THE UNIVERSITY OF MANITOBA

AN ECONOMIC PLANNING APPROACH TO WESTERN CANADA RAILWAY CAPACITY EXPANSION IN THE 1980s

by

© THOMAS P. PACI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE
STUDIES IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER
OF ARTS

DEPARTMENT OF ECONOMICS

WINNIPEG, MANITOBA

APRIL, 1988

ISBN 0-315-47880-2

AN ECONOMIC PLANNING APPROACH TO WESTERN CANADA RAILWAY CAPACITY EXPANSION IN THE 1980s

BY

THOMAS P. PACI

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

ABSTRACT

This thesis revolves around the planning problems and issues related to railway system capacity expansion in Western Canada during the 1980s. Since the late 1970s, Canada's two national railways--CN Rail and CP Rail, have implemented long range investment plans to increase the productive capacity of their respective rail networks in the West. The overall objective of this study is to provide a detailed overview showing why these efforts developed, how they proceeded, what their impacts were, and what the rationale behind them was.

To develop the background necessary to utilize this approach, the thesis conducts a review of the theoretical literature dealing with the economics of railway capacity planning. This review identifies the key factors, elements, and structure related to the railway capacity planning process. The economic and physical interrelationships existing between the factors that impact upon capacity are also discussed at length.

From the theoretical discussion, this study develops an economic approach to outline some important aspects regarding the Western capacity expansion programs of both railways. This study looks at the state of the Western railway network during the 1970s, and identifies a number of relevant issues that impacted upon railway planning.

The planning elements of both railways are examined in detail to determine their impact upon network capacity and the economics of the carrier. The study then assesses the rationale of these plans from an economic planning approach.

This study found that the railways were reluctant to make long range investment plans to expand capacity when uncertainty existed as to the plans' future economic viability. The existence of the statutory Crow rates for grain, because of their adverse effects upon the economic viability of the railways, impeded the long range planning efforts of CN and CP for many years.

The thesis found that when both CN and CP did lay out their Western capacity expansion programs, it was done in a series of logical steps. The first options to be executed were those that cost the least and could be implemented in the short-run. The last options to be taken were extremely costly and only took effect over the long-run.

The thesis concludes that the rationale behind the planning efforts of the two railways to expand the capacity

of their respective Western networks fitted into the concept of economic planning. The resolution of the Crow rates problem signified that revenues from the movement of grain, now at compensatory rates, would help ensure the viability of large scale investments in Western railway capacity. Both CN and CP were now free to initiate and carry out long range planning that was truly economic. Operating and investment decisions to increase network capacity followed a logical sequence that was analagous to the long-run expansion path from production function theory.

ACKNOWLEDGEMENTS

I would like to express my thanks to Professor Ralph Harris for all of the valuable assistance and encouragement he provided as advisor to this thesis. His extraordinary efforts to keep me on track are certainly very much appreciated.

Special thanks to my sister Darlene who spent many a long night entering my hastily scrawled thesis drafts into a word processor. Thanks, also, to my father who was always there to encourage me to finish what I had started.

I would also like to express my gratitude to those friends and colleagues who carefully listened to me, and cheerfully gave advice, on many aspects of this work.

TABLE OF CONTENTS

Chapter 4 An Oveview of the Railway Capacity Problem in Western Canada in the 1980s

Chapter 5 Western Canada Railway Capacity Planning in the 1980s

Chapter 6 An Economic Planning Approach to Western Canada Railway Capacity Expansion by CN and CP in the 1980s

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

The issue of railway system capacity in Western Canada continues to be of great interest to the national railways, shippers, the federal government, and the various Western provincial governments. The movement of much of Canada's international trade in grain, coal, potash, sulphur, and other important bulk commodities originates on rail lines in the four Western Canadian provinces of British Columbia, Alberta, Saskatchewan, and Manitoba. Since the mid 1970s, the railways and the other interested parties have recognized the potential for a shortfall in the capacity of the existing rail system in Western Canada to meet projected demands.

The two national railways, CN Rail and CP Rail, have invested large sums of capital in order to increase the capacity of their mainline tracks running through the mountainous regions of British Columbia to the West Coast. Some of this expansion has been achieved through the installation of better signalling, controlling, and dispatching systems. The main thrust in expanding capacity, however, has been through double tracking, grade reductions, track improvements, and the building of longer and additional sidings to facilitate train meets.

The federal government has also helped to play a significant role in the expansion of railway system capacity in Western Canada. The railways have contended for many years that operating losses incurred in the movement of grain at statutory rates has inhibited them from making any significant investment in the expansion of railway system capacity in the West. The enactment of the Western Grain Transportation Act in 1983 has shifted a large part of the cost of transporting grain from the carriers to grain producers and the federal government. This has resulted in significant changes to the operating environment and financial well being of the railways.

The 1982-83 recession caused the railways to scale back their investment plans from those originally proposed in the early 1980s. The expected growth in rail traffic simply did not materialize because of the economic slowdown. Despite the slowdown, the railways did not overlook the likelihood that traffic growth would come quickly once economic recovery and expansion resumed in the international markets for Western Canadian commodities. However, they adopted a more cautious approach for forecasting growth and making investments in capacity.

1.2 Rail Capacity in the Context of Railway Investment Planning

In order to meet the demands of ever changing levels of traffic and to avoid future deterioration and disruptions of service, the planned investment of railway capacity is essential. Railway system capacity planning should be executed within a proper context of projecting traffic demand and mix, examining physical and operating strategies, and estimating the resulting economic and service impacts. See Figure 1.1.

Analysis of capacity investment alternatives are important in the context of railway system planning. Capacity analysis requires an understanding of the various dimensions of rail capacity, the complexities of capacity, and the associated trade-off analyses. In the context of the Western Canadian rail system, an analysis of capacity investment alternatives should focus on identifying the various options to affect an increase in railway system capacity to meet the projected demands upon the system. The capacity impact of each of these options must be weighed with the economic cost of their implementation. In a broader analysis, the socio-economic impacts of these options upon a variety of affected groups should also be

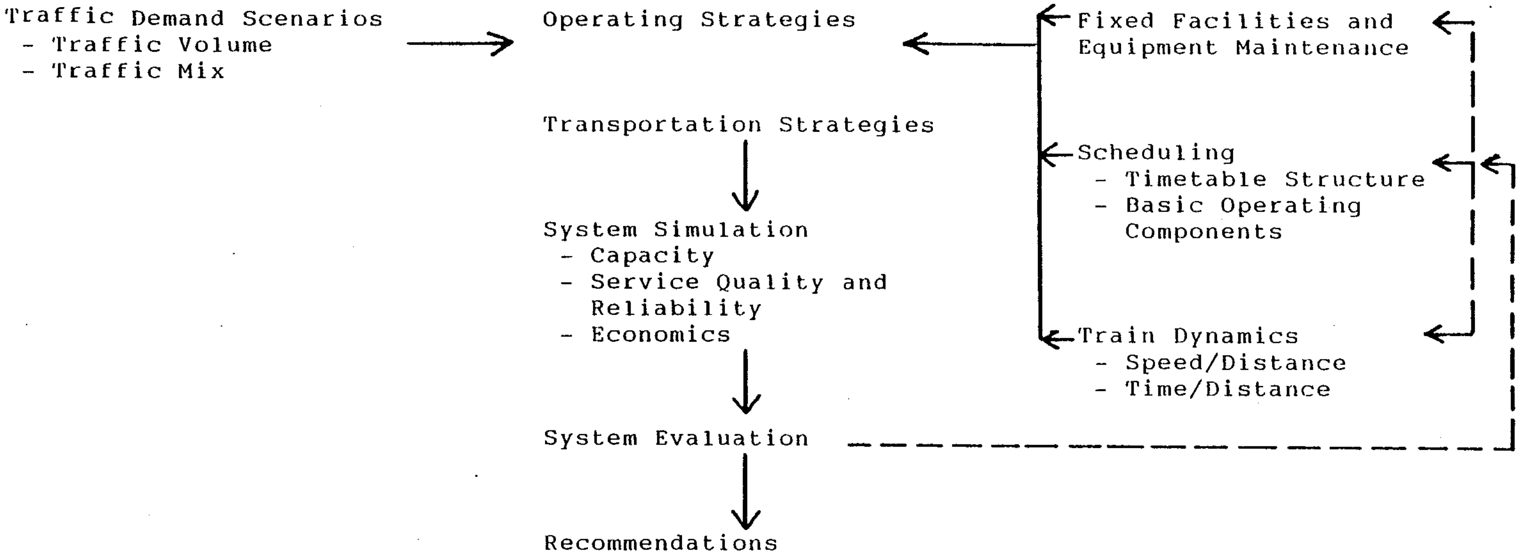


Figure 1.1: The Context of Railway System Capacity Analysis

(from Khan 1979, fig. 1.1)

taken into account.

1.3 Thesis Objectives and Scope

The general objective of this thesis is to provide an overview and evaluation of the various planning options that are available to increase railway system capacity in the short, medium, and long-run. Furthermore, the efforts made by CN Rail and CP Rail in the 1980s to increase the capacity of their respective systems in Western Canada will be evaluated within the context of an economic planning approach. This approach shall define the economic and physical interrelationships that exist between these options, as well as outlining their impacts upon system capcity.

Specifically, the objectives of this thesis are:

1. to analyze and evaluate the economics of various planning options with respect to increasing railway system capacity;
2. to provide an overview of the railway capacity problem in Western Canada for the 1980s;
3. to discuss the options that were available to increase railway system capacity in Western Canada in the 1980s;
4. to evaluate the options taken by the railways to increase railway system capacity in Western Canada in the 1980s, in the light of economic and other criteria.

This research is intended to focus upon the allocative efficiency of the various capacity investment options that were and are available to the railways to increase their system capacity.

1.4 Thesis Organization

This thesis is organized into two parts. Chapters 2 and 3 provide an overview of the various concepts, planning issues and definitions with respect to railway system capacity. Chapter 3 identifies the various factors that can affect railway system capacity. In the second part of the thesis, chapters 4 and 5 provide a description of the railway system in Western Canada and an overview of the various options available to increase capacity. Chapter 6 presents an evaluation of the CN and CP plans from the perspective of economic planning. Chapter 7 summarizes this evaluation and provides the conclusion to the thesis.

CHAPTER 2

CAPACITY AND THE ECONOMICS OF RAILWAY TRANSPORTATION

2.1 Introduction

The capacity of a railway system to move traffic may be affected by a myriad of factors. These factors, whether they occur by design or accident, can either increase or reduce the ability of a railway system to produce transportation services. These changes in the level of system capacity will impact upon the physical and economic parameters that govern the production of railway transportation services.

In order to identify the various factors that can affect system capacity, and to comprehend their impacts, it is essential to understand the structure, operations, and economics of the railway. Present day railway systems are comprised of a number of highly complex components that interact with one another to achieve the movement of passengers and freight. This chapter focuses on providing an overview of the structure, operations, and economics of the modern railway system. In addition, this discussion will also concentrate on describing the various definitions of capacity, as well as detailing the relationship between capacity and the economics of a railway system.

This chapter should help to illuminate the factors, definitions, and concepts relevant to railway system capacity analysis in general, and specifically to the decisions taken by CN Rail and CP Rail to increase their railway system capacity in Western Canada during the 1980s.

2.2 Concept of a Railway System in General

Modern and highly developed railway systems, such as those owned and operated by CN Rail and CP Rail in Canada, are a complex aggregation of a number of different components that are interdependent and interactive with one another. These components interact with one another to produce railway services. Railway services essentially involve facilitating the movement of freight and passengers to and from the various points served by the system. These services can be characterized as being extremely heterogeneous in regards to the type of traffic carried (e.g. passenger, freight, commodity type, etc.), and to their service attributes (Khan 1979).

Table 2.1 lists the major components that comprise the structure of the railway system. This table also presents the elements, and their associated characteristics, that

TABLE 2.1

Major Components of a Railway System

Major Components	Elements and Characteristics
1. Lines (Guideways)	Track and subgrade Type Sidings Crossovers Grade (maximum or ruling) Track structure (type, condition)
	Signals and communications (control) Type
2. Terminals	Loading and unloading facilities Type Location
	Holding facilities Type Location
	Service facilities Type Location Size
3. Yards	Storage facilities Number and length of tracks
	Inspection Facilities Type Location
	Classification facilities Number Location
	Repair facilities Location Size
	Service facilities Type Location Size
4. Equipment (load carrying system)	On-line vehicles and equipment (rail cars and locomotives), and off-line vehicles and equipment Type Number Service

Continued on next page

TABLE 2.1 - Continued

5. Repair and maintenance facilities	On-line track maintenance Type Capacity
	Off-line maintenance Type Number
	Testing Type
	Repairs (on-line and off-line) Type Number
6. Resources (labour, materials, funds)	Plant, operations, maintenance Quantity Availability
7. Management system	Load support services (fare collection, load processing, documentation, tracing)
	Marketing system (load procurement, sales force, advertising)
	Personnel (recruiting, training, management)
	Financial system (cash flow managment, billing, internal accounting and analysis)
	Planning and analysis system (corporate planning)
	Organizational structure (internal structure for control and accountability)

Sources: Khan 1979, table II.2; Manheim 1979, table 5.1.

make up each of the major components. Some of these elements are physical, while others are institutional. These components interact with one another to produce the movement of goods and passengers (i.e. railway services) over the railway system.

The major components of the railway system include lines, terminals, yards, repair and maintenance facilities, equipment (i.e. locomotives, rail cars, other equipment, etc.), resources (i.e. labour, materials, and funds), and management. The railway plant, or infrastructure, includes the following components: lines, terminals, yards, and repair and maintenance facilities. The management component includes all of the elements that form the overall management system of the railway. The elements of the management component are institutional rather than physical.

The production of railway services requires a wide array of resources. The input list of resources necessary to build and operate a railway system are shown in Table 2.2 (Manheim 1979). The construction, maintenance, and operation of a system that is able to provide proper service requires the consumption of different types of resources. Some of these resources are literally consumed in the building, maintenance, and operation of the system. These resources would be items such as land, labour, fuel, materials, etc. Other types of resource consumption,

TABLE 2.2

Resources Consumed in Building and Operating a Railway System

Resource	Activity
1. Labour	Train operations Fixed facilities operations On-line vehicle maintainance Fixed facilities maintainance Manangement system Vehicle fabrication Fixed facilities fabrication
2. Materials	Vehicle fabrication (metal, rubber, plastics, etc.) Fixed facilities fabrication (steel, cement, etc.) Non-energy consumables needed for system operations (paper, replacement parts, etc.)
3. Land	Fixed facilities for the system (guideways, transfer facilities, etc.) Fixed facilities for the fabrication of materials
4. Energy	Power for system operations Power for train operations Power for facilities fabrication
5. Environmental degredation	Air quality Noise level Water quality Odors
6. Ecological effects	Effects on animal life Effects on plant life
7. Social effects	System as a physical barrier Effects of displacement of homes and businesses Effects of community cohesion Effects on social stability
8. Aesthetic effects	View of the system from outside View of the system to users View of the environment from the system

Source: Manheim 1979, table 5.2

however, involve the effects of railway service production upon the environment, ecology, and society. These forms of consumption can be defined as the degradation of existing quality due to the provision of transportation services by the system. An example of this would be the contamination of air with pollutants from railway diesel locomotives. This type of consumption involves the use of resources in the sense that the existing level of air quality is reduced (or consumed) by the operation of the railway locomotives (Manheim 1979, 171).

The provision of railway services involves two key types of operations: line haul and terminal/yard activities. Line haul operations deal with movement of rail cars (loaded and unloaded) between terminals. Terminal and yard operations involve the reception, sorting, classification, and assembly of individual rail cars and locomotives into trains. These operations occur through the interaction of the various components of the railway system. More specifically, trains interact with the existing railway system infrastructure to produce "movements" of freight and/or passengers.

The railway system infrastructure, or plant, consists of the following components: lines, terminals, yards, and repair and maintenance facilities. Breaking these components down further, would yield such elements as track and subgrade, signals and switches, terminal facilities, yard trackage and buildings, etc.

The line component of the railway system is made up of the following elements: track and subgrade, signals and communications, and power distribution system. The function of the railway line component is to provide the guideway for traffic to move to and from all points that are served on the system. In this context, the track and subgrade act as the support system which transmits the load from the traffic, which is carried in rail cars pulled by locomotives (i.e. trains), to the ground. Railway line characteristics include location and number of tracks (i.e. single, double, multiple), number and length of sidings, location and number of crossovers, grade and alignment, and track structure.

The purpose of the signal and communication element is to provide directional guidance and control for the trains that move over the system lines. Signals are an important part of the railway and are utilized not only for increased safety, but also to step up the efficiency and capacity of a railway line to handle traffic. Decisions regarding train movements over the railway system must be made on a moment by moment basis. The control system impacts upon quality of decisions, the time needed to implement them, and the delay in train transit time caused by them. Types of signals include the following (Khan 1979, 35):

1. Manual block signalling

2. Automatic block signalling (ABS)
3. Automatic train control (ATC)
4. Centralized traffic control (CTC)

These types of control systems will be discussed in greater detail in a following chapter.

The power distribution system describes the propulsion system that is employed to move trains over the railway. This element can be characterized by the type of power distribution system that is utilized by the railway system (e.g. overhead power lines or electrified track sections for electric railways, railside fuel depots for conventional diesel locomotives, etc.).

The railway equipment component represents the locomotives and cars that carry the traffic loads over the railway system lines. These vehicles are an essential part of the overall railway system. Locomotives function as the power and propulsion system of the train. Locomotive characteristics include such things as horsepower, traction, adhesion, efficiency, and reliability. Rail cars provide the load containing system (both passengers and freight) for the train. There exists a variety of different rail car types (e.g. hopper car, boxcar, flatcar, etc.). They are designed to carry and protect different types of cargo.

The major elements of the railway system terminal component include facilities for the loading and unloading, holding, and servicing of trains. Facilities for loading and unloading include such equipment as cargo belts, cranes, fork lifts and other related types of machinery. Holding facilities consist of tracks located at the terminal that are used for the purpose of surging (or holding) rail cars. Terminal service facilities contain equipment, personnel, and materials that are employed to service railway locomotives and cars. These service functions include vehicle maintenance inspections, cleaning, and fueling.

Railway yards are comprised of the following elements: facilities for storage, inspection, classification, maintenance and repair, and service. Railway yard storage facilities are comprised of trackage that is used for the purpose of storing vehicles and their cargos for both short and long term. Service and repair facilities include equipment, personnel, spare parts, and other materials that are employed for the purpose of maintaining locomotives and cars. The main function of the railway yard, however, is to sort or classify incoming trains, or "blocks" of cars, that are brought into the yard by incoming or through trains, and then assemble these into outbound trains.

Railway terminals and yards are commonly referred to as the terminal system. According to Khan:

> An overall definition of terminals and yards ... encompasses: main marshalling yards, storage yards,

> holding (surge) yards, satellite yards, industrial support yards, by-pass yards, and intra-terminal track.
>
> Terminal and yards are quite often unique in configuration and traffic handling patterns. Also, yards are essentially networks of subsystems which are interdependent and interactive and are designed to perform various functions or operations. The individual yard or terminal system itself, being a part of the railway network, is also interactive with other terminals and with the mainline. Furthermore, such external forces as weather or equipment breakdown affect the operation of the terminal and yard system. (1979, 21).

Typical rail-yard operations can be represented by the following five types of activities:

1. Receiving and inbound inspection
2. Classification and sorting
3. Reclassification and/or wait for connection
4. Train marshalling or assembly
5. Outbound inspection and departure

Rail cars arrive at the receiving trackage of a railway yard as blocks of cars that are transported and deposited by a through train. When these cars enter the yard they are first of all inspected by personnel in order to check their operating condition. These cars are then shunted to the receiving track where they are sorted and moved to the classification tracks to be classified. In the next step, the classified cars are assembled into blocks and then into outbound trains. These trains are assembled by moving blocks of classified cars from the classification tracks to the departure track. Here, locomotives are attached to the train and a final outbound inspection is conducted. The outbound train then leaves on the departure track for the next yard. Again, referring to Khan:

> It is of interest to note that a typical rail car spends the majority of its time at a yard terminal with only a fraction of the total trip time spent on between-yard movements. This is due to the unique feature of the rail mode--its combination of individual cars, going from various origins to various destinations, into trains moving in common. However, considerable economies are achieved by shared use of motive power, fuel, and operating staff while the shipment is moving on a train. A disadvantage is the need for extra car handling, sorting, and storage operations at the yard. (1979, 23).

The system repair and maintenance component includes those elements that are essential for keeping the existing railway infrastructure in proper operating condition. The elements of this component would include facilities and equipment for on-line track maintenance, off line maintenance, testing, and repairs. An example of on-line maintenance equipment is the track machine. The state-of-the-art version of this machine combines several maintenance functions into one. In one operation the track machine can lift the old track, remove the ties and rails, cleans the ballast, lays and spaces the ties, lays continuous welded rails, and finally, fastens the rails to the ties. Other examples of on-line maintenance equipment include: work trains for ballast cleaning, replacing, and tamping (compacting); spiking machines; mobile rail welding equipment; rail resurfacing equipment; and portable track maintenance equipment that doesn't take up track space (R. S. Wallace and Associates 1981, 117-8).

Off-line maintenance facilities usually includes shops and other related structures that are used to support repair and maintenance operations. These facilities would contain personnel, tools, equipment, parts, and other materials used to support these operations.

Testing is also an important element of the repair and maintenance component. This element includes a variety of machines that are used to detect and measure faults in the track. These machines would include such things as track geometry cars, which are used to measure track radius of curvature, superelevation, and irregularities in horizontal and vertical alignment. Another example of testing equipment are railhead wear machines. These machines are pulled over sections of track in order to determine the presence of any corrugations in the contour of the railhead (R. S. Wallace and Associates 1981, 118).

Labour, materials, and funds define the resources component of the railway system. The quantity and availability of these resources impact upon the plant characteristics, operations, and maintenance of the railway system.

The management component of the railway system is comprised of the following elements, or sub-systems: load support services, operations systems, marketing system, communications and control system, personnel system, financial system, planning and analysis system, and organizational structure. These elements are essential for the operation and maintenance of railway services by the system in the short, medium, and long term.

2.2.1 The Railway Network

The railway network is a representation of all of the components of the railway system that have spatial or temporal characteristics. The spatial structure of the railway is reflected in the physical characteristics of the supply components of the railway (e.g. location and length of track, geographical points that are served by the railway, etc.). The temporal structure of the railway system relates to the variation of the physical performance characteristics of the railway system over time (e.g. hour, day, week, etc.). As such, it is an extremely important consideration for the purposes of capacity analysis.

According to Khan (1979), for the purposes of capacity analysis, railway networks should be analyzed on a sub-network (or system) basis. That is to say, that some well defined portion of the overall railway network should be isolated and then further analyzed.

Railway networks can be classified in terms of nodes and links. A node represents some location where railway activities occur (i.e. terminals where traffic originates and terminates, yards that assemble and classify trains, etc.). Nodes, therefore, can be classified according to their function(s). In a railway network, the links are defined as the routes that connect the various nodes with one another. The interaction of train flows over a railway network is influenced by the physical performance characteristics and conditions of the existing links. Therefore, the overall physical performance of the railway network is defined not only by the performance of each component but also by their interaction with one another.

2.3 Railway System Capacity: Concepts and Definitions

For any railway system, sub-system, or component thereof, the capacity is the maximum throughput (i.e. traffic) that can be moved by it, or through it, during some given period of time. Railway capacity, in this context, can be thought of as an actual, finite number.

Figure 2.1 graphically illustrates this concept for some railway system component (e.g. a particular segment of railway line). The horizontal axis represents the total throughput that is moved through this component. The vertical axis depicts the average time that it takes to move a unit of throughput over the component (in hours).

This system component has a throughput capacity of X_C units. For any level of throughput between the range of 0

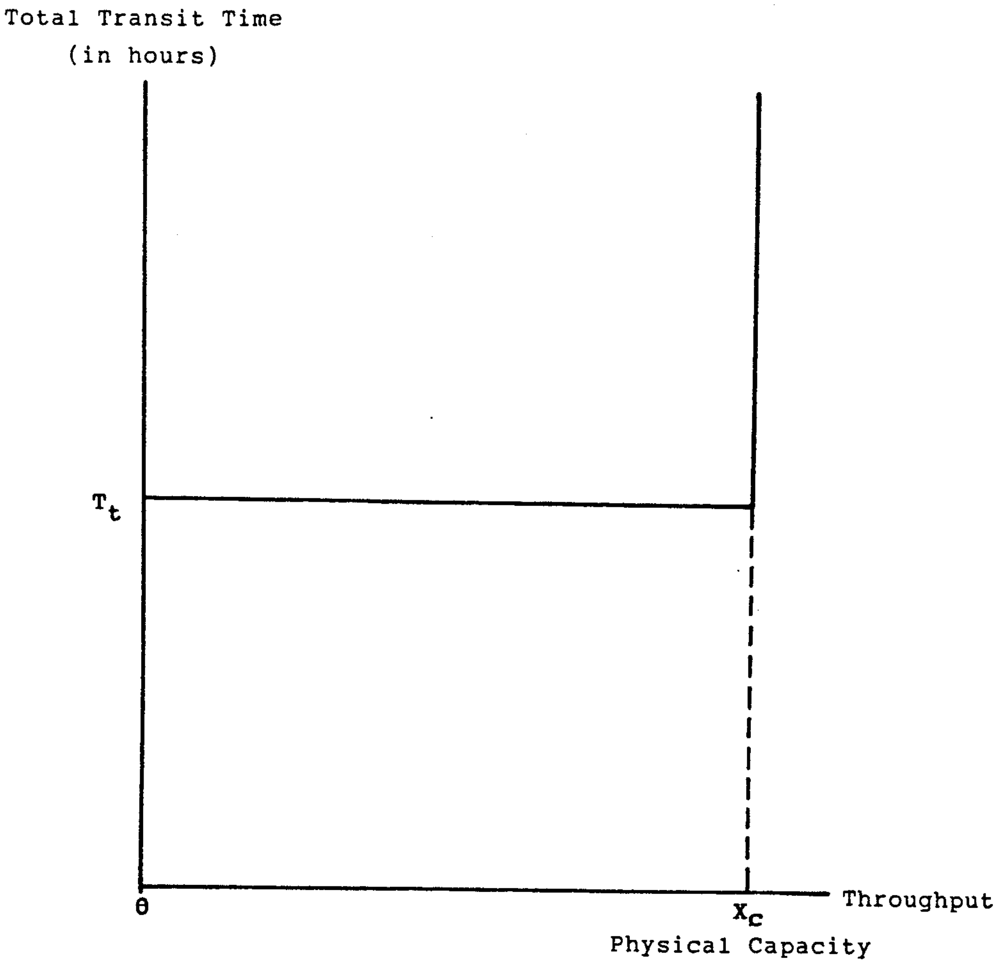


Figure 2.1: System Component Capacity and the Total Transit Time of Throughput

and X_c, it takes an average of T_t hours to move through the component. Any throughput that occurs beyond the value of X_c will not be able to move through the component. This implies that the average transit time of any unit of throughput that occurs ofter the capacity of the component has been reached is infinity.

When the level of throughput exceed the capacity of the component it is moving through, there are three separate conditions that may arise:

1. The system component breaks down because there is a complete jam up of traffic and no further movement is possible.
2. A waiting queue forms and grows ever longer.
3. The queue that builds up is eventually dissipated in some future period when the level of throughput is less than X_c

It is virtually impossible to satisfactorily define the capacity of a railway system or sub-system using any single measure. The concept of railway capacity is beset with many complexities. According to Khan:

> railway capacity, at the broadest possible (corporate) level can be regarded as the ability of the carrier to supply as required the necessary services within acceptable service levels and costs so as to meet the present and projected demand for such services. (1979, 41).

Khan went on to say that railway capacity can be viewed in more quantifiable terms at the system and component level, and that at each level there are a number of possible definitions of capacity that may be applied (1979, 41).

The concept of railway capacity generally relates to the maximum traffic handling capabilities of a particular rail line over a given period of time. Broadly speaking, there exists two wide definitions of railway capacity--physical and economic--for any given railway system or sub-system. These definitions are complicated by a number of supply and demand side factors. The physical or economic capacity of a railway system or component can be significantly affected by a number of supply side elements. The following are among the most important: conditions and configuration of the railway right-of-way (i.e. track and related infrastructure), type of signal and communication (control) system, the size, location, and type of railway yard/terminal facilities, number and type of locomotives and rail cars available for service, number and type of railway repair and maintenance facilities/equipment, availability of financial capital, and railway operating/management policies. Among the more significant demand side factors affecting railway system capacity are: volume of traffic, mix of traffic, and the seasonal variations of traffic flows (i.e. peaking).

Physical capacity is usually defined as the maximum throughput (i.e. total volume of traffic that moves over a railway line) that can move over a particular segment of railway line during a given period of time. Either all traffic movement over the railway line will cease because of a congestion induced breakdown of the operation or the additional throughput will form into a queue until it can move over the railway line at a later time. This concept of railway capacity defines some absolute maximum in a physical context. As the absolute physical capacity of a railway is reached, marginal cost will begin to increase rapidly and approach infinity.

There is, however, a problem with defining capacity in this manner. As the capacity of the component is approached by the level of throughput, the average delay per unit of throughput grows very large. Thus, physical capacity represents the maximum level of throughput that can be squeezed through the system, sub-system, or component per unit of time. Alternatively, a practical capacity can be chosen at a lesser point of throughput so that the delays in the movement of throughput are still in the tolerable range. See Figure 2.2.

The other measure of capacity is one defining the economic capacity of a railway. Economic capacity defines the level of throughput at which any additional throughput that moves over the rail line will move through at an ever increasing cost per unit (i.e. marginal cost). This approach defines capacity as the point on the range of possible throughput levels where marginal cost (MC) intersects the short-run average cost curve (SRAC) at its minimum. This concept is useful because it has significant

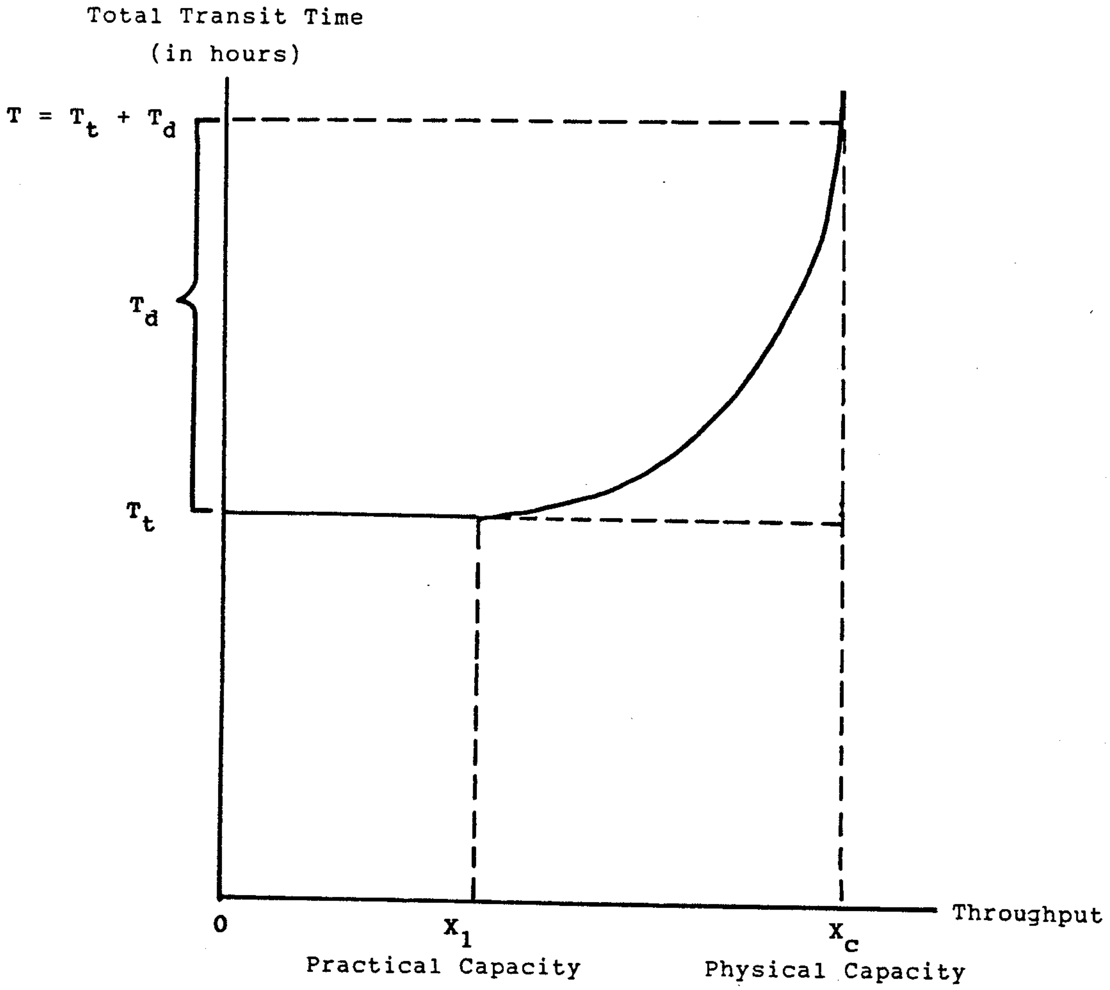


Figure 2.2: System Component Capacity and the Total Transit Time of Throughput in the Presence of Congestion

implications for efficiency in resource allocation. The constraint imposed by the restriction that capacity is the point where short-run cost is minimized produces a useful indicator of economic efficiency. See Figure 2.3.

The distinctions in the definition of capacity from an economic and physical perspective also reflect differences in their respective approachs to planning. Engineering planning primarily deals with changing the physcial aspects of the railway network to affect changes in the level of physical capacity. The engineer is concerned with identifying tangible options that are technically feasible given existing constraints. Although these constraints include economic cost, this process can by no means be considered economic planning.

The concept of economic capacity is relevant to the economic planning perspective. Economic planners are not only interested in providing an adequate supply of railway capacity for the future, but they are also concerned with identifying the potential revenues occuring from an increase in the demand for that capacity. An increase in economic capacity will result in reduced costs per unit of output, while increasing the ability of the railway to carry more revenue traffic. Greater streams of revenue will be generated from this additional traffic. The purpose of economic planning is to anticipate whether these revenue increases, occuring from an investment in capacity, will

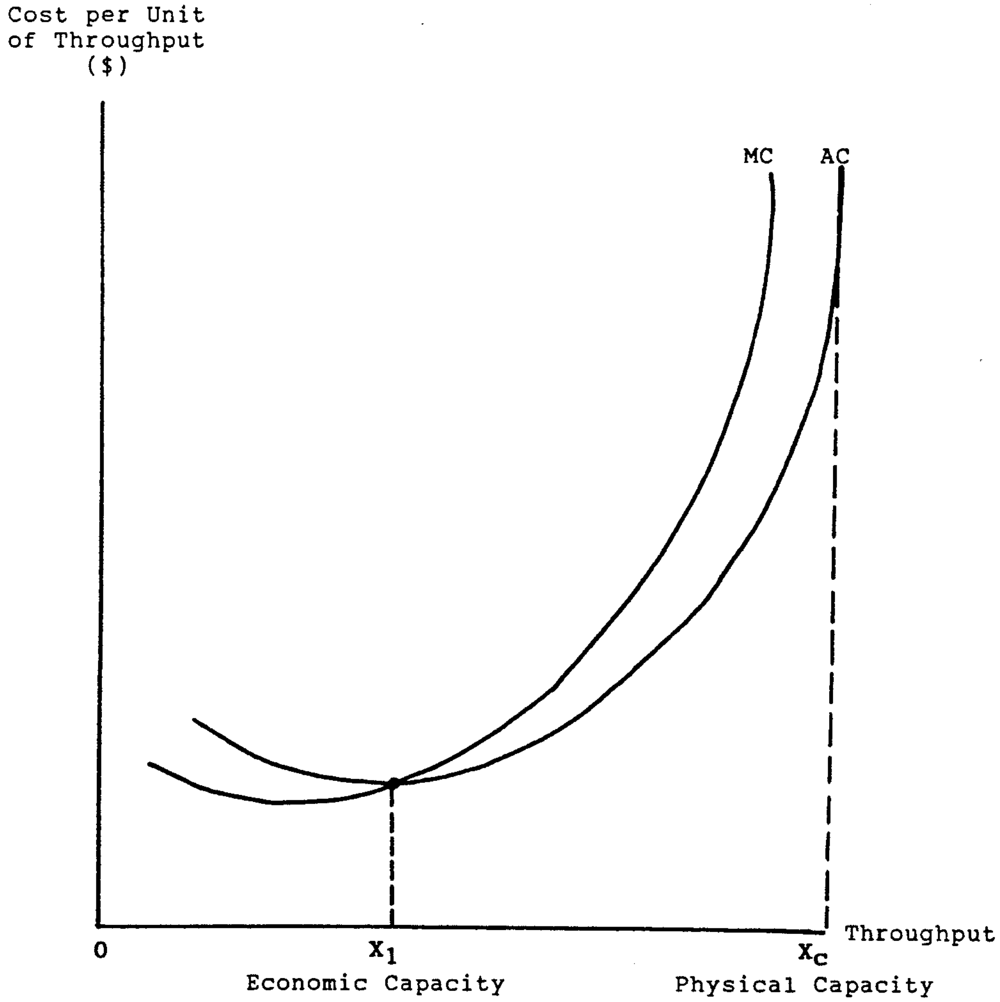


Figure 2.3: Economic Capacity

ensure the viability of that investment over time. Planning is not truly economic unless it anticipates the economic viability of an investment over time.

