones. However, as is shown later, this is not necessarily optimal when there are more than one congested link. When restricting the attention to the case where the endpoints of the congested link are allocated to different zones, there are 8 different zone allocations in the example. They are all exhibited in Figure 7-3.

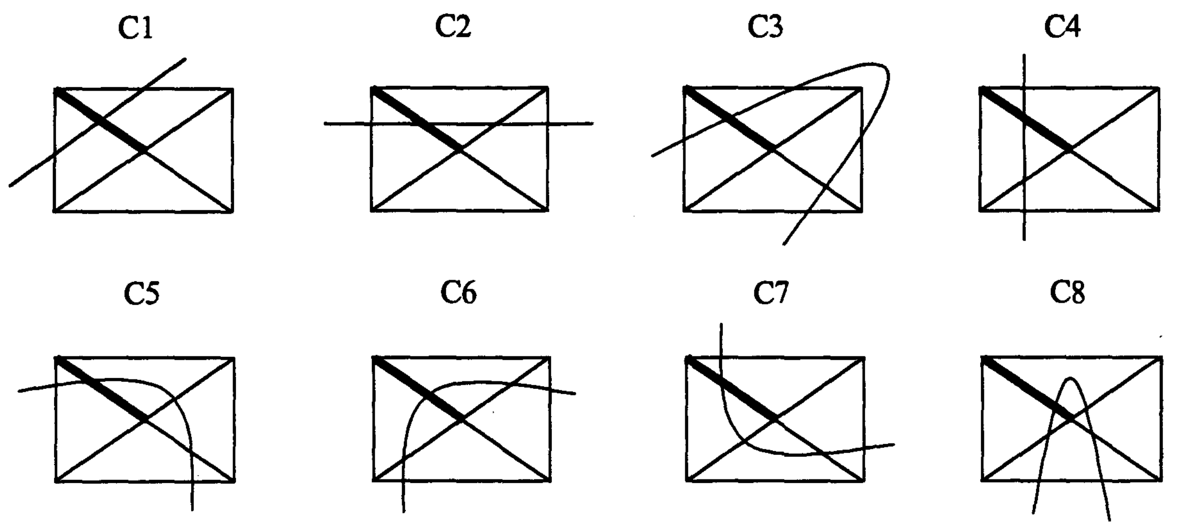


Figure 7-3 Zonal Allocations

Generally, if we consider a single congested line in an n - node network, and if we assume that the endpoints of the congested link are to be allocated to different zones, the number of allocations to two zones is equal to[45]

$$\sum_{i=0}^{n-2}\binom{n-2}{i}.$$

It may be questioned whether all these cuts are meaningful, if not, this is an "at most" number. For instance, as regards cut C3 the zone containing nodes 1 and 4 is not connected, so it can be argued that the network has in practice 3 zones and should be treated accordingly.

[44] Consider for instance the WEPEX (Western Power Exchange) proposal discussed by Stoft [76].

In the given example, introducing 3 different prices would increase total social surplus from 3439.552 to 3536.556. The grid revenue (equal to the merchandizing surplus (3-7)) would increase from –86.111 to 88.762.

Assuming that the network can be represented by a planar graph G, we construct the dual D of this graph by placing exactly one vertex in each region of G. Moreover, an edge of D intersects exactly one edge in G. In the example considered, D is given by the 5 vertices ⊗ and the 8 dotted edges in Figure 7-4.

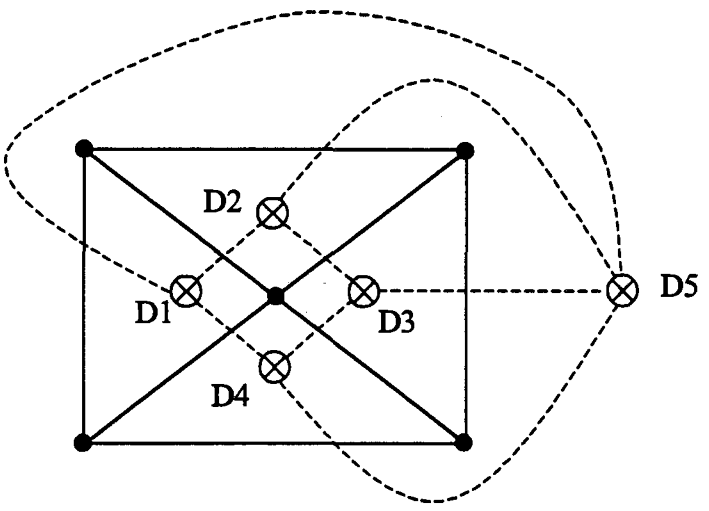


Figure 7-4 The Dual Graph

A connected zone in the original graph then corresponds uniquely to a cycle in the dual graph. For instance, the zone that contains node 1 only, corresponds to cycle D1-D2-D5-D1.

In Table 7-2 and Table 7-3 we show the results for the different zone allocations C1-C8. In addition, we show the results for the constrained (OD) and unconstrained (UD) optimal dispatch. In the first part of Table 7-2 total social surplus and grid revenue are exhibited. The next parts show the results for the different nodes, i.e. prices *p*, quantities (generation *q(s)*, consumption *q(d)*, and net injection *q*) and surpluses (*S(s)* for the producers and *S(d)* for the consumers with a total of *S* to the region as a whole). Table 7-3 shows how individual line

[45] Allowing nodes 1 and 2 to be in the same zone would add 7 more possibilities in the example. Generally, the total number of different allocations to two zones would be equal to $\sum_{i=0}^{n-2}\binom{n-1}{i}$.

flows vary with different zone allocations (a negative entry in row $i-j$ implies that power flows from node j to node i). The highest and lowest zonal surpluses are in boldface types, and so are also the maximal and minimal (zonal) line-flows.

Table 7-2 Prices, Quantities, and Surpluses

	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
Total	3606.557	3550.954	**3537.568**	3503.220	3439.552	3506.243	3424.065	3498.269	**3422.521**	3470.630
Grid	0.000	90.092	**88.741**	59.194	-86.111	-78.663	-61.359	-59.881	**-262.753**	-119.092

NODE 1	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	14.892	14.531	14.957	14.516	14.897	15.102	15.369	15.018	15.693
$q(s)$	163.934	148.925	145.311	149.569	145.160	148.972	151.020	153.689	150.184	156.933
$S(s)$	1343.725	1108.927	1055.764	1118.544	**1053.567**	1109.638	1140.359	1181.015	1127.767	**1231.395**
$q(d)$	72.131	102.151	109.378	100.862	109.681	102.055	97.959	92.622	99.631	86.134
$S(d)$	130.073	260.869	299.089	254.329	**300.746**	260.382	239.899	214.471	248.160	**185.479**
q	91.803	46.774	35.933	48.707	35.479	46.917	53.061	61.067	50.553	70.798
S	1473.797	1369.797	1354.853	1372.873	**1354.313**	1370.020	1380.259	1395.486	1375.927	**1416.873**

NODE 2	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	17.695	17.001	17.573	17.852	17.493	18.710	18.126	18.588	19.576
$q(s)$	32.787	35.391	34.001	35.145	35.703	34.985	37.420	36.252	37.175	39.152
$S(s)$	268.745	313.124	**289.025**	308.800	318.682	305.989	350.067	328.557	345.499	**383.220**
$q(d)$	72.131	46.094	59.985	48.546	42.967	50.149	25.799	37.477	28.248	8.480
$S(d)$	130.073	53.116	**89.956**	58.918	46.153	62.873	16.639	35.114	19.949	**1.798**
q	-39.344	-10.703	-25.984	-13.401	-7.263	-15.164	11.622	-1.225	8.927	30.672
S	398.818	366.240	378.981	367.718	364.835	368.862	366.706	**363.670**	365.448	**385.018**

NODE 3	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.494	17.001	14.957	17.852	17.493	15.102	15.369	18.588	15.693
$q(s)$	81.967	82.470	85.004	74.784	89.258	87.463	75.510	76.844	92.938	78.466
$S(s)$	671.862	680.138	722.562	**559.272**	796.706	764.974	570.180	590.507	**863.748**	615.697
$q(d)$	72.131	70.118	59.985	100.862	42.967	50.149	97.959	92.622	28.248	86.134
$S(d)$	130.073	122.914	89.956	**254.329**	46.153	62.873	239.899	214.471	**19.949**	185.479
q	9.836	12.352	25.018	-26.078	46.292	37.314	-22.449	-15.778	64.690	-7.668
S	801.935	803.052	812.518	813.601	842.859	827.846	810.079	804.979	**883.696**	**801.176**

NODE 4	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.894	17.001	17.573	14.516	17.493	15.102	18.126	15.018	15.693
$q(s)$	56.645	56.315	56.669	58.576	48.387	58.309	50.340	60.420	50.061	52.311
$S(s)$	447.908	475.707	481.708	514.666	**351.189**	509.982	380.120	**547.594**	375.922	410.465
$q(d)$	72.131	62.110	59.985	48.546	109.681	50.149	97.959	37.477	99.631	86.134
$S(d)$	130.073	96.441	89.956	58.918	**300.746**	62.873	239.899	**35.114**	248.160	185.479
q	-17.486	-5.795	-3.316	10.030	-61.294	8.160	-47.619	22.943	-49.570	-33.824
S	577.981	572.148	**571.664**	573.584	**651.935**	572.855	620.019	582.708	624.082	595.944

NODE 5	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
p	16.393	16.494	17.001	17.573	17.852	14.897	18.710	15.369	15.018	15.693
$q(s)$	27.322	27.490	28.335	29.288	29.753	24.829	31.183	25.615	25.031	26.155
$S(s)$	223.954	226.713	240.854	257.333	265.569	**184.940**	**291.722**	196.836	187.961	205.232
$q(d)$	72.131	70.118	59.985	48.546	42.967	102.055	25.799	92.622	99.631	56.134
$S(d)$	130.073	122.914	89.956	58.918	46.153	**260.382**	**16.639**	214.471	248.160	185.479
q	-44.809	-42.628	-31.651	-19.258	-13.214	-77.227	5.385	-67.007	-74.601	-59.979
S	354.027	349.626	330.810	316.251	311.722	**445.322**	**308.361**	411.307	436.121	390.711

Table 7-3 Line Flows under Different Zone Allocations

FLOW	UD	OD	C1	C2	C3	C4	C5	C6	C7	C8
1-2	33.515	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000
1-3	20.036	6.724	1.022	17.990	0.322	-3.132	**23.670**	14.495	**-5.439**	19.181
1-5	38.251	25.050	19.911	15.717	20.157	35.049	**14.392**	31.572	**40.992**	36.618
2-3	-13.479	-8.276	-13.978	2.990	-14.678	-18.132	**8.670**	-0.505	**-20.439**	4.181
2-4	2.914	2.523	-1.917	-2.108	17.258	-2.081	18.560	**-2.292**	18.374	**19.874**
2-5	4.736	10.050	4.911	0.717	5.157	20.049	**-0.608**	16.572	**25.992**	21.618
3-4	16.393	10.799	12.061	**-5.098**	31.936	16.051	9.890	-1.787	**38.813**	15.693
4-5	1.821	7.527	6.828	2.824	-12.101	**22.129**	**-19.168**	18.864	7.617	1.744

Variations in Total Social Surplus

As can be seen from Table 7-2, the zonal allocations show considerable variations when it comes to total social surplus. C1 is best with total surplus of 3537.568, only 13.386 below optimal dispatch (a difference of 0.376%). The poorest allocation is C7 with a surplus of 3422.521, which is 115.047 below C1 or 3.616% below optimal dispatch.

Allocation of Surplus to Individual Agents

For individual agents the outcome is heavily influenced by the allocation to zones. In the example, the greatest difference is experienced by the grid-company, which would prefer C1 with a merchandizing surplus of 88.741, which is 351.494 greater than the –262.753 of C7. For the individual producers and consumers the difference in surplus between the best and worst allocation can be several hundreds, for instance the surplus of producer 3 in C7 is 304.476 greater than the surplus attained in C2. Likewise consumers 3, 4 and 5 preferring C2, C3 and C4 respectively, will be better off by more than 200 compared to their least favorable allocation. It is also evident from the tables that between producers and consumers there is a conflict of interest, the allocation preferred by the producer is the allocation *least* favored by the consumer, and the contrary.

Based on these observations it is questionable that the grid-company shall have this power to effect the distribution of surplus among the participants in the market. Also since the selection of zone boundaries affects the surplus allocation to the grid-company, there might be a conflict of interest between the grid-company and the market participants. One way to handle this conflict of interest could be to specify a rule for selecting zone boundaries, for instance a regulation specifying that the grid-company shall select zone boundaries in order to maximize total social surplus. This would in the example mean C1. However, such a regulation is dependent on well-behaved players, i.e. suppliers and consumers must truthfully reveal to the system operator their cost and demand schedules. As pointed to by several authors (ref. section 3.1) there are strong incentives for players not to behave in this way.

If the zonal pricing system is to be based on pre-specified zones, which is now considered introduced in the Norwegian system, one could base the zone allocation on some form of bargaining mechanism based on results from a typical load flow situation. This approach is something that we are currently investigating.

Line Flows

As displayed in Table 7-3, line-flows vary greatly from one zone definition to another. In some cases lines may be heavily loaded while other allocations leave the links practically unused. In addition, the direction of the line-flows depend on which cut is considered. This may have the effect that lines that are not congested in optimal dispatch may be congested in the zonal solution, i.e. additional limitations may be introduced.

Consider for instance the case where there is a flow limit of 15 on line 2-5. This constraint does not bind, neither in the unconstrained solution nor in (the constrained) optimal dispatch. Choosing a zone definition corresponding to C4 however (or C6, C7 or C8) activates the constraint. Holding on to zone definition C4, it is not possible to find *two* zonal prices that clear the zonal markets and induce a feasible flow in the given example. Adding a third zone by separating nodes 1 and 5 solves the problem, and the partition of the network with new zonal prices is shown in Figure 7-5 part A. Due to the new constraint requiring 3 zones, social surplus has increased, also compared to C1. This also illustrates that in a system with fixed zones, it may be more difficult to find prices that resolves the capacity problems, and if not allowed to increase the number of zones, other methods for relieving congestion must be used, for instance counter purchases as described in section 7.3.

Figure 7-5 part B illustrates that the degree of improvement depends partly on the system operator being allowed to make an efficient redispatch. As is also discussed by Stoft [74], restricting the system operator to redispatch only until congestion is relieved (implying

$q_{25} = 15$) might reduce social surplus. Moreover, line-flows that are left at its limit may constitute a security threat.

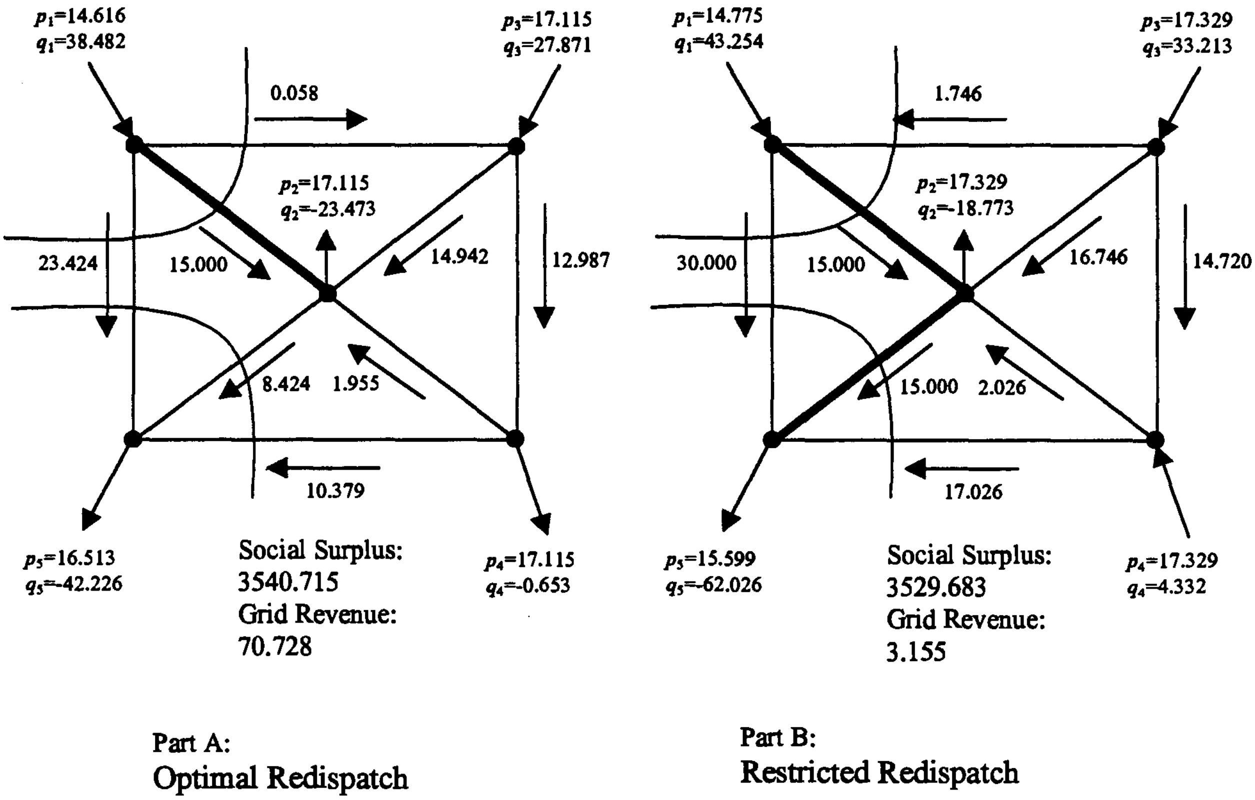


Figure 7-5 Secondary Constraints

Merchandizing Surplus

In Table 7-2 there are several examples of negative merchandizing surplus. This is closely related to line-flows varying as a consequence of choosing different zone allocations. Consider for instance C7, letting area I consist of nodes 1, 4 and 5, while nodes 2 and 3 belong to area II. In unconstrained dispatch, area I is a surplus area with a combined net injection of 29.508 which is exported to area II. In C7 however, flow over the zone-boundary from area I to area II has been reduced to –73.618, i.e. there is net flow from the high price area to the low price area, with the result that the revenue from the grid is equal to $-73.618 \cdot (18.588 - 15.018) \approx -262.816$.

Changing the parameters of the example, for instance by reducing the capacity of line 1-2 to 1 unit, gives a negative merchandizing surplus even for the best cut (which is still C1). Even if so in Table 7-2, the best cut does not necessarily give the maximal merchandizing surplus. This is illustrated in Figure 7-6 where we assume two congested lines, 1-2 with a capacity of 15 units and 4-5 with a capacity of 5 units. The figure displays optimal nodal prices as well as zonal prices in the case of three (upper part) and four (lower part) zones. Only zone allocations corresponding to maximal social surplus and maximal grid revenue are exhibited.

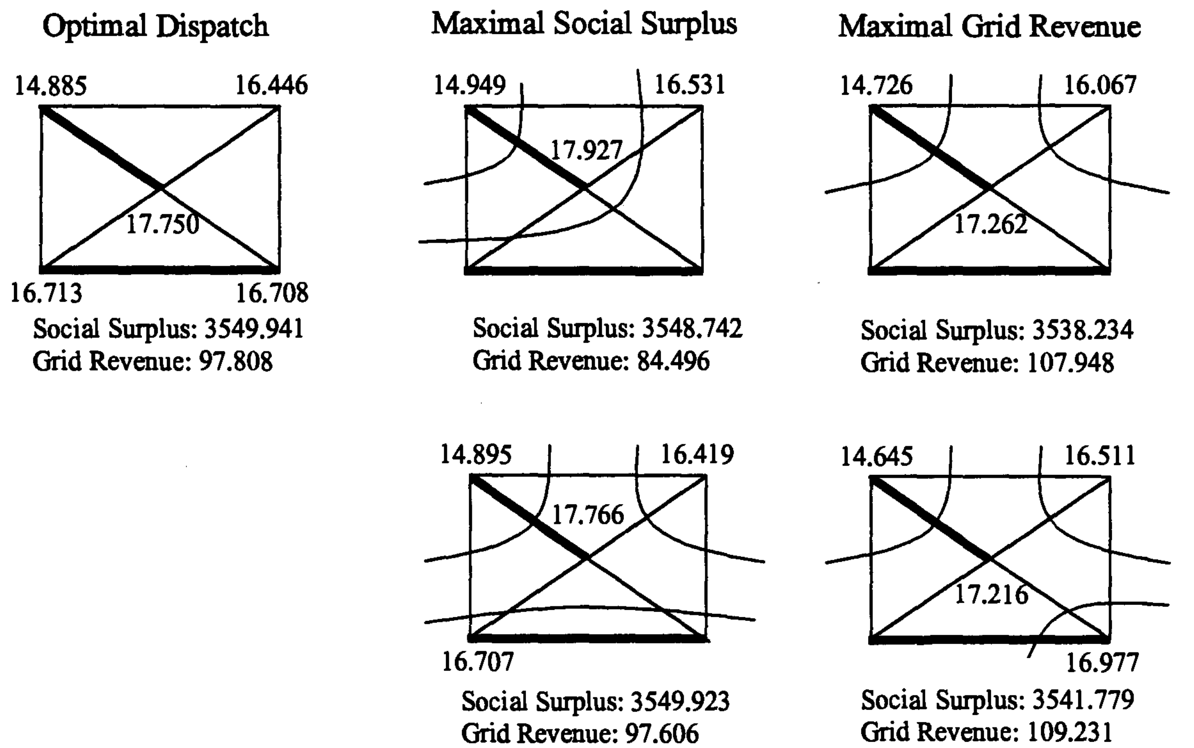


Figure 7-6 Two Congested Lines With Three and Four Zones

Practical Implementations

The results of Table 7-2 and Table 7-3 are based on optimizing the zonal prices. A question that may be raised is whether the practical procedures outlined in the beginning of the chapter will converge to these prices. We can think of two possible problems.

Adjusting Prices

An algorithm relying on price adjustments may run into problems because prices are not to be changed in the expected direction. In the example of Figure 7-7 (with input data in Table

7-4), line 1-2 has a capacity of 20 and is overloaded in unconstrained dispatch. Assuming nodes 1, 3 and 4 to be in one zone and node 2 to be in the second, node 2 is a surplus area and therefore defined to be the "low price area". Choosing the optimal zonal prices corresponding to this partition however requires node 2 to have the highest price. This implies that a procedure of adjusting prices in the supposed direction from the system price will not converge. Since the surplus area does not necessarily have the lowest price, the interpretation of Statnett on positive and negative charges is not general. However, if the validity of this interpretation is used as a criterion for choosing zone allocations, it will guarantee a positive revenue from the grid.

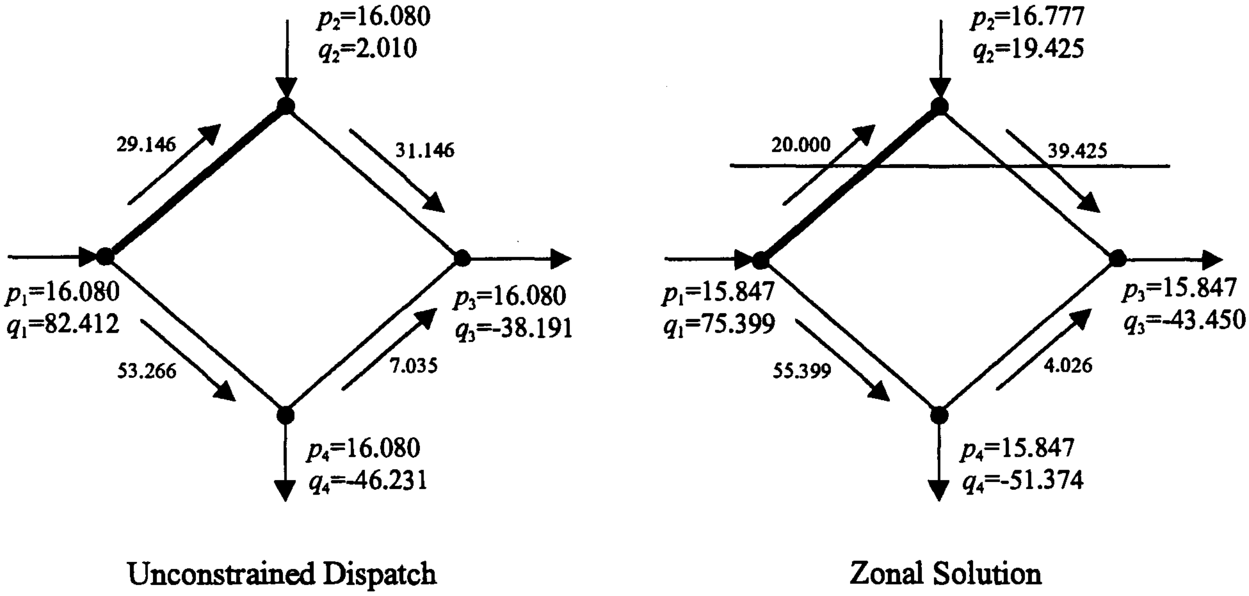


Figure 7-7 "Low Price" Becomes High Price

Table 7-4 Input Data for 4-Node Example

NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.2
3	20	0.05	0.4
4	20	0.05	0.5

Curtailing Flow

A procedure based on curtailing flow over the zone-boundary may also run into problems. If for instance a zone contains both the congested link and the nodes adjacent to it, the procedure must curtail *intra*-zonal flows to make progress. We have already stated that practical implementations typically place the endpoints of a congested line in different zones. However, as can be seen from the example of Figure 7-6, this may not be optimal. Both in the three- and four-zone optimal solutions, nodes 4 and 5 are in the same zone and the congested link 4-5 is intra-zonal.

7.3. The Correct Basis for Zone Allocations

As already mentioned in the beginning of the chapter, both Stoft [74] [76] and Walton and Tabors [84] focus on nodal price differences when evaluating zonal proposals. As the examples of Stoft clearly illustrate, if two nodes have different prices in optimal dispatch, they should in principle belong to different zones. It is however also stated that if a zonal approach renders significant simplification it is no doubt worth some loss of efficiency. The question is then how to allocate nodes to zones such that the loss of social surplus is minimal.

The statistical methods used by Walton and Tabors aim at identifying zones that should be split or combined from means and variances of optimal nodal prices within zones. More specifically, it is reported that a difference-of-the-means test is applied to examine the probability that two zonal samples are, in fact, part of a single sample. Moreover, within each zone outliers are identified (having values further than two standard deviations from the mean), i.e. Walton and Tabors are comparing nodal prices with the average nodal prices in the zones.

Returning to the example at the beginning of section 7.2, assuming line 1-2 is congested, we have varied the line capacity and changed supply and demand data such that line 1-2 is still congested[46]. In this case it seems like the best zonal division, C1, is quite robust to changes. Also, C1 corresponds to allocating nodes based on absolute price differences, i.e. placing

[46] Without being in any way exhaustive.

nodes 1 and 2 in different zones and then allocating node i to zone 1 if $|p_i - p_1| < |p_i - p_2|$, and to zone 2 otherwise.

If line 1-2 is the only congested line, it follows from nodal price theory that p_1 would be the lowest price and p_2 the highest, and that p_i can be found as a weighted average of p_1 and p_2 (Stoft [76] or Wu et al. [90]). In the "DC" approximation the exact weights are constants depending on network characteristics only (though p_1 and p_2 depend on the exact capacity and cost and benefit data). Introducing the dual price μ_{ij} of capacity on line ij, prices can be related by the Chao-Peck rule of point B) in section 3.3. Since $\mu_{12} > 0$ and $\mu_{ij} = 0\ \forall ij \neq 12$,

$$p_j = p_i + \mu_{12}\beta_{12}^{ij}$$

where β_{12}^{ij} is the load factor of line 1-2 of a trade from i to j.

In the example

$$|p_i - p_1| = p_i - p_1 = p_1 + \mu_{12}\beta_{12}^{1i} - p_1 = \mu_{12}\beta_{12}^{1i}$$

and

$$|p_i - p_2| = p_2 - p_i = p_2 - (p_2 + \mu_{12}\beta_{12}^{2i}) = -\mu_{12}\beta_{12}^{2i}$$

i.e.[47] $|p_i - p_1| < |p_i - p_2|$ if $\beta_{12}^{1i} < \beta_{12}^{i2}$. Choosing node 2 as the reference point, the condition can be written $\beta_{12}^{1} < 2\beta_{12}^{i}$, meaning that whether the price of node i is closest to p_1 or p_2 depends only on network characteristics and can be decided before any bids are received. Since $\beta_{12} = (7/15, 0, 1/5, 2/15, 1/5)$ it is easily seen that according to a rule based on smallest (absolute) price differences, nodes 3, 4 and 5 would be allocated to zone 2.

[47] Remember that $\beta_{kl}^{ij} = -\beta_{kl}^{ji}$, and with node m being the reference node, $\beta_{kl}^{ij} = \beta_{kl}^{im} - \beta_{kl}^{jm} = \beta_{kl}^{i} - \beta_{kl}^{j}$.

The interpretation is that since price differentials are based on electrical distances and influences (through distribution factors), it is natural that node 2 has a stronger influence on nodes 3, 4 and 5 than node 1 has. Similarly, if only line 4-5 is congested (in direction 4 to 5, i.e. $\mu_{45} > 0$) we find that $|p_i - p_4| < |p_i - p_5|$ if $\beta_{45}^4 + \beta_{45}^5 < 2\beta_{45}^i$ (node 2 still being the reference node). Since $\beta_{45} = (-1/15, 0, 1/15, 4/15, -4/15)$ and $\beta_{45}^4 + \beta_{45}^5 = 0$, node 1 is allocated to 5, node 3 is allocated to 4, and node 2 can be allocated to either, which is also expected since node 2 is equally "far" from both nodes 4 and 5.

Consider now both lines 1-2 and 4-5 to be congested, with $\mu_{12} > 0$ and $\mu_{45} > 0$. Now the general Chao-Peck-expression for relating prices is

$$p_j = p_i + \mu_{12}\beta_{12}^{ij} + \mu_{45}\beta_{45}^{ij}.$$

Examining the relationship between prices of nodes 1 and 2 now give $p_2 - p_1 = \mu_{12}\beta_{12}^{12} + \mu_{45}\beta_{45}^{12}$, and since $\beta_{12}^{12} > 0$ while $\beta_{45}^{12} < 0$, the size and also the sign of $p_2 - p_1$ depend on load factors *and* the size of μ_{12} and μ_{45}, and therefore on the specific input data, i.e. line capacities and cost and benefit data. By considering the other pairs of nodes, some qualitative statements can be made, and they are given in Table 7-5. For instance, the entry of row p_3 and column p_5 is <, implying that $p_3 < p_5$. A question mark indicates that the relationship cannot be decided without knowledge of shadow prices. The given relationships could possibly be used to assess zone definitions, at least in a heuristic sense.

Table 7-5 Price Relationships

	p_2	p_3	p_4	p_5
p_1	?	?	?	<
p_2		>	>	?
p_3			?	<
p_4				?

Allocating nodes to zones based on optimal nodal prices requires clustering techniques. An overview over cluster analysis is given by Hansen and Jaumard [31], in which the steps of a clustering study are discussed. In particular, different criteria for evaluating homogeneity and separation are reviewed, together with algorithms for different clustering types. In our setting, the criterion to base zone allocations on, is of special interest.

In the example of Figure 7-1 it seems useful to allocate nodes to zones based on price differentials. In general however, optimal nodal price differentials are in themselves not indicative of the best zone allocation. As exhibited in the example of Figure 7-8 the best zone allocation (Z1 or Z2) varies with the capacity of line 4-5 (all the other parameters are fixed)[48]. When the capacity is equal to 4.2, Z1 is the best partition, allocating node 5 to nodes 3 and 4. Reducing capacity by 0.1 to 4.1, Z2 is best, allocating node 5 to node 2. A capacity of 4.14858 makes Z1 and Z2 equally good when it comes to total social surplus, although, as exhibited in Table 7-6 and Figure 7-8, the allocation of surplus to individual agents vary considerably.

This switch of best zone allocation occurs even if price differentials are almost identical in the two cases. Both with capacity equal to 4.1 and 4.2, p_5 is closer to p_2 than p_3 and p_4 are, and more so when capacity is 4.1 than when capacity is 4.2. However, in both cases p_5 is closer to p_3 and p_4 (or their average) than to p_2.

[48] Note that the Z2 solutions are identical under the exhibited capacities of line 4-5, the reason being that this constraint is not binding in Z2.