The concept of congestion is very closely related to the notion of capacity. Congestion occurs when the average transit time of a unit of throughput moving through some component of the railway begins to increase. The presence of congestion is a function of two factors. The first one is that any given component of a railway system is designed to provide only a fixed amount of throughput capacity. The second factor is that the movement of railway traffic is a process that has a stochastic character; that is to say, there is some degree of randomness in both the demands placed on the process to move the traffic, and the ability of that process to meet those demands.

Figure 2.2 represents the significance of congestion for the railway system component discussed above. The throughput range between 0 and X_1 is the uncongested area of production. For any unit of throughput occurring between 0 and X_1, it takes an average of T_t hours to move through the component. However, any unit of throughput that occurs beyond X_1 will take an ever increasing amount of time to move through the component.

In general, the total time (T) it takes to move a unit of throughput through the component can be defined as the total transit time (T_t) plus a total delay time (T_d).

This can be shown as:

$$T = T_t + T_d$$

Where some degree of randomness exist in both the demand for and the ability of the railway system component to move traffic, T_d is an increasing function of the level of throughput $X_1$to system capacity X_c:

$$T_d = f(X/X_c)$$

Referring back to Figure 2.2, congestion occurs when the average transit time per unit of throughput increases because of the level of demand for the system component. That is to say, congestion occurs when T_d is greater than 0, or alternatively, when T is greater than T_t.

The capacity of a railway system, sub-system, or sub-system component, can be best comprehended by referring back to Figures 2.1 and 2.2. As the level of throughput X (i.e. demand) increases, approaching the level of capacity X_c, the delay time T_d increases. Any level of throughput beyond X_c cannot be moved, and the queue will build up towards infinity if the demand level remains constant. Should the demand level drop in a future period, the queue will start to disperse when the level of throughput X drops below X_c.

There are two different types of congestion that can occur in railway systems (Manheim 1979). Load-independent congestion occurs when transport system performance is negatively affected by the interaction of the railway system components, even if the system is not utilized. Railway rolling stock that moves along a railway line can experience congestion even if there is no freight being carried on the cars. In this situation it is the demand for infrastructure capacity by the railway vehicles that causes congestion rather than the amount of freight.

Load-dependent congestion describes the situation where the performance of the railway system is degraded by the flow volume of the freight loads. Therefore, if the amount of freight carried is zero, no degradation of the railway system's performance will occur.

In the case of load-independent congestion there are two specific types of congestion that can occur--vehicle-facility congestion and vehicle-schedule congestion. The former occurs when some railway facility, whether guideway or terminal, exhibits the effects of congestion. Railway terminals are designed with some service rate at which they can process incoming and outgoing vehicles. Similarly, railway right-of-ways have specific service rates at which railway rolling stock can move over them. Therefore, as the volume of rolling stock attempting to move through a facility, or over a railway line, approaches capacity, the interaction between vehicles will cause reductions in the average transit time and thus increase the liklihood of delays. Vehicle-facility congestion will occur regardless of whether or not the railway vehicles are full or empty.

In this case, the demand is caused not by the amount of freight carried, but by the number of vehicles.

Vehicle schedule congestion is the type of congestion that arises when the number of trips specified in the schedule is high relative to the amount that can be provided by the existing fleet of railway rolling stock.

Where load-dependent congestion exists, there are two specific types of railway congestion that may occur - load-vehicle congestion and load-schedule congestion. Load-vehicle congestion arises when a stream of railway vehicles (i.e. a train) are moved through a terminal where freight is waiting to be loaded. The amount of time it takes for the freight to be loaded has two separate components; the waiting time for the train to arrive (after the arrival of the freight at the terminal) and the additional time needed to await an available rail car with enough space to take the load. The delay in moving the freight load is a function of the probability of finding an available empty rail car. The delay may also be due to the fact that competing demands exist for available vehicle capacity.

Load-schedule congestion arises when the time allowed for loading takes more time than originally scheduled. The time that is allowed for loading and unloading freight from railway cars at terminals is a key element of railway scheduling.

2.4 Capacity, Congestion, and the Economics of a Railway System

System capacity and congestion impact significantly upon the economics of railway system operation. Railway capacity and quality of service are joint products that yield significant cost implications for both the railway operator and user. Examples of user costs, excluding the obvious freight rates and fares, would include such items as incurring the additional cost of holding inventory that cannot be moved to market because of inadequate railway capacity, or of the costs they pay because of delays that occur during the shipment of their goods. These costs are usually not easily determined. Railway system capacity analysis, therefore, tends to explicitly consider the costs that are relevant to the railway operator along with some minimum level of service that is acceptable to the user (Khan 1979).

For any definable railway system, sub-system, component, or element, physical performance will vary with the level of throughput. The concept of capacity, from the perspective of railroad engineering, can be defined either as the maximum throughput possible, or a more realistic "practical" capacity. The practical capacity can be defined as the maximum level of throughput that is possible given some specified level of delay that is tolerable to the users (Khan 1979, 41-44).

It is generally recognized by railway analysts that capacity, cost, and level of service are interrelated with one another. Khan states that:

> It appears that railway analysts are moving towards defining capacity in economic terms as well as in physical terms. Reasons for such a development include the recognition that capacity, service, and cost have a logical interrelationship, and the desire to make the trade-offs explicit. (1979, 146).

An economic definition of capacity should include the trade-offs that are made between capacity, cost, and the level of service. It is clear that the more costly railway systems, sub-systems, components, and elements generally have a greater throughput potential than their less costly counterparts. For any given system or component that is rapidly approaching its maximum throughput capacity, any additional increase in throughput will result in an increase in direct operating costs plus additional costs incurred by a reduction in the productivity of the railway rolling stock.

The Canadian Transport Commission states that railways typically are large and highly complex enterprises whose cost structure is greatly affected by:

> traffic environment -- factors that determine the types of traffic carried and the conditions under which they are carried, including seasonal and cyclical fluctuations, and traffic imbalances
>
> economic environment -- the condition of supply of factors of production, including investment capital
>
> physical environment -- terrain, reflected in gradient and curvature of the track, tunnels, etc.,

and climatic conditions. (1983, 7).

Bonsor (1984) writes that railway costs are affected by these factors: variable costs; economies of scale; economies of density; joint and common costs; length, volume, and weight of haul; and the indivisibility and lifetime of capital.

Railway system costs can be clearly broken down between two separate categories--infrastructure and vehicles. The costs associated with infrastructure are for the most part, fixed costs (i.e. costs that do not vary with the level of output). Variable costs (i.e. costs that vary with the level of output that are associated with the maintenance and operation of the infrastructure tend to be small in relation to the capital cost incurred by its construction (Thomson 1974).

Infrastructure costs are much more relevant to the long-run cost function. Short-run cost functions exclude capital costs, but include current costs. Current costs for railway infrastructure include a high proportion for overheads. Overhead, or fixed costs, represents those costs that are incurred for the purpose of maintaining the railway system's ability to function and are not related to the volume of traffic.

The fixed cost of railway infrastructure has two components--the interest cost on the fixed capital and the fixed cost of maintenance. According to Bonsor:

> Keeler (1974) and Harris (1977) have shown that such costs interest costs of fixed capital represent less than 20 percent of total fixed costs. Second, the presence of track and structures implies some fixed maintenance expenditures for track, bridges, yards, switches, and so forth. (1984, 94-95).

Bonsor goes on to say that he does not imply that railway maintenance expenditures are independent of traffic volumes, but only that there is some minimal level of expenditure that is not related to volume. He goes on to say:

> There is general agreement that the percentage of short-run (non-capital) costs that are variable with output depends on the relationship between available capital and utilized capacity. A railroad - or a specific rail route - which is operating at only a very small fraction of available capacity will have a higher percentage of total noncapital costs fixed than will a railroad - or specific route - which is operating at near full capacity. Thus, the short-run marginal cost of increasing the volume of traffic will be lower for the former than for the latter. (1984, 95).

The literature on railway costs has traditionally suggested that the industry was characterized by decreasing cost and, therefore, increasing economies of scale in the production process. Conventional wisdom also concluded that railway operations were characterized by high fixed costs and low variable costs per unit output. These assumptions have been challenged by the results of a number of recent studies. These studies have attempted to determine whether or not the railroad industry is characterized by increasing, constant, or decreasing returns to scale. Harris (1977) showed that the long-run average cost curve of a railway firm decreases (i.e. increasing returns to scale) as the total number of track miles operated increase. However, Keeler (1974) and Caves, Christensen, and Tretheway (1978) indicated that the railway industry is characterized by constant, and not increasing, returns to scale. Bonsor concludes that all of the recent studies conducted on the subject have shown that economies of scale are either absent or very modest is size (1984, 93).

Railway system economies of scale defines the cost behaviour related to its overall size. System economies of scale are due to the following factors, according to a Canadian Transport Commission (CTC) report on competition and regulation in the Canadian railway industry:

> (i) economies of system-wide support facilities or services; (e.g., equipment maintenance and repair shops; computerized traffic and car control, etc.);
>
> (ii) economies of pooled resources; (e.g. better utilization of larger car fleets or locomotives employed in a number of related but distinct markets);
>
> (iii) advantages due to an ability to provide a wide range of services, or services to a number of market, or single-carrier handling of customer's needs;
>
> (iv) lower costs of access to capital markets,
>
> (v) economies in dealing with government agencies. (1983, 14).

The presence of economies of density is another determinant of railroad costs. This concept is the essence of the economic definition of railway system capacity. Economies of density, in contrast to economies of scale, occur when the average cost per unit decreases while output increases, holding all other factors constant (including firm size). Several studies, based on U.S. data, have shown that the railway industry is characterized by significant economies of density. Keeler determined that U.S. railways in the late 1960s exhibited:

> substantial unexploited economies of traffic density for most railroads, but constant long-run returns to scale . . . (1974, 207).

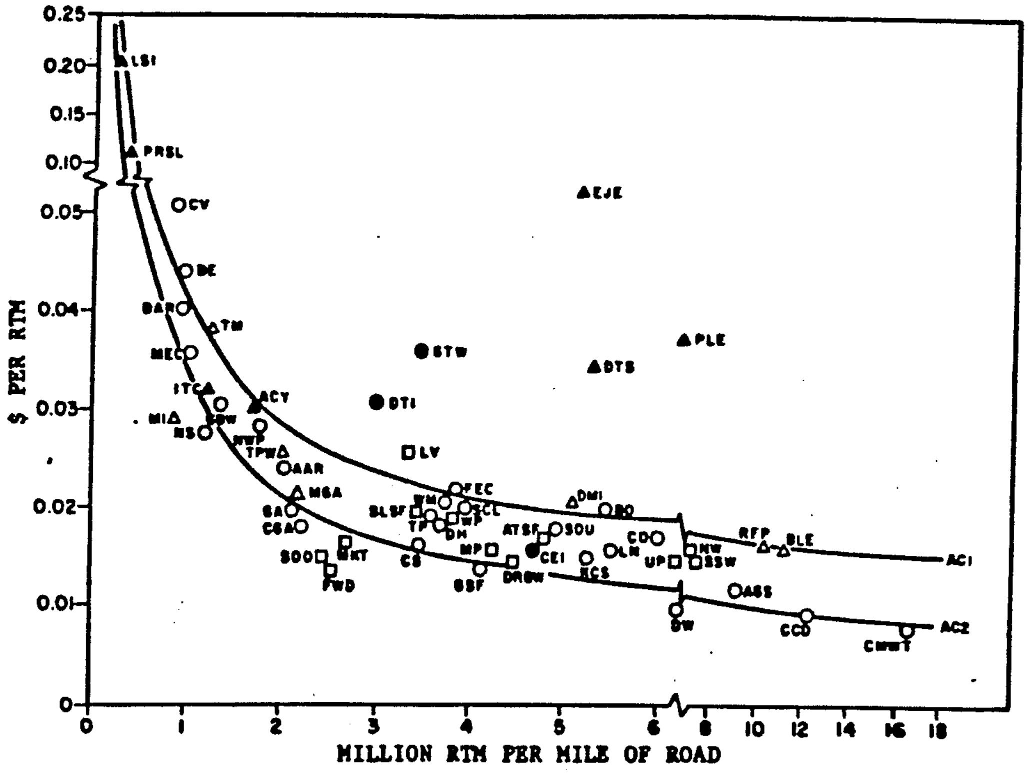
Harris found that the evidence strongly suggests that significant economies of density exist for U.S. railroads (1977, 556).

Figure 2.4 is a reproduction of the important diagram in the Harris study (1977, 562). His regression analysis showed that about two thirds of the economies of density are due to high fixed operating costs per mile of road. This cost can be attributed to the necessity of providing the installed capacity along with some minimal amounts of vehicles, crews, other labour, etc. According to Harris, the maintenance of way and structure expenses category accounted for some 15 percent of the economies of density. The transportation expense category accounted for 39 percent of the economies of density.

Levin also reached conclusions similar to Harris. He states that:

> the burden of excess capacity is a primary source of

THE AVERAGE COST DENSITY RELATIONSHIP
WITH 1973 DATA POINTS



AC$_1$: AC curve based on Reg. No. 3.
Nonurban, ALH = 100 miles.

AC$_2$: AC curve based on Reg. No. 3.
Nonurban, ALH = 300 miles.

△ Firms with ALH < 100 miles.
○ Firms with 100 < ALH < 300 miles.
□ Firms with ALH > 300 miles.
▲● Urban firms

Figure 2.4: Economies of Density in the U.S. Railroad Industry
(reproduced from CTC 1983, fig. 1)

> the unprofitability and the sluggish performance of the U.S. railroad industry. Analysis of recently compiled D.O.T. line segment data revealed that nearly two-thirds of the nation's Class I railroad mileage is below or barely above the break even level of traffic density at current rail rates, and a substantial fraction of these lines fail to produce revenues sufficient to cover variable costs. (1977, 40).

The Canadian Transport Commission claims that their research has come up with similar results, and:

> indicates significant differences in cost structure between high and low density lines and supports the results of the U.S. research. (1983, 13).

This CTC report also states that the following factors explain the existence of economies to traffic density are due to the following factors:

> (i) larger and specialized cars (up to the technically determined limits;
>
> (ii) larger and heavier trains (better utilization of motive power and crews; better line capacity utilization);
>
> (iii) savings achieved in classification through humping (hump yards are more efficient but require higher investments which only a higher volume can justify);
>
> (iv) savings due to terminals' mechanization;
>
> (v) heavier rail and ballast permit higher weights and speeds, but require high investments per km of line; and,
>
> (vi) better utilization of structures and supporting systems. (1983, 13).

The CTC goes on to say:

> System economies of scale and economies of traffic density may, but need not, co-exist (a relatively small and specialized railway may enjoy benefits of high density traffic). The current trend to railroad consolidation in the United States appears to confirm

> the existence of system economies of scale (at least, railway managements believe it to be so), but it can also be explained by the marketing or competitive advantages of comprehensive networks capable of providing one-carrier services to a greater number of destinations, or to the potential for the exercise of monopolistic power. (1983, 14).

The report quotes Fuss and Waverman that in the case of the multi-product railway firm:

> there exists no unambiguous measure of output-specific returns to scale except in the case of non-joint production since separate cost functions cannot be constructed when common costs exist. Thus we appear to be left with only the overall measure of returns to scale. Unfortunately, this measure is of no value when one is attempting to evaluate the possible efficiency gains from increasing the scale of production of one of the outputs in a multi-product production process. (1978, 10-11).

The process of producing transportation services by the railway system yields joint and common costs. For example, the movement of a rail car from point A to point B also produces an additional return movement of the car from point B back to A. Thus, the total cost of the trip from point A to B includes the additional cost of moving the car back to A. It is not possible to exclude the cost of the return movement from the total cost of the original movement from A to B, thus giving rise to joint costs. Joint costs arise when the production of a good or service results in the production of another good or service. Common costs occur in the situation where facilities that are employed to produce one good or service can be employed to produce other goods and services. The railway lines of the system carry a large number of different trains, all carrying different cargos from and to different origins and destinations. An example of this would be the use of the railway infrastructure to produce movements of different products. Such bulk commodities as wheat, coal, sulphur, etc. that are moved by the railway share the same track, terminal, and yard facilities as such finished products as furniture, machinery, automobiles, etc. For some parts of the railway system, line haul costs can be directly attributed to a specific type of service (e.g. grain dedicated branchlines in Western Canada). In contrast, common linehaul costs can only be associated with the total of all the services provided by that part of the system. It is not possible to measure with any precision the magnitude of either the common costs or joint costs of any specific service provided by the railway system (Bonsor, 92-93).

The CTC research report on competition and regulation in the Canadian railway industry states that:

> costs incurred in one part of the system may be related to the degree of utilization of another one, or to the pressure on the system as a whole.
>
> The pricing implications of 'related' (joint and common) costs have been recognized by economists for the better part of this century, with the majority view being that 'related' costs should be allocated according to elasticities of demand of the services in question. (1983, 15-16).

More recently, attacking the problem from the point of view of modern welfare theory, W. J. Baumol and D. F. Bradford states that:

> the social welfare will be served more effectively not by setting prices equal or even proportionate to marginal costs, but by causing unequal deviations in which items with elastic demand are priced at levels close to their marginal costs. The prices of items whose demands are inelastic diverge from their marginal costs by relatively wider margins. (CTC 1983, 16).

It has been established that for any type traffic carried over the railway system, the average cost per revenue ton-mile of transporting it declines as the line haul distance increases. Figure 2.5 illustrates the relationship between average cost and distance. This relationship is explained by the fact that for every movement, fixed costs are incurred for certain related "activities". These activities include loading, unloading, and switching operations that occur in the terminals and yards. As the line haul distance increases, the portion of total costs that are fixed becomes smaller.

Similarly, important economies exist with respect to the weight and volume of the shipment. Shipment weight refers to the tonnage loaded upon a single rail car, while shipment volume is usually defined as the total number of loaded cars. Economies with respect to shipment weight and volume exist in terminal, switching, and line-haul operations. Bonsor claims that the existence of these economies have led to increased usage of unit trains. He

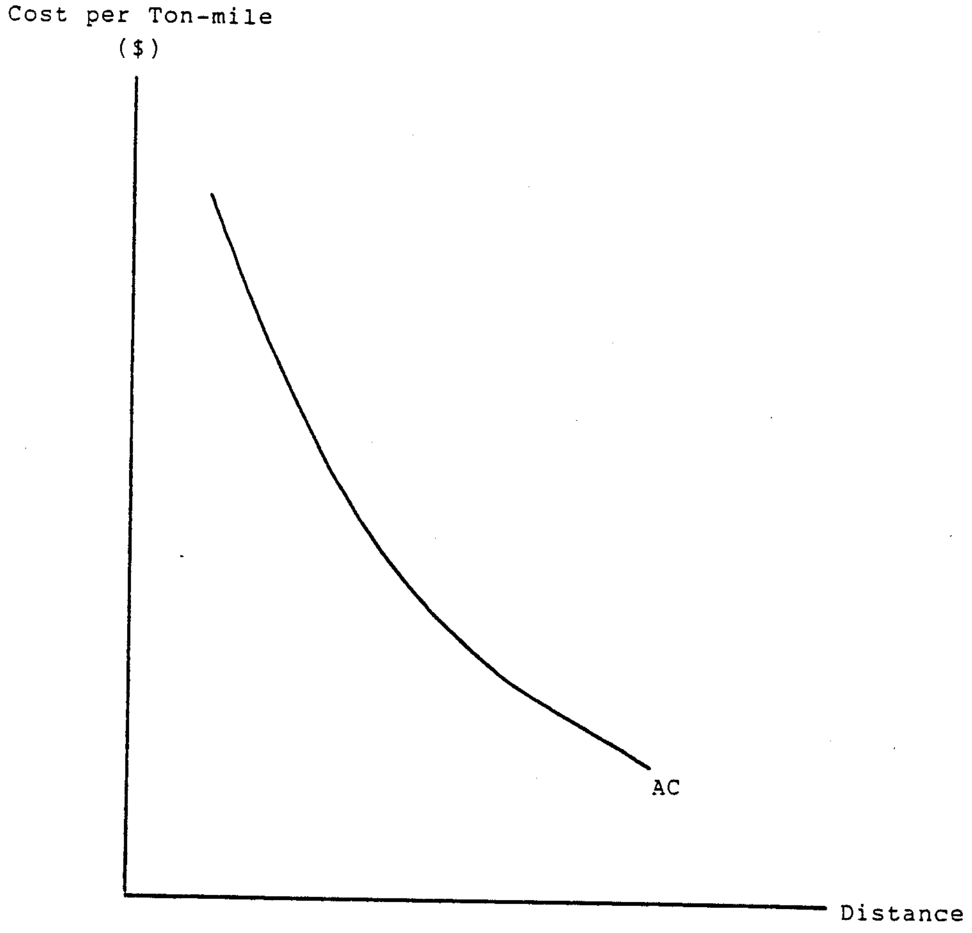


Figure 2.5: Average Cost and Length of Haul
(from Bonsor 1984, fig. 6-2)

defines a unit train as:

> a set of cars and engines which move from A to B loaded and from B to A empty, on regulary scheduled trips. The train is not "broken up" at any given time, and thus there are substantial economies in switching and other operations. In general, unit trains are employed in the movement of commodities such as coal, iron ore, and grain. (1984, 96-97).

Another determinant of the structure of railway costs is the indivisibility and lifetime of capital. Railway infrastructure is characterized as being indivisible in nature. Railways, unlike the trucking industry, must provide the right-of-way and the associated track and structure in order to provide railway service between different points. With the technology available at present, these ways and structures are not divisible (i.e. if 500 km of track and associated structures is required to provide railway service between two points, it is not possible to provide this service with anything less than 500 km of infrastructure). The provision of railway infrastructure is extremely expensive and requires high volumes of traffic to economically justify its construction. These constraints provide an effective barrier to the entry of new firms into the market. Bonsor writes that:

> because rail carriers provide their own right of way and because efficient operation requires high volumes of traffic, the rail industry is not one we would expect to be typified by a competitive market structure. (1984, 92).

Similarly, Kneafsey suggests that the U.S. railroad industry does not act in a competitive market, and that the market structure of the railroad industry is one of oligoply (1975, 12).

The capital stock of the railway, which consists of infrastructure vehicles, is characterized by having a long lifetime relative to the capital stock of other modes. According to Bonsor:

> It is generally assumed that the track and terminal components of the capital stock have a life-length of approximately 60 years, and the rolling stock a life-length of 20 to 35 years. The capital stock of the industry thus has a very long life-length, especially when compared with the producer-provided capital stock in other modes. (1984, 92).

2.4.1 The Demand for Railway Transportation

The demand for railway transportation services can be characterized as being a "derived" demand. By this we mean that the demand for railway services is derived from the demand for the products that are carried by the railway. For example, the demand for the railway movement of coal from northeastern B.C. to terminals in Vancouver is a result of the offshore demand for coal from export markets.

The elasticity of demand for railway services defines the magnitude of change in demand that occurs when the price of the service changes. Because the demand for transportation is a derived demand, the elasticity of demand for railway transportation is dependent upon the elasticity of demand for the product that is carried. The elasticity of demand for different commodities that are transported by rail vary across a wide range. Bonsor states that:

> the price elasiticity of demand of rail services will vary across commodities and across markets. With respect to manufactured goods, all recent studies of transport demand have shown that the demand for rail services, given existing levels of trucking rates is relatively inelastic over a wide range of prices.
>
> Across individual commodities and transport markets, Levin (1978) reports price elasticities of demand ranging from a low of -0.06 to a high of -5.97. On average, however (at average rates and average service differentials), the price elasticity of demand was a low -0.4. Oum (1979), using Canadian data, reports elasticities in the range of -0.458 to -1.19. (1984, 97).

Manufactured goods exhibit a great deal of demand inelasticity for railway transportation services over a wide range of freight rates. This situation, however, does not reflect the characteristics of transport demand for such relatively low value bulk commodities as coal, potash, sulphur, etc. These commodities are characterized as having a high elasticity of demand with respect to transport. The reason for this, according to the CTC is that:

> the elasticity of demand for transport depends on the proportion of transport costs to total costs; the higher this proportion, the higher the elasticity of demand for transport. This has been well understood by the railways' management as is evidenced by 'commodity rates' (lower rates on 'lower value' commodities). (1983, 2-3).

The transportation costs of such commodities is usually about 30 to 40 percent of the final delivered price (Levin 1981; Bonsor 1977). The transportation cost of manufactured goods, in contrast, is typically a very small percentage of the price of the product.

The high degree of demand inelasticity for the transportation of manufactured goods implies that when railway freight rates increase, there would not be a large decrease in the volume of these goods carried by rail. We would expect, in contrast, that a significant share of the rail traffic for bulk commodities would shift to other lower cost modes in the event of an increase in railway freight rates. However, the railways enjoy something approaching monopoly power with respect to the transportation markets for a number of bulk commodities. This is due to the fact that for many of these commodities there exist no economically viable alternative modes of transportation. In Western Canada, for example, it would not be economically feasible to make long-haul movements of wheat from Saskatchewan to ports in B.C. by truck. The volumes and distances involved in these movements make railways the only viable way of moving these commodities to market. In the absence of intermodal competition, there are two factors that act as constraints upon the freedom of the railway to set rates. The first factor is that many bulk commodities face a very elastic curve. Any significant increase in the price of transporting these commodities to market will result in a significant increase in their market price. These price increases, given elastic demand, will force consumers to find substitutes, thus reducing the demand for the commodities. The reduction in the final demand for these commodities will lead to a decrease in the demand for railway services to move them. The second factor is that the producers of many bulk commodities are large companies that often dominate the market. These individual companies contribute significantly to the total revenues of the railway and, as such, are often able to negotiate an acceptable mix of rates and services.

In a situation where there is competition from other modes, the shipper often chooses to pay more for a more expensive carrier to move his goods. This is due to the fact that the direct cost of transportation (i.e. the freight rate) may not reflect other costs incurred by the shippers and receivers related to the movement, handling, and storage of the product. According to the CTC:

> The extent to which these indirect costs are minimized by a transport service is often reflected in the notion of service quality, which refers to such features as: (i) loss or damage in transit (both direct and consequential costs); (ii) handling (including packaging, as well as costs of receiving or shipping); (iii) documentation; (iv) warehousing and storage (slower and/or less frequent or reliable schedules increase storage requirements), etc. (CTC 1983, 4).

The relative speed of transit is not an important determinant of the choice of mode for low-value bulk commodities (Oum 1979). However, the transit time for higher value products is often a more crucial determinant of modal choice than price (Levin 1978). This is true for most higher value products, and is particularly important for products that have a limited lifetime (e.g. fruits, vegetables, fresh fish, etc.). We would expect these products to exhibit a relatively high degree of demand elasticity for transportation with respect to the time of transit. Thus, greater quantities of these products would be carried by the mode(s) that could achieve the greatest reductions in transit time.

The notion of service quality affects the demand curve for railway services. Bonsor writes:

> The demand for railroad services at any given price and set of service characteristics - speed and reliability - is conditional on the prices and service characteristics of other transport modes and the total size of the transportation market. An improvement in the speed and reliability of rail movements, holding all other factors constant, will cause the demand curve for competing modes to shift inward. An increase in the price charged by other modes will have a similar - qualitative - effect. An increase in the speed and reliability of other modes will cause the railroad demand to shift inward. (1984, 99).

2.5 Summary

The preceeding discussion has higlighted the structural, operational, and economic factors that impact upon railway system capacity. It is clear that the structure of modern railway systems are complex in nature, and are characterized by a number of components that interact with one another to produce transportation services. These components all have separate functions within the system, and are all comprised of a number of different elements. The type, number, location, size, availability, and condition of these elements all impact on the capacity of the component to function. The capacity of the component will, in turn, affect the overall capacity of the railway system or network to move traffic. The physical capacity of the railway system is defined not only by the capacity and performance of each component, but also by the interaction of the components with each other.

This chapter has also explored the concept and definition of railway system capacity. The physical definition of capacity relates to the ability of the railway system to facilitate the movement of throughput over it during some given period of time. The physical capacity of the system can be quantified by the relationship between the level of throughput and its total transit time. The level of throughput where the total transit time approaches infinity defines the level of physical capacity for the system or component.

This level of physical capacity can also be affected by the presence of congestion. Congestion will occur at some point on the feasible range of throughput where total transit time of throughput begins to increase. This point defines the level of pratical capacity for the system or component, and occurs before the level of physical capacity.

Economic capacity is defined by the relationship between average and marginal cost and the level of throughput. The level of throughput that occurs where marginal and average costs are equal is the point of economic capacity. This point always occurs before the level of physical capacity (i.e. where average cost per unit of throughput approaches infinity).

The factors affecting the relationship between capacity, congestion, and the economics of a railway system were deliberated at some length. Capacity and congestion were determined to have significant impacts upon the economics of the railway. These impacts relate to the characteristics of the railway cost function. The railway cost function is characterized by significant economies of density due to high fixed costs. These high costs reflect the resources needed to build and maintain the fixed railway infrastructure. The provision of this infrastructure is extremely costly, and requires large volumes of traffic to economically justify its existence. Thus, for a system where economies of density can no longer be achieved, it would be necessary to expend large amounts of capital to upgrade or expand the existing infrastructure in order to increase capacity.

Capacity also has impacts on the demand side of the equation. Capacity constraints lead to a reduction in the quality and level of service. This results in a demand shift away from the railway to move certain products. Lower value bulk commodities tend to have a more elastic demand for transportation, but are not easily shifted to other modes.

The next chapter provides an overview of the specific factors and options that are available to change the level of system capacity, within the overall framework of railway planning.

CHAPTER 3

ECONOMICS, PROBLEMS, AND OPTIONS IN RAILWAY INVESTMENT PLANNING

3.1 Introduction

The previous chapter specified the structure, operations, and economics of the railway system, and outlined the concepts and definitions of capacity. This chapter will take a closer look at the individual factors that can affect railway capacity. This overview, however, will be within the context of overall railway investment planning. Planning is a process which interrelates decisions taken over time in order to achieve desired future outcomes. In order to understand the factors that affect the level of railway capacity, and the implications of the options available to change it, one must recognize the economics and issues implicit in the planning process.

The next two sections of this chapter provide an outline of the economic and other issues relevant to investment planning for railway capacity. This discussion will point to the various economic and investment problems that must be taken into account by railway planning. The final sections describe the specific factors that affect railway capacity, and the various types of planning options available to increase capacity. These options are grouped according to their cost, benefit, impact on capacity, and consumption of time needed to implement them.

The purpose of this chapter is to clarify the issues and options that have to be considered within the context of general investment planning for railway capacity. Specifically, this discussion should provide insights into how and why CN Rail and CP Rail invested billions of dollars for railway capacity expansion in Western Canada during the 1980s.

3.2 Economic Considerations for Railway Capacity Investment Planning

It is recognized that congestion and capacity have significant impacts upon the economics of a railway operation. Railway capacity and quality of service are joint products that have important cost implications for both the railways and the users. User costs tend to be extremely difficult to quantify, and therefore it is desirable in capacity analysis to explicitly consider costs relevant to the railways and a minimum service level that is acceptable to the users.

For any definable railway system, sub-system, or component, physical performance will vary with the level of throughput. The concept of capacity, from the engineering perspective, can be defined either as the maximum physical capacity or the practical capacity. The practical capacity is defined as the maximum throughput level which results in some tolerable level of delay, or alternatively a minimal level of quality of service to users.

Railway analysts recognize that capacity, cost, and level of service are interrelated. Therefore, there has been a trend towards defining capacity in economic terms. Such a definition would make the trade-offs between capacity, cost, and level of service explicit.

It is obvious that costlier systems, sub-systems, or components have higher throughput potential. For any given system or component that is rapidly approaching its maximum physical capacity, any increase in throughput will result in disproportionately higher service deterioration (i.e. delays). Service disruptions and delay will, in turn, cause an increase in operating costs and a reduction in productivity.

It is generally recognized by railway analysts that train delays will result in increased operation costs due to increases in the direct cost of operations and additional costs that are incurred due to reduced productivity of the rolling stock.

Capacity expansions through capital investments in railway plant can reduce operating costs. Gains in throughput may also occur by changing the operating environment (e.g. traffic mix).

The costs of a railway system can be broken down between two separate categories--infrastructure and vehicles. The costs associated with infrastructure are for the most part, capital costs. Current costs that are related to maintenance and operation of the infrastructure tend to be relatively small.

Infrastructure costs are much more relevant to the long-run cost function. Short-run cost functions exclude capital costs, but include current costs. Current costs for railway infrastructure include a high proportion for overheads. Overhead represents those costs that are incurred for the purpose of maintaining the railway system's ability to function and are not related to the volume of output.

In the long run, the railway operator must determine what are the infrastructure costs that must be incurred in order to meet future traffic demands. Generally speaking, economies of scale exist in railway infrastructure. However, it is useful to make some distinction between scale and growth. Economies of scale are related to alternative scales of investment that are made at a specific time. They are not necessarily related to changes in the scale of investment over time. The expansion of railway capacity is invariably more costly than if the same capacity were provided from the start. This is due to the fact that the construction of additional railway capacity, even if conducted in carefully planned stages, involves the input of additional labour, materials, and time. The economic justification of expanding capacity in stages lies in the capital saving due to the postponement of investment.

In the short run, because little can be done to alter the characteristics and capacity of the infrastructure, the marginal costs of railway transportation consist mainly of the costs of providing and moving rail cars. These costs are greatly influenced by the characteristics of the railway infrastructure. Specifically, the design, quality, and capacity of the infrastructure impact upon the cost of moving railway freight traffic.

The type of railway locomotives and cars used, their performance and operating costs, are as much a reflection of the design of the existing railway infrastructure as the latter are of the design of vehicles. Because so much of the railway system in Canada was laid down many decades ago, constraints exist on the type and number of rail locomotives and cars that can be utilized. The gauge, height limits, and curvature standards were determined a very long time ago. The CN rail line that runs to

Churchill, for example, cannot move the more modern, and heavier grain hopper cars that could make this route more economically viable. This constraint is due to the fact that the Churchill line was constructed of narrow gauge rail line over 60 years ago, and the heavier grain hopper can not move along it.

Vehicle operating costs are also dependent upon the quality of the existing infrastructure. The two important characteristics of quality are speed and vehicle maintenance. Both of these important factors are crucial for vehicle operating costs, and are much affected by the alignment, track condition, and management of the infrastructure. As the quality of the existing railway infrastructure deteriorates, one would expect that the vehicle operating costs would increase due to increased depreciation, maintenance, and fuel costs.

Figure 3.1 shows the economics of expanding the capacity of a railway system. The cost functions on the graphs illustrate the difference in total, average, and marginal costs for different levels of system capacity expansion. The throughput level that minimizes average total cost is the level of optimum output, from an economic point of view. According to Khan:

> Capacity expansions should be carried out such that the maximum desirable throughput for a given facility is higher cost, and lower than the physical ultimate capacity. That is, capacity expansions are desirable at locations in (Figure 3.1) where average total cost

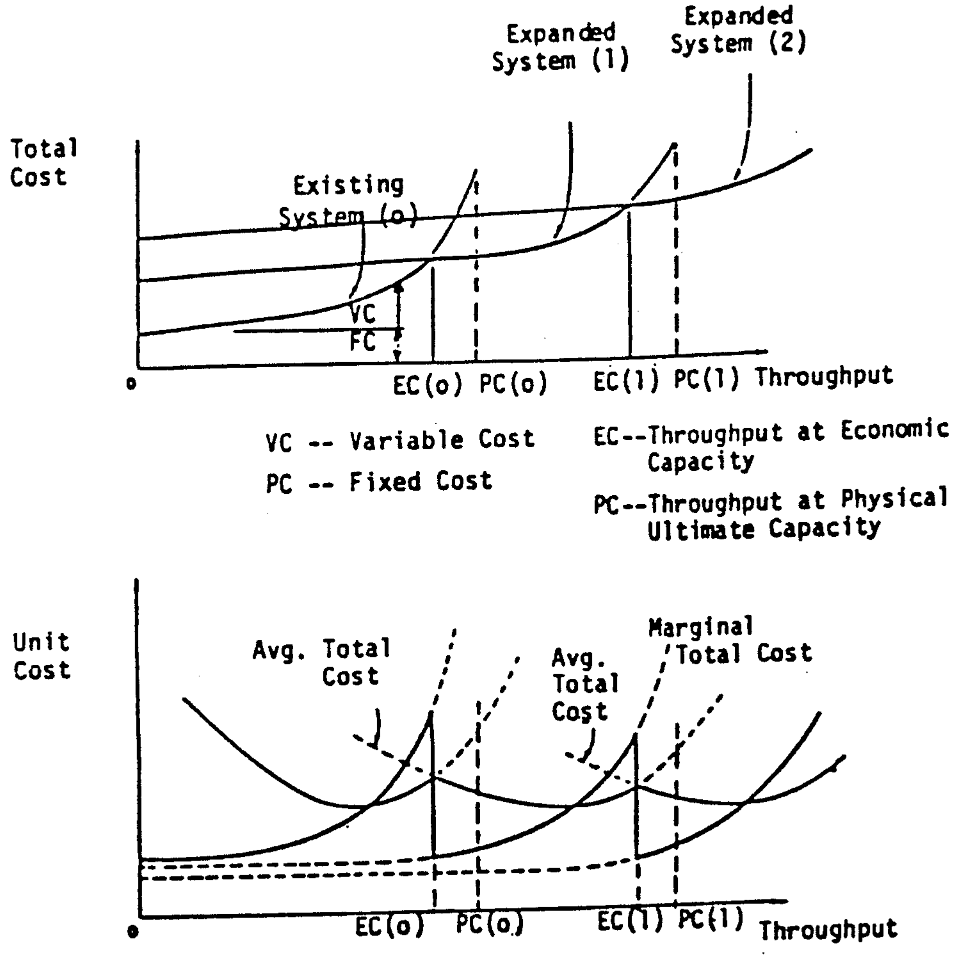


Figure 3.1: Short Run Cost Functions for Different Capacity Levels (reproduced from Khan 1979, fig. X.1)

> functions intersect. It should be noted that each facility is capable, in physical terms, of more throughput than those locations suggest. The effect of loading beyond these points in terms of increased unit costs is, however, obvious. (1979, 150-51).