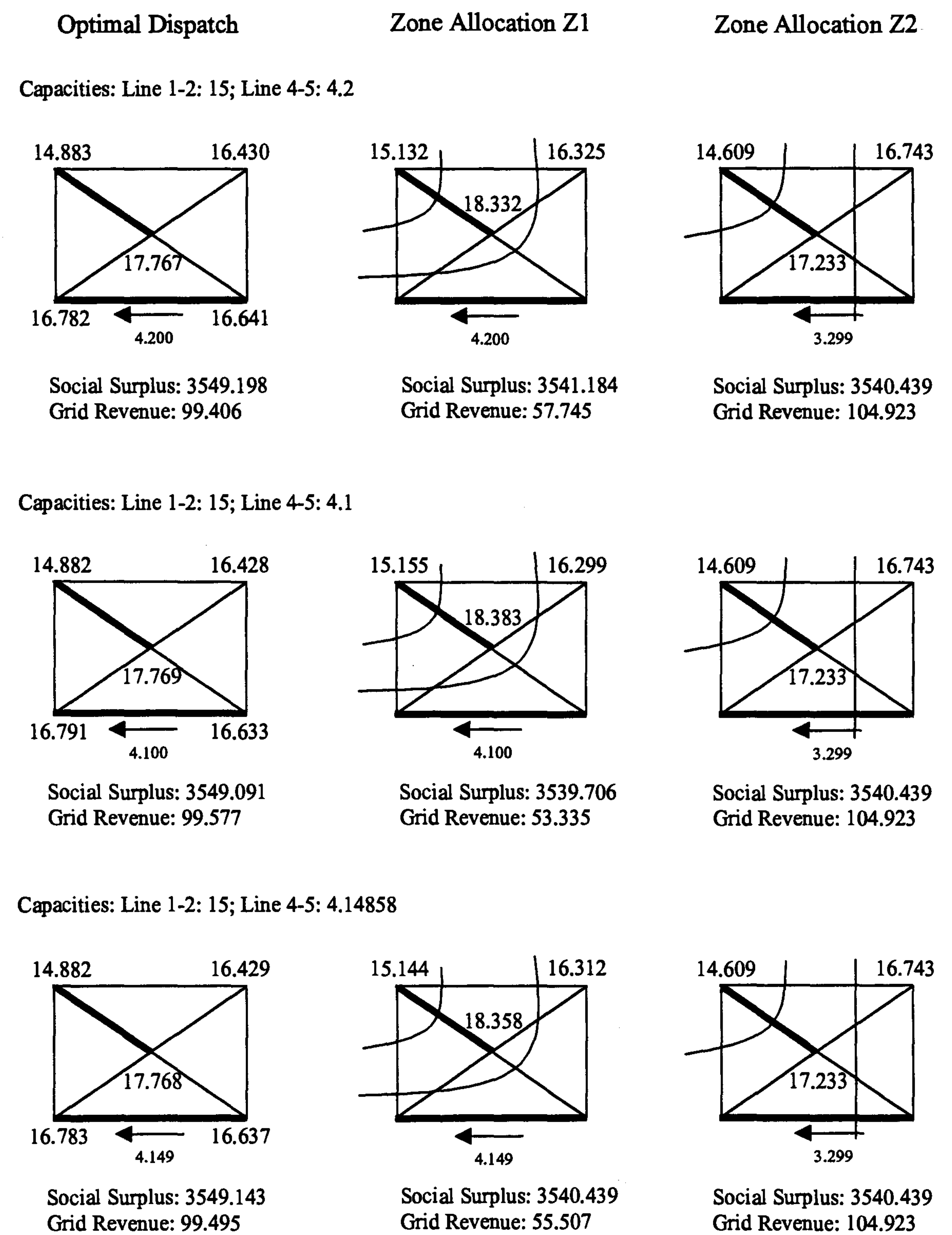


Figure 7-8 Capacity and Zone Allocation

Table 7-6 Quantities and Surpluses

Node	q(s)			q(d)			q		
	OD	Z1	Z2	OD	Z1	Z2	OD	Z1	Z2
1	148.823	151.436	146.086	102.353	97.128	107.828	46.470	54.308	38.258
2	35.536	36.716	34.466	44.642	32.841	55.345	-9.106	3.875	-20.879
3	82.146	81.558	83.715	71.417	73.766	65.140	10.728	7.792	18.575
4	55.457	54.372	55.810	67.257	73.766	65.140	-11.800	-19.394	-9.330
5	27.977	27.186	28.721	64.270	73.766	55.345	-36.293	-46.580	-26.623

Node	S(s)			S(d)			S		
	OD	Z1	Z2	OD	Z1	Z2	OD	Z1	Z2
1	1107.421	1146.642	1067.057	261.904	235.847	290.671	1369.325	1382.489	1357.728
2	315.698	337.015	296.968	49.823	26.963	76.575	365.521	363.978	373.544
3	674.791	665.177	700.819	127.511	136.037	106.082	802.302	801.214	806.900
4	461.324	443.451	467.212	113.088	136.037	106.082	574.412	579.489	573.294
5	234.822	221.726	247.474	103.266	136.037	76.575	338.088	357.763	324.049

The rationale for using price-differences when evaluating zone allocations can be illustrated by comparing the unconstrained and constrained dispatch in a two-node example with consumption in node i and production in node j, the nodes being connected by a line of limited capacity. For simplicity we still assume linear supply and demand functions.

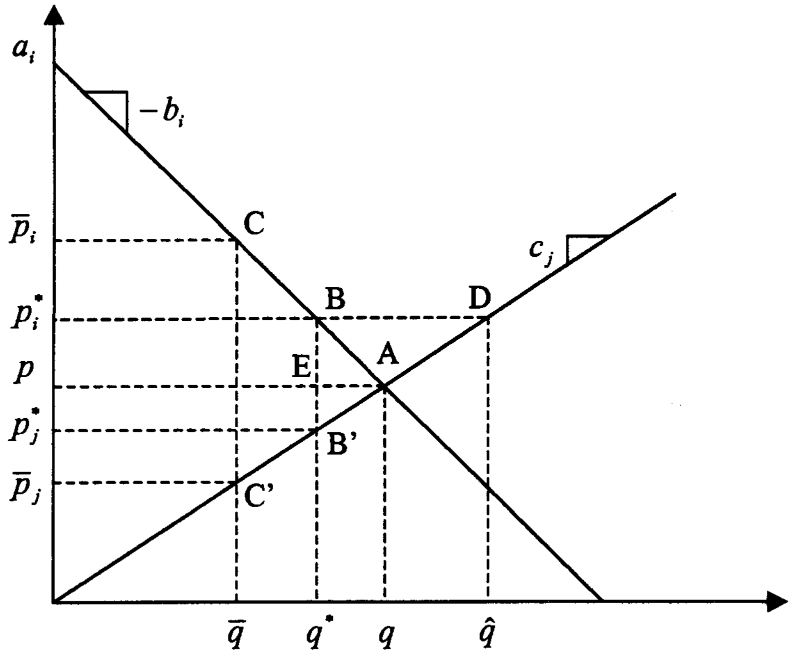


Figure 7-9 Dead-Weight Loss

In Figure 7-9 unconstrained dispatch is given by point A. Because of the limited capacity of the line connecting nodes i and j, this cannot be attained, and optimal dispatch is given by B and B', p_i^* and p_j^* being the optimal prices inducing a production of q^* equal to the capacity of the line. The reduction of social surplus resulting from the capacity limit is equal to the area of triangle ABB'. Points C and C' correspond to deviating from the optimal prices. In the single line case considered, the resulting equilibrium does not fully utilize the capacity of line *ij*. In a larger and more general network involving loop flow, the congested line could still be fully utilized even if prices are not optimal.

The reduction of social surplus, or dead-weight loss, resulting from the price error can be expressed as a function of prices. Comparing with unconstrained dispatch, the reduction corresponds to triangle ACC' and is equal to

$$
\begin{aligned}
\tfrac{1}{2}(\overline{p}_i - \overline{p}_j)(q - \overline{q}) &= \tfrac{1}{2}(\overline{p}_i - p)(q - \overline{q}) + \tfrac{1}{2}(p - \overline{p}_j)(q - \overline{q}) \\
&= \tfrac{1}{2}(\overline{p}_i - p)\left(\frac{a_i - p}{b_i} - \frac{a_i - \overline{p}_i}{b_i}\right) + \tfrac{1}{2}(p - \overline{p}_j)\left(\frac{p}{c_j} - \frac{\overline{p}_j}{c_j}\right) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p)^2 + \tfrac{1}{2c_j}(p - \overline{p}_j)^2,
\end{aligned}
$$

showing that the reduction of social surplus in a node due to the capacity limit *and* price error is proportional to the square of the difference between the (unconstrained) system price p and the prevailing price.

Alternatively we could consider the reduction in social surplus from choosing non-optimal prices only. This is equal to the area of trapezium BCC'B', i.e.

$$
\begin{aligned}
&\tfrac{1}{2}(\overline{p}_i - \overline{p}_j + p_i^* - p_j^*)(q^* - \overline{q}) \\
&= \tfrac{1}{2}(\overline{p}_i - p_i^*)(q^* - \overline{q}) + \tfrac{1}{2}(p_j^* - \overline{p}_j)(q^* - \overline{q}) + (p_i^* - p_j^*)(q^* - \overline{q}) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p_i^*)^2 + \tfrac{1}{2c_j}(p_j^* - \overline{p}_j)^2 + \tfrac{1}{2b_i}(p_i^* - p_j^*)(\overline{p}_i - p_i^*) + \tfrac{1}{2c_j}(p_i^* - p_j^*)(p_j^* - \overline{p}_j) \\
&= \tfrac{1}{2b_i}(\overline{p}_i - p_i^*)(\overline{p}_i - p_j^*) + \tfrac{1}{2c_j}(p_j^* - \overline{p}_j)(p_i^* - \overline{p}_j).
\end{aligned}
$$

This is also a function of price-differences, involving optimal nodal prices and the prevailing prices $\bar{p}_i$ and $\bar{p}_j$.

The example also illustrates why a uniform market price is not possible without rationing producers or consumers. Raising the price from p in the example reduces consumption, and at point B the total quantity demanded can be handled by the capacitated line. However, at this point suppliers prefer quantity $\hat{q}$, which is greater than q^*, and production must be curtailed or rationed by some mechanism. Using price p_j^* in node j is of course one alternative, the price constituting the rationing mechanism, but this implies that prices are no longer uniform.

Another alternative is *counter-purchases*, buying off some production by compensating producers with the difference between p_i^* and their cost of production. The cheapest way to implement this would be to compensate the costlier producers. In the given example this would imply a cost equal to the area of triangle BB'D. In general, consumers could also take part in this process, in which case optimal dispatch is attainable by transferring ABE to the least valued demand and AB'E to the most expensive supplies. Moreover, in a network involving loop flow we should take into account the effect of individual agents on the congestion considered. For producers generating counter-flows and consumers relieving congestion this could imply being compensated for *increasing* output. In general, finding the least cost redispatch involves solving an optimization problem of the same possible complexity as the optimal dispatch problem (Fang and David [22] [23] and Singh et al. [67]).

In principle, this arrangement corresponds to the Swedish system of managing congestion, where the cost of counter-purchases is recovered through the fixed network charges[49]. Also the (real time) regulation power markets of both Norway and Sweden manage congestion by redispatching based on incremental and decremental bids. The exact curtailment procedure determines the allocation of social surplus to individual agents. In the discussion above we assumed competitive markets, however, as is illustrated by Stoft [79], a counter-purchase arrangement is vulnerable to gaming.

Generally in a meshed network, possibly containing both production and consumption in each node, some agents may loose while others are better off due to price errors. However, the dead-weight loss can also in this case be expressed as a function of prices. If p_i^* is the optimal nodal price of node i and $\bar{p}_i$ is the zonal price resulting from a given zone allocation, the difference between surplus in optimal dispatch and in the zonal solution is equal to

$$\sum_i (\tfrac{1}{2b_i} + \tfrac{1}{2c_i})(\bar{p}_i + p_i^*)(\bar{p}_i - p_i^*),$$

assuming linear demand and supply functions. Each part of the expression can be positive or negative depending on the sign of $(\bar{p}_i - p_i^*)$, which again depends on the exact zone allocations that determine $\bar{p}_i$. It is far from obvious how to construct zones from this expression.

The best allocation of nodes to a given number of zones K in the presence of a capacity constraint on line kl (in direction k to l) can be formulated as a non-linear mixed integer program.

(7-7) $$\max \sum_i \sum_j \left[\tfrac{1}{2}(a_i + p_j)q_{ij}^d - \tfrac{1}{2}p_j q_{ij}^s\right]$$

(7-8) s.t. $$p_i = c_i \cdot \sum_j q_{ij}^s \qquad i = 1,\ldots,n$$

(7-9) $$p_i = a_i - b_i \cdot \sum_j q_{ij}^d \qquad i = 1,\ldots,n$$

(7-10) $$q_{ij}^s \le M_1 \delta_{ij} \qquad i = 1,\ldots,n; j = 1,\ldots,K$$

(7-11) $$q_{ij}^d \le M_2 \delta_{ij} \qquad i = 1,\ldots,n; j = 1,\ldots,K$$

(7-12) $$\sum_j \delta_{ij} = 1 \qquad i = 1,\ldots,n$$

[49] Note that attaining optimal dispatch by counter-purchases involves a *cost* on the hands of the grid-company, whereas under optimal nodal pricing a positive revenue (merchandizing surplus) is collected.

$$p_j \le (1-\delta_{ij})M_3 + p_i \qquad i=1,\dots,n; j=1,\dots,K \tag{7-13}$$

$$p_j \ge p_i - (1-\delta_{ij})M_4 \qquad i=1,\dots,n; j=1,\dots,K \tag{7-14}$$

$$\sum_i \sum_j (q_{ij}^s - q_{ij}^d) = 0 \tag{7-15}$$

$$\sum_i \beta_{kl}^i \cdot \sum_j (q_{ij}^s - q_{ij}^d) \le C_{kl} \tag{7-16}$$

$$q_{ij}^s, q_{ij}^d \ge 0 \qquad i=1,\dots,n; j=1,\dots,K \tag{7-17}$$

$$\delta_{ij} = 0/1 \qquad i=1,\dots,n; j=1,\dots,K \tag{7-18}$$

Here, δ_{ij} is a binary variable, which is equal to 1 if node i belongs to zone j and 0 otherwise. It can be interpreted as an indicator of whether node i is allocated to zonal market j. Production in node i when allocated to zone j is q_{ij}^s and total production in node i is $\sum_j q_{ij}^s$. Consumption q_{ij}^d is similar, and M_1-M_4 are arbitrarily large positive constants ("big Ms").

Assuming linear cost and demand functions, the objective function (7-7) expresses the difference between consumers' willingness to pay and the cost of production. Constraints (7-8) and (7-9) define the price in node i, (7-12) allocates each node to exactly one zone, and (7-10) and (7-11) guarantee that only q_{ij}^s and q_{ij}^d corresponding to $\delta_{ij} = 1$ can be strictly positive. (7-13) and (7-14) set $p_j = p_i$ if $\delta_{ij} = 1$, otherwise they put no restriction on the relationship between p_i and p_j. Constraint (7-15) balances total supply and demand. (7-16) is the capacity constraint, where the left hand side, multiplying load factors and net injections, is equal to the flow over link *kl* in direction from *k* to *l*. Finally, (7-17) and (7-18) specify q_{ij}^s / q_{ij}^d and δ_{ij} as non-negative and binary variables, respectively.

Due to the non-linear and discrete nature of the problem, it is difficult to solve. However, the non-linearity occurs in the objective function only. Hansen et al. [32] have studied zonal pricing in relation to facility location and developed solution methods for this related problem.

7.4. Conclusions and Future Research

From the analyses of this chapter, it is evident that

- Zonal pricing is second-best
- Zone allocations affect the surplus of individual agents, thus possibly emphasizing conflicts of interest
- Merchandizing surplus may be negative even if zone allocation is optimal
- The best partition may not have the maximal merchandizing surplus
- The best zone allocation may or may not separate the endpoints of congested lines
- Optimal nodal price differences may or may not be indicative of the best partition

It has also been demonstrated that zonal pricing is difficult if it is to be optimal. This raises the question whether a zonal approach to managing congestion is really a useful simplification of nodal pricing. It may be so if the main point of managing congestion is to obtain feasibility, or if it can be established that the disadvantages of not finding optimum is outweighed by the perceived simplicity of having only a few prices.

This chapter has also identified a number of interesting topics for future research, including developing solution methods for the non-linear mixed integer program given by (7-7)-(7-18). Moreover, there may be a need for further investigation of whether it is useful to base zone allocations on optimal nodal prices, in which case the specific clustering criteria must be identified. In this context we should look for possibilities of making judgements on the error resulting from using optimal nodal prices as the basis for allocating nodes to zones.