Capacity costs are defined as the portion of capital fixed costs that are expended for the purpose of providing a certain level of production capacity. Khan states that:

> capacity expansion involving incremental additions to the plant according to a minimum total cost criterion defines the long run cost curve. That is, the long run total cost curve connects the points of optimal capacity for various levels of fixed cost. According to this definition, these cost curves describe optimal rather than actual behaviour of the railway . . . (1979, 150).

The capacity of the infrastructure greatly impacts upon the cost of moving railway traffic. We shall discuss the problems of railway capacity planning in greater detail in the following discussion.

3.3 Problems of Railway Capacity Planning

Two separate definitions of capacity were considered in the previous chapter--engineering and economic. These two definitions, to a large degree, reflect the differences between the perspectives of engineering planning and economic planning.

Railway engineering capacity planning is primarily concerned with affecting change in the level of physical capacity through the alteration of tangibles. Engineers, given the task of increasing the capacity of a railway system, will identify the technical options that are available to achieve the objective. This objective, reflecting the engineering definition of railway system capacity, will be specified in physical terms (e.g. a certain number of trains running over a section of track in a day). Engineering plans explicitly consider the physical and/or operating changes required in the tangible elements of the railway to affect an increase in capacity (i.e. the amount of concrete needed, the weight of the rail to be used, train operating speed, etc.). Thus, the purpose of capacity planning from an engineering perspective is to determine the physical capacity objectives, isolate the physical options capable of achieving these objectives, deciding on the appropriate options, and then implementing these options over time.

Economics does play a role in the engineering planning process. The engineering plan must not only identify the physical options that are possible from a technical point of view, but are also feasible within existing cost constraints. Thus, engineering planning must weigh the cost implications of an particular option against its technical merit. This, however, should not be confused with economic planning.

Economic planning to expand railway capacity must, like engineering planning, define objectives, options, and a course of action. However, its purpose is to affect change to a number of different economic parameters. Although these economic parameters exist in the abstract, unlike the hard and precise tangibles defined in engineering, they reflect the true purpose of the activities undertaken by the railway. Railways do not solely exist to move a certain number of trains over a specific section of track at a given time. They exist because the transportation services they provide to society meet an existing demand. This demand produces an economic return to the capital invested by the railway's owners.

Within the context of economic planning, expansions in railway capacity are undertaken for the purpose of achieving economic objectives. Capacity expansions are made in order to meet increases in expected demand. These increases in demand will translate into additional revenues to the railway. These additional revenues, coupled with declines in average cost per unit of throughput, will move the railway toward profit maximization. The economic definition of railway capacity is relevant to this planning process. Increases in the level of economic capacity will result in lower costs per unit of output, while increasing the ability of the railway to carry more revenue traffic. Over time the railway increases its profitabilty as total revenues grow, and average cost declines.

Capacity expansions, from an economic planning approach, are only undertaken if the present outlook of the future indicates a likelihood that traffic demand, and thus revenues, will increase enough to justify the initial investment required. However, capacity expansion decisions made on the basis of overly optimistic traffic demand forecasts are costly. Railways find themselves burdened with excess capacity, resulting in a negative impact on profitability.

In a market driven economy, the viability of an investment depends on its ability to economically sustain itself over time. There is no purpose in undertaking an expansion of railway capacity when there is little or no likelihood that it will generate the additional revenues and profits needed to pay for itself. The whole point of economic planning is to choose the options and investments that will meet the objective, earn a return, and remain economically viable over time. Making choices that are viable with respect to existing and potential economic forces is the essense of economic planning.

3.3.1 Time Horizons

Railway capacity planning is also concerned with determining the degree to which each of the various options available to affect capacity can be implemented over the time horizon of the planning period. Different options are variable over different time frames, ranging from days and weeks for scheduling changes, to years for the construction of new railway infrastructure.

The various options that are available to increase railway capacity can be separated into three different categories that relate to the time it takes to implement them in the railway system. These categories are as follows:

1. Short-run options: The short-run options generally reflect the aspects of the railway system that can be altered in a short period of time. These generally concern changes in the operation of specific services of the railway. They would include such things as changes in service frequency, times, prices, routes, etc. It should be noted, however, that regulatory constraints may cause a delay in the implementation of some or all of these types of options.

2. Medium-run options: Some options fall between the short and the long run. These options would generally involve changes in things such as the purchase of new and/or replacement rail cars and locomotives, changes in vehicle characteristics, fleet size, or fleet mix.

3. Long-run options: Options for change in some parts of the railway system may occur very slowly because they require a great deal of time for planning, investment, and construction. These type of options generally involve such items as the construction of new fixed facilities - such as multiple trackage, lines, and yards. These type of options often take several years to plan, design, and construct. (Manheim 1979).

Figure 3.2 shows the economic implications of these time frames for the cost function of a particular railway system. In part A of Figure 3.2, we assume that the short-run, medium-run, and long-run options are fixed at O_s, O_m, O_L, respectively. This graph illustrates how average total cost (ATC) varies with the level of volume.

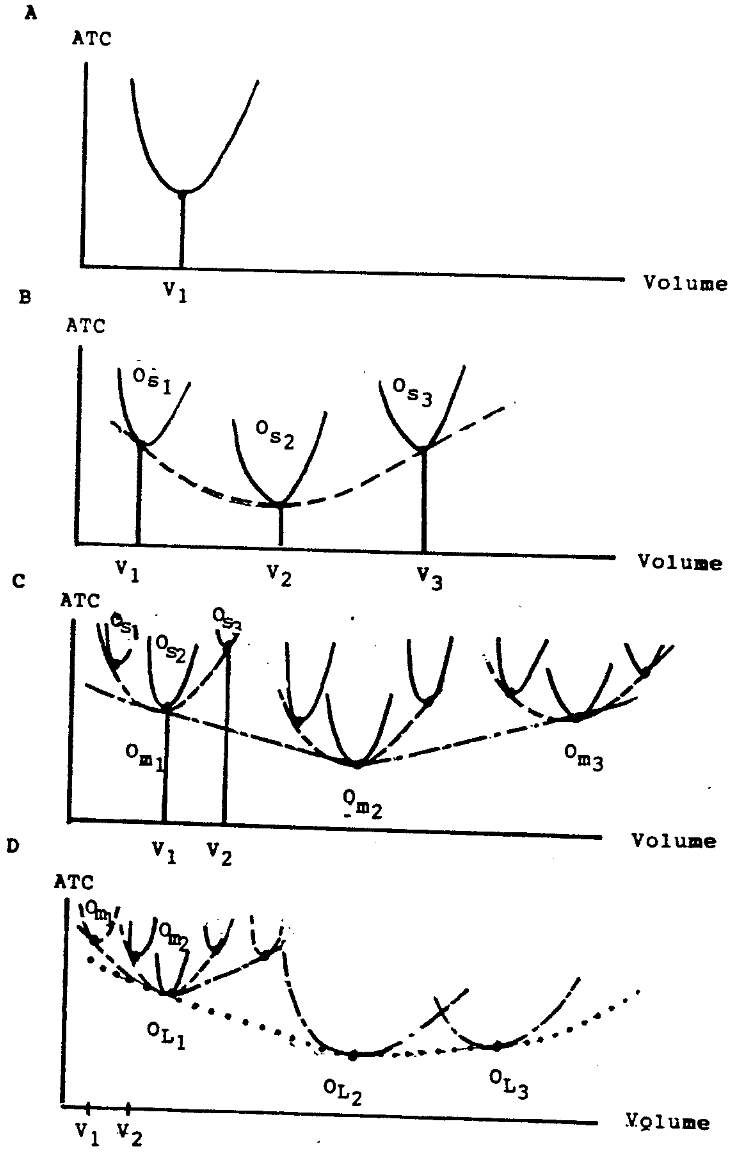


Figure 3,2: Time Frames of Capacity Expansion Options
(from Manheim 1979, fig. 7.6)

In this case, we assume the situation of both economies and diseconomies of scale. In part B, we assume that the short-run options have three different values, O_{s_1}, O_{s_2}, and O_{s_3}. In this case, the medium and long-run options are held constant. For each particular short-run option, a different average total cost curve is generated (ATC). Should the short-run options O_s be varied over a wide enough range, we can find each level of volume (V) and determine the short-run option O_s for which average total cost is minimized. Thus, a curve for average total cost is generated, illustrated by the dashed line in the figure. This curve gives the lowest values of average total costs possible for the possible values of the fixed medium and long-run options, O_m, and O_L.

In Part C the long-run options are fixed but now the medium-run options take the values O_{m_1}, O_{m_2}, and O_{m_3}; and we now let the short-run options vary over a range. For each particular point of the medium run options, O_m, the variation of the short-run options O_s generates an envelope curve of average total costs. This curve is the result of setting the points of the short-run options O_s for each volume level which yields the lowest average total cost for that volume, dependent on the value of the medium and long-run options, O_m and O_L. By varying the medium and short-run options O_m and O_s, we can generate the envelope curve for average total cost as shown by the dashed and dotted line in part C.

In part D the long range options O_L are variable as well. The dashes represent the curve generated by varying the short-run options O_S for the fixed medium and long-run options, O_m and O_L. The dot and dash curve mark the envelope curves generated by varying both short and medium-run options O_S and O_m, and keeping the long-run options O_L fixed. The dotted curve marks the envelope curve generated by varying all three sets of options.

The difference between the short, medium, and long-run options allows us to develop a set corresponding to average total cost and volume curves. However, this distinction is a relative one. That is to say that what is short-run in one context may be long-run in another. It is important to establish over any predetermined time horizon which options are variable and which are fixed.

The inferences that can be drawn from this are fairly clear. Part B of Figure 3.2 illustrates the short-run average total cost function. The cost function can be moved to meet demand only over a range. To alter the medium and long-run options takes some time. Should volume V_2 occur, the railway operator is at a favorable point as the short-run options to O_{S_2} can be varied to yield the minimal average total cost for O_{m_1} and O_{L_1}. However, should volume be much more than is anticipated, the average total cost of the railway may be much greater than what it might have been if the medium and long-run options could have been changed. In a competitive environment, it is possible that other railways or modes with lower average total costs will capture a larger market share because they will be able to offer reduced rates. In the case that the railway is publically owned, net revenue and level of service decline because greater costs occur than might otherwise have been the case

It would seem that the relationship of options with different time horizons has a great impact on reducing the efficiency of railway transportion. This is due to the existence of time lags necessary to implement medium and long-run options. With less than complete information about the expected level of demand, it is often difficult to obtain the combination of options that is optimal (in terms of average total cost). In the case of the Canadian railroads, this has been reflected in periods when a great deal of excess capacity was installed (and, as a result, higher costs incurred) because demand turned out be less than had been forecast in an earlier period.

3.3.2 Indivisibility of Options

Railway networks, systems, and sub-systems are characterized by some degree of indivisibility; that is to say, some options cannot be varied over a continuous range of output but instead, must be installed in discrete values. The following are examples of indivisibility:

1. Operating policy: Operating policy is not necessarily discrete and is usually continuously variable (frequency, timetables, prices, etc.). An exception is when a railcar or locomotive must return to a maintenance base at a fixed interval.
2. Railcars and locomotives: The number of railcars and locomotives owned or operated by a particular railway must be integral.
3. Fixed facilities: The number of tracks, yards, signals, etc. can only take integral values. (Manheim 1979).

The indivisibility of capacity options is a key determinant of overall railway costs. The inference from this is that the options cannot be varied so as to yield the exact amount of capacity to meet some expected demand level. In practice, there will always be too little capacity or too much capacity. Indivisibilities, and the existence of various time frames needed to implement different options, make it virtually impossible to select the right combination of options to provide the optimal level of capacity.

3.4 Factors Affecting Railway System Capacity

There are a number of different factors that can affect railway system capacity. On a system basis, the capacity of a railway is a function of several components. These include: lines; terminals; yards; equipment; repair and maintenance; resources; and, management. From the demand side, system capacity is also affected by such traffic characteristics as type, mix, peaking, and priorities. The capacity of these components of the railway system can be affected not only by their physical and traffic characteristics, but also by external forces.

Many studies have been conducted to determine the effect of capacity on a number of relevant factors. Khan writes that:

> much has been learned from studies about the factors that affect capacity at the (macro) system level as well as at the lines and yard/terminal level. Through parametric analyses sponsored by the Federal Railroad Administration (FRA, U.S.A.) and simulations carried out by numerous investigators in Canada, U.S.A. and elsewhere, the effect on capacity of all relevant factors has been investigated. These include: physical characteristics of the lines (i.e. siding and cross-over spacing, siding capacity and length, signal block length, proportion of multiple tracks, track maintenance, grade and alignment, line outs and degree of signalization), physical operating characteristics of trains (i.e. power to weight ratio, average speed), and the interaction between trains (train speed distribution, train priorities or mix, meets and overtakes, number of trains, traffic imbalances, peaking, non-productive trains, local train service and slow orders). (1979, 61).

Table 3.1 lists the various factors that can effect the capacity of the railway line. The following subsections discuss these factors in greater detail.

3.4.1 Physical Characteristics of the Railway Line

The physical characteristics of the railway line are important parameters in determining the capacity of a railway system. The number and location of track (track configuration) is an important determinant of capacity. As the proportion of double or multiple tracks to the total track mileage of a railway system increases, there will be

TABLE 3.1

The Railway Line

1. Line characteristics	Track construction Rail weight and metallurgy Ties (types and spacing) Fastening (type, depth) Track condition Rail, ties, ballast, subgrade, switches Track alignment, surface, cross level Track configuration Double track Sidings, spacing, length, switches Crossings (railway, highway) Alignment Maintenance resources Work equipment fleet, shops, mechanics, operators Line maintenance personnel Snow plows, location and condition
2. Signals and communications	Spacing, signal aspects, number, type, car or wayside, automatic features, radio, talkback, and speakers
3. Line operations	Train length, train mix, train load, train speed, dispatching logic, fleeting, line out strategies, queuing, peak leveling, spacing, control systems, information systems, work trains
4. Motive Power	Number of units Condition of units Type of units (locomotive, aux., snow) Horsepower Tractive effort Adhesion characteristics Efficiency of units Reliability of units (maintenance requirements)

Continued on next page

TABLE 3.1 - Continued

5. Rail cars	Number and types of cars
	Condition of cars
	Degree of specialization
	Capacity
	Reliability (maintenance)
	Loading unloading, speed
	Gross/light weight ratio
	Drawbar strength

Source: Khan 1979, table II.4

an increase in the capacity of that system. These increases occur because the multiple track sections allow more trains per day to run, and decrease the average delay per train. The location of the multiple track sections within the system is also a crucial factor affecting capacity. Khan writes:

> Double tracking increases capacity, but is an expensive measure. However, double tracking confined to selected sections of single-track lines could be an appropriate step towards increasing capacity. The placing and length of such sections, however, is crucial. (1979, 62).

Figure 3.3 illustrates the average delay relationships for alternative configurations of 100 mile rail line. It is clear from this graph that a single running double track exhibits superior potential capacity in terms of the number of trains per day carried, and the average delay (in hours) per train.

The spacing length of railway sidings are also an important factor in overall system capacity. Siding spacing and length impacts on the number of trains per day that can run over a particular section, average train delay, and the length of train that can be accommodated. Crossover sidings affect the number of trains per day and the average train delay parameter of double or multiple track sections. See Table 3.2. According to Khan:

> Siding spacing affects train delay. Parametric study of rail line commissioned by the FRA (1975) found an average elasticity for siding spacing to be 0.5 (i.e. a 1% increase in siding spacing resulted in about

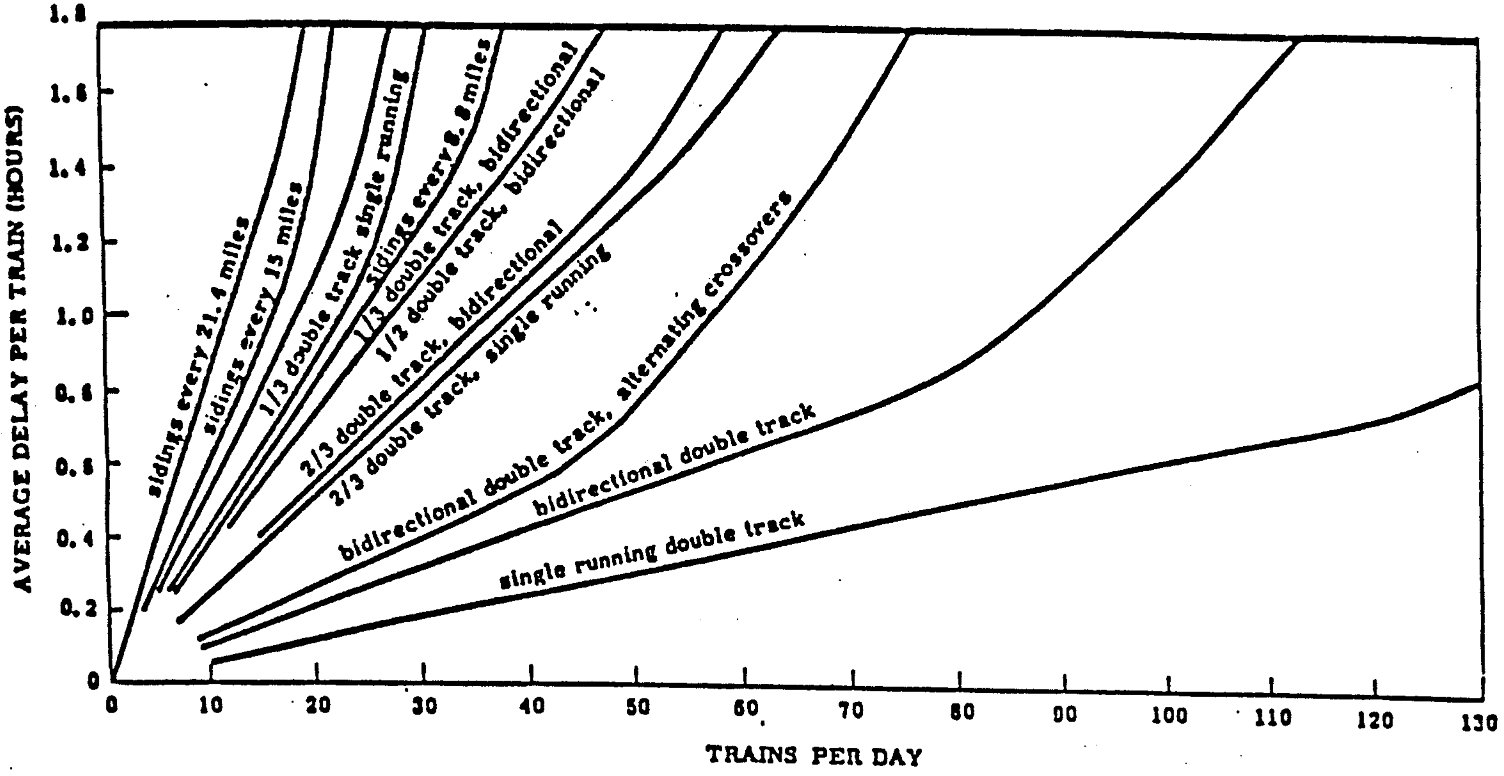


Figure 3.3: Train Volume - Average Delay Relationships for Alternative Configurations of 100-Mile Rail Line

(reproduced from Khan 1979, fig. III.1)

TABLE 3.2

Expected Average Train Meet Delay
(Minutes per Train per 100 Miles)

Siding Length (Miles)	Average Distance Between Sidings (Miles)		Average Trains Per Day				
		20	30	40	50	60	70
2.5	5	-	-	35	55	79	*
2.5	7	-	-	43	73	138	*
2.5	9	-	-	65	*	*	*
2.5	11	22	49	79	*	*	*
2.5	16	32	83	*	*	*	*
2.5	22	38	*	*	*	*	*
5.0	5	-	-	27	40	53	*
5.0	7	-	-	31	49	73	*
5.0	9	-	-	34	52	74	*
5.0	11	-	-	46	*	*	*

* - indicates track configuration unable to handle traffic volume

These average train meet delay times are subject to the following conditions:

1. All trains are 110 car coal trains with 4 3,000 hp locomotives and 100 net tons per car.
2. All loaded trains operate in one direction and all unloaded trains operate in the reverse direction.
3. An equal number of trains operate in both directions.
4. All trains have the same priority.
5. The signal system is CTC with 35 mph turnouts.
6. No trains have a predetermined schedule; therefore, all trains enter the line segment randomly.
7. The speed limit is 50 mph for all trains.
8. The distance between sidings varies randomly up to + or - 2 miles for average distance of 9 miles or less and + or - 2.5 miles for average distances greater than 9 miles.
9. Only one train is allowed to operate in a 2.5 mile siding; two trains are allowed in a 5 mile siding.
10. The number of trains per day varies randomly about the average.
11. The first train to reach the location of a meet takes the siding.
12. The allowable delay is the time available for meet, overtakes, and random delays while the train is running. Scheduled delays such crew changes and fueling are considered separately.
13. Expected delay is only the time lost due to meets and maintaining minimum headway. Time lost due to factors such as weather, track maintenance, or mechanical problems would be additional time that must be charged against allowable delay.
14. Minimum headway between trains is the same direction and is 10 minutes.
15. All trains fit in all sidings.

Source: Khan 1979, fig. III.3

> 1/2% increase in train delay). It was also observed that the greater the average siding capacity, the more sensitive capacity was to siding spacing. Average siding spacing, as compared to uniformity of siding spacing, was observed to be more important (FRA 1975).
>
> As far as the length of siding is concerned, it obviously affects the length of the train to be accommodated. There are, however, limits to the length of trains since other considerations such as grades, yard lengths and probability of failure rates of longer trains come into play.
>
> Crossover spacing is of importance in determining the capacity of double track line. Cross-overs become essential in moving trains when failures occur during critical periods. (1979, 63).

Table 3.2 shows the expected average meet delay (minutes per train per 100 miles) for a combination of different siding spacings and lengths.

The control characteristics of a railway line are also an important determinant of system capacity. Signals and communication (control) play a vital role in the operations of the railway system. The function of signal system is twofold: to step up the efficiency and capacity of a railway line to carry traffic; and, to enhance the safety of line-haul operations. Signal and communication systems affect the time to implement control decisions and the delays that may be caused by the exercise of control. The characteristics of control impact upon direct train delays through the quality and speed of control decisions. Burlington Northern Railways (1977) concludes that the major factors that affect control characteristics are signal systems, the nature and location of turnouts, and operating rules.

The different types of signals include: manual block signalling; automatic block signalling (ABS); automatic train control (ATC); and, centralized traffic control (CTC).

Railway lines that employ the manual block signalling system are divided into signal segments (blocks), and only one train may enter a block at any one time. Signals located at the block division indicate the presence of a train in the block ahead. Automatic block signalling systems perform the signalling function automatically. The length of the block would impact on the number of trains per day that could safely pass over a section of railway line. Peat, Marwick, Mitchell and Co. in a study for the FRA (1975) also concludes that effects upon capacity were observed in parametric research not from average block length but also from the variability of block lengths.

Automatic train control is a system automatic supervision of train operations. Implemented with "continuous-coded" automatic train control, this system insures that the speeds of trains running over the lines are lowered or increased in accordance with the signals.

The centralized traffic control system monitors and controls train movements over the entire system by controlling signals from a central command location. The movement ot trains through the system is monitored electronically and switches can be set from a central location. The CTC system impacts upon the number of trains per day that can move over the system, as well as improving the quality, speed, and exercise of control. Decreases in train delays are also achieved through improvements in control. Khan states:

> For heavily-travelled single-track lines or congested sections of multi-track routes, CTC can be regarded as desirable. (1979, 37).

Table 3.3 compares the estimated engineering capacity of both single and double track rail line sections using the ABS amd CTC systems.

Grades and alignment can have significant effects upon the capacity of a railway system due to the impact it has on average train speed. Grades also can impact on train load and length. Khan writes that:

> On some of the extreme grades the train load is limited by the strength of the coupler or drawbar. Therefore, in such cases there is a need to add "slave locomotives". Improvements to grades and alignments can be attractive in specific cases as a method of increasing line capacity. (1979, 62).

Thus, the effects of grades upon a train show up through the number of locomotives required to move the train over them. There is a relationship that exists between grades and power vs. distance.

It is also true that track alignment can affect system capacity. Alignment has an impact upon train performance

TABLE 3.3

Rail Line Segment Engineering Capacity

	Signal System			
	Automatic Block Signal (ABS)		Centralized Traffic Control (CTC)	
Number of Tracks	Trains per Day (1)	Gross Tons per Year (2) (Millions)	Trains per Day (1)	Gross Tons per Year (2) (Millions)
Single	40	62	60	93
Double	120	186	160	250

Sources: Khan 1979, fig III.2

(1) Total both directions.

(2) Millions of gross ton miles per route mile; total both directions.

because trains that run over the system are required to make numerous changes in their directions. These changes in direction can be both vertical and horizontal. Alignment also affects the performance and life of railway track. It has been estimated that the standard 132 lb/yd has a lifetime of about 800 million gross ton-miles over straight sections. Curved track, on the other hand, has a life of only 80-200 million gross ton-miles per mile of track.

Limitations on system capacity also occur due to the effects of track repair and maintenance operations. The movement of large volumes of traffic in heavily loaded rail cars at high speeds results in a much shorter life span for the track. Trends indicate that trains will increase in length, weight, and speed. Thus, greater demands are being placed upon track time for maintenance and repair operations. As the time for repair and maintenance operations increases, the potential capacity of the railway system is decreased. The time needed for these operations becomes a major capacity limitation, particularly if the track that is being repaired is not part of double track section with several cross-overs.

However, railway technology in the past few years has considerably improved track maintenance methods. According to R. S. Wallace and Associates:

Track maintenance equipment will likely evolve into

> two types. Large changeout machines capable of removing track and re-laying new in one operation, will be used increasingly for major track rebuilding projects. Smaller highway/rail portable maintenance units, each performing only one function, will be used to carry out periodic spot maintenance between major overhauls. As the time available between train passes becomes shorter, the emphasis on maintenance equipment design will be on increased speed of operation, reliability, simplicity, and automation. (1982, 121).

3.4.2 Physical and Operating Characteristics of Trains

Train speed is a crucial determinant of railway system capacity. Khan writes that:

> For both single and double track lines, speed is the most important single parameter for capacity analysis. For a one percent decrease in speed, a 2 percent increase in average delay was observed in simulation studies (Peat, Marwick, Mitchell and Co., 1975). This relationship was stated to hold for both uniform and non-uniform speeds. (1979, 64).

Khan goes on to say that the uniformity of speed is another important factor for capacity analysis. Increases in capacity occur for both single and multiple track railway lines when trains are operated at higher uniform speeds. These increases result from declining interactions amongst trains running over the system. Priority trains can also have a significant impact upon system capacity. Interaction between trains running over a line are minimized when they move at equal speeds and the same priority. This phenomenon can be ascribed to something called the "channel theory" (Leppan 1978; Khan 1979). Appendix A provides a full explanation of this theory.

The power to weight ratio of trains affects system in that it is a crucial determinant of average train speed. Khan states:

> The power to weight may affect the average speed especially where speed limits are greater than the train capabilities on the grade. Speed as noted above is a strong determinant of capacity. Power to weight ratios also affect the acceleration of the trains. Delays therefore could occur if train stopping is involved.
>
> Train and line capacity may be increased by increasing the load per car and per train. Care must be taken, however, as trade-offs are involved between power to weight ratio, maintenance, and the constraints set by grades and drawbars. Conceptually, a multi-dimensional relationship exists between power, weight, grades, speed limits, maintenance, and probability of failure and capacity. (1979, 64).

3.4.3 Other Factors Affecting Railway System Capacity

The capacity of a system has often been characterized as being a function of the component with the lowest capacity. Thus, a factor affecting the capacity of the railway yards has a potential to affect the overall capacity of the system. There are a number of factors that can affect the capacity of a railway yard. Khan states that with respect to yard capacity:

> the efficiency (or inefficiency) of various functions/operations in the yard (e.g. train make-up, train break-up, servicing, etc.) will have a direct effect on yard capacity. It has been suggested that severe congestion within a yard occurs due primarily to the classification and assembly operations. The waiting of rail cars to make connections on an outbound train generally does not result in congestion. Peaking of traffic, usually handled by "surge" yards, could also have capacity effects. (1979, 81).

Yard capacity is affected by such physical characteristics as the length and number of tracks located in the receiving, classification and departure areas. Shortage of trackage in these areas can affect the "processing" and/or delay time of the various yard functions. The number of engines and switching facilities can also impact upon the classification activities of the yard. According to Khan:

> Classification operation is often performed by one engine which moves a number of cars back and forth over the switch lead and the switching crew setting switches so as to move cars into appropriated classification track. In order to increase capacity, more than one switching lead is made available, with a different engine operating on each lead. (1979, 83).

Other factors that impact upon system capacity include interruption of capacity due to such factors as weather, accidents, labour disputes, equipment failure, etc. These events are usually rare and unscheduled, and imply reductions in the amount of time available for the system to provide railway services.

3.5 Options Available to Expand Railway System Capacity

The railway planner has a number of different categories of options available to effect increases in the overall capacity of the system. There exist four distinct groups of options that can be employed to affect changes in system capacity. These groups are:

1. operating solutions
2. traffic solutions
3. equipment solutions

4. plant solutions

These groups can be characterized by their costs, divisibility, and the time needed to implement them. Generally, operating solutions require the least amount of expenditure and time to implement, whereas plant solurions are the most expensive and take the most time to install. Table 3.4 summarizes the impacts and costs of these various options.

3.5.1 Operating Solutions

Operating solutions generally require the least expenditure of time and funds to implement. These solutions usually involve making changes to the operating plan of the railway. Railway operations typically involve the utilization of rolling stock (i.e. rail cars and vehicles) over a network of rail lines that connect all the points that are serviced. The efficient use of the railways' vehicles, labour, and facilities require the specification of an operating plan (Manheim 1979). the structure of such a plan should include consideration of the several type of resources that are required for railway operations. These resources include facilities, vehicles, and labour.

The specification of an operating plan involves the following elements:

1. An inventory list of all available railway facilities, rolling stock, and labour;

TABLE 3.4

Railway System Capacity Expansion Options: Costs and Benefits

OPTIONS	CAPACITY PARAMETERS AFFECTED	ECONOMIC BENEFITS	ECONOMIC COSTS	TIME HORIZON
1. Operating Solutions				
Options Faster Trains Runthrough Trains Schedule Revisions Schedule Adherence	Time Decrease in average train transit time Decrease in average train delay	Revenues Increased revenue to railway and lower costs to users due to reduction in delays and improved service Operating and maintenance costs Decrease in local and switch crew wages Decrease in time portion of rail car costs Decrease in most non-distance cost categories	Operating and maintenance costs Insignificant for railway systems operating below capacity Increased no. of trains increase total operating and maintenance costs for railway systems operating beyond capacity, due to greater interaction between trains	Short-run
Options Longer trains **Heavier trains** Increased frequency of trains running	Time Increase in average train transit time for railway systems operating beyond capacity Increase in average train delay for railway	Revenues Increased revenue to railway and lower costs to users due to reduction in delays and improved service	Operating and maintenance costs Increased no. of trains increase total operating costs of the railway system Heavier and longer trains increase operating and	Short-run

Continued on next page

TABLE 3.4 - Continued

Railway System Capacity Expansion Options: Costs and Benefits

OPTIONS	CAPACITY PARAMETERS AFFECTED	ECONOMIC BENEFITS	ECONOMIC COSTS	TIME HORIZON
	systems operating beyond capacity Volume Increase in the average load carried per train Increase in the no. of trains		maintenance costs for railway systems operating beyond capacity, due to higher cost of moving more throughput (e.g. greater depreciation of plant and vehicles)	
2. Traffic Solutions				
Options Unit trains Pooling of traffic Shipped owned and loaded cars Uniform scheduling of customers' shipments	Time Decrease in average train transit time due to reduction in average yard/ terminal processing time for trains Decrease in average train delay	Revenues Increased revenue to railway and lower costs to users due to reduction in delays and improved service Operating and maintenance costs Decrease in local and switch crew wages Decrease in time portion of rail car costs Decrease in most non-distance cost categories	Operating and maintenance costs Insignificant for railway systems operating below capacity	Short-run

Continued on next page

TABLE 3.4 - Continued

Railway System Capacity Expansion Options: Costs and Benefits

OPTIONS	CAPACITY PARAMETERS AFFECTED	ECONOMIC BENEFITS	ECONOMIC COSTS	TIME HORIZON
3. Equipment Solutions				
Options More rail cars Heavier rail cars Commodity specific rail cars More locomotives More powerful locomotives	Time More and faster locomotives will lead to decrease in average train transit time Decrease in average train delay More trains increase average train transit time for railway systems operating beyond capacity, due to greater interaction between trains Increase in train delay for railway systems operating beyond capacity Volume Increase in the average load carried per train Increase in the no. of trains	Revenues Increased revenue to railway and lower costs to users due to reduction in delays and improved service Operating and maintenance costs More efficient new equipment reduces operating and maintenance costs	Operating and maintenance costs Increased no. of trains increases total operating and maintenance costs of the railway Heavier and longer trains increase operating and maintenance costs for railway systems operating beyond capacity, due to higher cost of moving more throughput (e.g. greater depreciation of plant and vehicles) Capital costs Increase in capital costs for new equipment	Medium-run

Continued on next page

TABLE 3.4 - Continued

Railway System Capacity Expansion Options: Costs and Benefits

OPTIONS	CAPACITY PARAMETERS AFFECTED	ECONOMIC BENEFITS	ECONOMIC COSTS	TIME HORIZON
4. Plant Solutions				
Options Longer sidings More sidings Improved signals and communications Double tracking Alignment and grading Improvements to yards and terminals	Time Decrease in average train transit time due to reduction in average yard/terminal processing time for trains Decrease in average train delay	Revenues Increased revenue to railway and lower costs to users due to reduction in delays and improved service	Capital costs Significant increase in overall capital costs	Long-run
	Volume Longer trains Increase in the average load carried per train Increase in the no. of trains	Operating and maintenance costs Decrease in fixed plant operating and maintenance costs		

Sources: Manheim 1979; CN 1985.

2. A list of all of the points on the railway network that are to be serviced;

3. A schedule of points that are to be serviced along with a corresponding service timetable;

4. Assignment of rolling stock to the schedule above such that the service points are covered and the service timetable is adhered to; and

5. assignment of labour to operated the facilities and trains.

The operating plan of a railway system is a description of how vehicles are to be utilized in order to provide transportation service to the various points in the network. The detailed specification of this plan includes the facilities, vehicles, and labour to be used, the points to be serviced, the timetable of services over the network, and the assignment of facilities, vehicles, and labour to the schedule of services.

Changes in the operating plan of the railway can affect changes in capacity by altering the time and/or volume parameters of the existing system. Table 3.4 lists some of the major operating alternatives that can be used to affect changes in system capacity. The first part of the table groups those operating alternatives that will affect the time parameter of system capacity. This group includes such alternatives as faster trains and increased schedule adherence (i.e. reduced variation of actual service times from scheduled service time). The second group of alternatives can affect system capacity by altering the volume parameter, and include such options as longer, heavier trains and increased frequency of trains.

Operating solutions are generally low cost in nature and can be put into effect in the short run. As such, they are usually the first to be considered by railway planners. Traditionally, capacity improvements achieved by Canada's major railways have been due to the availability of excess capacity coupled with advancements in railway technology (Khan 1979). Only small amounts of investment were required to make gains in capacity. It is clear that in the future, however, large capital investments will have to be made in order for the railways to have an adequate amount of capacity available to deal with increased traffic demands.

3.5.2 Traffic Solutions

Gains in system capacity can be achieved through changes in the characteristics of railway traffic. These options can usually be implemented in the short run with a minimum of cost. Examples of these options include terminal pooling of traffic, unit trains, shipper-owned rail car, and uniform scheduling of customers' shipments. With respect to Canadian railways in the 1970s, the pooling of traffic such as Canadian Wheat Board grains greatly reduced costly switching and the associated bottlenecks for grain at Vancouver (CN 1985). The advent of heavier, shipper-owned cars produced significant increases in capacity in Western Canada in the late 1970s. The railways encouraged this development by the use of incentive pricing to shippers. This resulted in a large reduction in the workload at certain points on the Western Canadian rail system, thus increasing capacity. Also contributing to gains in capacity was the trend from the single-car shipment to blocks of cars and unit trains. Unit trains increase primarily through enhanced movement through terminals and yards.

An oft cited option to increase railway system capacity (CN 1985) is changing the shippers' shipping and receiving patterns through uniform scheduling of shipments and rate incentives. However, there are certain factors that may prevent this option from having any serious potential. In a market where competition exists, attempts to change the service schedules for shippers may result in a significant decline in market share for the railway as shippers turn to the competition. In addition, in a market that is characterized by numerous types of traffic, origins, and destinations, no degree of railway control over scheduling could have much of a positive impact on capacity, given the uneveness of the flow of traffic and the variety of movements.

Discouraging selected traffic through rate increases is not thought to be an effective or controllable option (CN 1985). The logic behind this option is that selectively constraining low profit traffic can provide additional capacity for other, higher-profit types of traffic. However, government legislation under Section 262 of the National Transportation Act has provided a very effective means to prevent the railways from refusing to move certain types of "captive" traffic. Rate increases were further constrained by Section 23 of the Act. Captive shippers are often able to exercise enough power to negotiated good service by the railways for low-profit traffic by threatening to reduce the amount of their higher-profit that moves over the system. The railways have always been able to turn away competitive traffic, but according to CN:

> Shippers having competitive options have generally exercised great power in removing from us our highest profitability traffic when we have sought to restrict the lower profitability traffic. (1985, 14).

3.5.3 Equipment Solutions

Equipment solutions can be more costly and take more time to implement than either operating or traffic solutions. Examples of equipment solutions include heavier rail cars, commodity specific cars, more powerful locomotives, and better on-line repair and maintenance equipment. Increasing use of 100-ton rail cars has provided additional capacity for Canadian railways (CP Rail 1985; CN Rail 1985). CN Rail decided, however, that:

> the use of more horsepower per train in order to shorten running times would have entailed very heavy investment in motive power with only modest improvement in capacity and was, accordingly, not a favored option. (1985, 13).

It should be said that equipment investment and plant investment tend to be highly interrelated. Improvements in plant capacity often result in the increased utilization of rail cars, and in a decreased requirement for equipment maintenance. Increases in car size and weight does cause a requirement for additional plant investments to upgrade track structure, and a corresponding increase in track occupancy for construction work.

3.5.4 Plant Investment Solutions

Generally, railway plant investment achieves one of the following objectives:

1. It replaces worn out elements of the plant
2. It upgrades elements of the plant
3. It expands the capacity of the elements of the plant

The economic benefits achieved by investing in plant replacement or upgrading are from the extended economic lives of the renewed elements, more efficient maintenance, ability to move faster and heavier trains, and the ability to provide competitive services to a particular market. Investment in railway plant capacity not only adds traffic capacity to the system, but it also improves the systems' operating efficiency. Improvements in operating efficiency lead to reduced operating costs.

Investment in plant capacity becomes inevitable once the potential from the lower cost options mentioned previously are exhausted. Plant investment takes place in a hierarchy of expansion steps. This approach allows the greatest value to be realized from the investment. Railway plant expansion programs are associated with providing additional capacity by the installation of a number of options. These options are usually carried out in the following sequence:

1. In the first step, additional capacity is provided by building longer train sidings and, therefore, increase the capability of the line to handle longer trains.
2. In the next step, the creation of more such sidings on the systems' mainlines permit trains greater opportunity to pass and ability to achieve faster meets.
3. The third step includes the installation of more, and improved, signals in order to reduce the headway between trains. This insures a reduction in train transit time and delay while permitting increases in the number of trains running on the line.
4. Where extreme congestion exists, the fourth step would be the installation of double track to alleviate the situation. Double tracking allows large increases in train density over a given segment of line. We would expect that delays in train running times would be further reduced. Reduction of grades and changing alignments would also be a further, albeit expensive, step to increase capacity.
5. Finally, improvements in yard and terminals are implemented. These changes have the greatest potential to achieve increases in railway system capacity since most railway cars spend about two-

> thirds of their time within terminal and yard areas. Investments in terminals and yards help achieve increases in capacity onlines by enhancing the ability to achieve the optimum design of trains. Also, the faster equipment repair and maintenance time achieved within the improved yards will lead to quicker put-through times and fewer equipment failures occurring on the track.

3.6 Summary

This chapter has given an overview of the specific factors that affect railway system capacity, and the options that are available to expand it. These factors and options have been presented within the context of economic and investment issues relevant to the railway planning process. Economic analysis implies that the level of throughput that minimizes average total cost for a particular railway system is optimal. Capacity expansions should occur when the existing average total cost for a particular level of throughput is equal to, or greater than, the average total cost that would occur with the expansion, for the same level of throughput. These expansions should be carried out along a curve that connects the point of optimal capacity for various levels of fixed costs.

The planning process must also take a number of other factors into account before initiating a particular investment. The economic planning perspective is different from that of engineering planning. Engineers consider capacity planning expansion in terms of implementing options to increase the physical capacity of the railway system. Engineering plans identify and implement options that are optimal in a technical sense, and feasible given existing physical and cost constraints. Economic planners, on the other hand, look at options in terms of their economic impact and viability. Expansions are undertaken to meet expected demand, and options are considered in terms of their effects on future revenues, costs, and profitability.

The planning process recognizes that the various options that are available to expand capacity will take different amounts of time to implement. The indivisibility of these options will usually make it impossible to install just the level of capacity necessary to meet expected demands.

The preceeding discussion has outlined the numerous factors that affect railway system capacity. The physical and operating characteristics of the railway system, and of the trains, impact greatly on capacity. In terms of the overall railway system, capacity can be characterized as being a function of the component or element with the lowest capacity. Thus, the overall capacity of the system may be defined by its weakest link.

There are a number of options identified by the planning process that are available to increase capacity.

These options can be classified according to their impact on capacity, costs, benefits, and the time needed to implement them. The railway planning process, logically, would first implement the options requiring the least cost and time. At the point where the lower cost options are no longer effective in increasing capacity, it would then become necessary to implement the more expensive options.

The purpose of this chapter, as well as the previous one, has been to provide an insight into the complexities of the factors that affect railway capacity. This insight should assist in comprehending the planning and investment decisions taken by CN Rail and CP Rail during the 1980s to increase railway capacity in the West. The next section of this thesis will outline and evaluate the circumstances and planning process that led to those decisions.

CHAPTER 4

AN OVERVIEW OF THE RAILWAY CAPACITY PROBLEM IN WESTERN CANADA IN THE 1980s

4.1 Introduction

The first section of this thesis described the structure, operations, and economics of the railway system, and their relation to capacity. A number of options available to expand capacity were also outlined within the overall context of economic planning. All of these points were discussed at length to build a theoretical background for the following discussion of railway capacity expansion in Western Canada.

This chapter summarizes the issues and circumstances that led CN Rail and CP Rail to plan and implement a series of investments to increase railway capacity in the West during the 1980s. The issue of how the statutory "Crow" rates for grain acted to impede the long range planning efforts of the railway to increase Western capacity is discussed. The supply and demand dynamics that acted upon the railway network in Western Canada during the 1970s are also detailed. This chapter then looks at the major issues and concerns in the 1970s relating to the ability of this network to meet expected demands. Details of the traffic demand forecasts facing the railways in the late 1970s are outlined, as are the critical capacity contraints that existed in the Western railway network during this time.

4.2 The Crow Rate Issue and the Problem of Railway Capacity in Western Canada

The major impediment facing the planning efforts of the railways to expand railway capacity in Western Canada was the existence of the statutory Crow rates for grain. The costs associated with moving this grain escalated during the 1970s, while the statutory rates remained fixed. These losses adversely affected the financial position of the railways. As a consequence, the railways were reluctant to invest the large sums of capital required to expand their network capacity in the West. From the perspective of the railways, there was little point in planning to spend billions of dollars to increase capacity over the long-run when there existed serious concerns regarding the viability of these investments. Although capacity expansion would lead to more revenue from such profitable traffic as coal and potash, it was felt that grain would have to be able to pay its own way. Grain carried at compensatory rates would increase the probability that investments made to expand capacity would remain economically viable over time.

The railways and the federal government recognized by the late 1970s that there would have to be some solution to the problem of noncompensatory grain rates. This issue had implications not only for the railways, but also for the overall development of the Western Canadian economy.

The statutory Crow rates for grain were part of a deep seated subsidy policy that had its orgins in the latter years of the nineteeth century. Statutory rates for grain were enacted in the spirit of the National Policy. Policy makers at that time realized that the embryonic wheat economy of the West would never attract sufficient private capital to expand and diversify. Subsidies would have to be provided in order to ensure the development and prosperity of Western Canada.

At the time of the Crow's Nest Pass agreement, these rates were more than sufficient to cover the cost of moving grain. It has been argued that the statutory rates benefited Western Canada for the first half of this century. Gilson writes that the original agreement and subsequent changes served the west and the railways well for many years (1982, II-5).

Problems, however, began to emerge in the 1960s. Grain prices began to strengthen in the face of increasing demand in export markets. As a result:

> unprecedented demands were placed on the grain handling and transportation system. By the late 1960s, evidence began to emerge that the system lacked the capacity to serve the needs of the grain industry. (Gilson 1982, II-5).

In the late 1960s and early 1970s, the railways began to incur losses due a combination of increasing grain shipments and increasing rail costs. While revenue per mile remained fixed at a level below that prevailing in the late 1890s, costs continued to rise. The rapid increase in labour and fuel costs, especially in the period after the 1973 energy crisis, forced costs and railway losses to increase.

By the 1970s the Western economy had become much more diversified since the time when the Crow rates were enacted. The ever increasing export demand for Western Canadian coal, sulphur, and potash affected the needs of railway transportation in the West. This traffic, which moved at rates set by the railways, was profitable to the carriers. The bulk shippers felt that the railways were turning to them to make up for revenue shortfalls from grain movement. The bulk shippers were particularly vunerable to the cross-subsidization of grain from their traffic because no economic alternatives to rail existed.

The question of what should be done with the statutory Crow rates became a major public policy issue in the 1970s. The federal government set up the

Snavely Commission to review the 1974 railway costs. Bonsor writes that:

> The statutory grain rates yield carriers an average of 0.5 cents a ton-mile. By comparison, the lowest nonstatutory rate for the movement of grain and grain products in 1980 was approximately 1.7 cents a ton-mile. (1984, 103).

The ratio of variable costs to the statutory rate for 1977 was determined by the Snavely Commission to be 3.1 to 1 (CTC 1983, 62).

The gap between the cost incurred to the railways by the movement of grain at statutory rates widened in the late 1970s. The 1976 Snavely Report concluded that the railways lost about $105.5 million in 1974 on grain shipments (Schweitzer 1984, 74). Based on figures published by Gilson in 1982, this loss had reached $214.9 million in 1980 (Schweitzer 1984, 74).

This burden had significant implications with respect to the ability, and willingness, of the railways to maintain, upgrade, or expand their existing Western networks. Snavely concludes that:

> the level of rate offers no incentive to the railways to maintain, upgrade, or modernize the road property or equipment they provide for the transportation of statutory grain. (CTC 1983, 64).

This situation was detrimental to the capabilty of the rail based grain handling system in Western Canada to move traffic. As the severity of the revenue short fall increased in the 1970s, the railways made little or no investment to maintain or upgrade the grain tranportation system. Gilson says that:

> The size of the box car fleet declined drastically from 34,000 in 1969 to 12,560 in 1980. . . . The branch line system continued to deteriorate. The railways had made no investments in rehabilitating grain rolling stock or branch lines since the early 1960s. Maintenance work was minimal . . . Train speeds were reduced and efficiency further declined . . . (1982, II-7).

The federal government attempted to alleviate some of the negative impacts of this situation through a series of ad hoc programs. These programs, however, were no more than "band-aid" remedies to a problem of enormous proportions. In 1970 subsidy payments began to be made to the railways, under a program authorized in the National Transportation Act of 1967, to operate grain branch lines retained in the public interest. Between 1971 and 1981, $971 million in subsidies was paid out to CN Rail and CP Rail under this program (Gilson 1982, II-8).

In an effort to upgrade and rehabilitate light-density grain lines, the federal government initiated a program for the Western branch line network in 1977. This program, which rehabilitated some 2,300 miles of prairie branch lines, ran over a period of seven years at a cost of $700 million (CTC 1983, 65).

Programs were also introduced to deal with the deficiencies in the rolling stock. A cost-sharing arrangement between the federal government and the railways to rebuid some 7,400 box cars for grain service was worked out in 1974 (Gilson 1982, II-9). This was in addition to a federal government progam begun in 1972 to lease and purchase new hopper cars for the grain car fleet. These programs, with a total cost of $280 million by 1982, provided over 50 percent of the rolling stock in the national grain car fleet by the early 1980s (CTC 1983, 62).

The statutory rates clearly denied grain traffic exposure to the full costs of using the railway system. Over the long-run, this endemic short fall of revenue was met by one or a combination of economic loss to the railways, cross-subsidization of grain from other traffic, and government compensation.

The ad hoc programs of the federal government certainly covered some of this short fall, but led to results that were less than optimal. Subsidies for branch line losses and rehabilitation did not encourage efficiency. The CTC states that:

> Compensation for branch line losses incurred does not encourage their most efficient operation, nor are the branch line rehabilitation or hopper car purchase decisions efficient. With the hopper cars, for example, the investment decision criteria is minimum capital cost per unit of capacity. This . . . is not necessarily the most cost effective choice of car. (1983, 65).

The federal government certainly did not have unlimited resources to subsidize inefficient railway operations. By the early 1980s, most governments were faced with the need to restrain expenditures. With no solution to the Crow rate problem, the cost of the railway subsidies certainly would have burgeoned.

Cross-subsidization of grain from other profitable traffic, over the long-run, acted to distort the allocative efficiency of the transportation market. This situation encouraged the inefficient use of rail capacity to move grain. The movement of grain used capacity beyond its real needs, and this impacted upon the ability of the system to move other commodities. The cross-subsidization of grain from profitable traffic is not efficient, and will result in reduced rail traffic and/or productivity (CTC 1983, 67).

Over time grain related losses adversely affect the rate of return of the invested capital of the railways. This resulted in the drying up of new and replacement investment capital. This affected the willingness and ability of the railways to invest in the modernization of the Western network and particularly in the grain handling system.

The statutory Crow grain rates had clearly outlived their usefulness by the 1970s. Investment in the railway infrastructure of Western Canada was drying up. The deteriorating railway network no longer possessed the capacity to move Western grain to growing export markets. Other commodities paid for the inefficiencies of the Crow through rate cross-subsidization, and loss of potential revenue due to insufficient transportation capacity. The ad hoc federal subsidy programs became increasingly less effective and more costly over time.

The federal government recognized the linkage between resolving the Crow issue and achieving their transportation objectives for Canada. In 1981 Transport Canada outlined their planning targets for the 1980s. These objectives included the reduction of existing transportation subsidies, while encouraging productive investments to ensure the modernization and expansion of Canada's existing transportation infrastructure. Resolution of the Crow problem, and the expansion of the Western railway network, tied directly into the these objectives. The issue of the Crow rates and Western rail capacity expansion was determined by Transport Canada to be:

> the major issue facing Transport Canada. . . . With no resolution of the "Crow", there will be increasing demands for subsidy payments related to grain, for refinancing CN, and supporting CP Rail Western infrastructure plans. With a resolution, an estimated $2 billion in expenditures can be expected to 1984-85, but eventual reduction will occur in subsidies for branchlines, hopper cars, and branchline rehabilitation. (1981C, 47).

The Crow rate stood in the way of the railways' long term planning efforts to undertake the necessary investments to expand Western rail capacity. The problem was finding the capital to finance these investments.

The federal government was running large deficits, and would not commit the billions of dollars of public capital that would have been required for these investments. Foreign capital, which had historically financed the expansions of the Canadian railways, was expected to be in short supply in the 1980s. The only reasonable alternative was to:

> improve the transportation industry's profitability, thereby, augmenting the amount of internally generated funds, which would enable firms such as CN, which depend\` on retained earnings for equity, to obtain additional financing in the market. (CTC 1985, 87).

The passage of the Western Grain Transportation Act in 1983 achieved this goal. It served to strengthen the potential profitability of the railways. They finally recieved a freight rate for grain that not only covered variable costs, but also contributed toward constant costs.

The removal of the Crow rate was a necessary element in the long term planning of Western rail capacity expansion. The railways, with their improved financial outlook, could now make effective long-term investment plans to expand the capacity of the rail network in the West. The federal government achieved its objectives of encouraging productive investment by the railways to ensure adequate Western rail capacity, while also controlling and eventually reducing the level of subsidies paid out to them. The emergence of a commercial, efficient, and self-sustaining railway system in Western Canada should serve the public interest well.

4.3 Background to the Railway Capacity Problem in Western Canada in the 1970s

The transportation for much of Canada's international trade in such important commodities as grain, coal, potash, sulphur, etc. relies heavily on the railways. Furthermore, almost all of this trade originates in Western Canada. The railway system provides the crucial link in the movement of these commodities from the Western provinces to international export markets, via the exit ports of Vancouver, Prince Rupert, Churchill, and Thunder Bay. Increasingly, the logical focal point for much of this export is Vancouver, and to a lesser degree Prince Rupert, given the ever growing importance of a number of large export markets in the nations of the Pacific Rim.

The national railways, producer groups, provincial governments in the West, and the federal government recognized the potential for capacity problems in the Western railway system as far back as the early 1970s. In 1976 the Western Transportation Advisory Council (WESTAC) surveyed the concerns of a number of individuals, firms, organizations, and governments who either used, supplied, regulated, or made policy for the Western transportation system. The capacity of the railway system in Western

Canada to meet the demands of the future was identified as a major element of uncertainty amongst those surveyed. According to this study:

> there is a credibility gap between the railroads on the one hand and shippers, ports, terminals, and provincial governments on the other hand, regarding the carriers' ability to adequately serve demand levels that are envisioned for the future. This contributes to uncertainty about the future adequacy of track capacity, supply and quality of rolling stock, and service levels in the Lower Mainland of British Columbia. (WESTAC 1976, 9).

WESTAC identified the individual concerns of a number of specific groups. Bulk commodity producers expressed their concern in regards to:

> track capacity and car supply for their products, and also expresses the need to know how the forecast growth of other commodities would affect the movement of their own shipments. (WESTAC 1976, 54).

Terminal and port operators stated that they were concerned about any capacity limitations, since possible impacts on demand flows affect their own facility planning. Some provincial governments were of the view that carriers should be "forced" into providing the required capacity (WESTAC 1976, 55).

During the 1970s the federal government was also expressing a disquieting interest in the potential for a railway capacity crisis in Western Canada by 1990. Transport Canada analyzed this problem in some detail in 1975. Their analyses concluded that a number of critical rail links in Western Canada would experience capacity limitations before 1990 (Transport Canada 1975, 37).

This analysis went on to say that:

> capacity limitations may be significant for most or all sections of the CN mainline between Vancouver and Thunder Bay and the CP mainline between Vancouver and Winnipeg. (1975, 37).

The Transport Canada report concluded that:

> investments required to provide needed increases in freight transportation capacity during the next 15 years will be larger, proportionately, than those required during the past 10 years because: - substantial new rail capacity must be added for the first time in many decades, and - increases in rail . . . productivity are beginning to level out as vehicle size and efficiency limits are reached . . . (1975, 52).

Both CN Rail and CP Rail were well aware of the possibility of capacity shortage in the 1980s. Even as early as 1972, it was becoming evident to the railways that capacity related problems were going to get worse on the mainlines running west through to the B.C. ports. At this time, there was virtually no double tracking on the CN Edmonton-Vancouver line, or on the CP Rail Calgary-Vancouver line. The CN track and subgrade on this line was constructed with jointed, medium weight rail resting on wood ties and sub-optimal ballast. Centralized traffic control was still incomplete over the route. According to a speech made by Ross Walker, the CN Senior Vice-President in Western Canada:

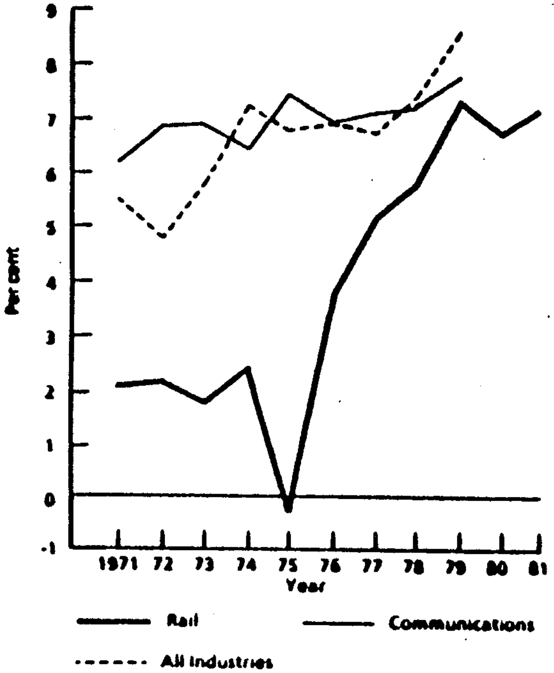
> In the early 70s it became obvious that we were facing capacity problems, particularily in the Western part of the system. It was equally obvious that the old standby of adding another siding or yard track or

a few signals was not going to be enough. (1980). CP Rail was also beset with capacity related problems. Capacity on the CP mainline to the south was constrained by a single track running over mountain passes with controlling grades as high as 2.2 percent. A one percent grade defines a vertical rise of one meter for every horizontal run of 100 metres.

However, the railways were reluctant in the 1970s to commit the hundreds of millions of dollars necessary to increase the capacity of their respective railway networks in Western Canada. The railways complained that it was not feasible for them to finance investments for new capacity given the low profitability of the railway industry, which was compounded by the low statutory Crow rates for hauling Western grain. According to the CTC:

> the rate of return on assets for the Class 1 railways - Canadian National Railway Company and CP Rail - varied between 1.7 per cent and 2.3 per cent in the years prior to 1975. In 1975, the rate of return was negative, mainly as a result of traffic erosion caused by the economic recession. (1984, 23).

Figure 4.1 illustrates the annual rate of return of Canada's Class 1 railways. The railways said that they were not able to achieve the higher rates of return necessary to facilitate large scale investments in railway capacity expansion because of the statutory grain rates. Until the mid-70s, policy makers had assumed that the railways received sufficient remuneration from the hauling

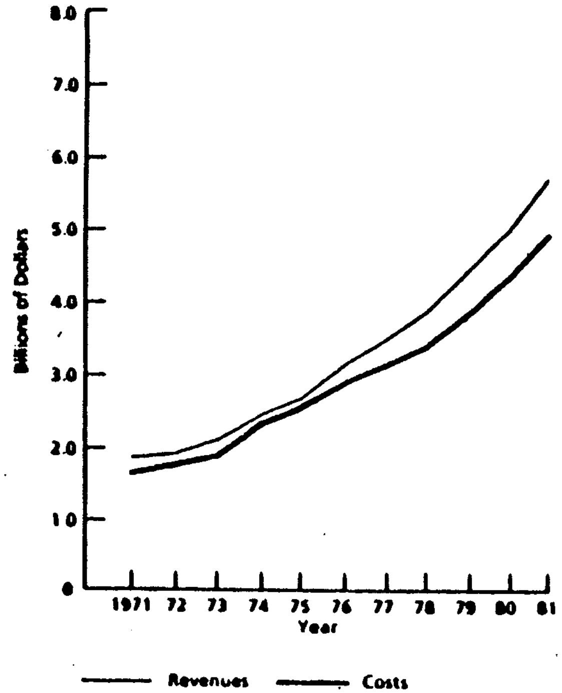


Sources: 1) Statistics Canada Catalogue No. 61-207.
2) Railway Transport Committee, Canadian Transport Commission.

*Does not include VIA Rail.

Figure 4.1: Annual Rates of Return - Class 1 Railways: 1971-1981

(reproduced from CTC 1983, p.23)



Source: Statistics Canada Catalogue No. 52-208.

Figure 4.2: Railway Operating Revenues and Costs: 1971-1981

(reproduced from CTC 1983, p. 23)

of grain to cover their overall costs. However, as was discussed in the previous section, the railways were losing hundreds of millions of dollars by hauling grain at the statutory rates.

By the late 1970s, the financial situation facing the operating environment and financial situation of the railways began to change for the better. The gap between railway revenues and costs began to significantly widen towards the end of the decade (see Figure 4.2). The CTC states that:

> since 1975, revenues have increased much faster than costs; the margin of 768 million dollars of revenues over costs in 1981 represents the highest figure achieved in the 1971-1981 time period. (1984, 23).

This was reflected in the increasing profitability of the Class 1 railways during the late 1970s. The CTC analysis of the rate of return for railways concluded:

> There was considerable recovery to much higher rates of returns in each successive year after 1975, the figure reached a record 7.4 per cent in 1979. The rate dropped to 6.7 per cent in 1980 but increased to 7.2 per cent in 1981. On a comparative basis only available to 1979, the rate of return for all corporations in Canada varied from 4.8 per cent to 8.6 per cent and for communications corporations the rate ranged from 6.2 per cent to 7.8 per cent (VIA Rail has not been included in this section although it is considered a Class 1 railway). (1984, 23).

The steadily improving financial picture, combined with a growing realization among policy makers that the statutory Crow rates would soon have to be abandoned, led the railways to start planning significant investments to expand railway capacity in Western Canada. By 1980, the railways were in the midst of planning enormous capital expenditures to install the capacity necessary for them to carry the growth in traffic volumes expected for the coming decade. According to the CTC:

> (CN and CP Rail predicted a 60 per cent increase in total traffic by 1980). Moreover, as much of the growth being expected involved Western Canada bulk commodities such as coal, potash, sulphur, and grain, traffic in this region was expected to rise by 70 per cent (compared with 30 per cent in Eastern Canada). (1985, 87).

By the beginning of the decade of the 1980s, the national railways had lost much of their earlier reluctance to invest in Western Canadian railway capacity. Now they were starting to implement their massive plans to change and expand Western Canada's railway infrastructure in order to meet the huge increases in traffic forecasted for the 1980s.

4.4 Projected Demand for Railway Capacity in Western Canada in the 1980s

The ability of the railway system in Western Canada to respond and adapt to projected changes in the volume, direction, and type of commodity movements, received considerable attention from both levels of government, producer groups, and the two national railways during the 1970s. Many studies conducted during this period focused on forecasting future commodity flows and the capacity of the railway system in Canada to meet these anticipated increases in demand.

The strong growth in the economies of the Western provinces during the 1960s and 1970s was reflected in their increasing share of domestic rail traffic. Transport Canada stated that in 1956, the Western share of total domestic rail traffic was 32 percent. By 1978, this percentage was estimated to have increased to 45 percent (Mulder 1980, 4).

According to Transport Canada, these increases were

> largely due to increasing demand for such bulk commodities as grain, potash, sulphur, forest products, and coal. (Mulder 1980, 4).

The demand for these products, and thus the derived demand for their movement by rail, was forecasted by Transport Canada to increase significantly during the decade of the 1980s. According to a report published by that department in 1981:

> flows of grain, coal, potash, and sulphur will provide the major stimulus to rail and marine traffic, and will also accentuate the continuing importance of Western provinces in the demand for freight transportation services. (1981A, 11).

This report also says that:

> during the forecast period the impetus for growth in rail traffic will come from grains and coal as well as potash and sulphur, . . . (1981A, 13).

The report concludes that:

> forecast growth in such resource commodities as grain, coal, and potash will affect the demand for marine rail transportation services in the West . . .

Growth will be strongest in the West. (1981A, 34).

In the late 1970s, the railways' own demand forecasts generally concurred with the ones conducted by Transport Canada. The railways, like the federal government, were well aware of the potential capacity crunch that could occur as a result of these increases. The Senior Vice-President for CN commented in 1980 that:

> in the first 40 years of its existence traffic on CN had grown to 80 billion gross tons miles per year. That figure doubled in the next 18 years. Our traffic this year is expected to exceed 160 billion gross ton miles.
>
> That sort of growth, while very gratifying, is not without its problems. In the early 70s it became obvious that we were facing capacity problems, particularly in the Western part of the system . . . The traffic that generates the need for this capacity is mainly bulk commodities - grain, coal, sulphur, and potash. At the same time we are facing growing volumes of forest products, petrochemicals, and intermodal traffic. We have to be prepared to handle it efficiently, expeditiously and competitively. (Walker 1980, 5-7).

CP Rail also foresaw greatly increased traffic volumes in the West, as well as their potential to cause capacity shortfalls. According to the Vice-President, Operations and Maintenance, for CP Rail:

> during the 1970s and early 1980s, it appeared that our Toronto to Vancouver mainline was going to run out of capacity . . . (Kelsall 1986).

It is evident that both national railways and the federal government, by the late 1970s, were forecasting huge growth in the demand for railway transportation in the West by 1990. In addition, it is clear that they agreed that the majority of this growth, based on recent trends, would come from increases in the movement of such bulk commodities as grain, coal, potash, and sulphur.

4.4.1 Forecasted Commodity Flows

Total railway loadings in the West were expected to post large increases in the decade of the 1980s. According to Transport Canada:

> the rapidly growing commodities originated primarily in the Prairie provinces and British Columbia, and are bound for Pacific Rim countries. Thus, the relatively high increases in rail loadings in the Prairies and unloadings in British Columbia of the last 20 years will continue. Prairie rail loadings will exceed 100 million tonnes in 1990 compared with 64 million tonnes in 1978, while rail unloadings in B.C. will reach 85 million tonnes in 1990 compared with 48 million tonnes in 1978. (1981A, 13-15).

Table 4.1 was produced by Transport Canada in the late 1970s (Mulder 1980, 5), and shows predicted rail traffic flows in Western Canada. It is clear that grain and coal were expected by 1990 to dominate commodity rail loadings in the West. Coal traffic, originating in Alberta and British Columbia, was expected to more than double from 1978 to 1990, reaching something around 35 million tonnes. Most of the coal traffic was destined for off-shore export markets via the west coast ports.

Grain traffic was also expected to increase significantly in the period from 1978 to 1990. Transport Canada forecasted that the traffic volume of grain would increase by over a third during this period. Figure 4.3 illustrates

TABLE 4.1

Rail Traffic Flows - Western Canada, 1978 and 1990
(Millions of Tonnes)

	Loadings			Unloadings		
Commodity	1978	1990	% Incr	1978	1990	% Incr
Grain	24.6	34.0	38.2	10.5	17.0	61.9
Fertilizer Materials	15.7	28.0	78.3	8.5	15.0	76.5
Coal	15.5	35.0	125.8	14.7	30.0	104.1
Forest Products	17.6	29.0	64.8	9.6	18.0	87.5
Other Commodities	25.0	37.0	48.0	24.9	40.0	60.6
Total	98.4	163.0	65.7	68.2	120.0	76.0

Source: Mulder 1980, p. 9

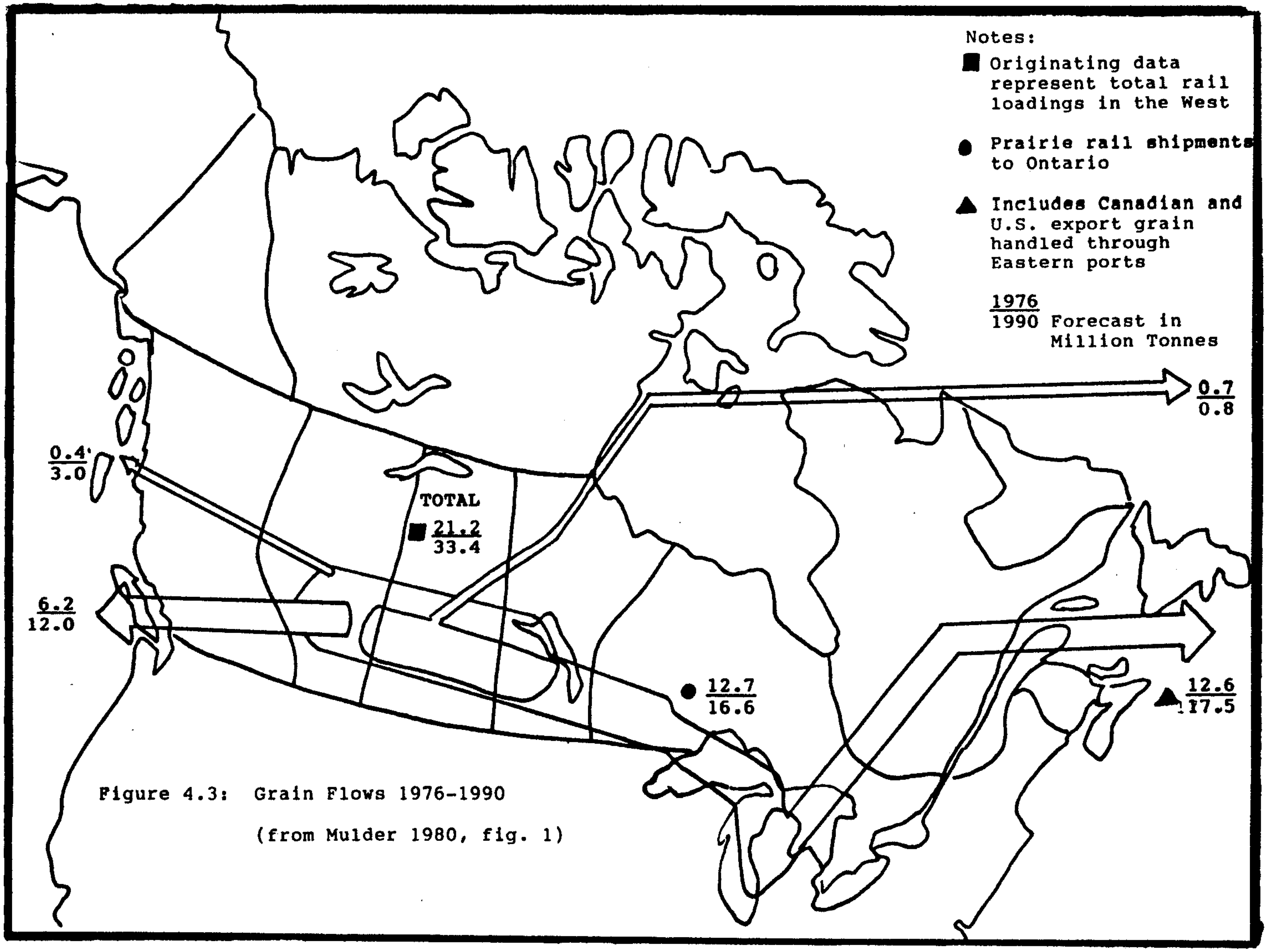


Figure 4.3: Grain Flows 1976-1990

(from Mulder 1980, fig. 1)

that about half of all Western grain destined for export markets in 1990 will move through Vancouver and Prince Rupert; the other half was expected to move through Thunder Bay and the Eastern Great Lakes/St. Lawrence Seaway system. Figures 4.4 and 4.5 show the expected direction and volume of Western coal, potash, and sulphur to be moved by the railways in 1990.

These traffic volume forecasts had a dramatic impact upon the policies of the federal government with respect to the regulation and subsidization of the railways. With billions of dollars for Western railway capacity expansion at stake, the federal government looked for policies that would achieve a balance between measures to improve railway profitability and efficiency, measures to ensure the fair pricing of railway services, and measures that would make carriers fulfill their obligations to common carriers.

From the perspective of the railways, both CN and CP Rail recognized that significant investments would have to be made to increase the capacity of their respective Western networks. These investments were essential to ensure the efficient movement of all Western commodities in the 1980s and beyond. The following section discusses the capacity related problems and issues that existed in the Western CN and CP Rail systems during the 1970s.