In Norway there has recently started a discussion on changing the current flexible zonal pricing system into a system with a few a priori determined zones. The findings in this paper indicate that it is very important to make a thorough investigation on the number of zones needed in a fixed zone system, if the fixed zones shall be the same in all load situations or different according to some pre-specified criteria. Given that a fixed zone system is to be used, there is also a need to investigate the redistribution effects the zone system has on the various market participants and take this into account when defining the fixed zones. Hence, even though a fixed zone system seems to be simpler to handle and may make it easier to develop a market for transmission capacity reservation trading, it is far from obvious that a fixed zone system would be efficient.

8. Paradoxes in Electricity Networks

In this chapter we address grid investments. In general, this question involves two main aspects, the first is that of detecting beneficial investments, and the second is how to induce them under the chosen market regime. Due to the special nature of electric networks, we will show that grid investments, that at first sight seem an improvement of the grid, may prove to be detrimental to social surplus (even without considering investment cost). Moreover, some agents will have incentives to advocate these changes.

8.1. Braess' Paradox and Generalizations

In user-optimizing traffic assignment problems where each individual user chooses the path with the lowest travel cost, it is well known that the equilibrium flow in a network is generally different from the system optimal flow, which minimizes total travel cost. In his original example, Braess [7] showed that adding a new link to a congested network may in fact *increase* travel cost for all, and this phenomenon is referred to as the Braess' paradox. Braess' paradox and variations of it have been the subject of several papers, like Murchland [54], Stewart [73], Frank [24], Dafermos and Nagurney [17], Steinberg and Zangwill [72] and Steinberg and Stone [71], among others.

More recently, Penchina [58] and Pas and Principio [57] have studied the classical Breass' traffic network configuration[50] with a single origin-destination pair and with fixed and variable user cost on the links, representing for instance travel and congestion cost respectively. Given the cost parameters, demand is varied and it is illustrated that the paradox typically occurs for intermediate traffic demand, whereas for low and high demand the additional link is beneficial. This means that when traffic demand increases over time, networks can "grow into" or "grow out from" the paradox region.

[50] In some papers, like Cohen and Horowitz [15] and Calvert and Keady [11], it is referred to as the Wheatstone bridge topology.

In relation to this, Penchina discusses different cures, including tolls and reversible one-way signs, showing that the "best" remedy depends on traffic, and although system optimum is achieved under marginal cost pricing, in some cases there is a trade-off between the optimality and complexity of the suggested cure. Similarly, Pas and Principio show that the paradox-region can be divided into two sub-ranges. In the first (for relatively lower demand) marginal cost pricing results in a flow pattern in which the additional link is used and the overall system performance is improved. In the second, marginal cost pricing results in the additional link *not* being used. This means that in this sub-range, not only will the additional link increase travel time in the user equilibrium flow pattern (Braess' paradox), the additional link is not warranted even under marginal cost pricing.

Yang and Bell [91] also study the classical Braess' network adding throughput capacities to the links and showing that at a given service level, a new link may reduce the throughput capacity of the network. Alternatively, at the same level of throughput queues may develop when the new link is introduced. The concept of reserve capacity, in the form of a flow-multiplier, is introduced as a means to detect and avoid capacity expansions that are detrimental to overall throughput capacity.

Hallefjord et al. [30] discuss paradoxes in traffic networks in the case of elastic demand. When travel demand is elastic, it is not evident what a paradoxical situation is, and in this case there is a need for characterizations of different paradoxes. An example is given where total flow decreases, while travel time increases due to adding a new link to the network. This is a rather extreme type of paradox. A different paradox is when the network "improvement" leads to a reduction in social surplus.

The reason for the traffic equilibrium paradoxes is the behavioral assumption that a traveler chooses the path that is best for himself, without paying attention to the effect this has on the other users (eventually including himself). In user equilibrium a user cannot decrease travel time by unilaterally changing his travel route, leading us to seeing the equilibrium as a Nash equilibrium of an underlying game. Korilis et al. [47] investigate the non-cooperative structure of certain networks, where the term non-cooperative emphasizes that the networks are "operated according to a decentralized control paradigm, where control decisions are made by each user independently, according to its own individual performance objectives". Nash equilibria are generally Pareto inefficient as demonstrated by Dubey [20], and Korilis et al. [47] use the Internet as an example while referring more generally to queuing networks.

Cohen and Horowitz [15] give examples of Braess' paradox for other non-cooperative networks like mechanical systems (strings) and hydraulic and electrical networks, and point to the need for specifications of conditions under which general networks behave paradoxically. This is partly provided by Calvert and Keady [11], and Korilis et al. [47] propose methods for avoiding degradation of performance when adding resources to non-cooperative networks.

In the following sections we will give examples of paradoxical situations that can occur in electrical networks due to electrons behaving "non-cooperatively". This behavior is reflected by the power flow equations. When computing the equilibria, we assume competitive markets.

8.2. Grid Investments in Electricity Networks

In Wu et al. [90] a 3-node example is given, showing that strengthening a line by increasing its admittance may lead to a larger minimum cost. The network and initial optimal dispatch is displayed in Figure 8-1 (assuming a linear lossless "DC" approximation of the power flow equations). In optimal dispatch the nodal prices will be related by $p_1 < p_2 < p_3$ since line 1-3 is congested (for an argument, see Wu et al.). When the admittance of line 2-3 is increased, the power flow equations change, and flow will increase on path 1-3-2 if injections are maintained. This will result in line 1-3 becoming overloaded and injection in node 1 must be reduced. If consumption is to be maintained, injection in node 3 must increase, leading to a larger minimum cost.

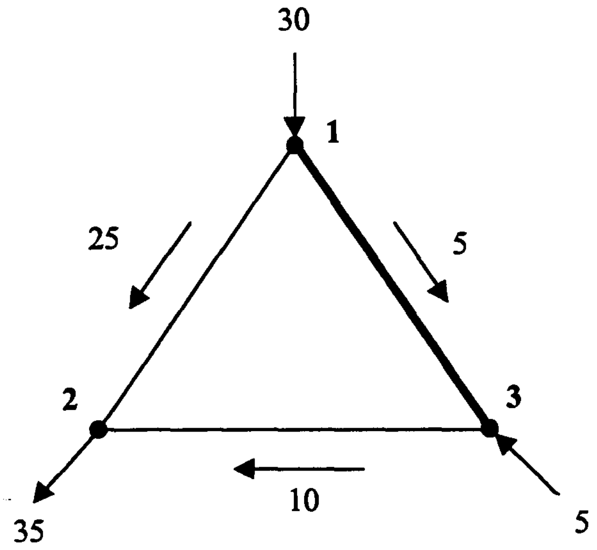
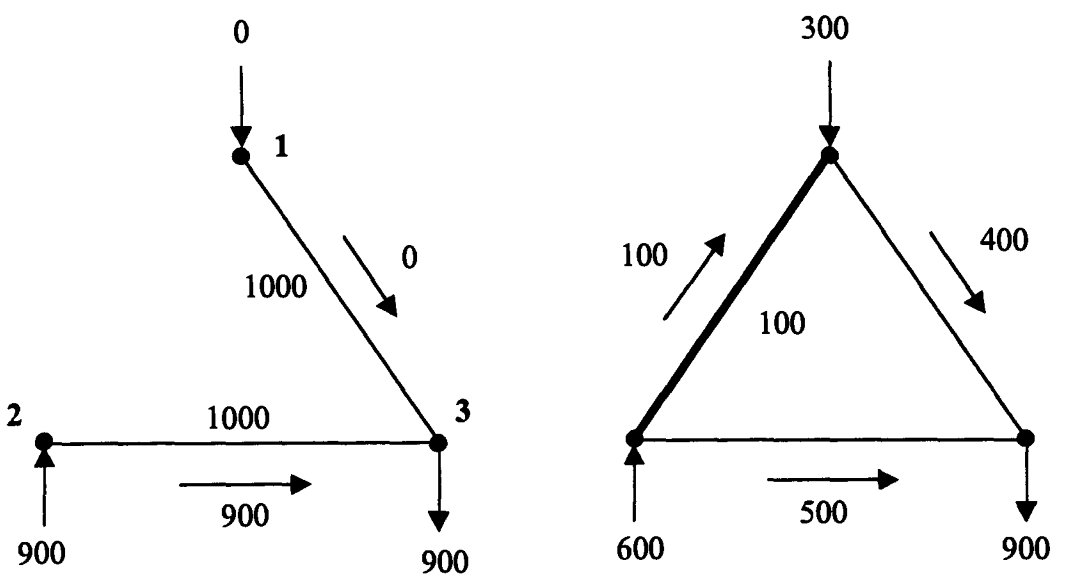


Figure 8-1 Increasing Admittance Increases Cost

In a similar 3-node example exhibited in Figure 8-2, Bushnell and Stoft [8] show that a new line hurts the network, but still collects congestion rent.



A: Initial Flows

B: Flows with the New Line

Figure 8-2 New Line Increases Cost

In this example there is high cost production in node 1 and relatively lower cost production in node 2. Consumption takes place in node 3 where there is a fixed demand equal to 900 MW. Initially, there are only two links, 1-3 and 2-3, each with a capacity of 1000 MW, and demand is supplied entirely by the low cost producers in node 2.

In part B of Figure 8-2, a new line has been built between nodes 1 and 2. This is a weak line with a capacity of only 100 MW, and it introduces loop flow, having as a consequence that the transfer capacity between nodes 2 and 3 is greatly reduced. Assuming reactances equal to 1 on every link, and no production in node 1 to generate counter flow on line 1-2, it is reduced from 1000 to 300 MW, since one third of the power injected in node 2 flows over path 2-1-3. In order to meet the demand of 900MW in node 3, injections at node 1 have to be induced, and supplying 900 MW to node 3 is obtained by injecting 600 MW in node 2 and 300 MW in node 1, which is obviously a more costly dispatch. The new line is congested in direction 2 to 1, and since $p_1 > p_2$, the new line receives congestion rent.

In the following we will give examples of paradoxes in a 4-node network with the Wheatstone bridge topology and with elastic demand and production in every node. We assume linear cost and demand functions, represented by $p_i = c_i q_i^s$ and $p_i = a_i - b_i q_i^d$ where p_i is the price in node i, q_i^s is the quantity produced in node i, q_i^d is the quantity consumed in node i, and a_i, b_i and c_i are positive constants. Net injection in node i is given by $q_i = q_i^s - q_i^d$. With input data given in Table 8-1 and a thermal capacity of 15 units on line 1-2, optimal dispatch and optimal prices are given in Figure 8-3. Part A shows the situation without line 2-4, while part B includes this line. We use a linear and lossless "DC" approximation with reactances equal to 1 one every line.

Table 8-1 Cost and Demand Parameters

NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.3
3	20	0.05	0.4
4	20	0.05	0.5

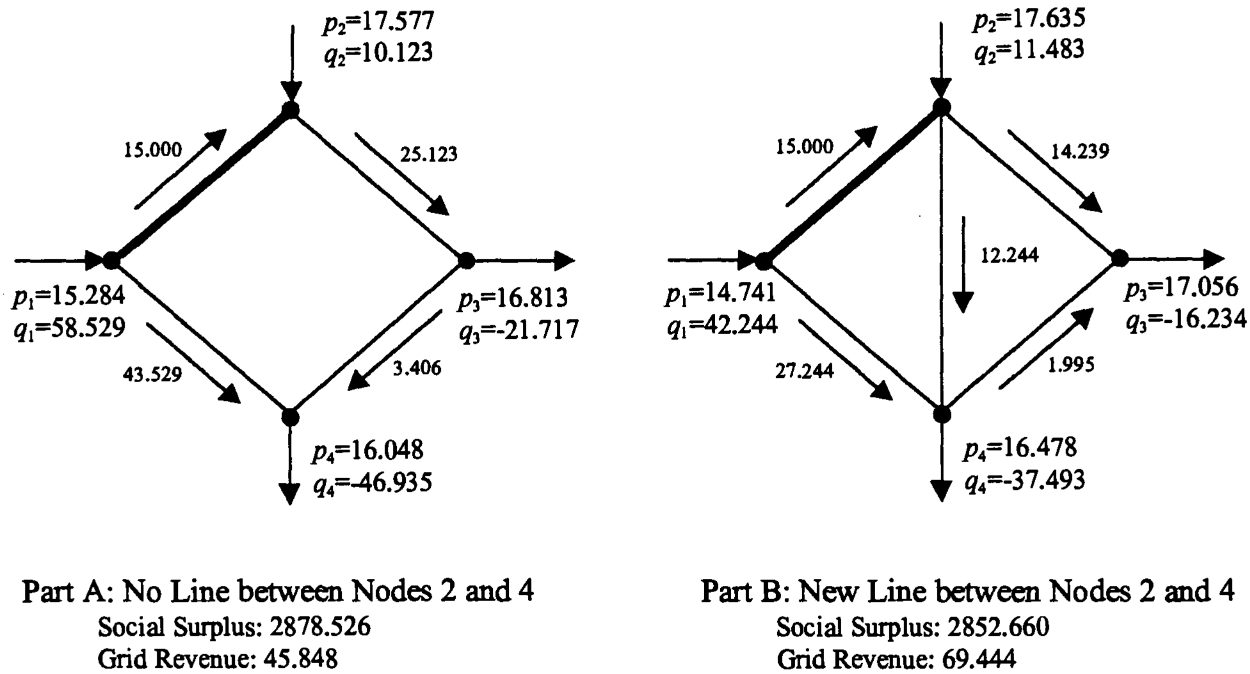


Figure 8-3 Optimal Dispatch Before and After Line 2-4

By introducing the new line, total production and consumption have been reduced together with the social surplus. On the other hand, grid revenue defined as the merchandizing surplus increases. The effect on individual agents varies, i.e. some agents loose while others are better off, as displayed in Table 8-2. A change in surplus for an agent results from a change in the nodal price that he faces. More specifically, if the price of node *i* increases as a consequence of the new line, producer *i* gains while consumer *i* looses. If the price falls, the opposite is valid.

Table 8-2 Allocation Effects of New Line

	Node 1		Node 2		Node 3		Node 4	
	Before	After	Before	After	Before	After	Before	After
Production	152.843	147.415	58.589	58.783	42.031	42.641	32.097	32.955
Consumption	94.314	105.171	48.466	47.300	63.749	58.874	79.031	70.448
Net Exports	58.529	42.244	10.123	11.483	-21.717	-16.234	-46.935	-37.493
Producer Surplus	1168.048	1086.554	514.901	518.321	353.328	363.646	257.552	271.511
Consumer Surplus	222.379	276.522	58.724	55.933	101.597	86.655	156.149	124.075
Surplus of Region	1390.427	1363.076	573.624	574.254	454.925	450.301	413.701	395.585

Considering the surplus of each region (i.e. the combined producer and consumer surpluses of each node), it is evident that in general, some regions are better off due to the new line while others loose. However, it is not difficult to construct examples in which every region looses because of the new line. For instance, changing the example above by letting $c_2 = 0.37$, makes every region worse off, while the grid revenue still increases when line 2-4 is introduced.

In the discussion so far, we have considered optimal nodal pricing as the means of managing congestion. Zonal pricing constitutes an alternative, which is treated in chapter 7. In the given example, assuming only two zones, there are four zone allocations that separate nodes 1 and 2. These are displayed in Figure 8-4. Different zone allocations affect social surplus, and for the parameters of our example, Z4 maximizes social surplus without the new line, while Z1 is best when the new line is included. This illustrates that modifications to the grid should lead to a reconsideration of zone allocations.