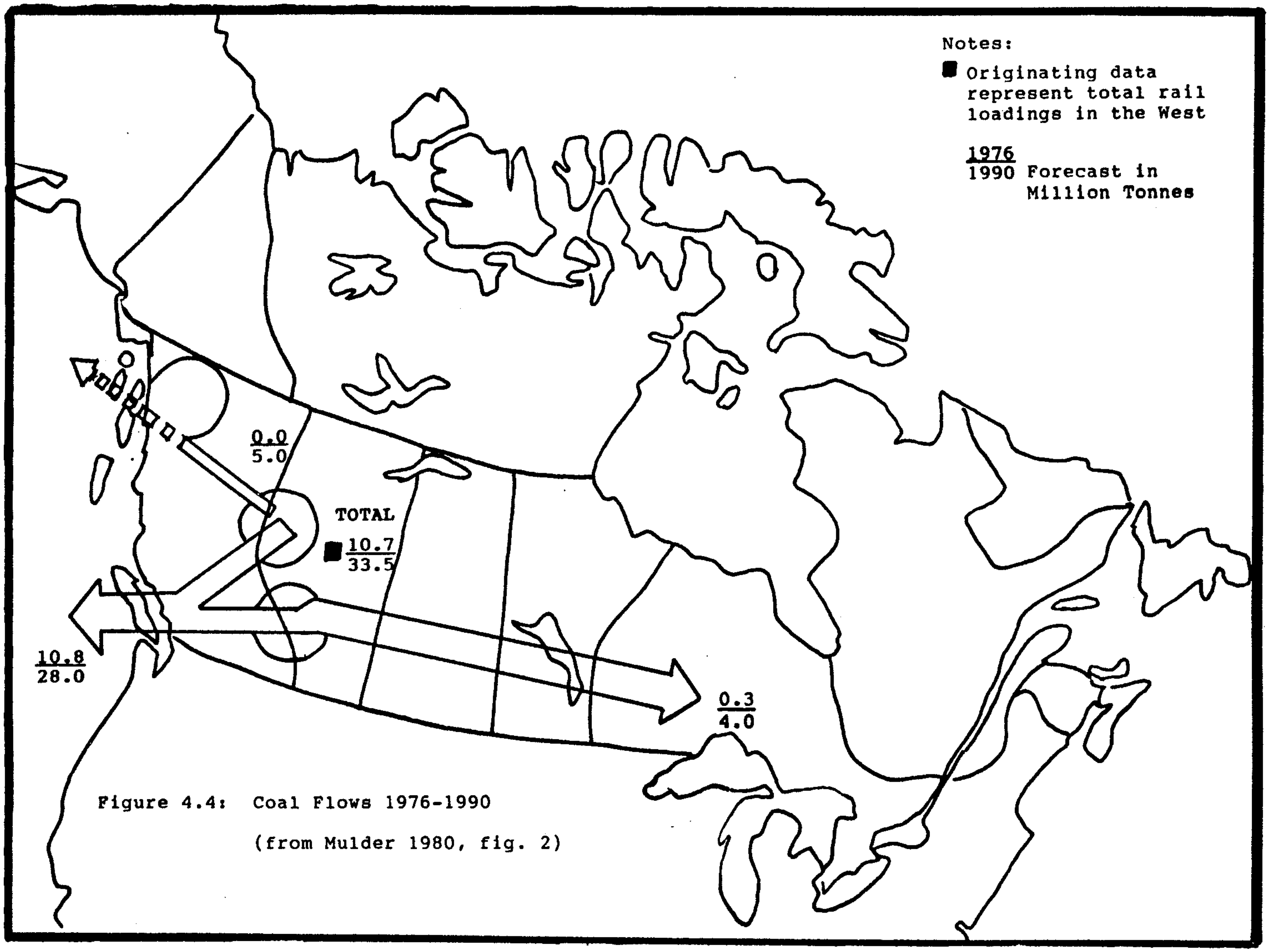
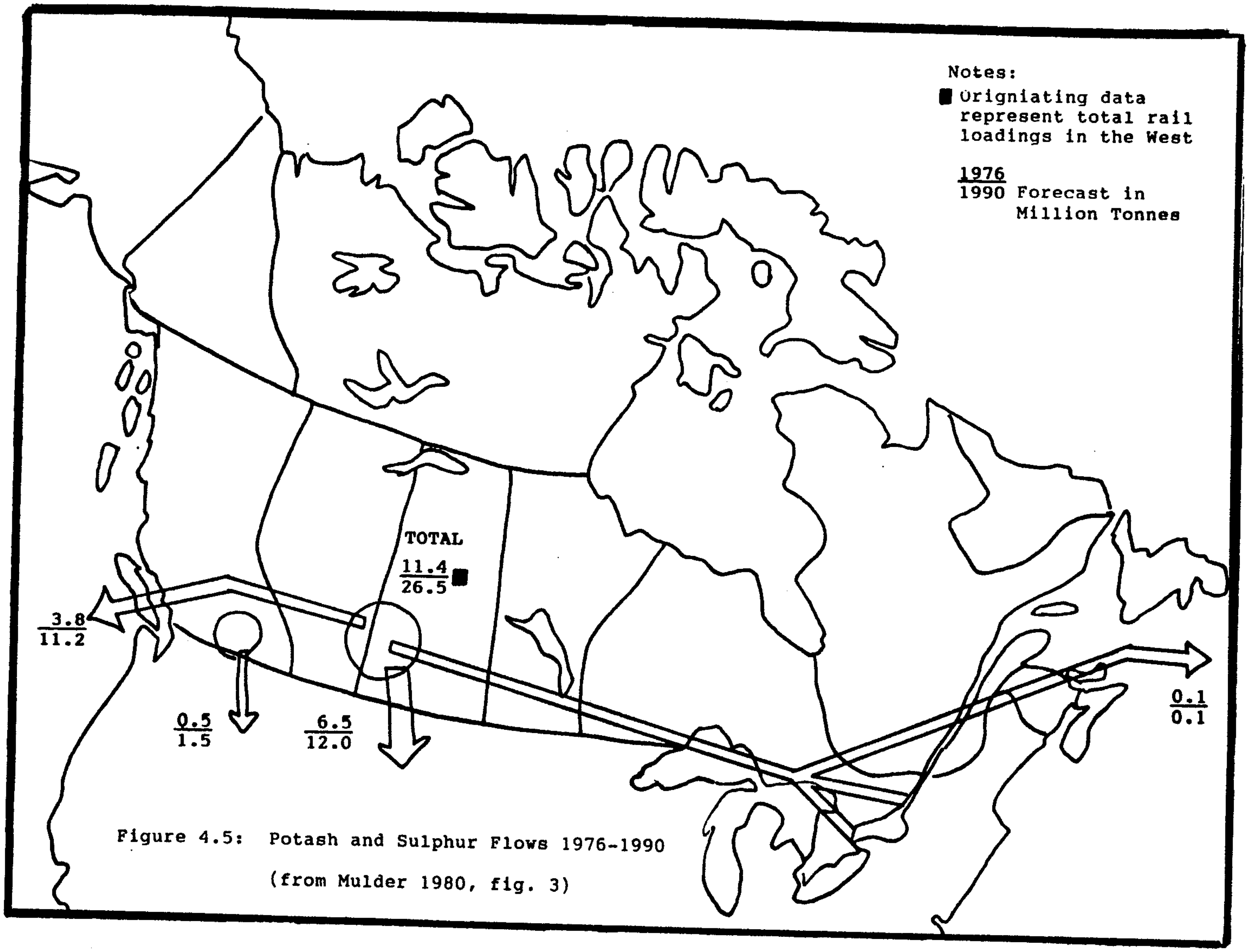


Figure 4.4: Coal Flows 1976-1990

(from Mulder 1980, fig. 2)



Notes:
■ Origniating data represent total rail loadings in the West

1976/1990 Forecast in Million Tonnes

Figure 4.5: Potash and Sulphur Flows 1976-1990

(from Mulder 1980, fig. 3)

4.5 Western Railway System Capacity in the 1970s: Capacity Constraints

Both CN and CP Rail own and operated an extensive railway system in Western Canada. In 1974, the combined mileage of first main track for both railways in Western Canada was 21,887 miles (Statistics Canada Cat. 52-209 1974). This figure represents almost 56 percent of the combined first main track mileage operated in Canada in 1974 by both CN and CP Rail. Figure 4.6 illustrates the location of the CN and CP Rail mainlines Western Canada.

The CN system, which operated 23,306 miles of first main track in the four Western provinces in 1974 (Statistics Canada Cat. 52-209 1974), starts at Thunder Bay and proceeds west to Winnipeg, with one line going to the north of the Lake of the Woods, and the other running south through part of the U.S. and then up to Winnipeg. From Winnipeg, the CN mainline branches into two, with one line proceeding northwest, and then northeast, to Churchill, while the other mainline runs northwest through Saskatoon to Edmonton. From Edmonton, the mainline goes towards the southwest to Red Pass Junction in the Rocky Mountains. Here the CN mainline again branches into two lines, with one running northwest towards Prince Rupert, while the other passes southwest through Kamloops to Vancouver. The major yard/terminal facilities of the Western CN system are located at Thunder Bay, Churchill,

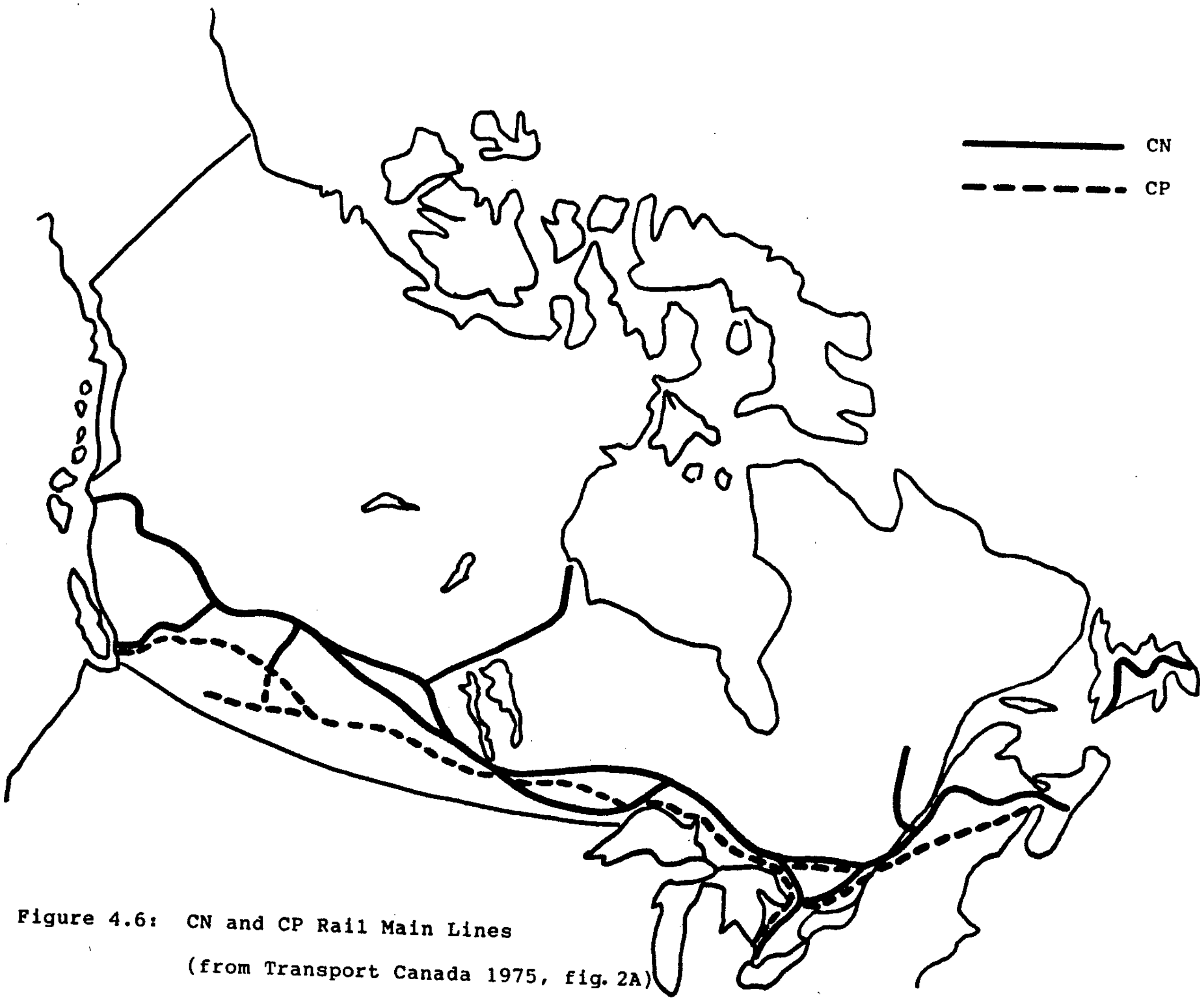


Figure 4.6: CN and CP Rail Main Lines

(from Transport Canada 1975, fig. 2A)

Winnipeg, Saskatoon, Edmonton, Prince Rupert, and Vancouver.

The Western CP Rail system also starts at Thunder Bay. The CP Rail mainline runs westward to Winnipeg, parallel to the northern CN mainline. From Winnipeg, the CP Rail mainline runs due west through Regina to Calgary. The mainline then runs west over the severe grades located at Rogers Pass in the Rockies, and then travels southwest towards Vancouver. The major yard/terminal facilities for the CP Rail Western system are situated at Thunder Bay, Winnipeg, Regina, Calgary, and Vancouver.

Figure 4.7 highlights the critical rail links in Western Canada that were indentified in an analysis conducted by Transport Canada in the mid-1970s (1975, 38). The purpose of the analysis was to identify the critical rail links which were expected to experience capacity limitations by 1990, and to determine the range of trains per day which would be able to move over those links. The analysis concluded that there would be serious shortages of capacity for:

> most or all sections of the CN mainline between Vancouver and Thunder Bay and the CP mainline between Vancouver and Winnipeg. (1975, 37).

The report went on to state that with respect to the CN system:

> The most critical links are between Edmonton and Red Pass Junction, presently carrying about 26 trains per day (both directions) and expected to carry 42 - 64

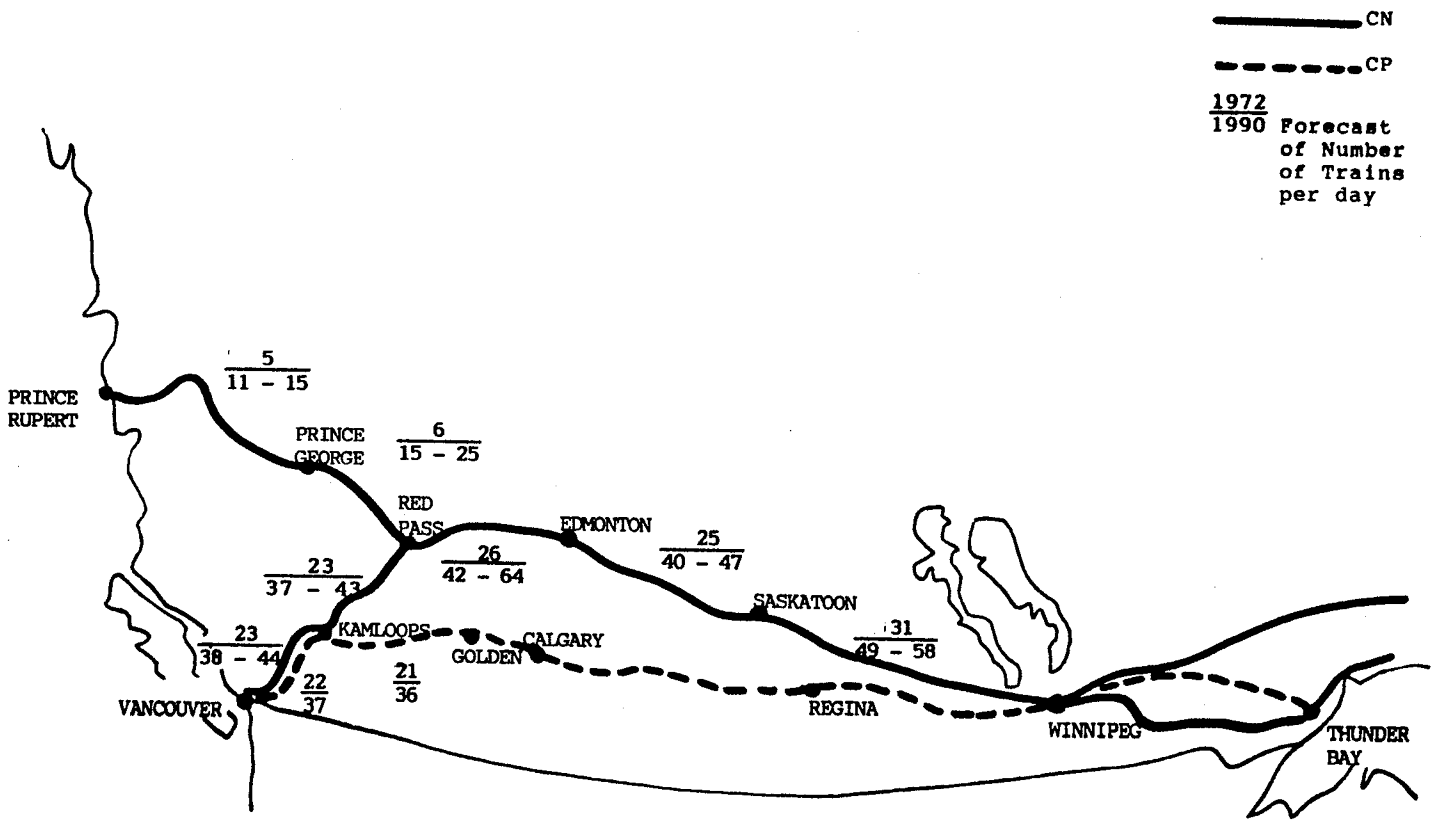


Figure 4.7: Critical Rail Links in Western Canada 1972-1990

(from Transport Canada 1975, fig. 22)

> trains per day by 1990; the link between Vancouver and Red Pass Junction (23 trains per day growing to 37 - 44); the link between Edmonton and Saskatoon (25 trains per day growing to 40 - 47); the link between Saskatoon and Winnipeg (31 trains per day growing to 51 - 58); the link between Winnipeg and Fort Frances (19 trains per day growing to 28 - 44); and link between Fort Francis and Thunder Bay (19 trains per day growing to 35 - 39). (1975, 37).

For the CP Rail system in Western Canada, the report stated that:

> the critical links are between Golden and Vancouver (24 trains rising to 39); between Golden and Calgary (20 trains per day rising to 33); Calgary to Regina (11 trains per day rising to 20); and Regina to Winnipeg (22 trains per day rising to 30). (1975, 38).

Transport Canada went on to say that capacity improvements in the CN and CP Rail mainlines running, respectively, west from Edmonton and Calgary, through the Rocky Mountains, and onto Vancouver would be very costly to undertake.

It was the mountain divisions of CN and CP Rail that presented the most difficult capacity problem for railway planners to solve during the 1970s. Any investment to increase capacity in the Rockies would be extremely expensive. In certain cases, the costs would be prohibitive. By the late 1970s, the railways had invested funds in upgrading the lines running through their mountain divisions. All of the less costly solutions for increasing capacity had already been undertaken. These solutions included longer sidings, reduced spacing between sidings, and double tracking where economically feasible. However, even with these measures, it seemed that capacity limits would soon be reached on the main- lines running through the mountains. Table 4.2 shows the projected rail traffic by commodity for the mountain divisions of CN and CP Rail, and Figure 4.8 shows the railways' forecast, from the early 1980s, for Western track capacity and demand. It is clear that the railways realized that they did not possess the track capacity to meet the traffic demands that they were forecasting to occur in the next few years.

By 1972, capacity problems were beginning to show in the Western CN system. This was particularily true on the section of mainline that ran from Edmonton to the ports on the west coast. According to CN:

> capacity problems on CN's mainline to the west coast were becoming evident. CN's line at that time had virtually no double track on the Edmonton-Vancouver route. Centralized traffic control over the route was incomplete. (1985, 9).

This report went on to say:

> At the same time, the basic plant consisted of jointed, medium weight rails on wooden ties, with less than optimal ballast and many wooden trestles over the route. There were numerous examples of deferred capital spending, the legacy of decades of prior years' annual financial losses for CN. (1985, 20-21).

On the segment between Edmonton and Jasper, trains were being queued on track sidings or yard tracks as far as Edson and as far west as Kamloops. The CN report claims problems on this segment were so bad that:

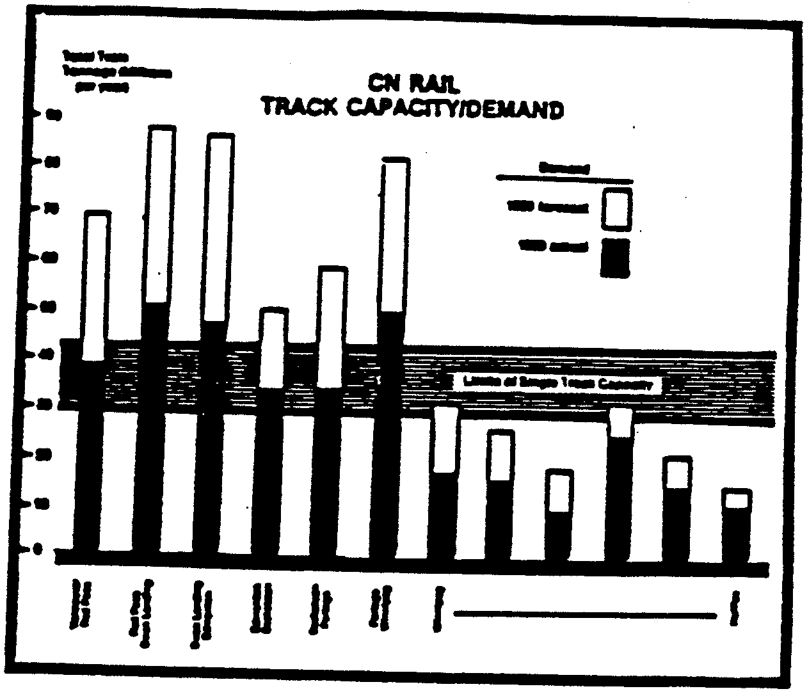
> many crew members worked day and night because of the

TABLE 4.2

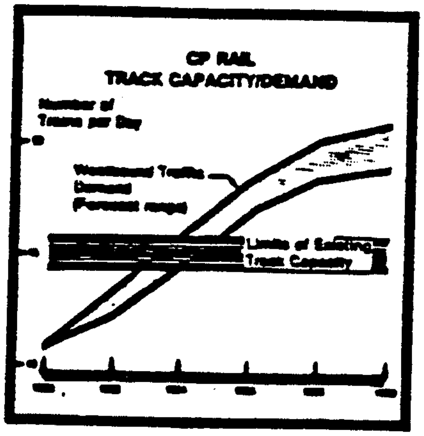
Existing and Projected Traffic on Western CP and CN Western Main Lines, 1980, 1986, and 1990
(Millions of Net Tonnes per Mile)

	CP Mountain Subdivision			CN Red Pass to Swan Landing		
Commodity	1980	1986	1990	1980	1980	1990
Coal	10.5	31.9	36.2	5.3	12.3	14.3
Grain	4.2	5.9	6.3	6.3	6.3	10.0
Potash	0.6	2.5	3.6	3.4	4.9	7.8
Sulphur	2.2	2.7	2.6	3.7	4.0	4.0
Forest Products	1.3	1.7	1.4	6.5	7.5	8.4
Other	5.8	8.1	9.3	2.9	3.5	4.9
Total - Net Tonne-Miles	24.6	52.8	59.4	28.1	38.5	49.4
- Gross Tonne-Miles	41.8	89.8	101.0	50.0	67.0	86.0

Source: Waters 1983, p. 34



Source: CN Rail



Source: CP Rail

Figure 4.8: CN and CP Rail Capacity Limits in Western Canada (reproduced from Waters 1983, exhibit 7)

> unproductive deadheading and tying up on line. Minor problems that would normally be of little significance, such as a broken air hose, could cause undue disruption to the flow of traffic. Trains were ordered at their respective terminals to depart at set intervals so as to fit the grid times (planned schedule for a given mix of trains, based on the "grid" of sidings and other factors), but bunched up after departure when the flow was interrupted. The result was one or more trains being set out short of destination because of crew time on duty, or being unable to clear the next day`s work block. (1985, 20-21).

Capacity problems on the Jasper to Vancouver route were just as severe. This line was controlled only partially by centralized traffic control. Siding length average 6,083 feet and were spaced, on average, 8.7 miles between one another (CN 1985, 22). Much of this section is located in the mountainous terrain of the Rockies. This terrain is characterized by natural barriers that constrain the number of sidings and sections of double track that could be constructed to relieve capacity problems. Tunnels, snow sheds, slide detectors, extreme curvature, etc. are all extraordinary characteristics that must be built into a railway line running through such difficult mountain terrain. With respect to the problems incurred by the movement of trains over this line in the early 1970s, the CN report provides some interesting insights as follows:

> Train stacking occurred frequently, mainly because of grain congestion at Thorton Yard in Vancouver. Potash trains were also frequently stacked because of delayed ship connections. Sidings were used regularly for storing trains and often trains were stacked as far east as Swan Landing and Solomon on the Edson subdivision -- over 550 miles from

> Vancouver. At that time, grain trains were marshalled by commodity and consignee so as to allow quick spotting after arrival at Vancouver. A coordinator was appointed at Thornton to work with the Chief Dispatcher at Kamloops to bring certain trains into the yard on a programmed basis. This eventually evolved into control of potash trains being dispatched from the lines on the Prairie Region. In spite of the tight control over traffic into Thornton Yard and efforts to have trains marshalled, stacking increased to the point that deadheading and running light (relocating locomotives by running them without a train) were too costly to continue. Track conditions deteriorated steadily and conventional maintenance procedures were not sufficient to keep up, let alone improve track conditions. (1985, 22).

In addition, congestion problems were beginning to appear at major CN yards in the West by the mid-1970s, particularly Vancouver and Edmonton. By 1973, the capacity of the latter began to be reached. The temporary solution to this problem was to proceed with the blocking of cars into trains at intermediate terminals located to the west and east of the Edmonton yards. However, according to CN, this created other problems:

> This practice, in turn, created delay at the intermediate terminals, since they were not designed as marshalling yards. A general problem developed with crews due to long hours on duty. (1985, 19).

CP Rail, like its competitor, began to experience capacity related problems over its Western Canadian mainline as far back as the early 1970s; and like CN, the section of mainline running over the mountains, west from Calgary towards Vancouver, was characterized by the worst problems. In the early 1970s, the major double tracked sections of the CP Rail Western system were located between

Thunder Bay and Winnipeg, with additional double track segments on the Prairies and for some 100 miles of the 625 mile mainline between Calgary and Vancouver (Transport Canada 1975, 7).

Capacity constraints on the CP Rail mountain division were mainly due to the limited amount of double tracking that could be undertaken on this route, along with the severe grades located in this region. Railway planners often regard the capacity of a railway line as the capacity of the weakest link along that line. In terms of railway operations, the weakest link is the "ruling grade". The ruling grade is the most severe grade a train must pass over while travelling over a particular railway line. The tonnage limitations on such a grade governs the length and weight of the trains that pass over it. In the case of CP Rail traffic moving west from Calgary to Vancouver during the 1970s, there were several severe grades which required up to ten 3,000 horsepower trains to move a 13,000 ton train. In comparison, it took only four such locomotives to move a similar train through the CN mainline from Edmonton to Vancouver (Duncombe 1975, 21).

Movement of commodities by rail to the west coast was hampered by the restrictions imposed by these severe grades. Critical ruling grades are those in excess of one percent. At this point, the slope becomes too severe to permit efficient unit train operation. CN's mountain division is relatively free of such critical grades, with the lines to the B.C. ports running in the 0.7 percent range (Duncombe 1975, 21). However, the CP Rail mainline running west from Calgary passed over four critical grades. These grades, located in the Rogers Pass area, ranged from 1.6 to 2.2 percent.

Westbound trains of both railways on their way to Vancouver must travel through the Thompson and Fraser River Canyons, west of Kamloops. The terrain and geology of this area make increases in capacity, through the expansion of the plant, extremely difficult. According to the CTC:

> it is through these areas (the Thompson and Fraser River Canyons) that double tracking is expected to be the most difficult, certainly extremely expensive, and perhaps impossible in parts. The difficulties are fundamentally geological and environmental - a narrow rocky canyon beside a fast flowing major river system with no room to construct a line on the land side, and no place to dump the debris from blasting. Moreover, much of the canyon is a zone of geological instability with landsides being a constant threat. The instability also threatens tunnels, especially long ones.
>
> Thus, it may be that short sections of the canyon lines can be double tracked, and sidings built, but there may well remain areas where these solutions cannot apply. Future research is needed to see if there are ways that will permit the railways to expand their traffic flows while avoiding the extremely high capital costs implied in full double tracking of both mainlines. One such possibility is the joint use of the existing tracks, with a common dispatcher controlling all traffic. (1984, 73).

Another concern of the railways during the 1970s was the issue of congestion of rail yards in Vancouver and

Thunder Bay. Studies were undertaken in regards to determining ways of alleviating the congestion problems, both real and potential, occuring at these yards. The CTC stated that in regards to Western railway capacity:

> there remains, however, the possibility of some problems westbound within Vancouver itself, and perhaps in the Fraser Canyon. In each case, work is underway or being considered, to improve the operating use of the existing infrastructure before major capital projects are undertaken. Railway capacity for commodities moving eastbound through Thunder Bay appears adequate, although very large increases in grain flows may well bring about operating changes to the entire system of grain handling and transportation, and not just the railway component. (1984, 76).

The limited life of track, and the subsequent efforts required to replace the track is another factor that affects Western railway system capacity. The standard 130 lb/yd railway track is estimated to have a life of some 800 million gross ton-miles per mile, over straight sections. this means that if the track is carrying 200 million gross ton-miles per mile of track every year, the rail would have to be replaced every four years. The more frequent the replacement of track, the more capacity over that line is limited. Double tracking with cross-overs, if possible, is one way that railway planners can deal with maintenance related capacity limitations for heavily travelled lines. From the perspective of the demand side, bottlenecks in railway system capacity in Western Canada often originated with the shippers. Shortage of rail cars result from the seasonality and peaking of demand. The volume of grains and other agriculture related products (including potash and other fertilizers) moved in the West by rail show great fluctuations, depending upon the time of year. According to Duncombe:

> grain and agricultural volumes in the West show peaks in the summer and fall months, livestock in the fall months, fertilizer (potash) in the spring and fall, wiht other products displaying a more constant flow rate throughout the year. The winter peaks of grain in the east reflect closing of the Great Lakes system and diversion of grain shipments to rail. In general the winter months; December, January, and February, represent the low point for both the rail and canal modes. Climatic conditions are important in this regard . . . (1975, 20).

Thus, substantial portions of overall system capacity may be unused at certain times, while at other times system capacity may be fully utilized due to the peaking of traffic. According to Transport Canada:

> the practical capacity of a transportation facility will be less . . . if traffic peaking is such that the facility is relatively unused for much of the time. Conversely, the capacity of a facility will be considerably higher . . . if a constant flow can be maintained for this facility on an hourly, daily, weekly and monthly basis. (1975, 11-12).

Figure 4.9 illustrates the practical implications of seasonality and peaking upon the average cycle time of hopper cars carrying grain to the different exit ports on the CP Rail Western system. Car cycle times in 1980 at Vancouver ranges from 18.7 days in January (weeks 1-5), to a high of 26.2 days July and August (weeks 28-31), and then back down to a low of 18.6 days in December (weeks 50-52).

TABLE 4.3

Canadian Pacific Hopper Car Cycle Times, 1980 and 1990
(Number of Days)

	Vancouver		Thunder Bay		Prince Rupert		Churchill	
Weeks	1980	1990	1980	1990	1980	1990	1980	1990
1-5	18.7	15.9	14.3	11.4	23.0	19.6	21.0	21.0
6-10	18.8	16.0	14.7	11.8	23.0	19.6	21.0	21.0
11-14	18.8	16.0	14.8	11.8	23.0	19.6	21.0	21.0
15-18	19.0	16.2	15.0	12.0	23.0	19.6	21.0	21.0
19-23	24.6	20.9	20.3	16.2	23.0	19.6	21.0	21.0
24-27	25.8	21.9	27.6	22.1	23.0	19.6	21.0	21.0
28-31	26.2	22.3	27.8	22.2	23.0	19.6	21.0	21.0
32-36	22.9	19.5	27.5	22.0	23.0	19.6	21.0	21.0
37-40	25.3	21.5	26.0	20.8	23.0	19.6	21.0	21.0
41-44	22.0	18.7	14.2	11.4	23.0	19.6	21.0	21.0
45-49	22.0	18.7	12.5	10.0	23.0	19.6	21.0	21.0
50-52	18.6	15.8	13.6	10.9	23.0	19.6	21.0	21.0

Source: Grain Transportation Authority 1981, p. 38

4.6 Summary

This chapter has outlined the concerns, issues, and problems regarding railway capacity in the West that faced CN and CP in the 1970s. It is clear that there was a number of different interest groups, including shippers and government, that focused attention on the problem of providing adequate rail capacity to meet the demands expected to occur in the 1980s.

The railways, although fully cognizant of the problem, were reluctant to commit the hundreds of millions of dollars needed to install additional capacity in the West. The existence of the statutory Crow grain rates was a significant impedement to the long range capacity planning efforts of the railways. The removal of the Crow rates was necessary to ensure the economic viability of the capacity expansion investments undertaken by the railways.

With their profitability strengthened, the railways began to undertake large scale investments to increase the capacity of their Western networks. These plans were predicated upon forecasts that foresaw significant increases for Western rail traffic demand. Coal, potash, grain, and forest products were the major commodities expected to post large increases in the demand for Western rail capacity during the 1980s.

The railways, shippers, and governments realized in the late 1970s that the existing railway network in the West would soon reach capacity. The operations and planning functions of both railways identified a number of critical capacity bottlenecks existing in their respective networks. The most critical of these capacity constraints were located on the CN and CP mainlines running west through the Rockies to Vancouver. Hundreds of millions of dollars would have to be invested by each railway for double tracking and grade reductions in order to eliminate these bottlenecks. Once these bottlenecks were eliminated, the production function for both CN and CP would shift dramatically, and Western rail capacity would be greatly increased.

The next chapter provides an overview of the investment plans made by CN Rail and CP Rail in the early 1980s to eliminate their critical network bottlenecks, and thus increase the capacity of their Western networks.

CHAPTER 5

WESTERN CANADA RAILWAY CAPACITY PLANNING IN THE 1980s

5.1 Introduction

The previous chapter described the circumstances which led CN Rail and CP Rail to invest in increasing Western railway capacity in the 1980s. During the mid 1970s, shippers and government expressed their concerns about the ability of the Western railway network to handle the traffic demands expected to occur in the next decade. The railways were at the point where they could no longer achieve significant increases in capacity without making large scale investments. In response to the impending capacity crunch, the railways devised plans to spend billions of dollars to eliminate the critical capacity bottlenecks that existed during the late 1970s.

The purpose of this chapter is to give an outline of the investment plans devised by CN Rail and CP Rail with repect to increasing the capacity of the Western railway system during the 1980s. Both railways began to implement their capacity expansion plans during the mid to late 1970s. These investments were large scale in nature and took several years to complete. This discussion will focus on the objectives, costs, benefits, and scope of the capacity investment plans initiated by each railway.

5.2 Overview of the CN Rail and CP Rail Investment Plans

Both CN Rail and CP Rail realized the necessity to install additional capacity in their respective Western systems in order to be able to meet the traffic demands expected in the 1980s. The two railways initiated long range investment plans to increase system capacity during the late 1970s and early 1980s. The objective of these plans was to build a level of capacity that would be adequate to meet the 60 percent increase in total traffic that was forecast for 1990 (CTC 1985, 87). Much of this growth was expected to occur in Western Canada, as the ever increasing export demand for that regions' bulk commodities began to place severe strains on the railways' ability to move them. As a result, CN and CP planned large scale investments to expand their railway infrastructure in the West. The CTC states that:

> consequent to this expected growth, the railways' plan for the 1980s, as indicated by the Corporate Vice-President for Canadian National at a Financial Post Conference in December 1980, included 11.5 billion dollars (1980 constant dollars) of capital expenditures, a substantial increase from the 2.4 billion dollars spent between 1975 and 1980. (1985, 87).

The railways identified several capacity bottlenecks in their respective Western networks. Their analyses identified the existence of a number of critical capacity bottlenecks on the mainlines that ran West over the Rockies towards the West coast.

CN focused its plans on double tracking the line running West from Edmonton to Vancouver. In addition, CN planned investments on the line to Prince Rupert to occur when additional traffic demands warranted an increase in capacity. The plans formulated by CP Rail concentrated on grade reduction efforts on the Westbound Calgary to Vancouver mainline.

However, the increases in Western rail traffic volumes that were predicted to occur in the 1980s did not happen. The CTC claims that due to:

> the 1982/83 recession, the expected growth in traffic did not materialize. Industries, in general, have adopted a more cautious approach to forecasting growth and planned capital expenditures, and the railways have scaled back their investment plans considerably from those predicted in the early 1980s. (1985, 88).