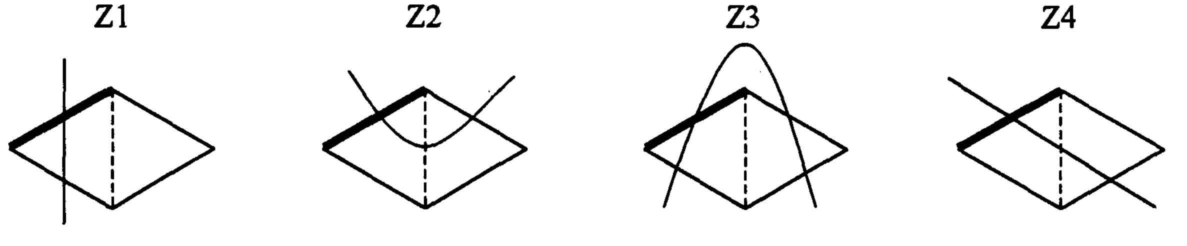
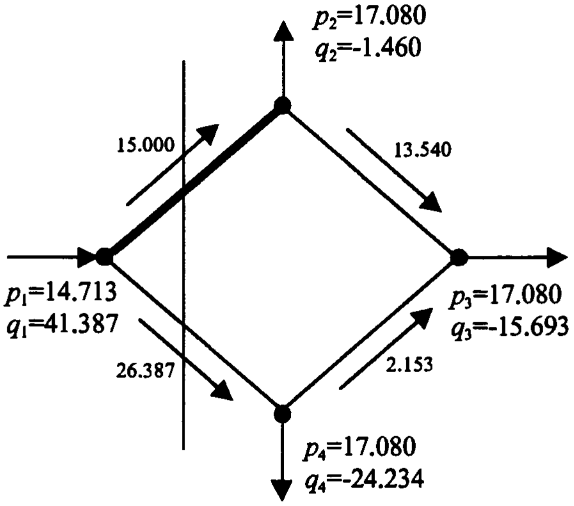


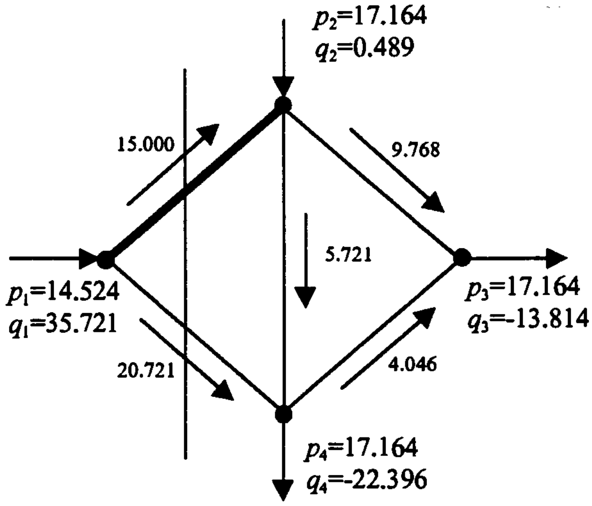
Figure 8-4 Allocations to Two Zones

Prices, net injections and power flows for Z1 and Z4 are displayed in Figure 8-5, together with total social surplus and grid revenue. As is evident from the numbers, also under zonal pricing total social surplus is reduced when the new line is built. This is so for fixed zone allocations (where the partition of nodes into zones remains the same after the new line is in place), but it is also valid even if the best zone allocation is chosen at every point. For fixed zone allocations grid revenue is reduced when building the new line. However, if the new line changes the partition of nodes from Z4 to Z1, grid revenue increases considerably, thus providing a strong incentive on the part of the grid owners to build the line.

Z1

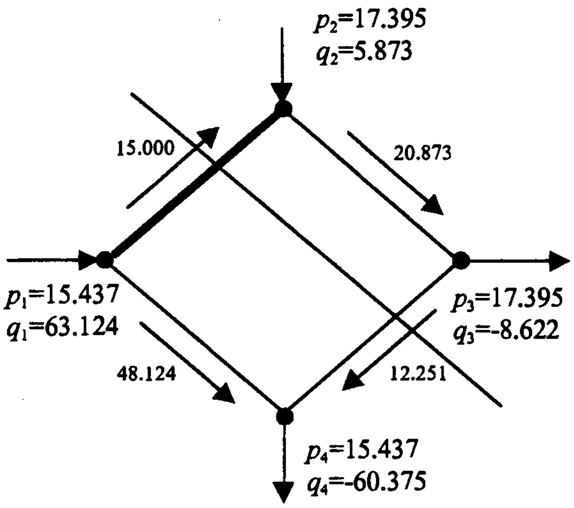


Part A: No Line Between Nodes 2 and 4
Social Surplus: 2858.235
Grid Revenue: 97.979

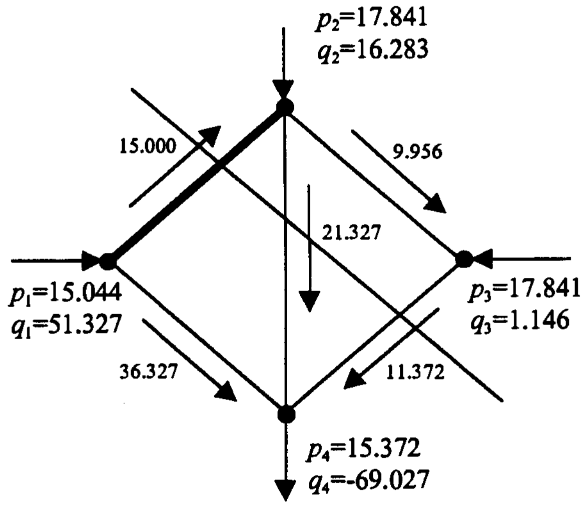


Part B: New Line Between Nodes 2 and 4
Social Surplus: 2844.051
Grid Revenue: 94.296

Z4



Part C: No Line Between Nodes 2 and 4
Social Surplus: 2869.871
Grid Revenue: 5.380



Part D: New Line Between Nodes 2 and 4
Social Surplus: 2821.270
Grid Revenue: -49.495

Figure 8-5 Zonal Solutions Z1 and Z4 Before and After New Line

In Table 8-3 we show the surpluses for each region. In general, the change of surplus for individual agents can be positive or negative. In Z1 every region surplus as well as the grid revenue decreases due to the new line. If parameters are changed so that $c_2 = 0.35$ and the thermal capacity of line 1-2 is 5 units, the effect of the new line on every region would be negative when choosing the socially beneficial zone allocations (i.e. switching from Z4 to Z1 when building line 2-4). Grid revenue on the other hand would increase.

Table 8-3 Region Surpluses

	Z1		Z4	
	Before	After	Before	After
Node 1	1361.881	1354.600	1399.745	1377.242
Node 2	571.474	571.434	572.168	577.110
Node 3	449.917	448.685	446.096	444.489
Node 4	376.983	375.036	446.482	471.924

In the examples cited so far, the reductions in social surplus are relatively minor. In the original example in Table 8-1 the reduction in total social surplus is equal to 25.866, or 0.9%. This is partly due to the assumption of identical demand functions in every node. By allowing more unequal distributions of consumption, the reductions can be of considerable size. For instance, increasing $b_i, i = 1,2$ to 0.25, i.e. the size of the markets in nodes 1 and 2 are assumed to be only 20% of the markets in nodes 3 and 4, social surplus in optimal dispatch is reduced from 2541.968 to 2394.397, i.e. by 5.8%, when the new line is built. This is more than 2.5 times the cost of the thermal limit itself, as social surplus in unconstrained dispatch is equal to 2600.506. If there is no consumption in nodes 1 and 2, social surplus is reduced from 2395.869 to 2129.125 (i.e. by 11.1%). Also when increasing demand by shifting the demand curves positively (for instance by raising the a_i's), the paradox becomes more severe.

The persistence of the paradox also depends on the cost parameters. Consider for instance varying c_2. When $c_2 \in [0, 0.080)$ the new line improves social surplus. When $c_2 \in [0.080, 0.102)$ the new line has no effect on social surplus because the thermal limit is not binding in optimal dispatch (neither with or without line 2-4). Finally, when $c_2 \geq 0.102$ the new line reduces social surplus, implying that the paradox also occurs when production in node 2 is so costly that it is not being used. The reduction of social surplus reaches a maximal value at $c_2 = 0.350$. Varying c_4 in the same manner, the thermal capacity is binding for all values of c_4. When $c_4 < 0.179$ line 2-4 improves social surplus, whereas the paradox arises for $c_4 > 0.179$.

From the treatment of the "DC"-approximation in section 2.2, we know that

$$q_{ij} = \frac{1}{x_{ij}} \sin(\delta_i - \delta_j) = Y_{ij} \sin(\delta_i - \delta_j)$$

where δ_i is the phase angle at node *i*, q_{ij} is the power flow over line *ij*, x_{ij} is the reactance of line *ij*, and the admittance Y_{ij} of line *ij* is equal to the reciprocal of the reactance of the line. Since the sine function has a maximal value of 1, we must have that $q_{ij} \le Y_{ij}$. Considering also the thermal limit C_{ij} of line *ij*, q_{ij} is bounded by $\min\{C_{ij}, Y_{ij}\}$. This means that "strengthening" a line has two interpretations: increasing the admittance or increasing the thermal limit.

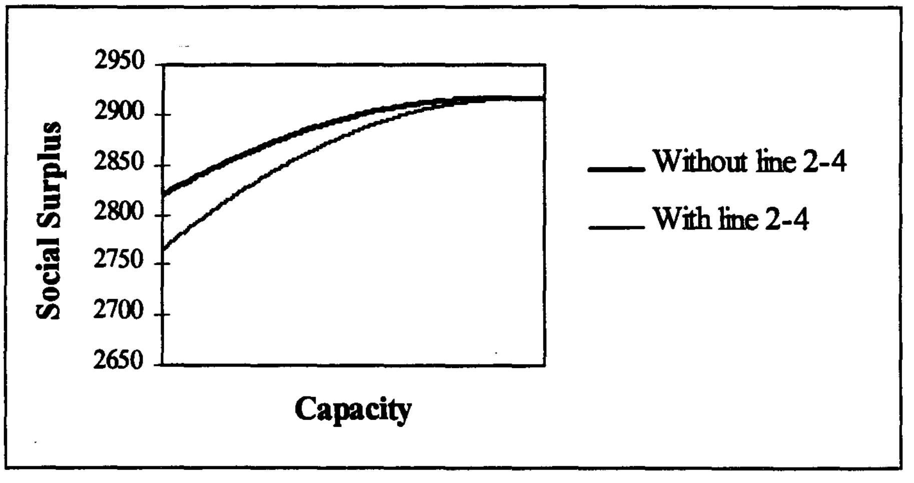
From the optimal dispatch problem in section 3.1, we know that the shadow price of the thermal limit is non-negative, i.e. $\mu_{ij} \ge 0$, which means that social surplus cannot be reduced by improving the thermal limit of any line. What we have shown by the previous examples is that whenever there is at least one binding thermal limit, say on line *ij*,

$$\frac{\partial \text{Social Surplus}}{\partial Y_{kl}}$$

may be negative for some line *kl*. I.e. by either increasing the admittance of an existing line or by building a new line[51], we may reduce social surplus.

Consider now varying the thermal capacity of line 1-2. In Diagram 8-1 social surpluses are shown as functions of C_{12}. The functions are concave and increasing, and the difference between the curves is the greatest for $C_{12} = \varepsilon$ and vanishes when C_{12} is so large that the thermal limit is no longer binding in any of the network configurations considered. This occurs at $C_{12} = 42.587$, which is the flow over line 1-2 in unconstrained dispatch assuming line 2-4 is included in the network. From this point, social surplus is constant and equal to 2916.525, and increasing the thermal capacity is not beneficial in either network configuration.

Diagram 8-1 Social Surplus and Thermal Capacity of Line 1-2



As is shown by Wu et al. [90], in optimal dispatch, the merchandizing surplus given by $\frac{1}{2}\sum_i\sum_j (p_j - p_i)q_{ij}$ is equal to the *congestion rent* defined by

$$CR = \sum_i\sum_j \mu_{ij}C_{ij}.$$

Since line 1-2 is the only congested line in our example[52], grid revenue is equal to $\mu_{12}C_{12}$, i.e. for a given thermal capacity C_{12}, the size of the grid revenue is determined by the value of

$$\mu_{12} = \frac{\partial \text{Social Surplus}}{\partial C_{12}}.$$

[51] I.e. increasing the admittance from the 0 level.

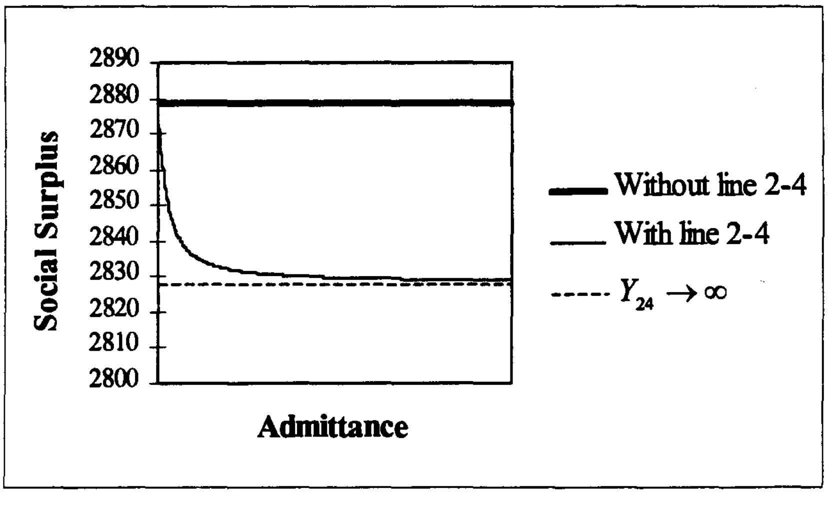
[52] Assuming $\mu_{12} > 0$ while $\mu_{ij} = 0$ for $ij \neq 12$.

As is indicated by the curves of Diagram 8-1, building line 2-4 will increase grid revenue since at every $C_{12} < 42.587$ the social surplus function *with* line 2-4 is steeper than the function depicting social surplus *without* line 2-4.

Note however that whether the grid revenue increases due to the new line is not indicative of whether the paradox occurs. Grid revenue may increase also when the new line is beneficial. For instance, letting $c_4 = 0.15$, total social surplus increases from 3448.992 to 3457.022 when the new line is built. Grid revenue increases from 58.969 to 64.530, i.e. total social surplus increases more than the grid revenue, leaving a net increase in benefit for the market participants due to the new line.

In Diagram 8-2 social surplus is shown as a function of the admittance of line 2-4. For reference, social surplus without line 2-4 (i.e. $Y_{24} = 0$) is also exhibited. We note that the difference between social surplus with and without line 2-4 increases with the admittance Y_{24}. When $Y_{24} \to \infty$, social surplus approaches the value 2828.161 asymptotically, signifying that the paradox becomes more severe the stronger is the new line, with a maximal degradation of social surplus equal to 2878.526-2828.161 = 50.365.

Diagram 8-2 Social Surplus and Admittance of Line 2-4



The load factors of line 1-2 for different trades can be expressed as functions of Y_{24}. When the new line is introduced with an admittance of Y_{24}, the power flow equations become the following (refer section 2.2, which includes an overview of the linear "DC" power flow approximation):

$$
\begin{aligned}
&\text{Kirchhoff's junction rules:} \quad && q_1 = q_{12} + q_{14} \\
& && q_2 = -q_{12} + q_{23} + q_{24} \\
& && q_3 = -q_{23} + q_{34} \\
&\text{Kirchhoff's loop rules:} && q_{24} = -Y_{24} q_{12} + Y_{24} q_{14} \\
& && q_{24} = Y_{24} q_{23} + Y_{24} q_{34} \\
&\text{Conservation of energy:} && q_1 + q_2 + q_3 + q_4 = 0.
\end{aligned}
$$

By solving the power flow equations for different trades, we find the load factor matrix

$$
\mathbf{B}_{12}^{Y_{24}} = \begin{pmatrix}
0 & \dfrac{3+2Y_{24}}{4(1+Y_{24})} & \dfrac{1}{2} & \dfrac{1+2Y_{24}}{4(1+Y_{24})} \\
-\dfrac{3+2Y_{24}}{4(1+Y_{24})} & 0 & -\dfrac{1}{4(1+Y_{24})} & -\dfrac{1}{2(1+Y_{24})} \\
-\dfrac{1}{2} & \dfrac{1}{4(1+Y_{24})} & 0 & -\dfrac{1}{4(1+Y_{24})} \\
-\dfrac{1+2Y_{24}}{4(1+Y_{24})} & \dfrac{1}{2(1+Y_{24})} & \dfrac{1}{4(1+Y_{24})} & 0
\end{pmatrix}
$$

where the entry of row k and column l is β_{12}^{kl}, which is the load factor of a trade from node k to node l on line 1-2 (in direction from 1 to 2). In the linear "DC" approximation, load factors are constants for given admittances, and $\beta_{ij}^{kl} = -\beta_{ij}^{lk}$. The negative numbers indicate that the corresponding trades generate counter flows on line 1-2.

When $Y_{24} \to \infty$, trades between nodes 2, 3 and 4 have no influence on line 1-2, which can be seen from

$$\mathbf{B}_{12}^{\infty} = \lim_{Y_{24}\to\infty} \mathbf{B}_{12}^{Y_{24}} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 \end{pmatrix}.$$

Nodes 2, 3 and 4 thus become one market with identical nodal prices. Net injection in node 1 on the other hand, distributes equally on lines 1-2 and 1-4 (load factors are equal to $\frac{1}{2}$), implying that the maximal export from region 1 is equal to $30 = 2 \cdot C_{12}$. An interpretation of this situation is that nodes 2 and 4 are electrically "the same", which is similar to a cost of zero on line 2-4 in a traffic equilibrium network. In the case of our electrical network, this makes the paradox maximal.