Despite the restriction on traffic growth caused by the recesssion, both CN and CP Rail's revised forecasts still predicted fairly significant traffic growth by the 1990s. For example, in the early 1980s CN had predicted that the total gross ton-miles carried over the Red Pass to Vancouver line would reach about 67 million by 1990 (see Figure 5.1). CN's revised forecast in 1984 predict-

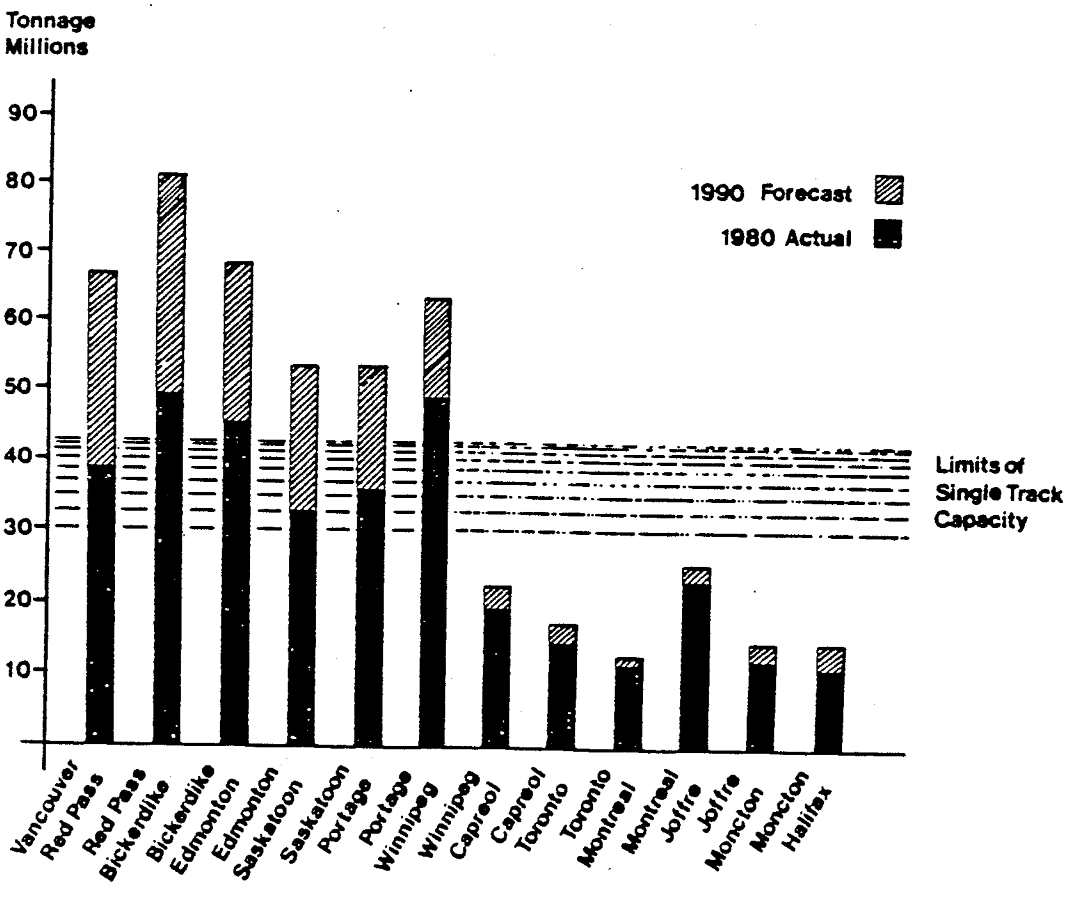


Figure 5.1: CN Rail Forecast Tonnages (Gross Tonne Miles per Mile)

(reproduced from Transport Canada 1984, fig. 7.2) ing that the total gross ton-miles carried over this route would reach about 60 million in 1994 (see Figure 5.2). CP made similar downward revisions to their traffic forecasts. Despite the reduced level of traffic growth expected for the 1990s, both CN and CP pressed their capacity expansion programs forward.

5.3 The CN Investment Plan to Increase Railway System Capacity in Western Canada during the 1980s

Capacity problems on the CN mainline to the West coast were becoming evident as far back as 1972. As a result, the railway began planning for the expansion of capacity of their Western system during the years of the early to mid-1970s. CN planned two consecutive programs to increase the capacity of their Edmonton to Vancouver route. The first one, dubbed the Plant Improvement Program, ran from the mid-1970s to 1981. This extensive plan was part of a national program to make improvements in all parts of the Canadian CN system. The Plant Expansion Program began in 1981, and was an effort to increase the capacity of the CN system in Western Canada through major investments in plant. However, it should also be noted that before CN attempted to increase capacity through expensive investments in their Western plant, they implemented a number of lower cost operating, traffic, and equipment solutions in order to achieve incremental

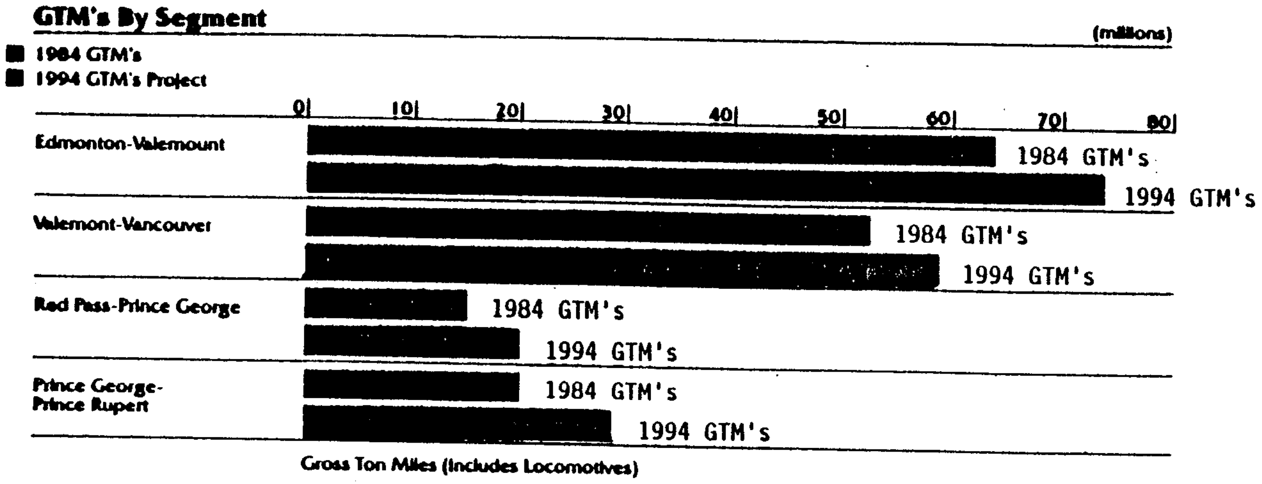


Figure 5.2: CN Gross Tonne Miles by Segment: 1984 and 1994 (projected) (reproduced from CN Rail)

increases in capacity.

In 1972 the majority of CN train movements running West from Edmonton passed southwest to Vancouver. These trains predominantly carried bulk commodities destined for export. Figure 5.3 shows the breakdown by type of traffic carried over the CN Edmonton to Vancouver route in 1972. Capacity was rapidly being approached on this line. According to CN:

> traffic growth to 1972 and short term expectations had already pushed some sections of the line to the accepted capacity limit. Yard congestion was becoming a severe problem at the major terminals, with serious effects on car cycles. Making this situation worse was the lack of satisfactory equipment maintenance facilities at a number of major locations, with outdoor inspection and servicing facilities providing particular problems for winter operations. This overall situation was all the more unattractive when viewed in the context of widespread forecast growth in the export of bulk commodities from Western Canada, which promised to put further extreme traffic growth demands on CN's Western line. (1985, 10).

Another capacity concern that was beginning to compound this problem was railway access to Vancouver. CN states that:

> the fact that the BC South Line terminates at the Port of Vancouver made the situation more difficult. Vancouver is a seaport with a large number of unavoidable bottlenecks for railway operation. Five railways serve the terminal area and there are numerous interchanges, bridges, and complex trackage. CN's primary traffic over this line is westbound and this means that the flow of traffic into the port area requires efficient unloading and cycling, or extreme congestion results. However, in the seventies (and to a lesser extent today) unloading facilities in the port were not efficient loop-track unit train facilities, but rather stub-end wharf tracks

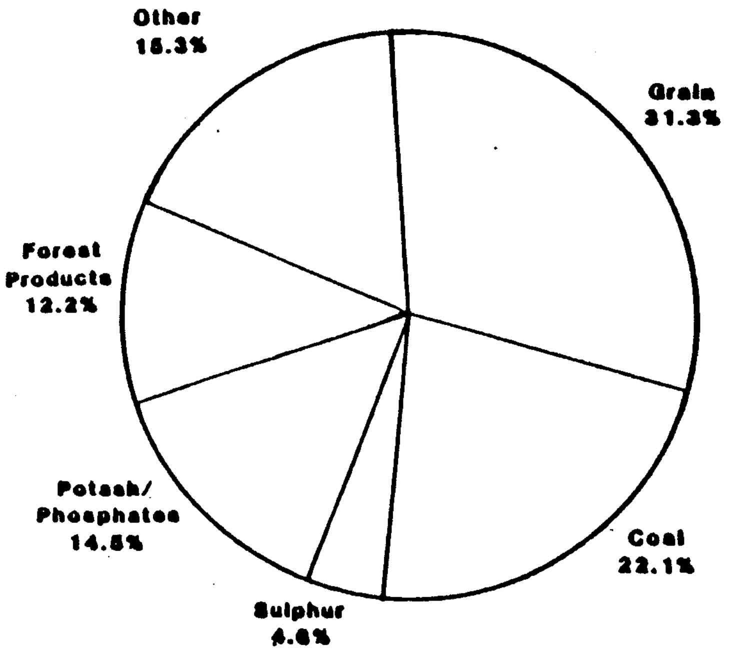


Figure 5.3: CN Rail Traffic Between Edmonton and Vancouver: 1972

(reproduced from CN 1985, fig. 5)

> which require multiple switches per train to place and pull cars, which are frequently 'held out' until ordered in by the terminals. The numerous sources of congestion within the Vancouver Port area created queuing problems and train set-outs for CN which backed up well eastwards along our mainline. (1985, 11).

CN studied a number of potential options to solve the growing problem of restricted limitations over their Western system during the 1970s. The first options to be seriously considered were certain operating solutions. These solutions were considered because they usually cost the least amount to implement. CN says that such solutions as:

> the use of longer trains was considered to be a means of building capacity, although longer trains led directly to a need to provide longer sidings and yard track. (1985, 11).

Another operating solution considered was the regulation of train speed at a more consistent level:

> Traffic over the line moved at a mixture of freight, passenger, and express train speeds, and the regulation of train speed at a consistent level for all was used more and more over the past decade to provide capacity. (1985, 11).

Solutions to congestion problems at yards located over the Edmonton to Vancouver line were also implemented:

> to reduce congestion in yards -- particularly in Jasper and Calder -- operating plans were developed which called for certain switching activities to be relocated, away from those yards. Trains were preblocked either east or West of the problem areas, or allowed to be mine run (i.e. cars in random order; cars left in the sequence in which they are received. (1985, 11).

The Canadian Transport Commission along with Transport

Canada and the governments of British Columbia, Alberta, Saskatchewan, and Manitoba sponsored a study of the Vancouver rail access problem. This study recommended several operating solutions to alleviate the problem. However, it noted that CN would be required to make substantial capital expenditures to upgrade the capacity of their Thornton Yard in Vancouver. The CTC states:

> while it is recognized that the railway plans will, when completed, postpone or prevent the impending crisis, the intergovernmental study has sought ways to enhance the system's capacity through closer co-operation between the railways, and through operating changes with minimum capital investment. Thus, it is felt to be a positive step to increase the capacity of the CP Rail Coquitlam Yard, and to divert traffic to it from the CN Rail Thornton Yard. Solid trains of bulk commodities destined to the south shore of Burrard Inlet are prime candidates for such diversion. Grain, potash, and sulphur in particular could be interchanged between CN Rail and CP Rail at Mission and run into Coquitlam in this way. . . . Eventually, additional capacity at the Thornton Yard would be needed, but this suggested diversion could make time available to construct the new plant and avoid traffic disruptions. (1984, 75).

CN also rejected a number of potential operating solutions. Train fleeting, the running of trains in batches, was considered, but then rejected on the basis that extreme problems would occur at terminals at both ends of the route (at Edmonton and Vancouver). Joint track usage by both CN Rail and CP Rail was considered for the Fraser and Thompson Canyons, but was rejected for the time being. CN claims that this plan was rejected because:

> for joint track usage in the canyon, parallel main tracks only provide an acceptable degree of capacity enhancement when there are crossovers every ten or twenty miles, and this would have been prohibitively costly to construct with the Thompson and Fraser Rivers separating the CN and CP lines. (1985, 12).

It should be noted however, that both railways and the governments of British Columbia, Alberta, Saskatchewan, and Canada have studied this option in greater detail. The CTC states:

> In 1983, the Province of Alberta approached the CTC's Research Branch to undertake a study of joint track usage in the Fraser/Thompson canyons. The project was to be carried out in a manner similar to the studies of Vancouver rail access and Thunder Bay rail capacity . . . The major conclusions of the study are that joint track usage is operationally feasible and that, relative to double tracking, environmental problems and capacity expansion costs can be reduced. However, forecasts now indicate that a total investment of 87 million dollars by both CN and CP under independent operation would provide sufficient capacty for at least the next 10 years. This contrasts sharply with previous forecasts of over 1 billion dollars and constitutes a strong argument against further governmental interest in the short term. Over the long term, should circumstances change such that forecast investment requirements again rise, joint track usage should again be addressed, especially in view of the long-term potential for savings in capital. (1985, 88).

Another category of options that were considered by CN to alleviate Western capacity problems were traffic solutions. CN sought traffic solutions to solve some of the capacity problems occuring at their yard facilities in Vancouver:

> To reduce costly switching and the associated bottlenecks for grain at Vancouver, the pooling in the terminal of Board grains (i.e. wheat, oat, and

> barley for export that is controlled by the Canadian Wheat Board) was achieved in the late 1970s. (1985, 12).

CN also encouraged the increased use of heavier, shipper-owned cars to produce increases in capacity through the use of pricing incentives. This solution worked particularly well in the movement of lumber products. CN states that:

> CN encouraged this development by the use of incentive pricing, particularly for lumber -- lumber was the number one revenue commodity for CN in the late 1970s and was a major user of capacity over the Edmonton/Vancouver line. Weighing of lumber shipments from the B.C. North division was a heavy workload, particularly at Jasper and Calder, and the implementation of weight agreements entailed a major reduction in workload at these points. (1985, 12).

Equipment solutions were also sought by CN to increase the capacity of their system. During the 1970s, CN recognized the potential of increasing system capacity through the utilization of heavier rail cars. Figure 5.4 illustrates the increasing use of 100-ton rail cars over the CN Edmonton to Vancouver line from 1972. Increases in the amount of horsepower per train was another option considered by CN. However, CN did not highly favour this option because:

> the use of more horsepower per train in order to shorten running times would have entailed very heavy investment in motive power with only modest improvements in capacity and was, accordingly, not a favored option. (1985, 13).

Despite the gains achieved through the implementation of the selected operating, traffic, and equipment solutions

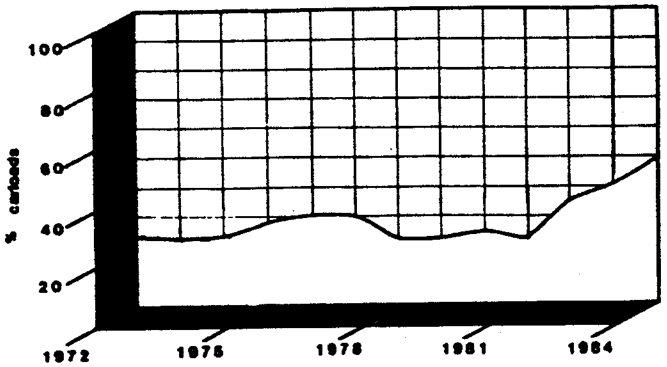


Figure 5.4: 100 Ton Cars as a Percentage of Total CN Carloads Between Edmonton and Vancouver: 1972 - 1984

(reproduced from CN 1985, fig. 6)

discussed above, it was clear by the mid-1970s that major investments would have to be made to expand the capacity of their existing railway plant in the West. CN foresaw that major capacity investments in plant were becoming increasingly necessary, and were due in part to the greater utilization of heavier rail cars:

> It should be noted that car equipment investment and plant investment are highly interrelated. Plant improvements aid in the improved utilization of cars, and in a decreased requirement for car maintenance. Increases in car sizes and weights create an additional requirement for plant investment (mainly upgrading of track structure), and a corresponding increase in the requirement for Engineering track occupancy for work blocks. (1985, 13).

CN recognized that potential for capacity improvements through implementation of the less expensive operating, traffic, and equipment solutions was becoming exhausted. At this point CN began making the major investments in their Western plant in order to insure adequate railway capacity to meet the demands of the 1980s.

The focus of the Plant Improvement Program was the improvement of the overall CN plant in Canada. According to CN, the Plant Improvement Program went to about:

> 1980-81 and was associated with providing capacity through longer sidings (therefore longer train capability), and improved signals and switches. It was a national, not strictly a Western program. (1985, 4).

Since 1981, CN has invested large sums of capital to increase the capacity of their Western plant, particularly over their Edmonton to Vancouver route. This plan is known as the Plant Expansion Program. The Plant Expansion Program:

> was associated primarily with increasing the number of full length sidings, the number of signals, and with double tracking; it was a Western program. (1985, 4).

The primary focus of this program was to increase capacity by permitting more, not larger trains.

From 1972 on, CN attempted to increase the capacity of the section of line running westbound from Edmonton to Jasper with investments in short double track stretches. In addition, CN constructed additional trackage for incoming and departing trains at Jasper. Capacity problems developed, however, as:

> Serious competition for track occupancy developed between the needs of Transportation to operate, and Engineering, to maintain the line and to accomplish new construction. (1985, 20).

Despite the difficulties created by the competing demands on track occupancy time, CN announced in 1981 their intention to double track all of the line running West from Edmonton to Valemount (see Figure 5.5). By 1990, they anticipated spending some $450 million on this double tracking project (CTC 1985, 87). CN also announced their plans to spend some $970 million on double track construction through the Fraser Canyon to Vancouver (CTC 1985, 87). This construction, which began in 1981, was expected to be complete in the 1990s. In order to alleviate capacity problems in the Rockies:

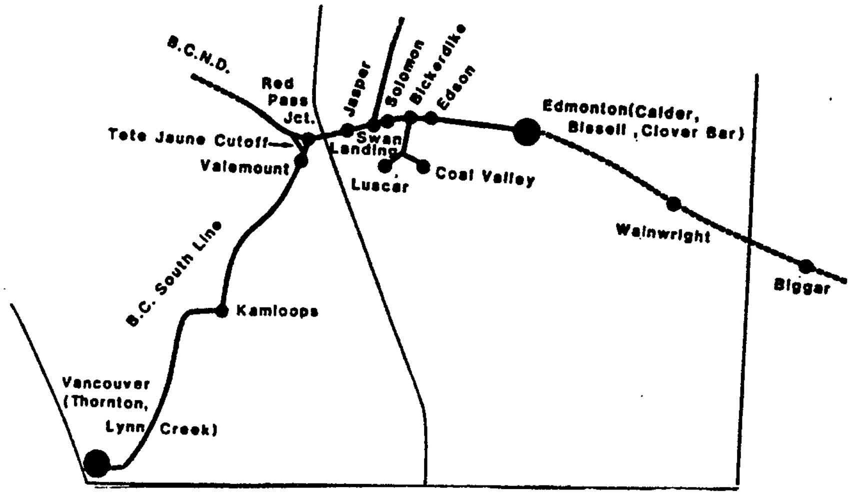


Figure 5.5: CN Rail Mainline Route Between Edmonton and Vancouver (reproduced from CN 1985, p. 23)

> CN indicated that by 1985, capital expenditures were expected to be of the order of 3 billion dollars. This sum did not include major investments on the line to Prince Rupert, although this would be required when coal movements began. (CTC 1985, 87).

CN had planned to double track some 40 percent of its entire mainline between Winnipeg and Vancouver by 1988, and was considering double tracking the entire route by the year 2000 depending on demand (Winnipeg Free Press 1980). The president and chief executive officer of CN, Dr. R. A. Bandeen, stated that the cost of building each mile of track west of Red Pass Junction would be about $2.7 million, and that each mile east of that point would cost about $1.3 million (Winnipeg Free Press 1980). CN expected that double tracking of the line from Edmonton to Valemount would increase its capacity to and from Vancouver, from 30 trains per day, to about 38 trains per day (Transport Canada 1980, 9).

The double tracking program from Edmonton to Valemount has been completed, and is continuing on certain sections of the line to Vancouver. According to CN, the double tracking of the Edmonton to Valemount line produced a number of significant benefits:

> By 1984, there were a number of indications of tangible results from the program to date. During 1984 record traffic was handled over the route, partly due to plant upgrading and partly due to a series of intensive short term operating and marketing actions. Line disruptions were significantly reduced. In 1984, CN changed to standard operation of two locomotives per train over this territory, instead of the former three. Not only was record traffic

> handled, but at an improving service level. CN achieved a 53 percent market share of WGTA grain, reversing a longstanding trend of CP outstripping CN's grain volumes. This was accomplished with CN having a smaller permanent fleet of grain cars than CP. Similarily, service helped CN to retain 82% of the potash market to the West coast during 1984. On-time performance of 200 series trains,CN's priority scheduled trains, rose from levels of below 50 percent, in the late seventies, to monthly averages of 80 percent by late 1985. (1985, 28).

Figure 5.6 illustrates the growth in traffic over the CN Edmonton to Vancouver route from 1972 to 1984.

With respect to the CN line running northwest from Red Pass Junction to Prince Rupert, CN began a capacity expansion program in response to two major developments. The first one was the initiative undertake by both the federal and B.C. governments to develop coal mining in Northeastern B.C. in the early 1980s. The other factor was the construction of a world class grain terminal at Prince Rupert in the mid 1980s. CN states that the:

> timing and scale of the expansion program were primarily determined by these external groups, not by CN; however, analysis shows that the venture has produced a net benefit for CN. (1985, 42).

Total investments made by CN for plant expansion on the Edmonton to Vancouver mainline totaled $457 million from 1972 to 1985. Of this total, double tracking accounted for $283 million, terminal and shop improvements $139 million, passing tracks $17 million, and fibre optics (communications) $7 million. CN is currently planning to make additional investments of some $280 million to the

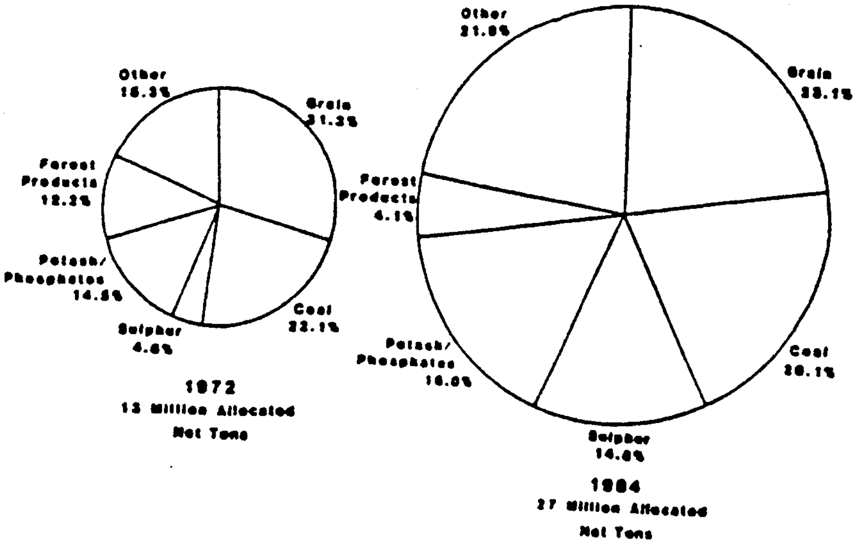


Figure 5.6: CN Rail Traffic Growth Between Edmonton and Vancouver: 1972 - 1984

(reproduced from CN 1985, fig. 9)

Edmonton-Vancouver plant in the period from 1986 to 1990 (1985, 49). CN states that:

> with certain exceptions, CN has now achieved on these lines a reasonable balance between traffic being handled and capacity available. Future expansion work will therefore be geared to market forecasts of traffic increases, and expected associated economic benefits. (1985, 49).

The level of future CN plant investments depend upon traffic growth. CN constantly monitors, and revises, its investment plans as part of the annual CN Capital Planning program. Current forecasts (see Figure 5.7) show that CN expects traffic over the Edmonton to Vancouver mainline to increase at a constant rate from 28 million ton-miles in 1984 to something around 33 million ton-miles in 1995. Figure 5.8 illustrates that an increasing share to total CN traffic will move over the mountain region of Western Canada.

Between 1978 and 1985 CN invested over $175 million for plant expansion over this line. Of these total, $41 million was for passing and back tracks, $80 million was for terminal and shop improvements, and $54 million for CTC signals (CN 1985, 42). Figure 5.9 graphically illustrates CN total investment over this line from 1972 to 1985. Specific improvements from 1978 to 1985 included increasing the total percentage of CTC controlled line from 0 to 57 percent; increasing the number of sidings from 29 to 45; decreasing the average spacing between

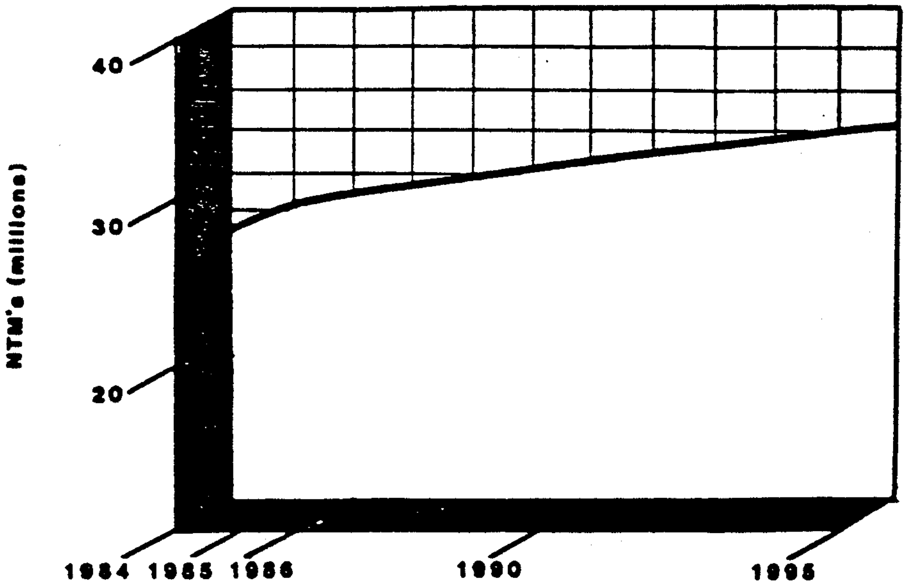


Figure 5.7: CN Rail Traffic Outlook Between Edmonton and Vancouver: 1985 - 1995

(reproduced from CN 1985, fig. 16)

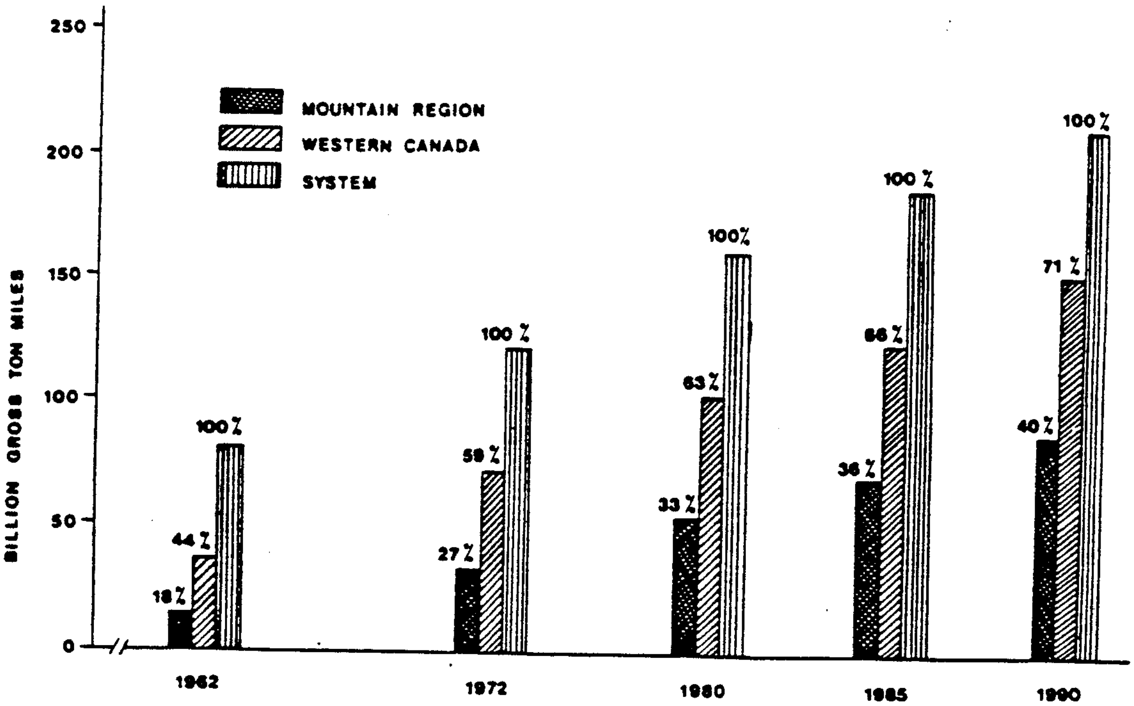


Figure 5.8: CN Rail Traffic by Region: 1962 - 1990

(reproduced from Transport Canada 1984, fig. 7.1)

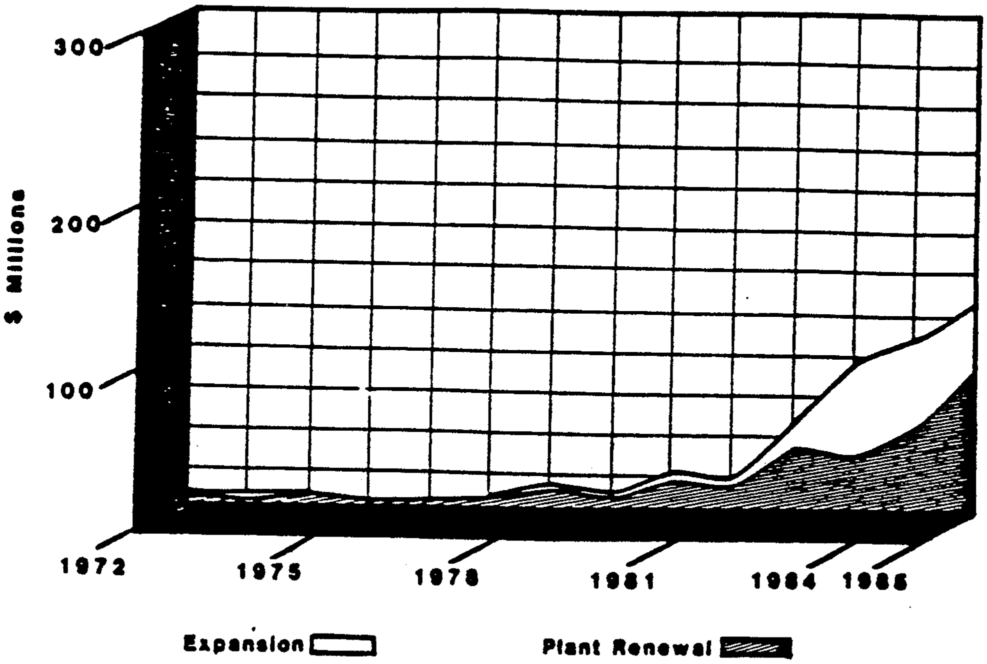


Figure 5.9: CN Rail Plant Investment B.C. North Division: 1972 - 1985

(reproduced from CN 1985, fig. 13)

sidings from 20 to 13 miles; increasing the average siding length from 6,040 to 6,231 feet; and increasing the total amount of heavier 132 lb rail on the line from 2 to 42 percent of total track mileage (CN 1985, 43). The majority of the total $175 million investment in the Red Pass-Prince Rupert line have been made on the section running West of Prince George - $167 million, in total (CN 1985, 43). Traffic volume from Red Pass Junction to Prince Rupert increased from 6 million net tons to over 13 million net tons in 1984. CN claims that this growth in traffic was:

> driven primarily by the commencement of coal traffic, and partly by growth in lumber. . . Coal traffic on the BCND (BC North Division) uses the Prince George - Prince Rupert segment only, while grain and lumber utilize the entire route. (1985, 43).

Figure 5.10 illustrates the growth in traffic over this line from 1981 to 1984.

CN has plans to invest around $94 million on the Red Pass to Prince Rupert line during the years from 1986 to 1990. Figure 5.11 shows the traffic growth projected to occur from 1985 to 1995. The actual amount of this investment, however, will be determined by future traffic growth that will be forecasted for the purposes of the CNs annual capital planning program (1985, 49).

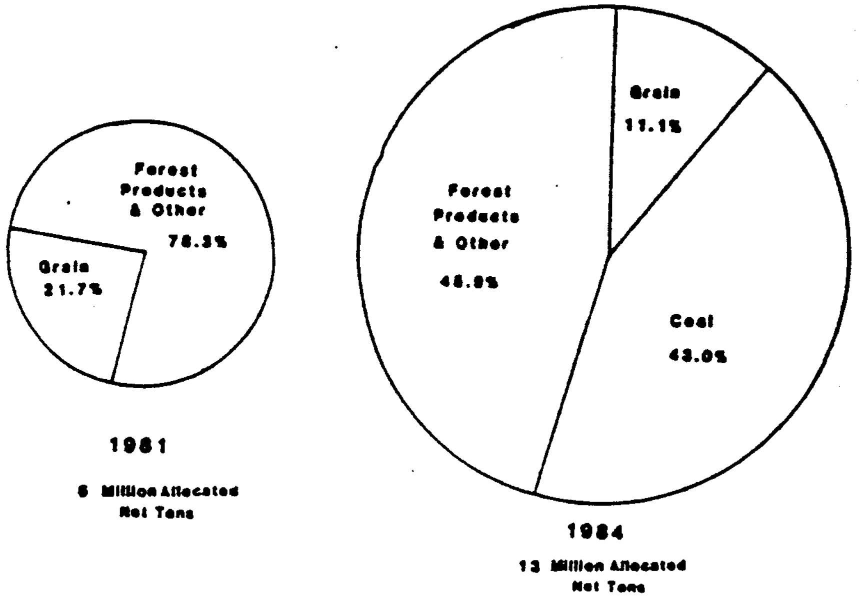


Figure 5.10: CN Rail Traffic Growth Over the B.C. North Division: 1981 - 1984

(reproduced from CN 1985, fig. 15)

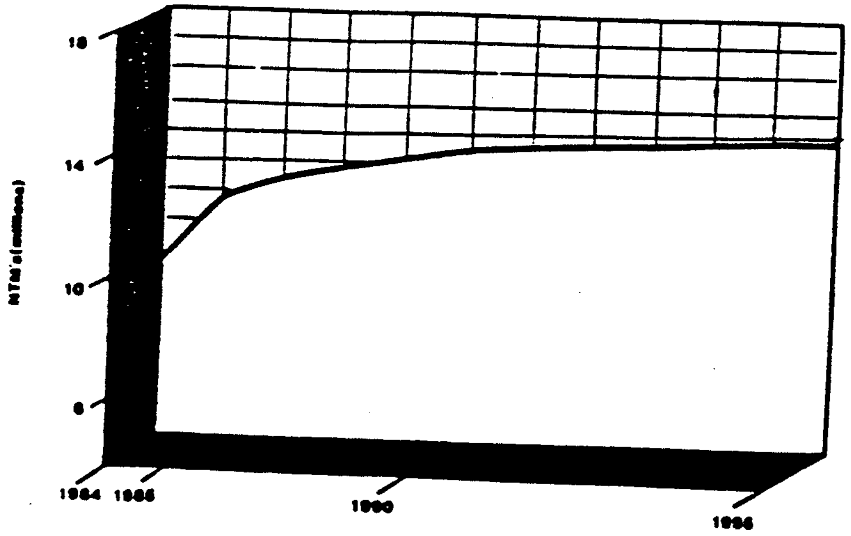


Figure 5.11: CN Rail Traffic Outlook Over the B.C. North Division: 1985 - 1995

(reproduced from CN 1985, fig. 17)

5.4 The CP Plan to Increase Railway System Capacity in Western Canada during the 1980s

The efforts of CP Rail to increase the capacity of their Western system began in earnest during the mid-1970s. The first stage of this capacity expansion occurred in the period from 1977 to 1981. During these years, CP Rail planning concentrated on investments to improve the railway plant, as well as spending money to acquire more and newer equipment. These plant improvements were specifically focused to expand capacity over the high traffic density CP mainline running west from Calgary, over the Rockies, to Vancouver. The second stage of this expansion program was planned to occur from 1981 and 1991. This plan focused upon the elimination of the capacity restricting grades located on the Calgary to Vancouver Westbound main line at Rogers Pass.

It was clear to CP Rail by 1975 that any further expansion in the capacity of their Western system, and specifically the capacity of the critical Calgary to Vancouver mainline, could no longer be achieved through such relatively inexpensive options as increasing train and siding lengths. According to an analysis by Transport Canada:

> the incremental approach to the provision of adequate line capacity to accomodate anticipated traffic growth in Western Canada is illustrated by recent CN and CP Rail capacity improvement programs.

> Having essentially exhausted possibilities for capacity improvements between Calgary and Vancouver through increasing train and siding lengths, adopting the use of mid-power and improving signalling systems, CP Rail is improving capacity on this section by removing three of the remaining four major bottlenecks between Calgary and Vancouver. All four of these sections which have severe Westbound grade restrictions associated with them require the construction of a second main track with less severe grade (in effect, providing sections of double track). (1980, 8).