The paradox of the example of Table 8-1 and Figure 8-3 can be interpreted in terms of the load factors. The load factor matrix without line 2-4 is equal to

$$\mathbf{B}_{12}^{0} = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ -\frac{3}{4} & 0 & -\frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix},$$

whereas the load factor matrix *with* line 2-4 (with admittance equal to 1) is equal to

$$\mathbf{B}_{12}^{1} = \begin{pmatrix} 0 & \frac{5}{8} & \frac{1}{2} & \frac{3}{8} \\ -\frac{5}{8} & 0 & -\frac{1}{8} & -\frac{1}{4} \\ -\frac{1}{2} & \frac{1}{8} & 0 & -\frac{1}{8} \\ -\frac{3}{8} & \frac{1}{4} & \frac{1}{8} & 0 \end{pmatrix}.$$

Considering optimal dispatch without line 2-4, $q_1 = 58.529$, $q_2 = 10.123$, $q_3 = -21.717$ and $q_4 = -46.935$. As is evident from matrices $\mathbf{B}_{12}^{0}$ and $\mathbf{B}_{12}^{1}$, the load factors of trades between net injection and net consumption nodes have developed unfavorably when introducing line 2-4. The positive load factors β_{12}^{13} and β_{12}^{14} stay the same or increase, meaning that the corresponding trades use the same or more of the capacity of line 1-2 under the new network configuration. The negative load factors β_{12}^{23} and β_{12}^{24} have increased, indicating that the trades that they represent, produce smaller counter flows on line 1-2, thus relieving the capacity constraint to a lesser extent. Under the new network configuration, the injection vector (58.529, 10.123, -21.717, -46.935) is no longer feasible. According to the characterization used by Bushnell and Stoft [9], the old dispatch belongs to the "newly infeasible region", and the "newly feasible" region that follows from the new line, provides no better dispatch, thus the paradox.

8.3. Market Integration

A consequence of the paradoxical characteristics of certain electricity networks is that in the presence of congestion constraints, social surplus can be reduced when markets are integrated. In Figure 8-6 market 1 consists of nodes 1, 2 and 3 while market 2 consists of nodes 4 and 5. We assume linear cost and demand functions, with parameters given in Table 8-4. We want to consider integrating the markets by building lines 2-4 and 3-5. Disregarding any thermal constraints we find that social surplus would increase from 3126.177 to 3157.895. The system price settles on 16.842, which is higher than the price of market 1 and lower than the price of market 2.

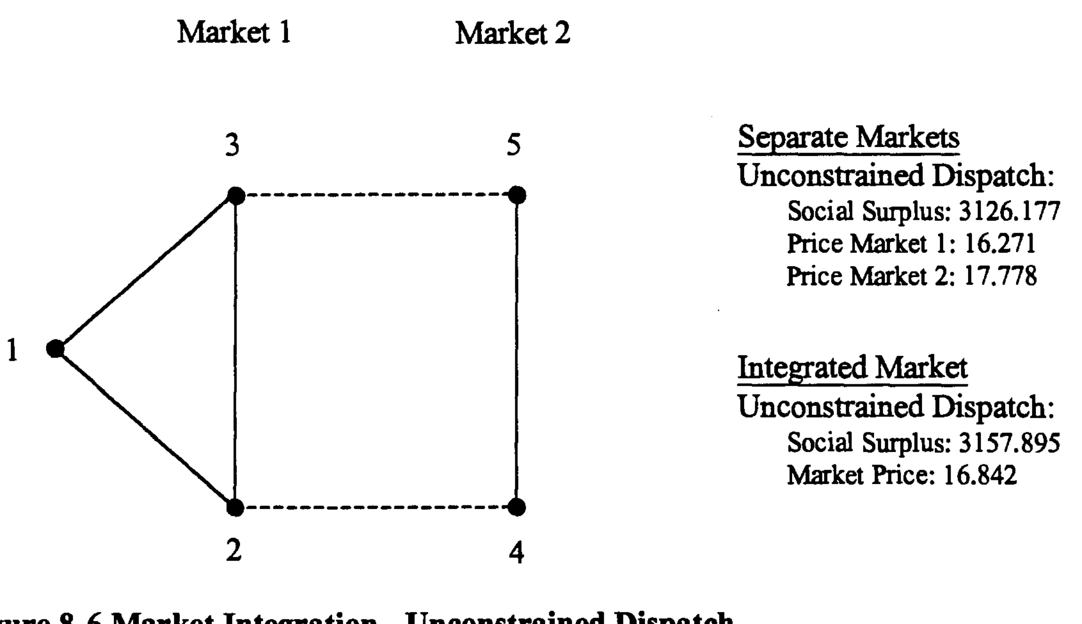


Figure 8-6 Market Integration - Unconstrained Dispatch

Table 8-4 Cost and Demand Parameters

NODE	CONSUMPTION		PRODUCTION
	a_i	b_i	c_i
1	20	0.05	0.1
2	20	0.05	0.8
3	20	0.05	0.4
4	20	0.05	0.6
5	20	0.05	0.3

Assume now there is a capacity limit of 10 units on line 1-2. In Figure 8-7 we show optimal dispatch without the connecting lines. Social surplus is equal to 3000.433. In Figure 8-8 the new lines have been built, and social surplus is reduced to 2988.241, implying that the thermal limit on line 1-2, which is internal to market 1, prevents the realization of potential benefits from market integration.

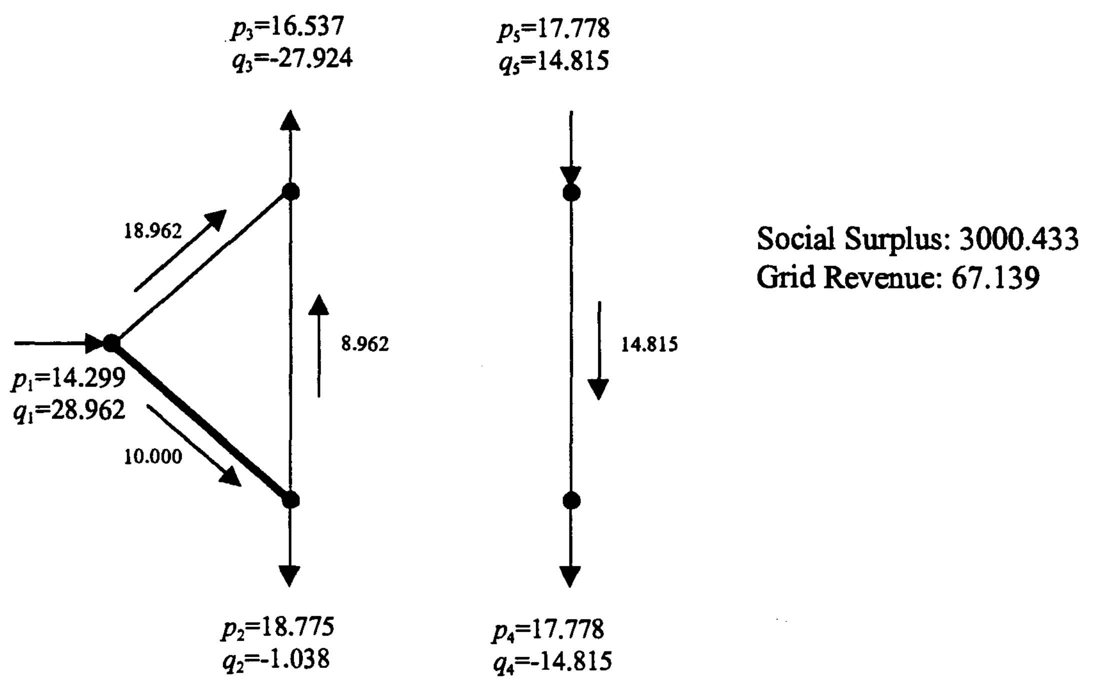


Figure 8-7 Optimal Dispatch - Before Integration

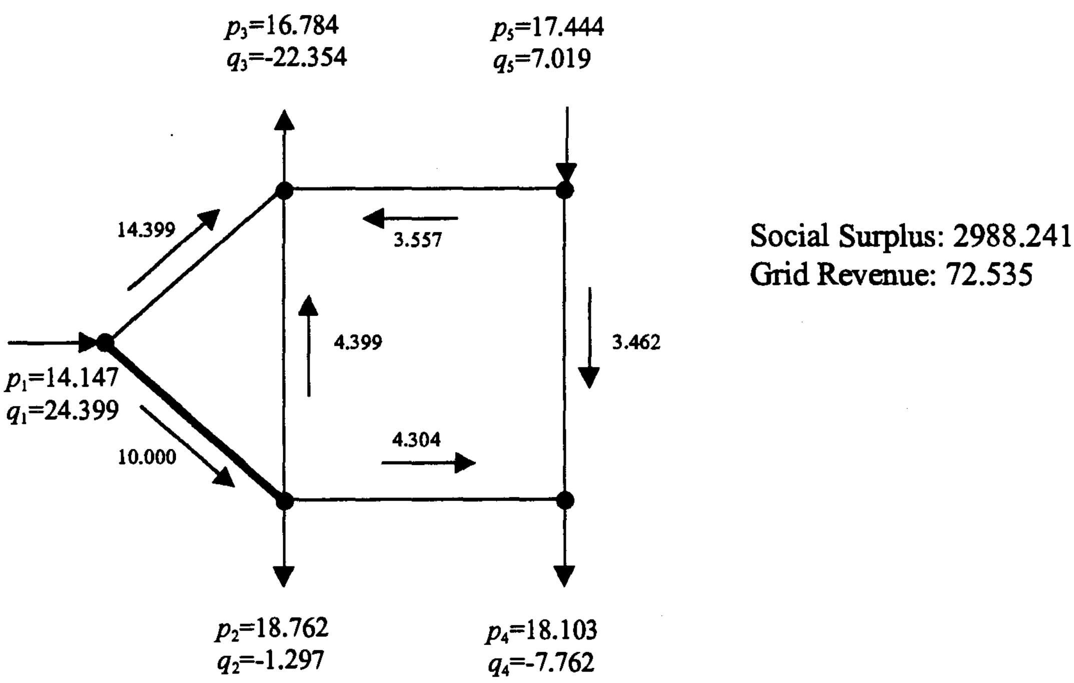


Figure 8-8 Optimal Dispatch - After Integration

8.4. Suggested Cures

Given that an investment has already been carried out, in traffic equilibrium networks marginal cost pricing can lead to improved overall system performance from the grid modification even when Braess' paradox occurs in user equilibrium (Pas and Principio [57]). In electricity networks there is no equivalent methodology, since electrons do not respond to marginal cost pricing. To alter flows for a given set of injections, we would have to alter line impedances.

Considering the investment decision itself, the obvious way to avoid the paradox in the example of Figure 8-3 is to build line 1-3 instead of line 2-4. This would resolve the capacity problem of line 1-2, but may be unacceptable for other reasons, for instance investment cost. Generally, the issue of how to encourage beneficial investments and discourage detrimental investments has been treated in the literature, for instance by Baldick and Kahn [3], Bushnell and Stoft [8] [9] [10] and Hogan [38]. As is also discussed in section 3.2, transmission congestion contracts (TCCs), where new contracts are allocated according to a feasibility rule, thereby internalizing the externality effects of detrimental grid investments, can provide at least a partial solution.

However, as is demonstrated by some of the examples in this chapter, and also pointed to by Bushnell and Stoft [9], the performance of a network depends on expected dispatch. This is influenced by future supply and demand conditions, which are constantly changing and subject to uncertainty. Thus, as market conditions change, so can the performance of the different network configurations considered. This is further complicated by typically long asset lifetimes and the lumpiness of the investment decisions, which sometimes make it desirable to expand the network in a manner that is not immediately beneficial but will be so in the long run. Ideally, we should compare different expansion *paths* rather than various fixed networks, as the investment problem is dynamic in nature.

8.5. Conclusions

Depending on the parameters of the problem considered (cost, demand, thermal capacity, and admittance) a new line may be detrimental to social surplus. In general, some agents are better off while others loose. In this chapter we provide examples where, in optimal dispatch, every region looses while the grid revenue increases. For fixed zone allocations there is the possibility that every region-surplus *and* grid revenue is reduced as a consequence of a new line.

To explain the paradox, we distinguish between strengthening the grid by improving thermal capacities and by improving line admittances. We also provide an explanation of the paradox in terms of the effect on load factors. Finally, it is demonstrated that a thermal limit, which is internal to a market, may result in market integration being disadvantageous. The possibility of such paradoxical effects and the incentives that they provide to different agents must clearly be taken into consideration both in the process of grid development and market development.

9. Competitive Effects of Congestion

In the electricity market there are (at least) two foundations for agents to exercise horizontal market power[53]. Firstly, some agents have a size that gives them a recognizable impact on the market price even in the complete market. Secondly, congested lines isolate parts of the network from competition, thus possibly increasing market power for agents with a favorable location compared to the transmission constraint. In this chapter we will consider this second source of market power.

Our starting point is the work of Stoft [75] and Borenstein et al. [6], which consider a single line network with a number of consumers taking price as given and a single producer in both endpoint nodes. The line connecting the two nodes has a thermal capacity of k units. In Stoft [75] both symmetric and asymmetric local markets are considered, and a number of lessons can be learnt even from these simple examples, including

- *Unused capacity may be needed*: For a line to support full competition, which in this case means duopoly, it may need to have a capacity that is much greater than the flow that will take place on it.
- *Increasing capacity is more effective on a small line*: If connecting two buses with a very strong line will reduce market power, then the first MW of connecting capacity will have the most impact and each additional MW will have less.
- *A congested line will cut a market into two non-competing regions*: In each region the generators will markup according to the elasticity of the demand in only their region.
- *A generator may reduce output in order to congest a line and thereby increase its market power*: This occurs when the line is large enough to support the duopoly line flow, but not large enough to stabilize the duopoly equilibrium.

The outcome of the simple market models depends on the size of k. Varying k from zero to infinity, the general pattern is that an interval $[0,k_1)$ supporting monopolies in the two nodes is followed by an interval $[k_1,k_2)$ with unstable prices, requiring mixed strategy solutions. This again is followed by values $k \geq k_2$ supporting full competition and a duopoly solution.

When finding the market equilibria, it is assumed that a nodal pricing regime applies, however other mechanisms that handle congestion efficiently may be used, including regulation or Chao-Peck prices as described in section 3.3. Moreover, it is assumed that competition takes place according to a Cournot model, i.e. producers choose quantities as their strategic variable and maximize profits given the quantity outputs of their competitors.

It has been questioned whether the Cournot model is appropriate, and alternative assumptions are investigated in some articles. For instance, Green and Newbery [27] assume supply curve bidding according to Klemperer and Meyer [44] while studying the British electricity spot market, as do Rudkevich et al. [60] for the case of the US market. Also von der Fehr and Harbord [83] study the British market while taking into account the stepwise nature of the supply function bids.

The consequences of loop flow are considered in Hogan [37], where the effect of a producer having plants on both sides of a congested link is investigated. In the recently published paper by Younes and Ilic [92], a three-node network is examined. Consumption is located in one node and is supplied by two different producers located in the other nodes. The agents are connected by a grid, of which various topologies are considered, including a triangle network with different links being congested. In this work, Bertrand competition is assumed, i.e. price is the strategic variable of the producers. Again market outcomes depend on the size of the thermal capacity and which link is congested. Also the triangle grid involving loop flow is applied in Oren [55] and Stoft [80] when investigating the effect of transmission rights.

9.1. Example with Three Identical Local Markets

As already mentioned we use the symmetrical 2-node/single-line examples of Stoft [75] as a starting point, with the same assumptions of nodal pricing and Cournot competition. Compared to this work, we will consider the effect of loop flow in a 3-node/3-line network with identical local markets in every node and a congested link between two nodes. Also we

---

[53] Market power may also work vertically, as generators own part of the transmission grid. This is a topic that has received much attention in the process of deregulation. It will not be considered in this work.

assume a linear and loss-less "DC" approximation of the network, considering real power only. In every market there is a single producer characterized by a marginal production cost of 1. In addition, every market has a group of consumers acting as price takers, the demand function being given by $p = 10 - q$, where p denotes price and q quantity.