By the mid-1970s, CP Rail was planning a limited amount of double tracking and grade reduction to eliminate the bottlenecks between Calgary and Vancouver, thus marking the first phase of their capacity expansion effort in the West. During the first phase of this work, from 1977 to 1981, CP Rail constructed some 34 kilometers of new second track on this line. This double tracking was installed over three separate sections located in the mountains (CP 1981, 9). From 1977 to 1979, CP Rail concentrated their efforts to eliminate two critical grades on the mainline running west from Revelstoke to Kamloops at the passes of Clanwilliam and Notch Hill. These two points had severe westbound grades of about 1.5 and 1.6 percent, respectively (see Figure 5.12). These bottlenecks were eliminated through the construction of some 15 kilometers of new second main track, with less severe westbound grades, at Clanwilliam and Notch Hill. These new second track sections were completed and operating by November 1979. During 1980, an additional 9

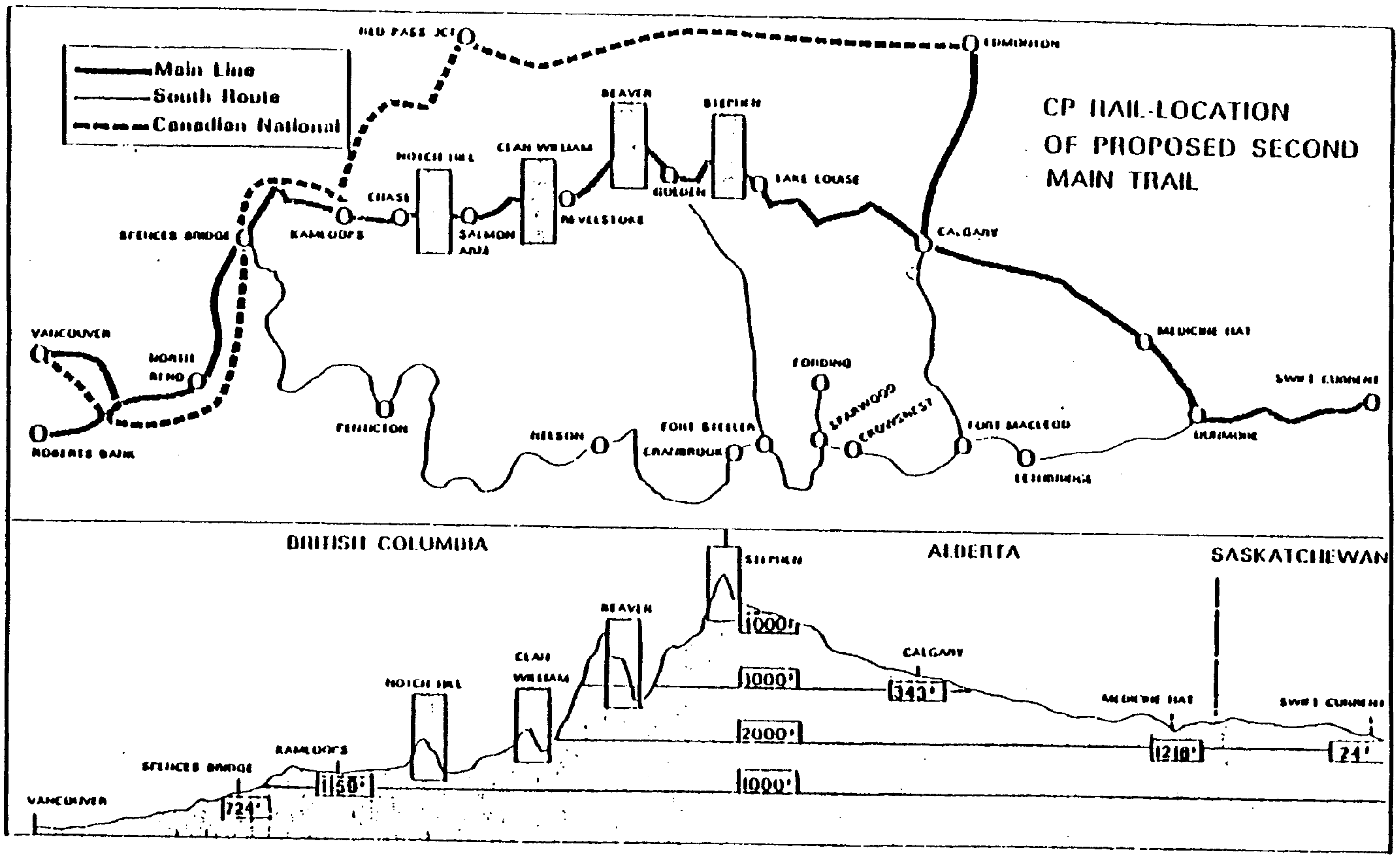


Figure 5.12: Critical Grade Locations on the CP Rail Mainline Between Calgary and Vancouver (reproduced from Transport Canada 1980, fig. II)

kilometers of second main track was constructed over the section between Lake Louise and Stephen on the Calgary to Vancouver line. The critical grade at Stephen, which had been in the area of 2.2 percent, was eliminated, and this new section of second main track became operational in July 1981 (CP 1981, 8-9).

The completion of these three grade reductions by CP Rail virtually eliminated grades in excess of one percent on its mainline west of Calgary by 1981, with the obvious exception of Rogers Pass. As a result CP Rail was able to use much less horsepower per train to move more tonnage without stopping to cut in and cut out extra locomotives to move over these formerly critical grades. Transport Canada estimated that the capacity of the CP Rail Calgary to Vancouver line in the early 1980s was about 30 trains per day, on a sustained basis (1980, 8). In 1975, CP Rail estimated that the total cost of double tracking the eastern slopes of Clanwilliam, Notch Hill, Stephen, and Rogers Pass (excluding tunnel construction) would be in the range of $60 - $70 million (Transport Canada 1975, 38). In 1981, CP Rail said that the cost of double tracking and grade reduction at the first three of the locations listed above was about $45 million (CP 1981, 8).

It was clear by 1981 that a limited amount of double tracking on the eastern slope of Rogers Pass would not, in itself, be enough to achieve the substantial gains in capacity needed to meet the demands expected to occur by 1990. Transport Canada forecasts at that time were predicting that something around 40 trains per day (both ways) would be passing over the CP Rail main line west of Calgary by 1990 (1980, 2). Transport Canada states:

> To meet these future increases, CP Beaver Tunnel project (at Rogers Pass), if proceeded with, would increase CP's capacity to 38 trains per day. (1980, 2).

In July 1981, CP Rail filed an application with the Canadian Transport Commission for the authority to undertake a $600 million railway grade improvement project in Rogers Pass. However, CP Rail stated that it would only proceed with the project if compensation would be forthcoming for the multi-million dollar losses it incurred by moving Western grains. With the political resolution of this issue imminent, CP Rail proceeded with preparatory work for the project in 1982. The Rogers Pass project was part of a more than $7 billion program planned by CP Rail to occur over a 10 year period from 1981 to 1991. The objective of this plan was to oversee the replacement, improvement, and expansion of plant and equipment during the 1980s in order to meet the traffic demands expected later in the decade.

The Rogers Pass project is enormous in terms of effort and cost. CP Rail considers it the largest single project they had ever undertaken since the transcontinental railway was built in the 1880s. This project involved the construction of 16 kilometers of new tunnels through two mountains, 8 new bridges, and some 34 kilometers of new second tracks. This project, began in 1982, and scheduled for completion by late 1988 (CP 1983, 2), was expected to provide jobs for over 800 on-site workers (CP 1981, 4).

This project was seen as critical to the efforts of expanding CP Rail's mainline capacity to Vancouver. Transport Canada stated that:

> until CP completes their Beaver tunnel project their capacity between Calgary and Vancouver is limited to 30 trains per day on a sustained basis. . . the fourth project (Rogers Pass) which, would permit CP to operate 38 trains per day over the Rocky Mountain segment, . . . (1980, 8).

The importance of this project in terms of Western system capacity was also recognized by the CTC, who states that:

> when complete, CP Rail will have eliminated all grades in excess of one percent on its mainline west of Calgary . . . Maximum train tonnage will rise from 11,000 tonnes to nearly 15,000 tonnes . . . (1984, 73).

The capacity expansion program of CP Rail also focused upon the acquisition of more newer, and larger, rail cars and locomotives. According to the CP annual report:

> the Company's transportation businesses are continually modernizing or adding facilities and equipment.
>
> Thus CP Rail is spending $77 million for 75 new diesel locomotives. Some 40 of these were ordered to provide additional power for moving grain, and the others are part of a program to reshape the existing locomotives fleet into a more powerful and flexible

> diesel force by the end of the decade. Mainly in order to handle potash traffic, CP Rail will take delivery of 500 rail hopper cars in mid-1981, at a cost of $27 million. (1980, 3).

This program was carried out in the early 1980s, and gave CP Rail greater flexibility and carrying capacity with respect to its locomotive and rolling stock fleet.

5.5 Summary

This discussion has outlined the plans of CN and CP to increase railway capacity in Western Canada during the '80s. It is clear by the mid 1970s both railways recognized the necessity to build additional capacity into their Western networks. From the mid '70s on, CN and CP devised plans to initiate large scale capacity investments over the coming decade. The push to install this additional capacity came largely from the demand forecasts of that time. These predictions called for significant increases in rail traffic volumes in Western Canada to occur by 1990.

In order to meet these demands, both CN and CP focused their attention on their mainlines running through the Rockies to the B.C. coast. The objectives of the CN plan were to increase existing capacity while eliminating many of the operating problems that were occuring in this area. In terms of cost, between 1972 and 1985 CN spent over $450 million towards expanding capacity of the Edmonton to Vancouver mainline. Another $175 million was invested between 1978 and 1985 for plant expansion on the line from Red Pass to Prince Rupert. The bulk of these expenditures went towards double tracking, longer and more frequent sidings, upgrading signals, and terminal improvements. These investments not only resulted in an expansion of CN's Western network capacity, but they also reduced or eliminated many of the operating problems that had previously existed.

CP investment plans revolved around the reduction and elimination of the critical ruling grades that acted as capacity bottlenecks on their Calgary to Vancouver mainline. From 1977 to 1981, CP made some limited efforts towards the elimination of these bottlenecks. In 1982 CP began construction on the massive Rogers Pass project. It is expected that CP will have spent well in excess of $600 million on this single project by the time it is completed in 1988. The planning objectives of this project are to facilitate a significant increase in Western network capacity, while achieving increased economies due to greater efficiency of operations.

The next chapter will take a closer look at the rationale for these plans from an economic planning approach. The cost and benefits of increasing railway capacity in Western Canada will be summarized within the context of production function analysis. This analysis will define the long range production expansion path of the railways, and its relation to economic planning.

CHAPTER 6

AN ECONOMIC PLANNING APPROACH TO WESTERN CANADA RAILWAY CAPACITY EXPANSION BY CN AND CP IN THE 1980s

6.1 Introduction

The previous chapters described the circumstances and planning related to the efforts of CN and CP to increase Western railway capacity during the 1980s. These capacity increases, planned during the mid to late 1970s, were initiated in response to demand forecasts that predicted large increases in rail traffic volume in the West by 1990. The rationale of CN and CP for increasing capacity was based on the premise that their operations and infrastructure had to change to meet these demands.

The recession of the early 1980s, and the subsequent downturn in export demand for major commodities produced in Western Canada, reduced the overall demand for railway transportation. It was clear to the railways that the large increases in rail traffic predicted in the late

1970s were not going to occur by 1990.

Despite this setback, CN and CP Rail are currently engaged in a extensive investment program to increase the mainline capacity of their respective systems in Western Canada. Hundreds of millions of dollars have been spent for plant improvements, and both railways have plans to invest more in the next few years. From a technical view, it is clear that these investments were planned and undertaken in the context of ensuring a level of network capacity that would be able to meet the increased level of traffic expected during the next few years.
However, there are other considerations that are quite separate from the purely technical issue of providing adequate Western network capacity. These considerations relate to evaluating these changes from an economic planning perspective. An economic planning approach details the economic and physical interrelationships that the various factors affecting railways capacity have with one another. This approach is also dynamic, and outlines the evolution of these factors, and their interrelationships over time.

In terms of understanding this evolution, it is important to recognize the economic considerations that led the railways to implement these developments. The following chapter will evaluate the operating and infrastructure changes, using an economic planning approach, that have been undertaken by CN and CP to increase capacity in the West. This evaluation should provide some sense of the underlying economics and objectives behind the planning and development of the railway system in Western Canada.

The notion of a production function provides a useful theoretical framework, from the perspective of the economic planning approach, to evaluate these developments. The following discussion focuses on the concept of a production function, and how it relates to the planning and development of additional capacity for the Western railway network.

6.2 The Railway Production Function

Microeconomic theory implies that the total level of output possible for a particular firm at any one time is a function of two production inputs - capital and labour. This relationship between output and inputs is commonly referred to as the production function. The production function can be expressed mathamatically as

$$Y = f(L, K)$$

where Y is total output, L is the total input of labour, and K represents the total input of capital stock (i.e. plant and equipment). The most well known form of the production function model is the Cobb-Douglas function. This function may be expressed in mathematical terms as

$$Y = AL^{a}K^{b}$$

or, using logarthmic form as

$$\ln Y = \ln A + a\ln L + b\ln K$$

The coefficients a and b measure the elasticities of output with respect to labour and capital. That is to say, the percentage change that occurs in output for a given percentage change in labour and/or capital. The sum of the coefficients a and b also yield information on the effect on output from a proportionate change in the level of the inputs. Should these two coefficients sum to one, then the production function exhibits constant returns to scale. In another words, when the level of inputs are increased by a certain percentage, the level of output will also increase by the same percentage. On the other hand, when the coefficients add up to to less than one, then the production function is said to display decreasing returns to scale. Thus, when labour and capital are increased by a certain percentage, output will also increase, but by a smaller percentage. Similarily, when a and b sum to greater than one, the production function possesses the property of increasing returns to scale. An increase in the inputs will lead to a proportionately greater increase in output.

Recent studies have shown, as was discussed in Chapter 2, that the railway firm is more likely to be characterized by constant, rather than increasing, returns to scale. Figure 6.1 depicts a two input production function with constant returns to scale. Capital and labour inputs are represented by K and L, and the isoquant and isocost lines are depicted by Z and C, respectively. An isoquant analysis shows the relationship between the long-run and short-run cost functions.

The isoquant line Z represent all of the possible combinations of labour and capital needed to produce a particular level of output. Similarily, the isocost line C indicates all of the combinations of labour of capital that yield a certain level of cost. The most efficient level of output occurs where the isocost line C intersects the minimum point of the isoquant Z.

The line OC represents the long-run expansion path. This path is determined by joining all of points on the various isoquants where efficient production occcurs.

This expansion path is of crucial importance in the economic planning that is involved to increase railway capacity. In the short run, the total capital stock of the railway is fixed. Increases in railway capacity can only occur through changes that can be implemented in the near future. Therefore, the only options that exist are such short-run alternatives as changing operating and/or

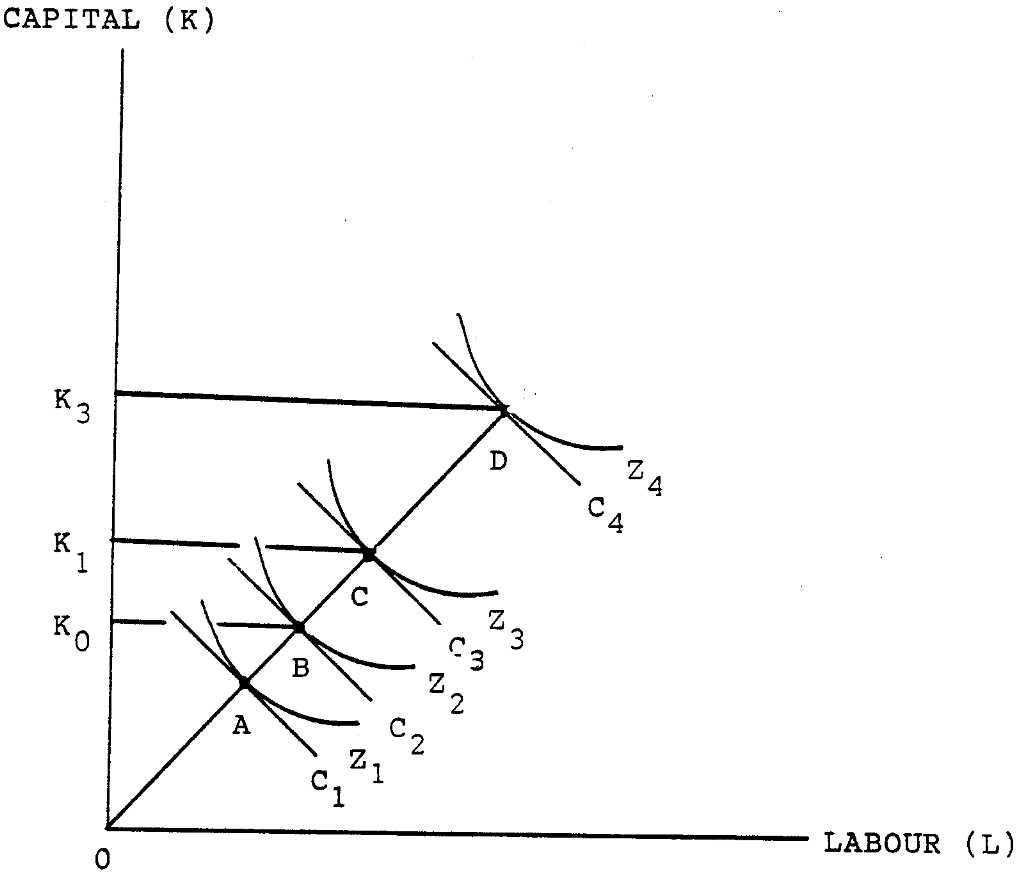


Figure 6.1: The Short, Medium, and Long-Run Output Expansion Path

traffic policies. This situation is depicted in Figure 6.1. For a railway that is operating at point A, the short-run expansion path is AB. The implementation of short-term options is represented by increasing the labour input. However, output beyond point B is impossible because the total capital stock of the railway is fixed at K_0 in the short-run.

In the medium-run, the output expansion path for the railway is OC. Increases in capacity can now only occur if the total capital stock of the railway is increased. Operating and traffic solutions hold no more potential to increase capacity, given the existing equipment and infrastructure of the railway. Additional equipment can be acquired during this period, thus increasing the captial stock of the railway from K_0 to K_1. This would include increasing the number of locomotives and rail cars available to haul freight. However, the infrastructure of the railway remains fixed. Thus, output can only increase to point C. Any output beyond this point is unfeasible no matter how much more railway equipment is acquired.

The most costly and time consuming options that are available to increase railway capacity involve making investments to expand the infrastructure. These options include such things as more sidings, double tracking, grade reductions, better signals, etc. These investments have the potential to achieve increases in output capacity that are much greater relative to the potential of the options available in the short-run and medium-run. The long-run expansion path in Figure 6.1 is OD. Investments in the railway infrastructure are indicated by the increase in the capital stock from K_1 to K_2.

It is clear that the production function has important implications relative to economic planning. The economic planning approach must identify the impacts that the options that are available to increase capacity have on costs and production. In addition, it is essential that economic planning consider the the potential of these options to effect increases in capacity over the short, medium, and long-run. These efforts should lead to the development of a long-run expansion program, as was detailed in the production function analysis above, to implement the operational and investment options necessary to increase capacity.

The following discussion evaluates, from an economic planning perspective, the expansion programs developed by CN and CP to increase railway capacity in Western Canada.

6.3 A Summary of the Impacts of the CN and CP Railway Capacity Expansion Programs

It was evident to CN and CP as far back as the early 1970s that the capacity of their networks in Western

Canada were inadequate to handle the traffic demands forecasted to occur over the next few years. In the case of CN Rail, lower cost operating, traffic, and equipment solutions were the first to be undertaken during the early to mid 1970s. These solutions were, logically, the first to be considered and implemented not only because of their relatively low cost, but also because they could be put into effect in the short and medium term.

CN operating solutions implemented in the 1970s included the use of longer trains, the regulation of train speed at a more consistent level, relocation of switching activities away from congested terminals and yards, and changes in scheduling. These solutions resulted in more trains, carrying heavier loads, being able to move over the system each day. The theoretical economic impacts of this should be that while the total cost of railway operations increase due to the greater number of trains running, average and marginal cost per gross ton-mile should decrease due to the existence of economies of density.

Similar efficiencies should occur when such traffic solutions as the greater use of shipper owned cars and unit trains are affected. The use of newer and heavier rail cars, combined with more powerful and efficient locomotives, will also increase the productivity of the railway system. These types of solutions, all of which were implemented by CN in the 1970s, are undertaken in order to increase system capacity while improving the economics of the operation. Revenues should grow as increased system capacity leads to lower costs and better service to shippers. Average and marginal costs will decline for each successive gross ton-mile of railway service produced as the average load per car increase, locomotive productivity increases, and the average train transit time decreases.

The potential for lowering the average and marginal cost of railway production using operating, traffic, and equipment solutions is limited by the capacity of the overall railway plant. Investment in the railway plant becomes inevitable once the potential from these other options become exhausted.

CN undertook a massive expansion program to increase the capacity of its Western Canadian railway plant when it became apparent in the mid-1970s that no significant increases in capacity could be achieved with the lower cost operating, traffic, and equipment solutions. In the period from 1972 to 1985, CN spent more that $632 million on expanding the capacity of its Western plant. The focus of this expansion was centered on the critical, and capacity limiting, mainlines running from Edmonton to Vancouver and Prince Rupert (CN 1985, 3). CN also has plans to spend an additional $278 million in the period from 1986 to 1990 on additional plant expansion in this area (CN 1985, 49).

Extensive double tracking in the CN Western network was considered the most important determinant of capacity expansion in this program. As a result, CN has planned to double track 40 percent of its Winnipeg to Vancouver mainline by 1990. Critical sections of track on the Edmonton to Valemount to Vancouver mainline were double tracked in order to provide significant increases in capacity. CN spent more that $456 million on expanding the capacity of the route (these investments included double tracking, terminal and shop improvements, passing tracks, signals and communications) from 1972 to 1985 (CN 1985, 7). A further $216 million dollars is expected to be spent by 1990 (CN 1985, 49). In addition, CN has spent some $175 million on increasing the capacity of its Red Pass to Prince Rupert line, and will spend and additional $62 million on it by 1990 (CN 1985, 42-49).

The economic rationale behind making investments in plant capacity expansion is that additional traffic will move at a increasingly lower average and marginal cost per unit. Thus, added capacity results along with improved economic operating efficiency. In light of this criteria CN has chosen to invest nearly one billion dollars to expand the capacity of its Western plant. For this investment, CN will have achieved significant increase in the capacity of this system. In 1980 the capacity of the CN line from Edmonton to Vancouver was about 30 trains per day on a sustained basis (Transport Canada 1980, 1). After the improvements to this line have been made, CN expects that it will be able to move an estimated 38 trains per day, an increase of over 26 percent. Significant increases have been also achieved over the Edmonton to Red Pass section. According to Transport Canada:

> As traffic destined for Prince Rupert also is handled over the Edmonton to Red Pass section (in addition to the traffic destined for Vancouver), CN has indicated that a major program of double tracking portions of this section would be initiated. With the completion of this program CN will be able to handle at least 46 trains per day between Edmonton and Red Pass Junction on a sustained basis. (1980, 10).

With respect to the Red Pass to Prince Rupert line, investments by CN Rail could increase the 1980 capacity of 10 trains per day to around 38 trains per day, if the full potential of the line is realized (Transport Canada 1980, 10).

Thus, for its nearly one billion dollars of capital expenditures, both actual and planned, CN will have achieved significant increases in the number of trains per day that are able to move over its Western system.

CP Rail also invested hundreds of millions of dollars to increase the capacity of its Western system in the late 1970s and 1980s. Like CN, these investments focused on the mountainous region of the CP mainline between Calgary and Vancouver. The first stage of this expansion occurred between 1977 and 1981. During this period, CP attempted to expand capacity of its westbound Calgary to Vancouver line through grade reductions and a limited amount of double tracking. These efforts, at a cost of about $45 million, eliminated three out of four critical grades that acted as major capacity constraints on this line. The economic impacts of this construction was that increased operating efficiencies were achieved on line haul operations over this route. These increased efficiencies were due to the fact that fewer locomotives were required to move trains over the reduced grades. This should have led to an increase in productivity, and therefore a reduction in the average and marginal cost of producing a gross ton-mile of railway service over this area. Decreases in fixed plant operating and maintenance costs should also occcur.

CP Rail also spent over $100 million in the early 1980s acquiring rail cars and locomotives as part of a program to increase the capacity and flexibility of its equipment fleet. The use of more efficient equipment reduces operating and maintenances costs as well as providing the potential for greater revenues through improved service and reduction in delays.

The grade reduction program at Rogers Pass began in 1982 is scheduled for completion in 1988. This project, which calls for the construction of two new tunnels and several kilometers of double tracking, is the most extensive single project ever undertaken by CP Rail since the construction of the original railway. When completed, CP Rail will realize increases in the capacity of its rail network in the West. The capacity limits existing on the mainline from Calgary to Vancouver will increase from the 1980 level of 30 trains per day to 38 trains per day. This represents an increase of 26 percent (Transport Canada 1980, 9). In addition, system capacity would significantly increase as the maximum train tonnage able to move over this line increases from 11,000 to 15,000 tonnes.

The grade reduction at Rogers Pass will also reduce the amount of time required for trains to move between Calgary and Vancouver. Trains no longer have to stop to cut in and out extra locomotives to move them over the passes. This will lead to a reduction in the transit time of trains running over this line.

These improvements in the CP Calgary to Vancouver mainline will lead to increased economic efficiencies in the movement of goods over their Western Canadian railway system. Fixed plant operating and maintenance costs will be significantly reduced. The cost of line operations will decrease as fewer locomotives, and thus less fuel, depreciation, etc., are required to move trains over this route. The productivity of CP railway operations should increase significantly as the amount of inputs needed to produce a gross ton-mile of railway transportation is reduced. Average and marginal costs of railway production will decrease as the level of throughput increases.

6.4 Summary

It is crucial, from the view of economic planning, to comprehend the notion of the production function. Production function analysis outlines the potential of various options to increase capacity over the short, medium, and long-run. The output expansion path that develops from this analysis outlines the points of efficient production over the long-run. The economic planning process that develops a program to expand railway capacity over the long-run must identify options that are feasible over various time frames. The CN Western expansion program that started in the mid 1970s focused on the mainlines running through the Rockies to the B.C. coast. CN has spent, or has plans to spend, over a billion dollars to increase its capacity in the West. In the initial phases of this expansion, increases were achieved through operating and traffic solutions. These not only impacted on the level of available capacity, but also reduced operating costs and increased service levels.

However, these options were available to increase capacity in the short-run. Their potential to affect additional increases in capacity was limited. It was necessary for CN to plan to implement investments in equipment and plant in order to increase the level of capacity to where it would be sufficient to meet expected levels of demand. These investments would have to occur over the long-run. The capacity of the mainlines running West of Edmonton to the B.C. coast has been the focus of this expansion, and will be able to carry significantly larger volumes of traffic at a much improved level of economic operating efficiency.

The net effect of CP Rail's $750 million investment in railway plant capacity is that the critical bottlenecks occuring over the westbound line from Calgary to Vancouver will be all but eliminated by 1990. This will allow an increase in the weight and number of trains able to run over this line. Economic operating efficiencies will result in increased productivity and greater traffic capacity.

CHAPTER 7

SUMMARY AND CONCLUSIONS

In the first part of the thesis, the economics of railway system capacity expansion were discussed. Chapter 2 showed that the railway system is comprised of a number of components that interact with one another to produce transportation services. The potential output of railway transportation is determined by the existing level of system capacity. Physical capacity defines the maximum level of throughput that can move over a railway system. Economic capacity, on the other hand, describes the situation where the average and marginal cost of throughput begins to rise. The physical interelationships between capacity and congestion were outlined, and several different types of congestion were identified.

Chapter 3 looked at the economics of increasing railway system capacity. Several issues, such as economic planning versus engineering planning, time horizons, and indivisibilities were looked at in some detail. This chapter also detailed a number of important factors that can affect railway system capacity. The physical and operating characteristics of the railway plant and vehicles (i.e. trains) are critical determinants of system capacity. Several categories of options that can affect changes in capacity were detailed. Operating, traffic, and equipment solutions were generally the first to be considered, as they are characterized by lower costs and can be implemented in the short to medium term. Once these solutions have exhausted their potential to increase capacity, investment in plant becomes inevitable. Plant investments tend to be relatively costly and can only be implemented over the long run.

The second part of this thesis outlines the railways' long term efforts to increase Western Canada railway capacity in the 1980s. Chapter 4 identified the problems related to the Western railway capacity issue that existed in the 1970s. The railways and others recognized that the existing Western system was inadequate to handle the traffic levels expected for the 1980s. The run down state of the Western railway system was in large part due to the problems related to moving grain at statutory rates.

Chapter 5 looked at the specific plans undertaken by the railways to increase the capacity of their respective networks. CN planning outlined investments of about one billion dollars to double track significant amounts of their Western system, with the focus primarily on increasing the limited capacity of their lines running through the Rockies to the West Coast. Similarily, CP spent some $750 million dollars for plant expansion from 1977 to 1988. Like CN, these efforts concentrated on increasing the capacity of the CP mainline running through the Rockies to Vancouver. Grade reductions, limited double tracking, and new equipment were the thrusts of this plan.

Chapter 6 detailed the impacts of the railways' plans from an economic planning approach. These plans were considered in view of the long range expansion path for a firm, derived from production function analysis. The railways' plans would lead to an increase in productivity due to reduced plant and line haul costs. These reductions would come from improved efficiencies that result from operating and traffic policy changes, and investments in plant and equipment. Productivity would grow as average and marginal costs decline, and throughput increases.

There are several conclusions that can be drawn from the efforts of CN and CP to increase their Western network capacity in the 1980s. First, it is apparent that by the mid 1970s the continuing diversification of the Western economy was increasing the overall demand for railway transportation. This major demand expansion was putting an enormous strain on the ability of the railways to provide adequate network capacity in the West. The federal and Western provincial governments, producer groups, and the railways all foresaw the potential for a shortage of network capacity in the upcoming decade. Forecasts that were produced during that time were calling for significant increases in the Western rail movements of grain, coal, potash, and forest products. Total rail loadings and unloadings in the West were expected to grow by 66 and 76 percent between 1978 and 1990 (Mulder 1980, 9). These forecasts were predicated on the assumption that the economy of Western Canada would continue to grow strongly in the 1980s, fueled by an ever increasing appetite for its products in export markets.

The Western railway capacity problem was compounded by the continuing deterioration of the existing rail infrastructure. Since the 1960s, both railways had been reluctant to maintain or upgrade the physical assets of their Western networks. The railways were reluctant to to invest the necessary funds because of the low rates of return generated by railway capital. These rates of return were extrmely low in the early 1970s--declining from just over two percent in 1971 to below zero in 1975 (CTC 1984, 23). Given these marginal rates of return, it is apparent why the railways hesitated to invest the large sums of capital necessary to increase Western network capacity. Better rates of return could be achieved by investing this capital in opportunities outside of the railway industry.

The statutory "Crow" rates for the movement of Western grain was a major factor behind the dismal profits earned by the railways. Both CN and CP were forced to incur losses in the hundreds of millions of dollars through the hauling of grain at the Crow rates. It can be concluded that the existence of the non-compensatory Crow rates was the largest impediment to the long range planning efforts of the railways to increase the capacity of their Western networks. The statutory grain rates not only caused the railways to lose enormous amounts of money in the present, but they also acted to distort the level of revenues expected in the future. From the perspective of the railways, there was little point in planning long range investments to increase Western network capacity as the existence of the Crow rates, and their associated losses, threatened the economic viability of these investments. The railways contended that grain would have to pay its own way before they would undertake the risk of investing the billions of dollars needed to expand Western rail capacity.

By the early 1980s, the federal government was looking for ways to ensure the adequacy of the Western rail network to meet expected demands, while also attempting to control and reduce its payout of transportation subsidies. The 1983 Western Grain Transportation Act was perceived to be the solution to these problems. The railways would now be paid compensatory rates to move grain, thus assuring that needed investments to increase Western railway capacity would be made. The elimination of the unprofitable grain rates would lead to a reduction in the level of subsidies paid out to the railways. It can be concluded that the resolution of the Crow issue finally freed the railways to initiate long range investment plans to expand Western railway capacity. The elimination of the statutory grain rates, and the subsequent improvement in expected revenues and profitability, was the critical factor in the decisions made by CN and CP to invest billions of dollars to expand the capacity of their Western networks.

It is clear that both railways approached the problem of capacity expansion using long range planning, but each proceeded with different options. From 1972 on, CN attempted to achieve increases in Western capacity, while improving operating efficiencies and service levels, through changes in operating and traffic policies. In terms of the economic planning approach, it can be concluded that CN implemented these options first becasue they had the potential to affect capacity in the short-run, and at relatively low cost. The logic of the output expansion path, as outlined by microeconomic production function analysis, would dictate that these type of options to increase capacity be first considered in the plan. However, it is also clear that from perspective of economic planning, that these options have a clear interrelationship with one another. The potential of operating and traffic solutions to affect increases in capacity is limited by the characteristics and level of the existing plant and equipment.

Thus, the potential of the operating and traffic solutions initiated by CN to increase its Western network capacity was limited by existing plant and equipment. There is little point to increasing the number of trains running on a section of track, for example, when there are not enough locomotives, or track capacity, to facilitate this increase. It is possible to conclude from the evidence that the CN network in the West had reached this point around the mid 1970s. At this stage, the planning process recognized that further increases in capacity could only come from high cost equipment and plant investments. The nature of these options is such that they can only be implemented over the medium to long-run. This implies that their implementation requires extensive consideration and plannning.

From the mid 1970s on, CN was forced to invest about one billion dollars to upgrade and expand its Western network. From an economic planning approach, it can be concluded that these investments were necessary in order to ensure that this network would be able to meet the traffic demands expected over the coming decade. The CN investment plan was clearly analogous to the long-run expansion path outlined in production function analysis. Lower cost operating and traffic solutions were the ones initiated in the short-run to increase network capacity. When these options had exhausted their potential to produce any further increases in capacity, CN found it necessary to upgrade the existing equipment and plant over the medium to long-run. The theoretical effect of these investments was to shift the CN production function towards a much higher level of potential output.

During the 1970s, because of the nature of the capacity restrictions existing on the CP Western network, operating and traffic solutions did not hold much potential to increase CP's Western capacity. The major bottlenecks in the CP system were due to the critical grades existing on the mainline running from Calgary to Vancouver. CP, therefore, was in a disimilar position from CN. Any significant increases in capacity would have to come from relatively large investments in plant and equipment.

However, like CN, the CP plan to increase capacity was laid out in stages. Before the extensive Rogers Pass project was undertaken in 1982, CP concentrated its efforts on reducing some less severe grades, and limited double tracking, on sections of its Calgary to Vancouver mainline. CP also invested some $100 million in the early 1980s to expand its fleet of rail cars and locomotives. Only when the potential for any further increases in capacity through these measures was exhausted, did CP undertake the expensive Rogers Pass project of grade reduction and double tracking.

It is clear that the efforts of CN and CP to increase their network capacity in Western Canada during the 1980s were not the result of a series of disjointed initiatives. Instead, it is apparent that these efforts were the result of a long-range plan to expand Western network capacity in a series of logical steps. It can be concluded that both railways integrated their short term expansion plans into their long range expansion plan. From an economic planning approach, it can also be concluded that the railways understood the physical and economic interrelationships of the various options. Planning objectives, options, logistics, and evaluation were considered in the light of physical and economic criteria. The results of this planning was that the capacity of the railway network in Western Canada was expanded along an long-run output expansion path. The definition of this path is the key aspect of economic planning, and is a critical determinant of the success or failure of these efforts.

APPENDIX A

EXTRACT FROM KHAN, 1979, pp. 65-75

IV.3.3. Distribution of Train Speed, Train Priorities, Meets and Overtakes

Least amount of interaction between trains on a line occurs when all trains travel at the same speed and with equal priority. Uniformity of speed and priority have therefore an important effect on line capacity. Simulation results indicate that on removing all considerations of priority, delays were reduced by a third, and consequent gain in capacity was 50 percent.[23]

This phenomenon can be explained by what the CN calls the "channel theory".[27] Simply stated, this theory says that since trains of various speed effect track capacity differently, it is convenient to refer to capacity as being the number of channels or slots available in a day.

A channel, associated with a particular train, equals the number of trains displaced by it from the base grid pattern. Figure IV.1 is an illustration of the base grid pattern of the standrad, lower class train (usually freight). Such a number can therefore be referred to as a relative interference magnitude set up by the superior train when interspersed into a grid pattern of lower class trains. See Figure IV.2 for the magnitude of interference of higher speed and higher priority train in the base grid.

For the above illustration, a simple stringline or time-distance plot is used and a double-track line was assumed for the sake of simplicity. The same factors would, however, apply to a single-track line. It is further assumed that there are terminals at each end of the track and no passing tracks are involved. In Figure IV.1, a single class of standard train is assumed which occupies one operating channel. The channel width is equated to the train's headway. Track capacity, in number of standard trains in each direction equals the number

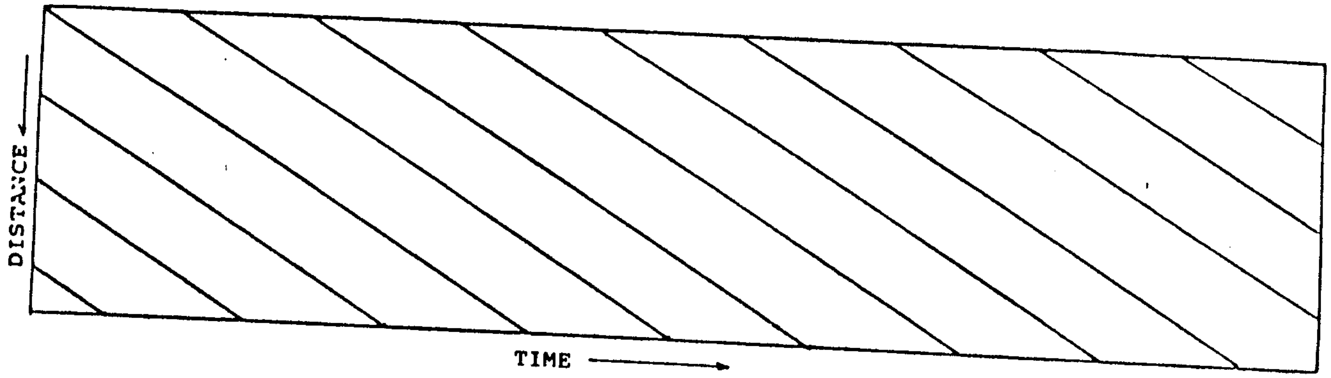


Figure IV. 1 Grid for Single Class of Train (Freight)

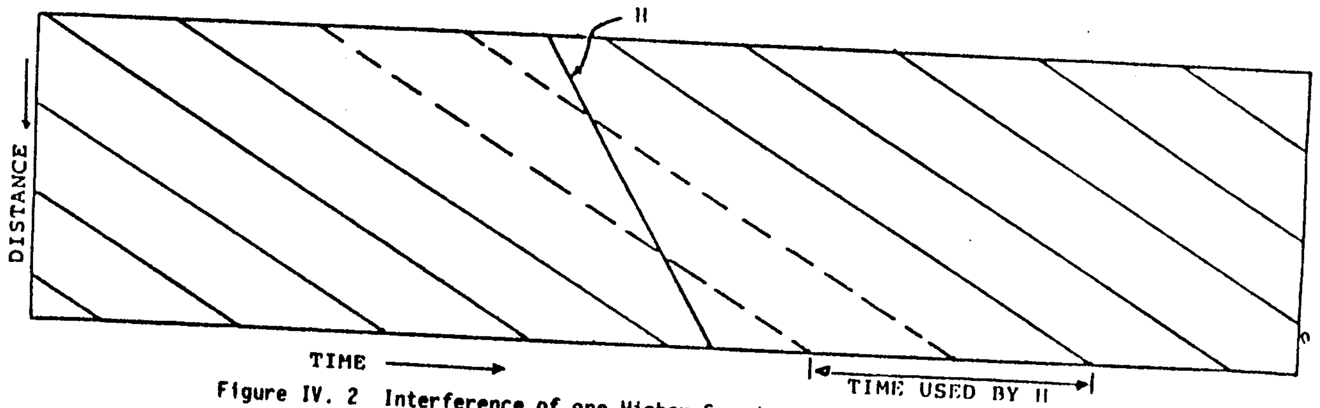


Figure IV. 2 Interference of one Higher Speed, Higher Priority Train in the Grid of operating channels which in turn is the ratio of time period to channel width (in terms of time).

Figure IV.2 indicates the effect of introducing one higher speed higher priority train, assuming no passing tracks. As can be examined, the interference by the high speed high priority train reduces the number of operating channels available to the standard train. The number of channels now available to the standard train is the reduced time available to run standard trains divided by standard channel width (time). The interference is being defined as a function of the running times, distances and headways.

Logically, if two higher speed, higher priority trains are introduced with no overlap between the time used by both together, they will use twice the occupancy of either train. Should there be the possibility of an overlap of times, as shown in Figure IV.3, then both trains together use less than twice the occupancy of either train alone since the superior trains have a shorter time occupancy of the track than the standard train.

The net effect of adding more high speed high priority trains is that fewer operating channels remain for standard trains, and there is a greater probability of overlap by higher speed trains. At a certain point, no more standard trains can be served even though additional higher speed trains can be accommodated until a new capacity/saturation point is reached. See Figures IV.4 and IV.5. The overall effect is illustrated in Figure IV.6.

The addition of still other classes of trains does not alter the basic concept. However it certainly adds complexity to the analysis of train interference. Additionally there is the possibility of a higher priority but slower train (e.g. a local passenger train) interference which ends up using more capacity. For an illustration of such a case, see Figure IV.7.

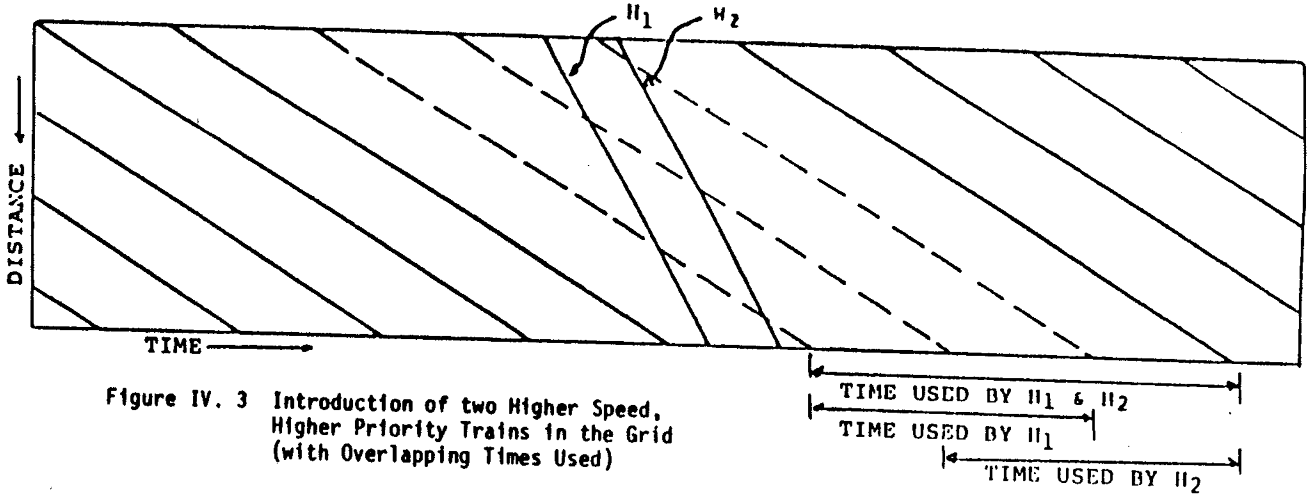


Figure IV. 3 Introduction of two Higher Speed, Higher Priority Trains in the Grid (with Overlapping Times Used)

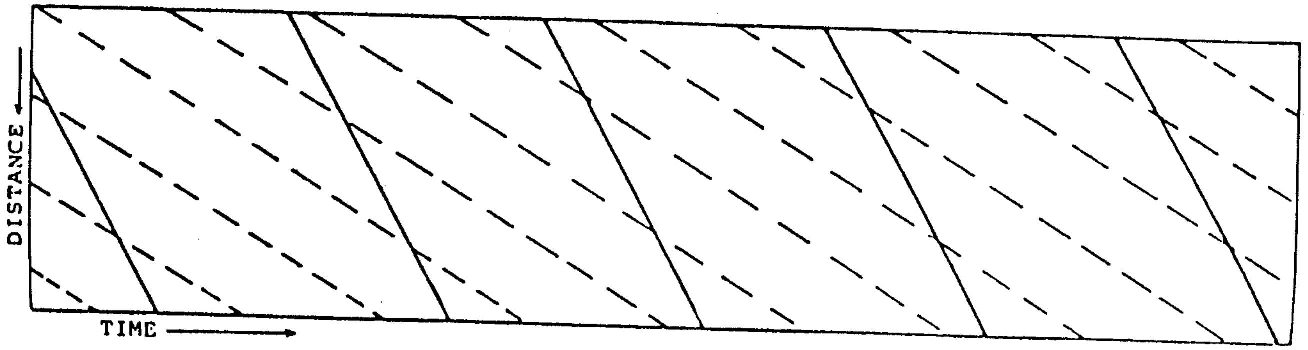


Figure IV. 4 Elimination of Standard Trains

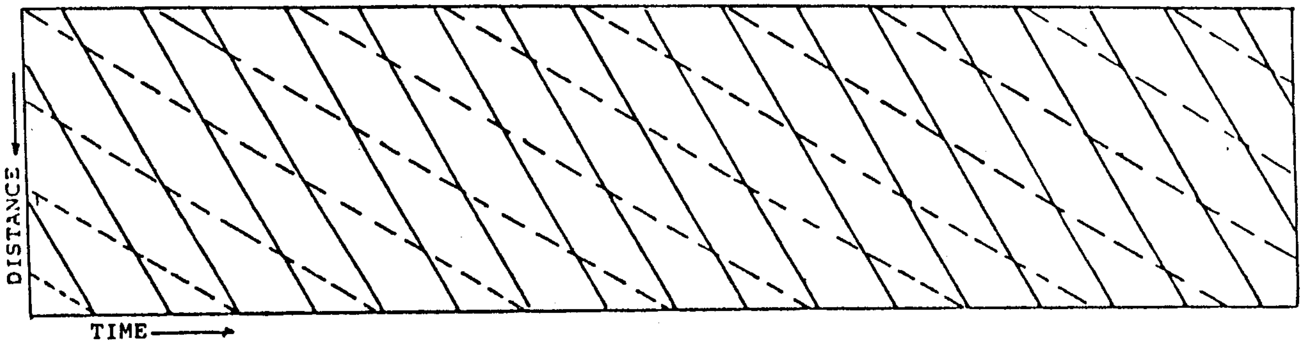


Figure IV. 5 Saturation of the Grid by the Higher Speed, Higher Priority Trains

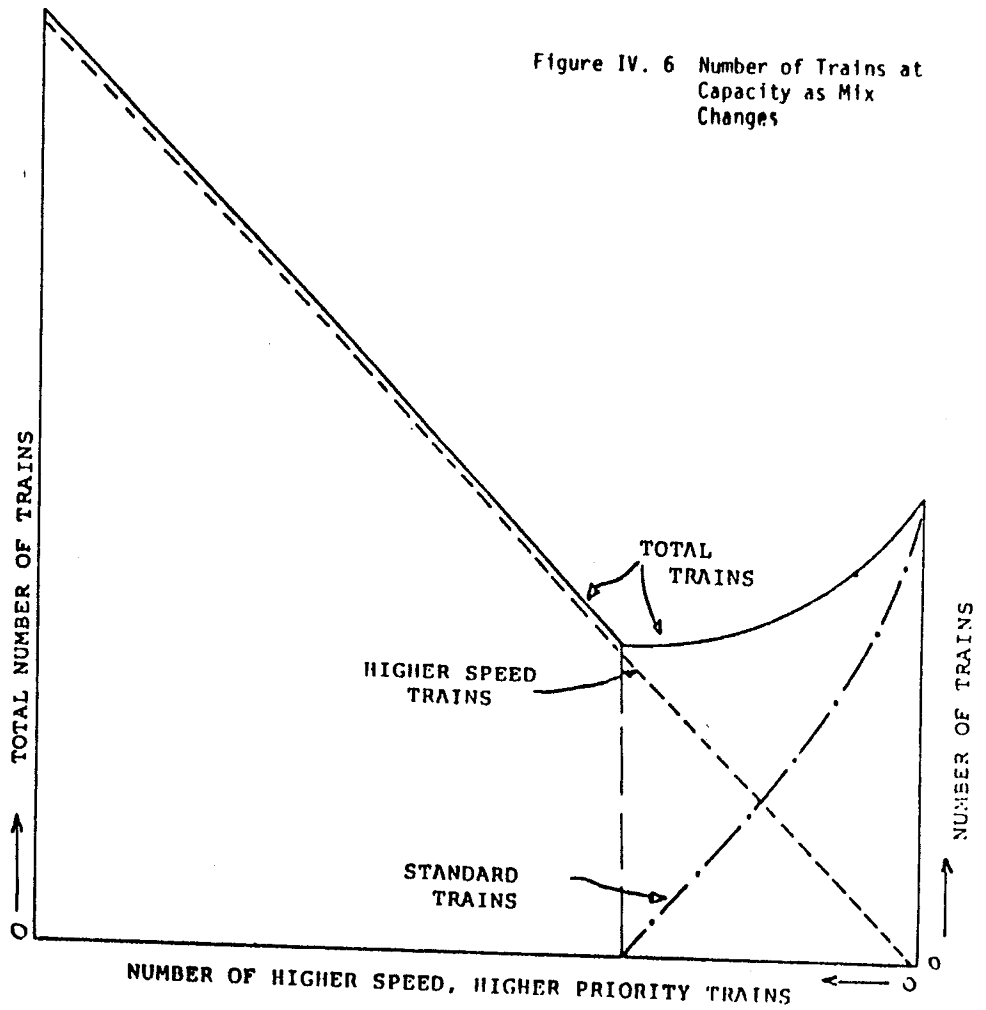


Figure IV. 6 Number of Trains at Capacity as Mix Changes

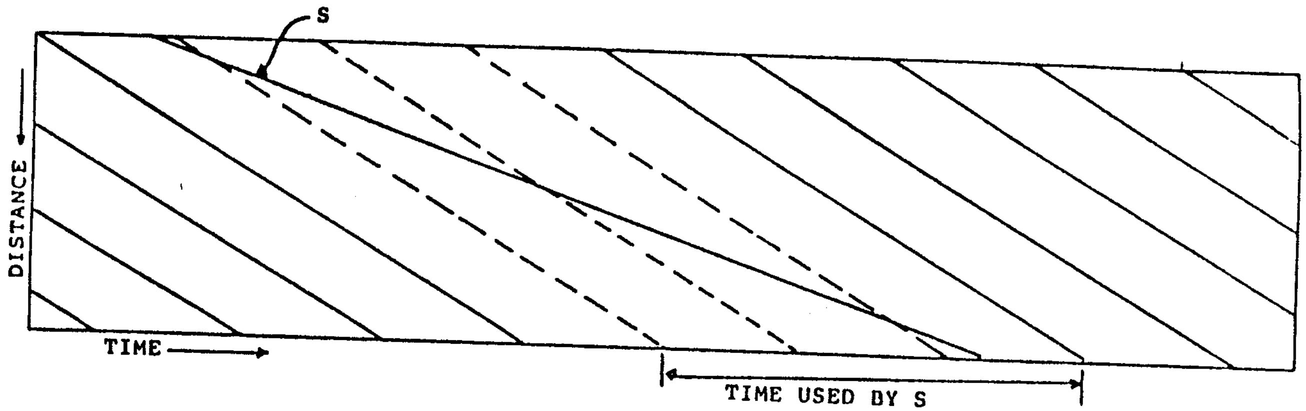


Figure IV. 7 Introduction of a Higher Priority, Lower Speed Train Into the Grid

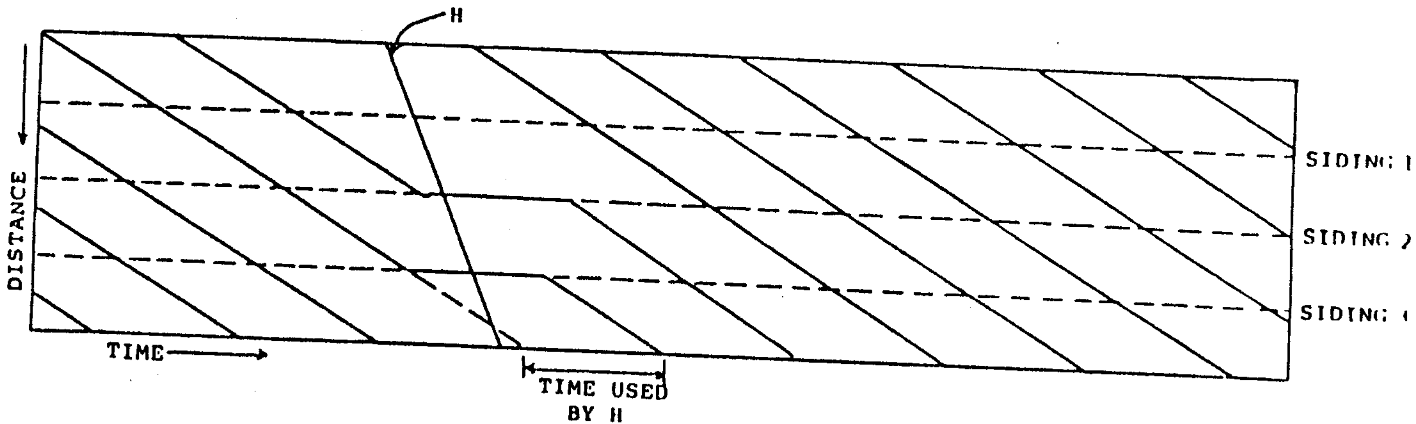


Figure IV. 8 Introduction of Passing Tracks in the Grid

It is possible to reduce the capacity consumption by the higher speed train if it is allowed to overtake the slower trains. This of course implies suitable speed differential and siding spacing. (Figure IV.8.)

Overtakes affect the slow train running time (Figure IV.9). As illustrated, the running times of freight trains vary widely depending upon the mix of trains. Additional illustrations are provided as Figure IV.10 in terms of total number of trains that a line can accommodate as it is affected by the service crew limitation and by the mix of trains. As shown in Figure IV.10, if there is a 10 hour limit on running time, this would limit the number of train combinations achievable.

An interesting observation can be made involving the reverse case, as shown in Figure IV.11. It can be noted that adding freight trains in a specific group of fast trains shows a relatively low influence on the fast train running time.

As is clear from the above, meets and overtakes are interferences which increase the minimum running times of trains and, consequently, affect line capacity. The time penalty due to meets and delays can be precisely estimated for most plant configurations,traffic level and mix of trains. Methodology exists for estimating such effects and for providing answers about line capacity under different demand supply strategies.

Addition of trains to a line, naturally, creates additional meets and increases the running times. Extension of running times, in turn, increases the probability of additional meets. Figure IV.12 illustrates the increase in meets per train and freight train running time due to increasing number of trains. Similarly, Figure IV.13 can be observed as an illustration of the effects of increasing train traffic on a line. The incremental increase in gross-ton-miles per day production follows a diminishing returns pattern.

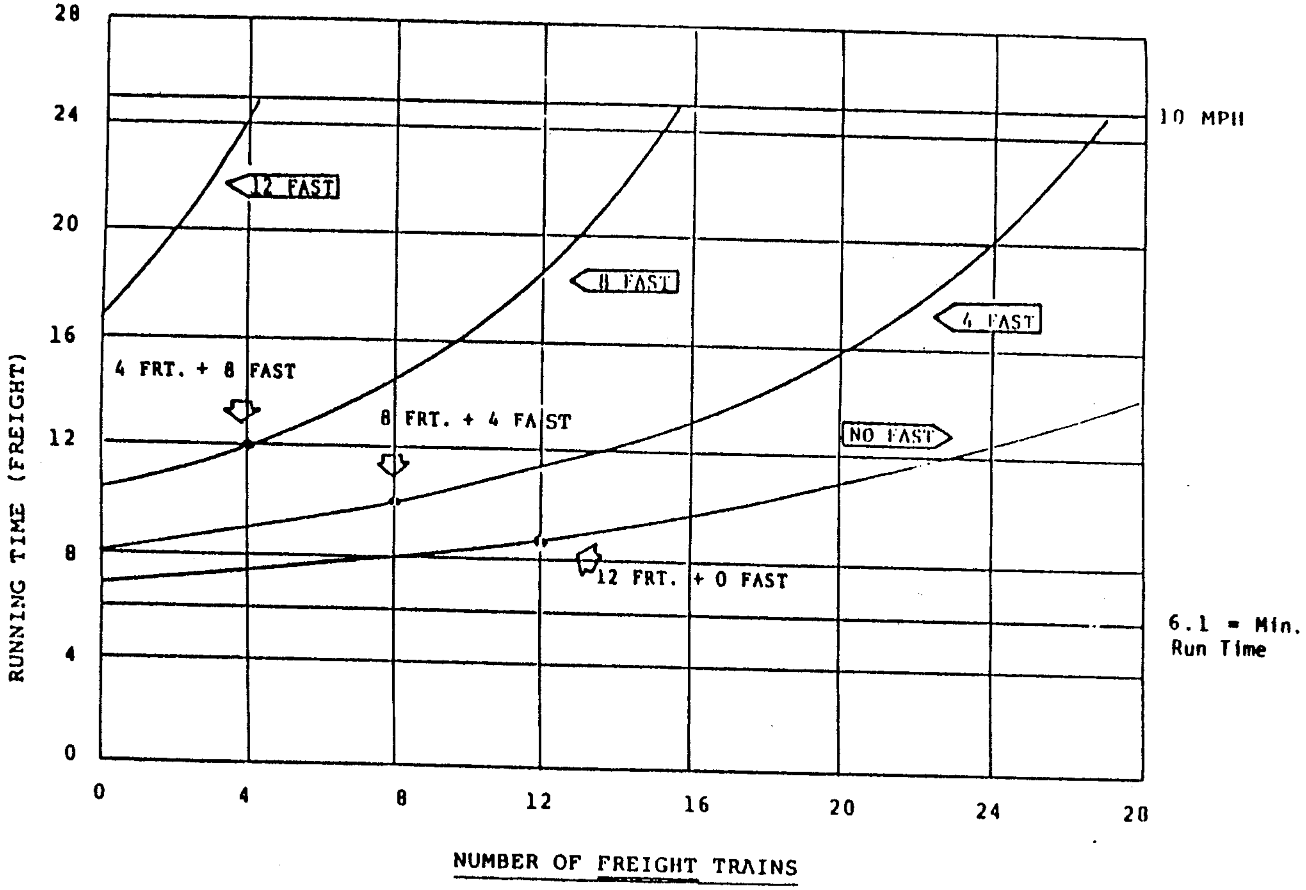


Figure IV. 9 Effect of Train Mix on Freight Times
(Ref. 25)

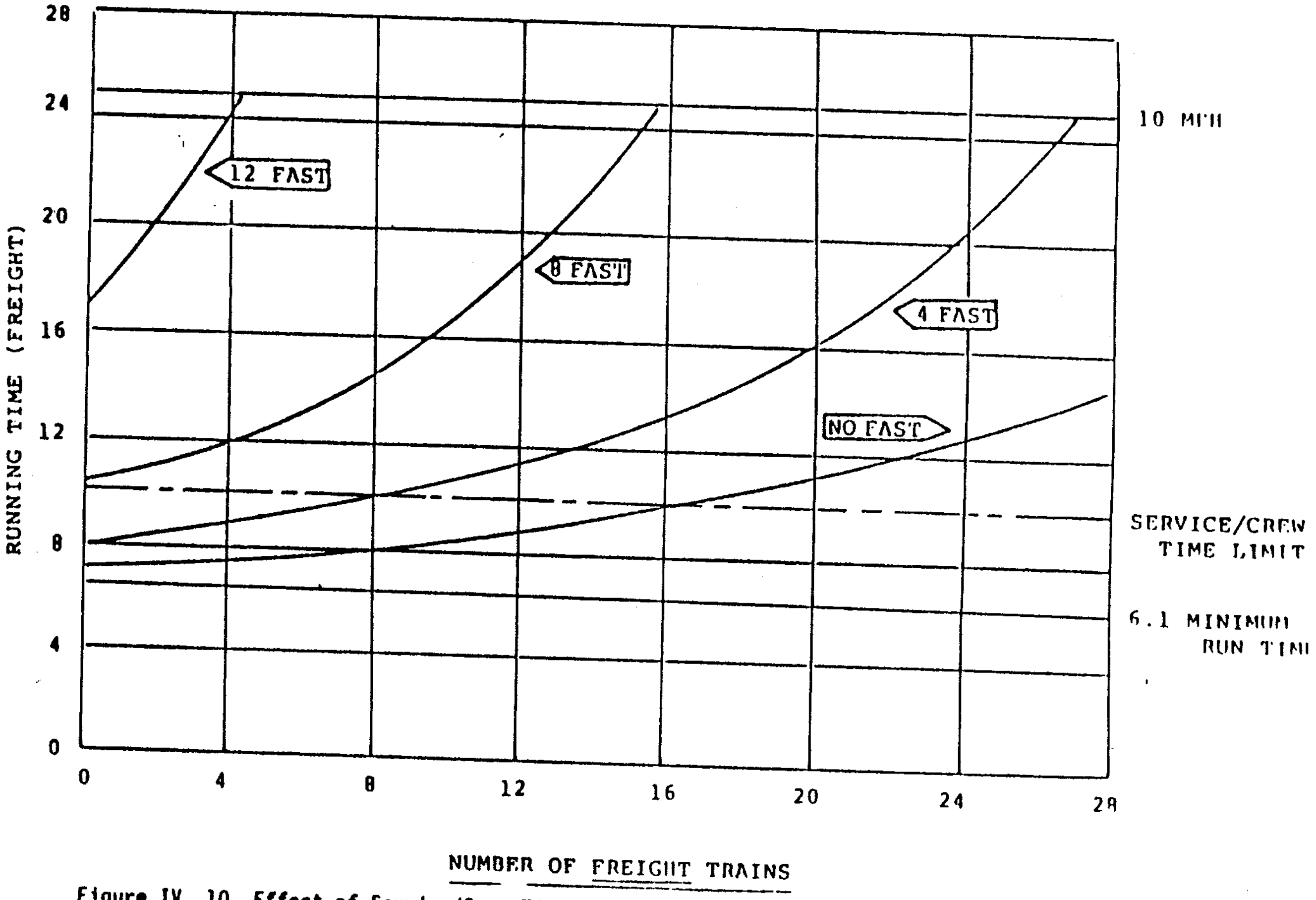


Figure IV. 10 Effect of Service/Crew Time Limitation and Train Mix on Total Number of Trains.

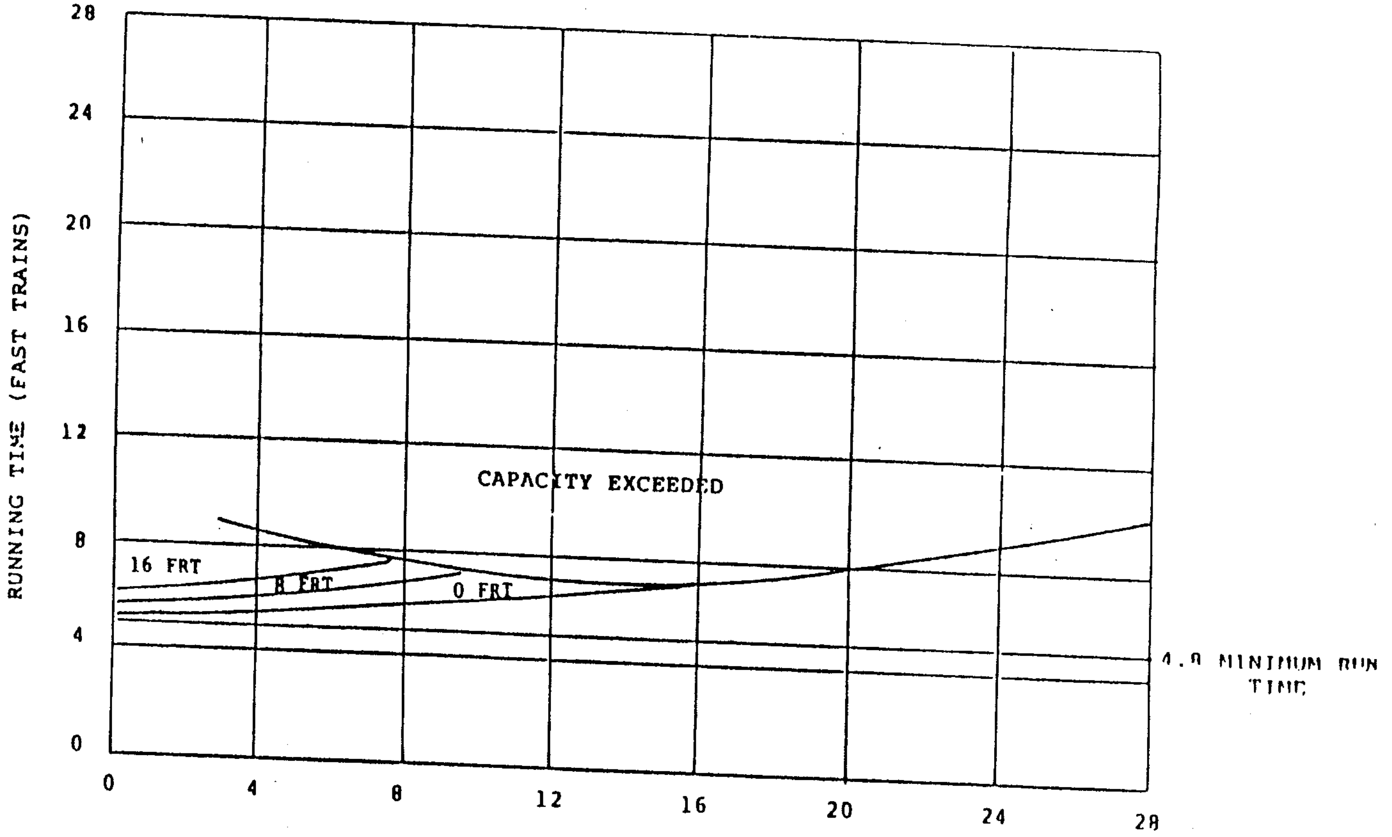


Figure IV. 11 Effect of Traffic Growth on Fast Train Times

BIBLIOGRAPHY

Anderson, F. J., and N. C. Bonsor. 1974. "The Optimal Pricing of Transport Services in the Presence of Congestion." Economica 41: pp. 424-31.

Bell, G. J., D. A. Blackledge and P. J. Bowen. 1983. The Economics and Planning of Transport. London: Heinemann.

Bonsor, N. C. 1977. Transportation Rates and Economic Development in Northern Ontario. Toronto: University of Toronto Press.

______. 1984. Transportation Economics: Theory and Canadian Policy. Toronto: Butterworths.

Burlington Northern Inc. 1977. Estimating Railroad Line Capacity. Industrial Engineering Section.

Canadian National Railway. 1979. Annual Report. Canada: CN.

______. 1980. Annual Report. Canada: CN.

______. 1981. Annual Report. Canada: CN.

______. 1982. Annual Report. Canada: CN.

______. 1983. Annual Report. Canada: CN.

______. 1984. Annual Report. Canada: CN.

______. 1985. Plant Investment on CN's Main Lines: Edmonton and West - 1972-1985. Canada: CN.

______. 1986. Going With the Grain. Winnipeg: CN Grain Communications.

______. 1987. Improving Productivity. Winnipeg: CN Grain Communications.

Canadian Pacific Ltd. 1979. Annual Report. Canada: CP Ltd.

______. 1980. Annual Report. Canada: CP Ltd.

______. 1981. Annual Report. Canada: CP Ltd.

______. 1982. Annual Report. Canada: CP Ltd.

______. 1983. Annual Report. Canada: CP Ltd.

______. 1984. Annual Report. Canada: CP Ltd.

______. 1985. Annual Report. Canada: CP Ltd.

Canadian Transport Commission. 1978. Railway Costing Study: Report on Phase 1. Ottawa: Canadain Transport Commission.

______. 1979. Railway Costing Study: Report on Phase 2. 5 vols. Ottawa: Canadian Transport Commission.

______. Research Branch. 1983. Competition and Regulation in the Railway Freight Industry. Ottawa: Minister of Supply and Services Canada.

______. Research Branch. 1984. Transport Review: Trends and Selected Issues. 1983. Ottawa: Minister of Supply and Services Canada.

______. Research Branch. 1985. Transport Review: Trends and Selected Issues. 1985. Ottawa: Minister of Supply and Services Canada.

Charney, A. H., N. D. Sidhu and J. F. Due. 1977. "Short-Run Cost Functions for Class II Railroads," The Logistics and Transportation Review 17: pp. 345-59.

DeSalvo, J. S. 1969. "A Process Function for Rail Linehaul Operations" Journal of Transport Economics and Policy. January: pp. 3-26.

Duncombe, G. 1976. Transportation Modes. Saskatoon: Dept. of Regional Economic Expansion.

English, G. W., and C. Schwier. 1977. Eastern Canada Railway Line Capacity Study. Kingston: Canadian Institute of Guided Ground Transport (Queen's University).

Fuss, M., and L. Waverman. 1978. *Multi-Product Multi-Input Cost Functions for a Regulated Utility: The Case of Telecommunications in Canada*. Toronto: University of Toronto.

Gilson, J. C. 1982. *Western Grain Transportation: Report on Consultations and Recommendation*. Ottawa: Minister of Supply and Services.

Grain Handling and Transportation Commission. 1977. *Grain and Rail in Western Canada*. 2 vols. Ottawa: Minister of Supply and Services.

Grain Transportation Authority. 1981. *System Design Parameters and Rail Car Requirements: 1981 to 1990*. Winnipeg: Grain Transportation Authority.

Hariton, G. 1987. *Railway Costing: A Review*. Ottawa: Minister of Supply and Services.

Harris, R. G. 1977. "Economics of Traffic Density in the Rail Freight Industry." *The Bell Journal of Economics* 8: pp. 556-64.

Hay, W. W. 1982. *Railroad Engineering*. New York: Wiley.

Holtgrefe, A. A. I. 1975. *An Optimizing Medium-term Planning Model for the Netherlands Railways*. Netherlands: Universitaire Pers Rotterdam.

Keeler, T. E. 1974. "Railroad Costs, Returns to Scale and Excess Capacity." *The Review of Economics and Statistics* 56: pp. 201-08.

Kelsall, J. P. 1986. "The Ins and Outs of Railway System Capacity." Paper presented at the Canadian Transportation Commission Seminar on Transportation Systems Capacity: Requirements and Financing, April 17, 1986, Ottawa.

Khan, A. M. 1979. *Railway Capacity Analysis and Related Methodology*. Ottawa: Canadian Transport Commission, Research Branch.

Kneafsey, J. T. 1975. *Transportation Economic Analysis*. Toronto: D. C. Heath and Co.

Levin, R. C. 1978. "Allocation in Surface Freight Transportation: Does Rate Regulation Matter?" *The Bell Journal of Economics* 9: pp. 18-45.

______. 1981. "Railroad Rates, Profitability and Welfare Under Regulation." The Bell Journal of Economics 11: pp. 1-26.

Manheim, M. L. 1979. Fundamentals of Transportation Systems Analysis: Basic concepts. Vol. 1. Cambridge: The MIT Press.

Mulder, N. 1980. "Western Canada's Transportation Issues in the 1980's: A System Overview." Paper presented at the Western Business Outlook Conference, May 28, 1980, Edmonton.

Oum, T. 1979. "A Cross-Sectional Study of Freight Transportation Demand." The Bell Journal of Economics 110: pp. 463-82.

Peat, Marwick, Mitchell and Co. 1975. Parametric Analysis of Railway Line Capacity. Washington: Federal Railroad Administration, DOT.

Rallis T. 1977. Intercity Transport: Engineering and Planning. New York: Wiley.

Schweitzer, T. T. 1984. Old Myths and New Choices: Railway Freight Rates and Western Economic Development. Ottawa: Economic Council of Canada.

Spector, A. N. 1987. An Assessment of Railway Line Capacity Models Available to the CTC. Ottawa: Canadian Transport Commission, Research Branch.

______. 1987. An Economic Evaluation of Capacity Issues in the Potential Joint Track Usage Territory Between Kamloops and Mission City, British Columbia. Ottawa: Canadian Transport Commision, Research Branch.

Thomson, J. M. 1974. Modern Transport Economics. Great Britain: Penquin.

Transport Canada. 1975. An Interim Report on Freight Transportation in Canada. Ottawa: Ministry of Transport.

______. News release "Minister Committed to Expansion of CN Rail Capacity", March 26, 1980. Winnipeg: Minister of Transport.

______. 1980. Western Railway Capacity. Transport Canada.

______. 1981A. Transport Canada Freight and Passenger

Forecasts. Transport Canada, Economic Analysis Directorate.

______. 1981B. *Capital Stock and Investment in the Transport Services Industry: Rail Transport*. Ottawa: Minister of Supply and Services.

______. 1981C. *Transport Canada's Planning Outlook to the Mid 1980s*. Transport Canada, Strategic Policy Directorate.

______. 1982. *Transport Canada Freight and Passenger Forecasts*. Transport Canada, Economic and Regional Analysis Branch.

______. 1984. *Canadian Freight Transportation System Performance and Issues*. Transport Canada, Policy Studies Branch.

______. 1986. *Freight and Passenger Forecasts 1986: A Summary*. Ottawa: Transport Canada, Economic Analysis.

Walker, R. Remarks to the Northern Regional Meeting of A.A.R. Communications and Signal Section, May 1, 1980.

______. A "Report to Shareholders" given to the Edmonton Chamber of Commerce, February 12, 1986.

Wallace, R. S. et al. 1983. *Rail Technology in the Twentieth Century*. Ottawa: Minister of Supply and Services.

Waters II, W. G. 1983. *Transportation, Transport Policies and the Future Development of Western Canada*. Ottawa: Economic Council of Canada.

Western Transportation Advisory Council. 1976. *Elements of Uncertainty in Western Transportation*. Vancouver: Western Transportation Advisory Council.

Wilson, G. W. 1982. *Economic Analysis of Intercity Freight Transportation*. Bloomington: Indiana University Press.

Wilson, W. A., and E. W. Tyrchniewicz. 1981. "Relationships Between Various Cost Components and Output for Railway Branch Lines in Western Canada." *The Journal of Transport History*. Third series. Vol. 2: pp. 29-38.

Winnipeg Free Press. "Double Track Seen by '88." March 27, 1980.

Wohl, M., and C. Hendrickson. 1984. Transportation Investment and Pricing Principles: An Introduction for Engineers, Planners and Economists. New York: Wiley.