Situation 1: No-Line Monopoly

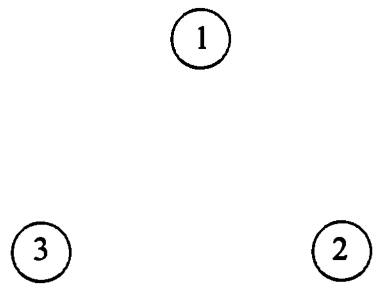


Figure 9-1 No Grid

Without a grid, there will be a monopoly prevailing in each market. The producers will maximize profit π and market equilibrium in market i is found by solving

$$\textbf{(9-1)} \qquad \max_{q_i} \pi_i = p_i \cdot q_i - TC_i = (10 - q_i) \cdot q_i - 1 \cdot q_i,$$

where TC_i is the cost of producer i. The first order condition is

$$\textbf{(9-2)} \qquad \frac{\partial \pi_i}{\partial q_i} = (-1) \cdot q_i + (10 - q_i) \cdot 1 - 1 = 9 - 2q_i = 0,$$

implying $q_i = 4.5$, $p_i = 5.5$ and $\pi_i = 20.25$. The consumer surplus in each market is equal to 10.125.

In contrast, the market outcome in the case of competitive consumption *and* production would be $q_i = 9$, $p_i = MC_i = 1$ and $\pi_i = 0$. The loss of production profit would be more than offset by the increase in consumer surplus, since 40.5 > (20.25 + 10.125). The competitive solution maximizes social surplus and is thus regarded the socially beneficial outcome. By lowering price and increasing quantity in every local market we approach the socially beneficial outcome. Consider now the effort of enhancing competition by establishing a transmission grid.

Situation 2: Oligopoly

We consider a grid connecting all markets and having no effective capacity constraints, for instance the tree shaped grid in Figure 9-2.

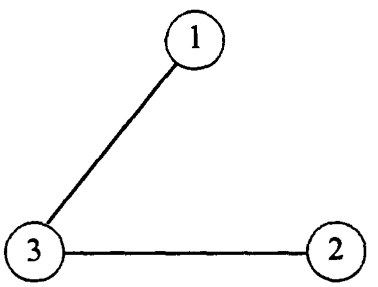


Figure 9-2 Radial Network - No Effective Capacity Constraints

With high capacity lines, where capacity is assumed to be larger than ever needed, the total market turns into an oligopoly with three identical producers serving a larger group of price-taking customers. Customers are located at three different spots, but since we have assumed a loss-less "DC" approximation of the network customers may be served by any producer at the same marginal cost. This is a good approximation if line capacity is large relative to the actual use of the line.

In this situation a producer must take into account the actions taken by the other producers while maximizing profit. The Cournot assumption specifies how this is done; a producer maximizes his own profit taking the competitors' quantities as given. Demand for the combined market is given by $q_{tot} = 3\cdot(10-p) \Rightarrow p = 10 - \frac{1}{3}q_{tot}$, where $q_{tot} = q_1 + q_2 + q_3$. The maximization problem of producer i is then

$$(9\text{-}3) \qquad \max_{q_i} \pi_i = p \cdot q_i - TC_i = (10 - \tfrac{1}{3} q_{tot}) \cdot q_i - q_i = (10 - \tfrac{q_1+q_2+q_3}{3} - 1) \cdot q_i,$$

with first order condition

$$(9\text{-}4) \qquad \frac{\partial \pi_i}{\partial q_i} = (-\tfrac{1}{3}) \cdot q_i + (9 - \tfrac{q_1+q_2+q_3}{3}) \cdot 1 = 9 - \tfrac{1}{3}(q_1 + q_2 + q_3) - \tfrac{1}{3} q_i = 0.$$

The Cournot assumption is reflected by $\partial q_i / \partial q_j = 0$ for $i \neq j$. Since the three producers are identical, $q_1 = q_2 = q_3$, and $q_i = 6.75$, $p = 3.25$, and $\pi_i = 15.1875$. The grid is not actually being used, but the *threat* of it being used enhances competition and increases social surplus as the consumer surplus rises to 22.78125 in each region. This illustrates that unused capacity may be needed.

Generally, when output increases by one unit, the price of *all* units sold will be reduced, thus reducing profits by a tiny amount on all units. On the other hand, the extra unit will generate a profit, and a profit-maximizing firm will increase output until the negative price effect exactly balances the positive effect of quantity. Referring to the optimization problems of the monopoly- and oligopoly-situations ((9-1) and (9-3)), it is obvious that the price effect in the oligopoly is only $\frac{1}{3}$ of the price effect under monopoly ($\partial p / \partial q_i = -\frac{1}{3}$ and $\partial p / \partial q_i = -1$, respectively). The reason is that, taking the other producers' quantity as given, if a producer increases output by 3 units, only 1 will be sold in his home market, thus reducing the price effect to $\frac{1}{3}$ of the effect in the monopoly situation. As a consequence equilibrium will occur at a larger quantity. Stoft [75] describes how this very same argument applies even to a weak line connecting two monopolies and thus creating a force to move to the duopoly-solution.

Following from the optimization problem (9-3) and its first order conditions (9-4), the oligopoly solution is a Nash-equilibrium, or more specifically a Cournot-Nash-equilibrium, referring to the assumption of the conjectural variations being zero. A Nash-equilibrium is a triple of strategies (q_1^*, q_2^*, q_3^*) in which each strategy is the best response to the other two strategies, i.e. $q_1^* = R_1(q_2^*, q_3^*)$, $q_2^* = R_2(q_1^*, q_3^*)$ and $q_3^* = R_3(q_1^*, q_2^*)$. The computed solution is a Nash-equilibrium in pure strategies since producer i's strategy is to "produce exactly q_i^*".

Situation 3: Loop flow

We now add a third line of limited capacity, k, between nodes 1 and 2. This line introduces the problem of loop flow into our network (Figure 9-3).

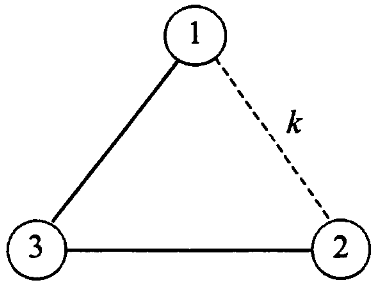


Figure 9-3 Loop Flow and Congestion

If k is small, the oligopoly solution is no longer a Nash-equilibrium. To see why, look at producer 1. What happens if producer 1 unilaterally switches back to the monopoly quantity $q_1 = 4.5$? The price in market 1 will change from $p = 3.25$ to $p_1 = 5.5$, a very profitable change as profits increase from 15.1875 to 20.25. However, we must expect agents to take advantage of this higher price, attempting to export from regions 2 and/or 3 to region 1.

Let x_2 and x_3 be the quantities exported to region 1 from nodes 2 and 3. Assuming identical impedances on every line, it follows from Kirchhoff's laws that $\frac{2}{3}x_2$ will flow along 2-1, $\frac{1}{3}x_2$ flows over 2-3-1, $\frac{2}{3}x_3$ flows over 3-1 and $\frac{1}{3}x_3$ flows over 3-2-1 (Figure 9-4). Confining our attention to the capacitated link 1-2, $\frac{2}{3}x_2$ and $\frac{1}{3}x_3$ flow over this link, both in direction 2-1. This induces a constraint on total export from nodes 2 and 3 to node 1, given by $\frac{2}{3}x_2 + \frac{1}{3}x_3 \le k$ implying that the maximal export possible is $3k$ from node 3 to node 1.

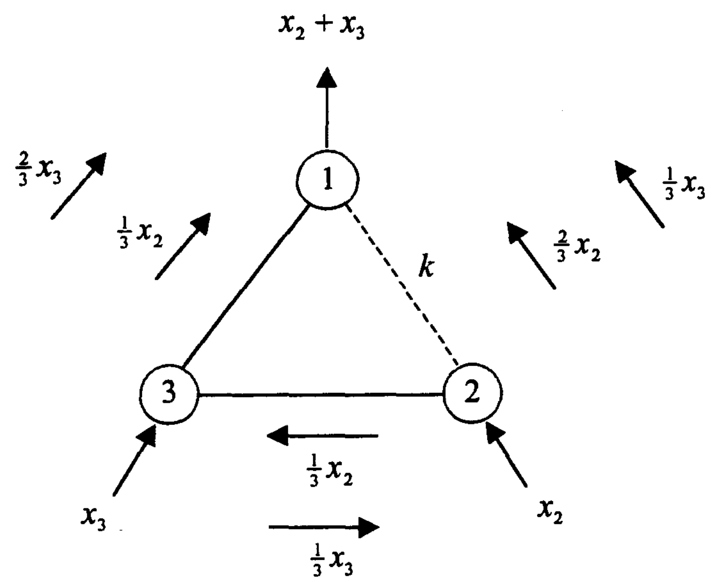


Figure 9-4 Exporting from Regions 2 and 3 to Region 1

The residual demand facing producer 1 is then given by (at least) $p_1 = 10 - (q_1 + 3k)$, and profit is equal to

$$\pi_1 = p_1 \cdot q_1 - q_1 = [10 - (q_1 + 3k)] \cdot q_1 - q_1 = 9q_1 - 3kq_1 - q_1^2.$$

The first order condition of the profit maximization problem is

$$\frac{\partial \pi_1}{\partial q_1} = 9 - 3k - 2q_1 = 0,$$

implying $q_1 = 4.5 - 1.5k$ and $p_1 = 5.5 - 1.5k$. Inserted into the profit-function,

$$\pi_1(k) = 20.25 - 13.5k - 2.25k^2,$$

and switching back to the (reduced) monopoly quantity is still profitable for producer 1 if $\pi_1(k) \geq 15.1875$, i.e. if $k \leq 0.401923788 \approx 0.402$. This means that for k small, the oligopoly solution is no longer a Nash-equilibrium.

The same logic applies to producer 2. He will also be better off by switching back to the monopoly quantity. What happens in the third market? Imagine that producer 3 generates his monopoly output at price 5.5. This will not be sustainable because power can be exported from producers 1 and 2 to market 3. The capacity constraint on link 1-2 will not limit total exports from 1 and 2 because the flow resulting from exporting from node 1 cancels the flow resulting from exporting from node 2. The capacity constraint only puts a limit on the difference between x_1 and x_2, that is to say $-k \le \frac{1}{3}x_1 - \frac{1}{3}x_2 \le k$, where x_1 and x_2 are now defined to be export from node 1 to node 3 and from node 2 to node 3, respectively, and $x_1 + x_2$ can be arbitrarily large.

So what will happen? Will there be monopolies in markets 1 and 2 while the three producers share the third market? The answer is no. A situation with monopoly prices in markets 1 and 2 will be exploited by exporting from producer 3. If the same quantity x_3 is exported from producer 3 into both markets 1 and 2, the flows over link 1-2 originating from these trades, will exactly cancel (Figure 9-5), and the monopolies of markets 1 and 2 are therefore not sustainable either.

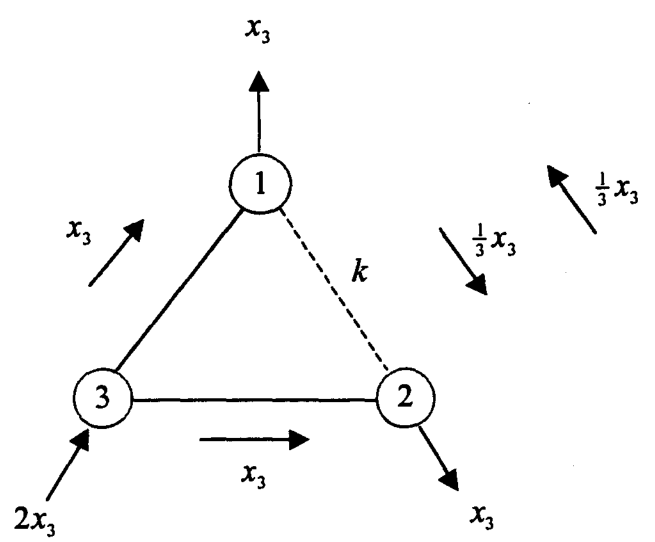


Figure 9-5 Flows Resulting from Exporting from Region 3 to Regions 1 and 2

Summing up, if $k = 0$ or k is large, the oligopoly solution is a Cournot-Nash-equilibrium. However, since neither the Cournot solutions nor the monopoly solutions are stable for small

k, there seem to be no Nash-equilibrium in pure strategies for $0 < k \leq 0.402$. To find a solution, we must consider mixed strategies.

In general, finding the mixed strategy equilibrium to compute expected prices, quantities and social surplus, is a difficult task, and we have to resort to numerical procedures. In Stoft [75] the continuous quantity variables are replaced by a discrete set of outputs and a "fictitious play"-algorithm is employed to estimate the mixed strategy equilibrium. If this procedure converges, it converges to a Nash-equilibrium. If it does not converge, more sophisticated and complex algorithms are needed. For the triangle network in Stoft [79], involving two producers, the Nash-equilibrium in mixed strategies is found by solving a linear complementarity problem. In our case, we have three players and finding mixed strategy equilibria for more than two players is exceedingly difficult.

9.2. Conclusions and Future Research

In the 2-node case, a weak line between two monopolistic markets increases competition. Thus, contrary to the 2-node/single-line case where *removing* the weak line disconnects the markets, *connecting* the weak line causing parallel flows in our 3-node example partially disconnects the markets. This is also a paradox pertaining to loop flow (ref. chapter 8), only that investing in a weak line now increases the probability of strategic bids, which negatively affect the operation of the markets. Thus, the new line possibly reduces the degree of competition, which again reduces social surplus.

An interesting approach for future research would be to develop a more general model allowing for asymmetric markets and different assumptions regarding the mode of competition. Especially, we would like to investigate supply curve bidding in a network involving loop flow.

10. Suggestions for Future Research

As is evident from the thesis, there are still many economic issues that are not completely resolved in electricity transmission. This concerns both the implementation of short-term marginal cost pricing and long-term considerations regarding grid investments and cost-recovery. Some topics for future research are given in the following.

A Price Adjustment Process in the Norwegian Scheduled Power Market

The Norwegian scheduled power market uses the zonal pricing approach for managing congestion. An interesting area of research is whether it is possible to find a good approximation of optimal nodal prices based on the uncongested system price and the loading vectors of congested lines. This can be interpreted as a CP+Hub pricing approach (ref. section 3.4), where the pool is located at the "hub" providing a price of energy at this location, and where the shadow prices of the congested lines must be estimated and used in a price adjustment process.

As the zonal division comes into practice only when congestion is expected to last for several days, the market responses of the initial nodal price estimates might be used to obtain better estimates. This procedure should be combined with some other mechanism such as curtailment or counter purchases to obtain feasible flows. The suggested approach also resembles the CMT-approach of Wu and Variaya [89] (ref. section 3.6), as we do not rely on private information, i.e. supply and demand curve bids. A similar approach of estimating cost parameters is studied in Glavitch and Alvarado [25].

Congestion and Market Power

The study of the exercise of market power accruing from network congestion is only in its beginning, and we would like to consider the possibilities of gaming the constraints of the Norwegian and/or Swedish system. In the Scandinavian market there are two different systems for handling congestion. These systems work simultaneously, and we would like to investigate whether market power implications differ for the two systems. In this context it is also interesting to examine whether the use of fixed versus flexible zone boundaries make any difference. Harvey and Hogan [33] have considered some market power questions in a zonal pricing system, however this system seems to work differently compared to the Norwegian system, thus for the Norwegian market we require a different analysis.

Analysis of the Network Hierarchy – Aggregation and Disaggregation

As shown in Figure 4-1, the Norwegian grid has a hierarchical structure. A practical issue has been to determine the extension of the central grid, and this is treated in Bjørndal et al. [5]. One suggestion is to determine the interface so that the sub-networks of the central grid have a radial structure. In the Norwegian system this occurs at 22kV. A completely different approach could be to replace the central grid by a "virtual central grid" as described in McGuire [52]. The question is what the efficiency and distribution effects of choosing different interfaces are.

The concept of "virtual networks" has similarities with the aggregation methods developed for computing equilibria in traffic assignment problems (see for instance Barton et al. [4]). We have also described other similarities between electricity networks and traffic equilibrium problems, and an integrated approach might yield new insights. For instance, load factors have proven useful when investigating electricity networks, and a corresponding term could ease the analysis of traffic equilibrium problems as well.

Long Run Perspectives

A question that has relevance in both electricity networks and traffic planning, where investment paradoxes can occur, is whether it is possible to find expansion paths that, within a budget, guarantee increased social surplus in each step, while attaining long run optimum.

Generally, long run efficiency is possibly the most challenging research topic at present. Hopefully, competitive markets will enhance efficiency in this respect as well. The investment disincentives from the fixed tariff of the current Norwegian system are considerable, as pointed to in chapter 4. Contrary to the Swedish system, the fixed tariffs per kW are not geographically differentiated, and the efficiency and fairness of this may be questioned. We are currently investigating the use of cooperative game theory to allocate fixed cost. Maybe this methodology can provide some answers.

11. References

[1] Agnagic, M., K. H. Abdul-Rahman, and J. G. Waight (1998), "Spot Pricing of Capacities for Generation and Transmission of Reserve in an Extended Poolco Model," *IEEE Transactions on Power Systems*, 13, 1128-1135.

[2] Allen, Eric, Marija Ilic, and Ziad Younes (1999), "Providing for Transmission in Times of Scarcity: An ISO Cannot Do It All," *Electrical Power and Energy Systems*, 21, 147-163.

[3] Baldick, R., and E. Kahn (1993), "Transmission Planning Issues in a Competitive Economic Environment," *IEEE Transactions on Power Systems*, 8, 1497-1503.

[4] Barton, Russell R., Donald W. Hearn, and Siriphong Lawphongpanich (1989), "The Equivalence of Transfer and Generalized Benders Decomposition Methods for Traffic Assignment", *Transportation Research-B*, 23, 61-73.

[5] Bjørndal, Mette, Einar Hope, and Hans Olav Husum (1994), "Sentralnettets utstrekning. En prinsipiell analyse," SNF-rapport 62/94.

[6] Borenstein, Severin, James Bushnell, and Steven Stoft (1998), "The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry," Working Paper PWP-040r, University of California Energy Institute, Berkeley, California.

[7] Braess. D. (1968), "Über ein Paradoxen aus der Verkehrsplanung," *Unterhehmensforschung*, 12, 256-268.

[8] Bushnell, James B., and Steven Stoft (1996), "Electric Grid Investment under a Contract Network Regime," *Journal of Regulatory Economics*, 10, 61-79.

[9] Bushnell, James B., and Steven Stoft (1996), "Transmission and Generation Investment in a Competitive Electric Power Industry," Working Paper PWP-030, University of California Energy Institute, Berkeley, California.

[10] Bushnell, James B., and Steven Stoft (1997), "Improving Private Incentives for Electric Grid Investment, " *Resource and Energy Economics*, 19, 85-108.

[11] Calvert, Bruce, and Grant Keady (1993), "Braess's Paradox and Power-Law Nonlinearities in Networks," *Journal of the Australian Mathematical Society, Series B Applied Mathematics*, 35, 1-22.

[12] Chao, Hung-Po, and Stephen Peck (1996), "A Market Mechanism for Electric Power Transmission," *Journal of Regulatory Economics*, 10, 25-59.

[13] Chao, Hung-Po, and Stephen Peck (1997), "An Institutional Design for an Electricity Contract Market with Central Dispatch," *The Energy Journal*, 18, 85-110.

[14] Chao, Hung-Po, and Robert Wilson (1987), "Priority Service: Pricing, Investment, and Market Organization," *The American Economic Review*, 77, 899-916.

[15] Cohen, Joel E., and Paul Horowitz (1991), "Paradoxical Behaviour of Mechanical and Electrical Networks," *Letters to Nature*, 352, 699-701.

[16] Cuthbert, Thomas R. Jr. (1987), *Optimization Using Personal Computers, with Applications to Electrical Networks*, John Wiley & Sons.

[17] Dafermos, S., and A. Nagurney (1984), "On Some Traffic Equilibrium Theory Paradoxes," *Transportation Research-B*, 18, 101-110.

[18] Dembo, Ron S., John M. Mulvey, and Stavros A. Zenios (1989), "Large-Scale Nonlinear Network Models and Their Application," *Operations Research*, 37, 353-372.

[19] Dolan, Alan, and Joan Aldous (1993), *Networks and Algorithms: An Introductory Approach*, John Wiley & Sons.

[20] Dubey, Pradeep (1986), "Inefficiency of Nash Equilibria," *Mathematics of Operations Research*, 11, 1-8.

[21] Enke, Stephen (1951), "Equilibrium Among Spatially Separated Markets: Solution by Electric Analogue," *Econometrica*, 19, 40-47.

[22] Fang, R. S., and A. K. David (1999), "Optimal Dispatch under Transmission Contracts," *IEEE Transactions on Power Systems*, 14, 732-737.

[23] Fang, R. S., and A. K. David (1999), "Transmission Congestion Management in an Electricity Market," *IEEE Transactions on Power Systems*, 14, 877-883.

[24] Frank, Marguerite (1981), "The Braess Paradox," *Mathematical Programming*, 20, 283-302.

[25] Glavitch, Hans, and Fernando Alvarado (1997), "Management of Multiple Congested Conditions in Unbundled Operation of a Power System," *IEEE Transactions on Power Systems*, 13, 374-380.

[26] Green, Richard (1998), "Electricity Transmission Pricing: How much does it cost to get it wrong?" Working Paper PWP-058, University of California Energy Institute, Berkeley, California.

[27] Green, Richard J., and David M. Newbery (1992), "Competition in the British Electricity Spot Market," *Journal of Political Economy*, 100, 929-953.

[28] Gupta, Alok, Dale O. Stahl, and Andrew Whinston (1997), "A Stochastic Equilibrium Model of Internet Pricing," *Journal of Economic Dynamics and Control*, 21, 697-722.

[29] Hallefjord, Åsa, Kurt Jörnsten, and Tore Nilssen, "A Note on Cost Allocation Rules for the Classical Transportation Problem," Christian Michelsens Institute, Bergen, Norway.

[30] Hallefjord, Åsa, Kurt Jörnsten, and Sverre Storøy (1994), "Traffic Equilibrium Paradoxes when Travel Demand Is Elastic," *Asia-Pacific Journal of Operations Research*, 11, 41-50.

[31] Hansen, Pierre, and Brigitte Jaumard (1997), "Cluster Analysis and Mathematical Programming," *Mathematical Programming*, 79, 191-215.

[32] Hansen, Pierre, Dominique Peeters, and Jacques-François Thisse (1997), "Facility Location under Zone Pricing," *Journal of Regional Science*, 37, 1-22.

[33] Harvey, Scott M., and William W. Hogan (2000), "Nodal and Zonal Congestion Management and the Exercise of Market Power," Center for Business and Government, John F. Kennedy School of Government, Harvard University.

[34] Harvey, Scott M., William W. Hogan, and Susan L. Pope (1996), "Transmission Capacity Reservations Implemented through a Spot Market with Transmission Congestion Contracts," *The Electricity Journal*, 9, 42-55.

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

[36] Hogan, William W. (1993), "Markets in Real Electric Networks Require Reactive Prices," *The Energy Journal*, 14, 171-200.

[37] Hogan, William W. (1997), "A Market Power Model with Strategic Interaction in Electricity Networks," *The Energy Journal*, 18, 107-141.

[38] Hogan, William W. (1999), "Market-Based Transmission Investments and Competitive Electricity Markets," Center for Business and Government, John F. Kennedy School of Government, Harvard University.

[39] Jörnsten, Kurt, Linda Rud, and Balbir Singh (1997), "Samspillet mellom nettariffering og energimarkeder: En illustrasjon," SNF Working Paper 55/97.

[40] Jörnsten, Kurt, and Balbir Singh, "Market Architecture and Efficiency: Transmission Pricing in a Competitive Electricity Sector," SNF Working Paper 60/96.

[41] Jörnsten, Kurt, Balbir Singh, and Mette Bjørndal (1997), "Transmisjonsprising – overføringstariffens fastledd," SNF-rapport 64/97 (confidential).

[42] Kahn, Edward, and Ross Baldick (1994), "Reactive Power is a Cheap Constraint," *The Energy Journal*, 15, 191-201.

[43] Kim, Balho H., and Martin L. Baughman (1997), "The Economic Efficiency Impacts of Alternatives for Revenue Reconciliation," *IEEE Transactions on Power Systems*, 12, 1129-1135.

[44] Klemperer, Paul D., and Margaret A. Meyer (1989), "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica*, 57, 1243-1277.

[45] Kolstad, Charles D. (1985), "A Review of the Literature on Bi-Level Mathematical Programming," Technical Report LA-10284-MS, US-32, Los Alamos National Laboratory.

[46] Korilis, Yannis A., Aurel A. Lazar, and Ariel Orda (1997), "Achieving Network Optima Using Stackelberg Routing Strategies," *IEEE/ACM Transactions on Networking*, 5, 161-173.

[47] Korilis, Yannis A., Aurel A. Lazar, and Ariel Orda (1997), "Avoiding the Braess Paradox in Non-Cooperative Networks," *Journal of Applied Probability*, 36, 211-222.

[48] MacKie-Mason, Jeffrey K., and Hal R. Varian (1995), "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications*, 13, 1141-1149.

[49] Marangon Lima, J. W., and E. J. de Oliveira (1998), "The Long-Term Impact of Transmission Pricing," *IEEE Transactions on Power Systems*, 13, 1514-1520.

[50] McCalley, James D., Sohrab Asgarpoor, Thomas Gedra, Mark Halpin, Narinder K. Saini, and Michael H. Scrameyer (1997), "Second Bibliography on Transmission Access Issues," *IEEE Transactions on Power Systems*, 12, 1654-1659.

[51] McGuire, Bart (1997), "Optimal Power Flow, Node Prices, and Transmission Tolls in a Number of Instructive Examples," Working Paper PWP-050, University of California Energy Institute, Berkeley, California.

[52] McGuire, Bart (1999), "Power-Grid Decentralization," Working Paper PWP-061, University of California Energy Institute, Berkeley, California.

[53] Migdalas, Athanasios, and Panos M. Pardalos (1993), "Editorial: Hierarchical and Bilevel Programming," *Journal of Global Optimization*, 8, 209-215.

[54] Murchland, J. D. (1970), "Braess's Paradox of Traffic Flow," *Transportation Research*, 4, 391-394.

[55] Oren, Shmuel S. (1997), "Economic Inefficiency of Passive Transmission Rights in Congested Electricity Systems with Competitive Generation," *The Energy Journal*, 18, 63-83.

[56] Oren, Samuel S., Pablo T. Spiller, Pravin Varaiya, and Felix Wu (1995), "Nodal Prices and Transmission Rights: A Critical Appraisal," *The Electricity Journal* (April), 24-35.

[57] Pas, Eric I., and Shari L. Principio (1997), "Braess' Paradox: Some New Insights," *Transportation Research-B*, 31, 265-276.

[58] Penchina, Claude M. (1997), "Braess Paradox: Maximum Penalty in a Minimal Critical Network," *Transportation Research-A*, 31, 379-388.

[59] Pérez-Arriga, I. J., F. J. Rubio, J. F. Puerta, J. Arceluz, and J. Marín (1995), "Marginal Pricing of Transmission Services: An Analysis of Cost Recovery," *IEEE Transactions on Power Systems*, 10, 546-553.

[60] Rudkevich, Aleksandr, Max Duckworth, and Richard Rosen (1998), "Modeling Electricity Pricing in a Deregulated Generation Industry: The Potential for Oligopoly Pricing in a Poolco," *The Energy Journal*, 19, 19-48.

[61] Rudnick, Hugh, Rodrigo Palma, and José E. Fernández (1995), "Marginal Pricing and Supplement Cost Allocation in Transmission Open Access," *IEEE Transactions on Power Systems*, 10, 1125-32. (Discussion pp.1133-1142).

[62] Samet, Dov, Yair Tauman, and Israel Zang (1984), "An Application of the Aumann-Shapley Prices for Cost Allocation in Transportation Problems," *Mathematics of Operations Research*, 9, 25-42.

[63] Samuelson, Paul A. (1952), "Spatial Price Equilibrium and Linear Programming," *The American Economic Review*, 42, 283-303.

[64] Schweppe, F. C., M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988), *Spot Pricing of Electricity*, Kluwer Academic Publishers.

[65] Shenker, Scott (1995), "Making Greed Work in Networks: A Game-Theoretic Analysis of Switch Service Disciplines," *IEEE/ACM Transactions on Networking*, 3, 819-831.

[66] Shenker, Scott, David Clark, Deborah Estrin, and Shai Herzog (1996), "Pricing in Computer Networks: Reshaping the Research Agenda," *Telecommunications Policy*, 20, 183-201.

[67] Singh, Harry, Shangyou Hao, and Alex Papalexopoulos (1998), "Transmission Congestion Management in Competitive Electricity Markets," *IEEE Transactions on Power Systems*, 13, 672-680.

[68] Statnett (1996), "Harmonisering av de svensk/norske overføringstariffer. Rapport fra den felles norsk/svenske tariffgruppen."

[69] Statnett (1997), "Tariffer for Sentralnettet 1997."

[70] Statnett (1997), "Tariffering i sentralnettet 1998: bakgrunn – status og utviklingstrekk," Notat.

[71] Steinberg, Richard, and Richard E. Stone (1988), "The Prevalence of Paradoxes in Transportation Equilibrium Problems," *Transportation Science*, 22, 231-241.

[72] Steinberg, Richard, and Willard I. Zangwill (1983), "The Prevalence of Braess' Paradox," *Transportation Science*, 17, 301-318.

[73] Stewart, N. F. (1980), "Equilibrium Vs System-Optimal Flow: Some Examples," *Transportation Research-A*, 14, 81-84.

[74] Stoft, Steven (1996), "Analysis of the California WEPEX Applications to FERC," Working Paper PWP-042A, University of California Energy Institute, Berkeley, California.

[75] Stoft, Steven (1997), "The Effect of the Transmission Grid on Market Power," LBNL-40479, Ernest Orlando Lawrence Berkeley National Laboratory, University of California, Berkeley.

[76] Stoft, Steven (1997), "Zones: Simple or Complex?" *The Electricity Journal* (Jan/Feb), 24-31.

[77] Stoft, Steven (1998), "Charging for Losses," White Paper, Office of Economic Policy, Federal Energy Regulatory Commission.

[78] Stoft, Steven (1998), "Congestion Pricing with Fewer Prices than Zones," *The Electricity Journal* (May), 23-31.

[79] Stoft, Steven (1998), "Using Game Theory to Study Market Power in Simple Networks," Federal Energy Regulatory Commission.

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

[81] Tabors, Richard D. (1996), "A Market-Based Proposal for Transmission Pricing," *The Electricity Journal*, 9, 61-67.

[82] Vicente, Luís N., and Paul H. Calamai (1994), "Bilevel and Multilevel Programming: A Bibliography Review," *Journal of Global Optimization*, 5, 291-306.

[83] von der Fehr, Nils-Henrik Mørch, and David Harbord (1993), "Spot Market Competition in the UK Electricity Industry," *The Economic Journal*, 103, 531-546.

[84] Walton, Steven, and Richard D. Tabors (1996), "Zonal Transmission Pricing: Methodology and Preliminary Results from the WSCC," *The Electricity Journal*, 9, 34-41.

[85] Wangensteen, Ivar, Ove S. Grande and Bjørn H. Bakken (1996), "Systemytelser," Teknisk Rapport A4441, EFI Sintef gruppen, Trondheim.

[86] Wangensteen, I., J. A. Holtan, O. B. Fosso, H.-H. Faanes and K. S. Hornnes (1995), "Avbildning av korttids grensekostnad i sentralnettstariffen," Teknisk Rapport F4313, EFI Sintef gruppen, Trondheim.

[87] Woo, Chi-Keung, Ira Horowitz, and Jennifer Martin (1998), "Reliability of Electricity Transmission," *Journal of Regulatory Economics*, 13, 277-292.

[88] Wood, Allen J., and Wollenberg, Bruce F. (1996), *Power Generation, Operation, and Control*, second edition, John Wiley & Sons.

[89] Wu, Felix F., and Pravin Varaiya (1995), "Coordinated Multilateral Trades for Electric Power Networks: Theory and Implementation," Department of Electrical Engineering and Computer Sciences, University of California.

[90] Wu, Felix, Pravin Varaiya, Pablo Spiller, and Shmuel Oren (1996), "Folk Theorems on Transmission Access: Proofs and Counterexamples," *Journal of Regulatory Economics*, 10, 5-23.

[91] Yang, Hai, and Michael G. H. Bell (1998), "A Capacity Paradox in Network Design and How to Avoid It," *Transportation Research-A*, 32, 539-545.

[92] Younes, Ziad, and Marija Ilic (1999), "Generation Strategies for Gaming Transmission Constraints: Will the Deregulated Electric Power Market be an Oligopoly?" *Decision Support Systems*, 24, 207-222.

[93] Young, Hugh D., and Roger A. Freedman (1996), *University Physics*, 9th edition, Addison Wesley.

[94] Yu, C. W., and A. K. David (1999), "Integrated Approach to Transmission Services Pricing," *IEE Proceedings – Generation, Transmission and Distribution*, 146, 255-260.